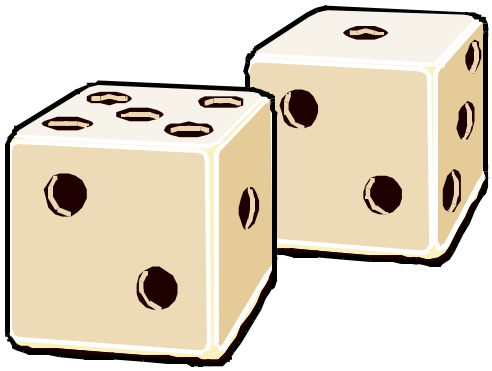
6

Teoria della probabilità



L'evoluzione delle conoscenze ha reso l'umanità consapevole che per controllare le forze temibili con cui convive è inutile sacrificare agli dei, ma le antiche paure sono ancora presenti e l'avanzare delle scoperte non fa che aumentare il mistero intorno a noi. Per non essere sopraffatti dalla sensazione di impotenza abbiamo inventato la teoria delle probabilità come tecnica di gestione dell'incerto e come respingente, almeno psicologico, del caos in cui siamo costretti a muoverci finché non sia svelato il meccanismo del fenomeno che incuriosisce e spesso sgomenta. Secondo A. Eistein la probabilità è anche un gesto di ribellione dell'umanità all'idea di essere un soggetto passivo in balìa dell'universo.

Nel primo paragrafo di questo capitolo discuteremo della casualità; nel secondo sarà presentato l'approccio assiomatico alla teoria elementare della probabilità come tentativo razionale di comprendere la natura di un pianeta che, con la sua indifferenza, ci è ancora ostile. Nel terzo paragrafo si porranno le basi di numerosi schemi sperimentali -basati sul calcolo combinatorio- in cui la casualità ha un ruolo facile da interpretare e con i quali riprendere diversi aspetti di statistica descrittiva lasciati in sospeso o volutamente trascurati; nel 4° paragrafo approfondiremo l'importante concetto della dipendenza stocastica e dei suoi risvolti operativi attraverso il teorema di Bayes.

La teoria della probabilità ha un duplice ruolo nello sviluppo del programma di Statistica. Innanzitutto, la selezione delle unità sulle quali effettuare una rilevazione parziale stabilendo le condizioni per poterne estendere i risultati all'intera popolazione e di questo si parlerà nel paragrafo 6.5. Un altro scopo, altrettanto ampio e interessante, è la predisposizione di modelli teorici che ripensano i fenomeni della statistica descrittiva in un coerente quadro probabilistico. Di questo però ci occuperemo nel prossimo capitolo

E' opportuno avvertire che il calcolo delle probabilità è ingannevolmente semplice: bastano pochi termini per proporre un problema dalla chiarezza palmare, ma la cui soluzione richiede pagine e pagine di calcoli simbolici e numerici.

6.1 Casualità e conoscenza

Secondo J. Watkins (1986) la conoscenza si mostra come un oceano che però non ha una profondità uniforme. In alcuni punti ci sono degli abissi: problemi insolubili con le conoscenze di oggi, problemi che non saranno mai risolti, problemi che nessuno ha mai posto. In altri punti c'è l'affioramento di microisole di certezza, ma solo per qualche attimo fugace ché subito sono sommerse dalle onde dell'incertezza. I fattori sconosciuti, involontari, imprevedibili, fortuiti sono un elemento essenziale e spesso prevalente in ogni problema. Su quello che deve ancora avvenire è possibile pronunciarsi solo in termini incerti, consapevoli della realtà nascosta in esso, ma che non è meno viva e palpitante solo perché ci è sconosciuta. "Tutta *la nostra vita è immersa nell'incertezza; nulla all'infuori di ciò si può affermare con certezza"* (B. De Finetti).

6.1.1 Gli esperimenti in senso statistico

Il termine "esperimento" evoca alambicchi, macchinari, protocolli e persone in camice bianco che si muovono e armeggiano in ambienti asettici; a qualcuno ricorderà le esperienze fatte nei laboratori della scuola superiore a conferma di quanto era loro insegnato. In Statistica, l'idea di esperimento è più ampia: il voto di un'elettrice/elettore, la percentuale di catrame ingerita per fumo passivo, il livello raggiunto da un indice di borsa, il diametro di un tondino di ferro, la preferenza per una marca, il gettito di una tassa, l'esito di un sondaggio elettorale, il rapporto di cambio euro/dollaro, la sentenza di un giudice. Ovunque si attivi un processo di osservazione e/o di misurazione -anche virtuale- di un fenomeno che può dar luogo a manifestazioni variabili là c'è un esperimento in senso statistico.

Casi unici e ripetibili

Un primo utile distinguo è tra fatti unici ed accadimenti ripetibili che possono cioè replicarsi spontaneamente oppure essere indotti artificialmente (fatte salve certe condizioni e garanzie). I casi unici riguardano un fatto volontario o involontario che non può riaccadere perché speciale e isolato o perché le condizioni ad esso antecedenti non possono essere ripristinate o perché si ignorano, perché non sono costanti e seguirne le variazioni comporterebbe costi impossibili da sostenere oppure perché non sono distinguibili da altre concomitanze notoriamente fuori controllo.

Esempi.

a) L'azienda che si dispone ad una fiera curerà l'efficacia dei messaggi espositivi. Il rapporto tra il numero di visitatori dello *stand* ed i visitatori della fiera sarà un indicatore dell'attrazione dell'allestimento, della collocazione, della scelta dei prodotti, del personale. La fiera è però un evento che non si ripete a piacere e tra una fiera e l'altra la comparabilità deve essere esaminata con attenzione.

b) Le azioni positive - adottate per rimuovere il *gap* tra uomini e donne sul lavoro - sono state incentivate dalla legge 125/1991 sulle pari opportunità. La verifica dell'impatto non può avvenire assumendo e licenziando come si desidera, dati i vincoli di legge e la rigidità delle realtà aziendali rispetto al fattore lavoro.

Rientrano negli avvenimenti isolati anche quei fenomeni che possono essere provocati, ma che è troppo costoso o pericoloso replicare deliberatamente: distruzioni di artefatti, incidenti su mezzi di trasporto, esposizione ad agenti inquinanti, attentati, manovre di politica economica (in questi casi ci si avvale di modelli e del computer per simularne il comportamento). In verità esistono anche fenomeni che non sono osservabili o la cui osservazione è difficile a causa di vincoli di morale (abitudini sessuali); di legge (audizioni riservate, processi a porte chiuse, commissioni di inchiesta); militari (piani di difesa, di evacuazione, basi operative, poligoni di tiro, armamenti nuovi). Su questi argomenti le possibilità di indagine sono subordinate all'accesso ai dati.

Quando gli eventi sono negati alla osservazione e non ci si può avvalere di rilevazioni indirette plausibili quegli eventi escono dalla Statistica (ma non dal problema).

Esempi:

a) Il principio della *Common Law* anglosassone è che casi simili debbano essere trattati in modo simile il che presuppone l'esistenza di una comune nozione di somiglianza tra situazioni diverse. Pubblici ministeri ed avvocati della difesa si impegnano in una ricerca strenua nella giurisprudenza dei casi simili giudicati secondo la tesi dell'accusa e di quelli, altrettanto prossimi al caso in esame, ma che risultano a sostegno della posizione della difesa.

b) Talvolta, osserva Curatolo (1980), si è in presenza di fatti singoli che tuttavia possono influenzare anche fortemente una popolazione di persone; ciò che possiamo rilevare in questi casi sono le reazioni e gli atteggiamenti rispetto ai fatti straordinari verificatisi.

c) Freund (1973, p. 2): *"Un altro handicap è la visione miope che l'incertezza sia causata dall'ignoranza e che quindi non ci sarebbe alcun bisogno di studiare la casualità se si potesse conoscere tutto di una data situazione. La miopia di questo punto di vista è che trascura il fatto importante che se a volte si è incerti su di un singolo evento, l'incertezza diventa certezza virtuale quando la stessa argomentazione è applicata ad elevato numero di quegli eventi. Non possiamo sapere se il sig. Brown, forte fumatore, svilupperà un cancro ai polmoni, ma è sicuro che il cancro ai polmoni è sviluppato maggiormente dai forti fumatori".*

Ad un esame superficiale nessun evento si ripete. Ogni fatto è unico ed è impossibile, nella realtà terrena, la sua replicazione esatta e completa: il sole che vediamo ogni giorno non è mai lo stesso e diversa è in ogni momento la luna; eppure esistono accurate tabelle delle eclissi di sole e di luna che consentono di prevedere esattamente quando questi eventi avverranno. Non sempre perciò ci si avventura nell'ignoto perché esistono schemi che possono dare conto adeguato di ciò che si sperimenta. Qui non si intende parlare della mera replicazione di un fatto ché ciò sarebbe inattuabile ed anche inutile: ammesso che fosse possibile la riproduzione esatta di tutte le condizioni -note, sospettate e sconosciute- anche i loro effetti sarebbero costanti e la Statistica non avrebbe materiale di lavoro. Piuttosto si intende il ripetersi di una versione scarna ed essenziale dell'avvenimento le cui caratteristiche rilevanti si mantengono intatte nella turbolenza delle varie manifestazioni: tutte le volte che si configura un insieme di circostanze determinate si può osservare uno spettro fisso di conseguenze (cfr. Scardovi,1996).

Esempi.

a) Una laureata in cerca di prima occupazione vorrebbe partecipare ad un concorso per una posizione molto interessante in un ente di nuova costituzione. Ci sono però costi di segreteria molto elevati. Varrà la pena prendere parte alla selezione?

b) E' in corso un ribasso generalizzato delle quotazioni azionarie. La tentazione è di vendere per evitare ulteriori perdite se avverranno nuovi ribassi; ma se questi non avverranno si perderà l'occasione di buoni guadagni se i titoli posseduti, per un effetto di rimbalzo schizzeranno verso l'alto. Gli esperti in questi casi consigliano nervi saldi, ma qualche nozione di Statistica sarà pure d'aiuto.

c) Una associazione di consumatori intende procedere contro un supermarket perché usa pubblicità subliminale. La proprietà smentisce dicendo che si tratta di una leggenda metropolitana nata con il libro di Vance Packard "i persuasori occulti" pubblicato alla fine degli anni '50 del secolo scorso. Come stabilire chi ha ragione?

d) Un ente locale ha allo studio la costruzione di una diga in un'area poco sviluppata, ma con discrete potenzialità. La popolazione risponderà positivamente aumentando la produzione agricola?

e) La direzione di un'impresa deve decidere se accrescere il *budget* destinato alle attività pubblicitaria riducendo le iniziative promozionali. Fino a che punto il rendimento di lungo termine dell'investimento in *spot* è superiore al beneficio immediato e breve di sconti e regali?

C'è chi sostiene che qui la Statistica non entri affatto trattandosi di situazioni nuove e senza precedenti assimilabili. E' una posizione drastica e potrebbe penalizzare lo sviluppo delle soluzioni perché la ricerca di somiglianze è comunque descrizione di un'esperienza, preludio all'accumulazione di sapere, a sua volta indispensabile per formulare diagnosi corrette e per classificare avvenimenti futuri anche potenziali e perciò, teoricamente infiniti. Intendiamoci, se si richiedono decisioni subitanee la Statistica non potrà dare alcuna risposta e cederà il passo all'intuito, alla disponibilità/avversità al rischio, alla capacità di lettura dei segnali deboli, ad esperienze soggettive, a spezzoni di emotività e ricordi che possono affiorare in una persona o in un gruppo di persone che deve dare una pronta risposta ad una situazione incerta ed incombente. Se il tempo c'è, si potranno cercare ripetizioni della stessa situazione, anche con riduzioni spinte e spericolate analogie formando un archivio di situazioni simili, magari già risolte. Ciò si avvicina ad una rilevazione statistica che aiuterà a decidere su di una base meno volubile.

Esercizio_TP01: *valutate se nei casi elencati sia possibile integrare la base informativa con delle "ripetizioni" ottenute sfruttando generalizzazioni e similitudini creative.*
a) Certi oggetti sono sottoposti a perizie che richiedono esami distruttivi o alteranti l'integrità del reperto (ad esempio far affiorare con inchiostro simpatico uno scritto).
b) Una gentile signora si presenta alla austera giuria del premio letterario "Primo libro" affermando che il libro vincente è suo e non del firmatario del testo. Sia la signora che il titolare contestato non hanno scritto e pubblicato altro.

Le rilevazioni ripetibili riguardano fenomeni che succedono e risuccedono in condizioni omogenee almeno per gli elementi ritenuti troppo lenti, di poco interesse, di poca rilevanza scientifica o di mero disturbo di modo che si possa ritenere senzaltro che sia la stessa situazione che si analizza nelle ripetizioni e non una situazione diversa in ogni ripetizione. Oggetto di studio sono le manifestazioni soggette a variazioni sensibili, sia dovute all'effetto di relazioni evidenti tra i fattori che influenzano il problema che all'effetto di eventi imprevedibili.

Esempio:
In uno studio sui titolari di licenza di caccia fu rilevata l'età, la residenza, le zone preferite, il tipo di arma, l'equipaggiamento, la polizza assicurativa, etc. La licenza era il fatto accomunante di tutte le rilevazioni che però potevano presentarsi con molti elementi variabili. Sono questi che interessano la Statistica.

Esercizio_TP02: *per i seguenti problemi verificate se è possibile la replicabilità (ed a quali condizioni) oppure se si tratta di eventi eccezionali non soggetti a ripetizione:*
a) Effetti di un trattato internazionale sul commercio di certi prodotti;
b) Casi si leucemia tra volontari e militari in zone dove è presente l'uranio impoverito;
c) Portata massima di un fiume;
d) Record del mondo in una gara sportiva.

Esperimento deterministico ed esperimento casuale

Un'altra importante distinzione è tra esperimenti deterministici ed esperimenti casuali. Il primo è un tipo di prova il cui esito è predeterminabile con certezza almeno alla luce delle conoscenze attuali.

Esempi:
a) Se si conosce il lato di un quadrato λ, la sua area sarà λ^2;
b) E' nota l'identità macroeconomica reddito-spesa: Y=C+I+G (consumi +investimenti + spesa pubblica);
c) Il valore attuale della rendita ad annualità costanti posticipate di 1 lira al tasso "r" è:

$$\frac{(1+r)^n - 1}{r(1+r)^n}$$

Fissato il tasso "r" e la durata "n" la formula determina -per un importo unitario- quale sarà il valore attuale.
d) Le orbite dei pianeti del sistema solare sono ellittiche.

Tutti questi enunciati sarebbero suscettibili di verifica empirica anche se ciò è ritenuto superfluo o comunque scarsamente utile dato che il risultato è scontato per l'accumulo di esperienze a conferma e che le discordanze tra teoria e realtà sarebbero attribuibili ad errori materiali o a mere ragioni di ordine pratico che ci impediscono di realizzare una misura scevra da imperfezioni. Questo implica che, nel mondo reale, non si possono effettuare esperimenti perfettamente deterministici, ma solo quasi-deterministici e cioè prove il cui risultato è conoscibile aldilà di ogni ragionevole dubbio, ma non con certezza assoluta a causa di errori strumentali trascurabili, ma presenti. Il progresso dell'umanità è dovuto alla conoscenza di leggi naturali sempre più numerose ed alla fiducia in esse riposta: sappiamo che se si sospende un masso sul piede di qualcuno e poi lo si lascia cadere qualcuno non ne rimarrà contento. Ci sono però altre situazioni in cui l'incertezza non è un innocuo rumore di fondo che disturba il passaggio dal mondo virtuale delle costruzioni teoriche al mondo reale delle prove concrete, ma è invece pervasiva.

Esempi:
a) L'ora, il luogo e le modalità con cui si verifica un incidente automobilistico dipendono da innumerevoli fattori ed una modifica, anche lieve, in qualcuno potrebbe evitare il sinistro: una partenza anticipata di qualche secondo, uno spostamento dello sterzo di pochi millimetri, un battistrada dal disegno diverso. Non è possibile stabilire, tra tutti coloro che si metteranno in macchina domani nei confini del territorio italiano, chi subirà un incidente, ma è praticamente certo che a qualcuno capiterà (si spera con solo lievi danni al mezzo).

b) L'uso delle carte di credito può ridurre i costi legati alla gestione delle note spese, prenotazioni, anticipazioni, valuta estera, fatturazioni. A fronte di un costo annuo permettono il pagamento dopo un certo numero di giorni. Sul mercato esistono diverse carte di credito. Come si orienterà il cliente? La scelta sarà in gran parte razionale, ma incideranno anche fattori come la pubblicità, la diffusione presso parenti e amici, il ricordo di punti vendita in cui vengono accettate, cioè fattori non predeterminabili.

c) Nel controllo di qualità di un prodotto soggetto ad usura l'esperimento consiste nel monitorare il tempo di funzionamento -in condizioni estreme- che precede il blocco o la distruzione del prodotto: la durata di una lampadina, di un pneumatico, di un sistema di raffreddamento. A causa nella naturale variabilità del processo, la vita dei prodotti cambia anche se ottenuti dallo stesso processo.

Requisiti per un esperimento casuale

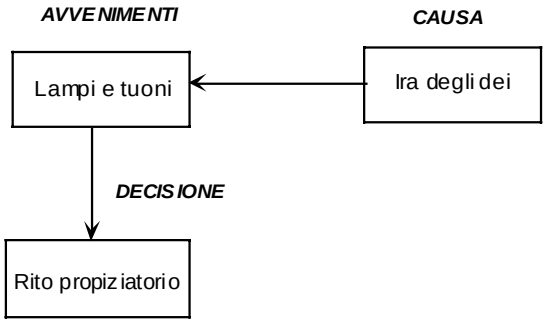
Le situazione proposte negli esempi non possono considerarsi esperimenti deterministici nemmeno in forma idealizzata in quanto le loro determinazioni sono incerte e non basta osservare che alcune sono più frequenti di altre. Un esperimento che non sia deterministico o quasi-deterministico è un esperimento casuale (o aleatorio) in cui cioè ricorrono le seguenti condizioni:
1) Tutte le possibili manifestazioni o esiti della prova sono note a priori.
2) In ogni prova è possibile stabilire quale esito si sia verificato e quale no.
3) La prova può essere riproposta -fisicamente o virtualmente- una, due, infinite volte nelle medesime condizioni senza che si possa prestabilire -dal solito esito della prova- quale sarà la prossima manifestazione o quale sia stata quella della prova precedente.

Compito_TP03:
a) Perché si possa parlare di prova o esperimento casuale esso non deve essere già stato effettuato. Vero o falso?
b) Considerate la seguente affermazione: "una sequenza di risultati è casuale fino a che non si possa dimostrare il contrario". Trovate almeno un argomento a favore ed uno contro tale enunciato.

Incertezza e casualità

Decidere significa stabilire un legame di causalità tra un insieme A ed un insieme B in condizioni di incertezza.

Esempio:
Là dove esistono lacune nella conoscenza entra il mistero ed il paranormale. I popoli primitivi terrorizzati dagli eventi naturali cercavano di placare gli dei con sacrifici propiziatori.



La valutazione del rapporto causa/effetto era legato alla causalità delle loro esperienze: in effetti, cronologicamente dopo la cerimonia, la furia degli elementi pareva quietarsi; ma anche alla casualità: non sempre ciò succedeva anche se la vittima sacrificale era della stessa natura, lo sciamano era lo stesso, stesse preghiere ed invocazioni al medesimo dio ed identica l'arma con cui era inferta la ferita mortale.

La scoperta della casualità deve essere avvenuta constatando che, a fronte di situazioni incerte, dopo aver setacciato bene i fatti, rimane solo il grado di fiducia sul verificarsi o no di un avvenimento. Anche sforzandoci di controllare ogni fattore influente sull'esito di una prova del tipo "lancio di una moneta regolare" non riusciamo a provocare con certezza l'evento "testa" ed ogni volta potremmo solo dire: è più (o meno) probabile che si verifichi testa oppure "sento" che uscirà croce senza che si possa spiegare l'origine della sensazione e del perché a volte sia rispettata ed a volte no; su questo non può incidere neanche l'accumulo di esperienza sullo stesso esperimento o su altri simili. D'acchito ciò appare inaccettabile: se potessimo conoscere il peso esatto della moneta, l'altezza precisa da cui è stata lanciata, l'angolatura con cui è tenuta tra le dita, la forza che il gesto ha impresso, la consistenza, l'estensione, le asperità, l'impatto con la superficie su cui è lasciata ricadere, la polarizzazione, la temperatura, la pressione, l'umidità dell'aria, le condizioni psicofisiche di chi esegue il lancio, insomma la posizione iniziale, la velocità e le traiettorie di tutte le particelle -note e ancora da scoprire- coinvolte nella prova nonché le relazioni che le collegano e ne governano il comportamento, allora dovremmo essere in grado di sapere con certezza come ricadrà la moneta. Questo concetto è il determinismo laplaciano dall'astronomo Laplace (1749-1827) che affermò: *le leggi della natura sono assolutamente deterministiche. La natura non si sbaglia mai; essa non gioca, non sceglie. Essa fissa la successione "necessaria" degli avvenimenti, per quanto piccoli essi siano. Il fine precipuo della scienza consiste nel precisare questa determinazione sottoponendola al calcolo, e solo l'analisi (deterministica) può contribuire a ciò.*

Successive riflessioni hanno fatto abbandonare questa aspirazione: anche nella prova più semplice è coinvolto un numero enorme di fattori talmente interrelati e in sistemi così complessi da scoraggiare ogni tentativo -umano- di predeterminazione. D'altra parte certe conoscenze sono impossibili (per quello che si sa).

Esempi:
a) Per il principio di intederminazione di Heisenberg non è possibile conoscere contemporaneamente la posizione e la velocità di un elettrone;

b) Operazioni che diamo per scontate come l'estrazione di una radice quadrata o il calcolo del logaritmo naturale sono in realtà successioni infinite di operazioni interrotte ad un punto arbitrario;

c) Il numero "π" si trova molto spesso nei calcoli soprattutto quelli che coinvolgono angoli e cerchi. Tuttavia "π" è un numero irrazionale e quindi dal valore indeterminato (fra le sue cifre decimali non è stata riscontrata alcuna struttura pur essendone note più di 6 miliardi).

d) Calcolare il perimetro di un poligono regolare non sembra difficile; lo diventa se si pretende una misura esatta dato che è impossibile per ciò che si è definito come il paradosso delle variabili continue. Anche con il computer più potente, la rappresentazione dell'asse reale presenta tanti vuoti i cui effetti progressivi rendono inaffidabili le operazioni di misura su vasta scala.

e) Pur conoscendo il tipo di relazione che lega due variabili non è possibile specificarlo completamente: è noto che all'aumentare del reddito aumentano i consumi, ma non si riesce a dirne l'ammontare. Talvolta si usa la retta, ma è appunto un'approssimazione.

Instabilità potenziale

Sforzandoci di rimanere nel mito meccanicistico di Laplace potremmo tollerare piccoli errori come inevitabili imperfezioni dovute alla debolezza dei nostri sensi o alle carenze dei nostri strumenti di misura e comunque senza conseguenze significative sull'esito della misurazione. Ma pure questo è contestabile. B. Mandelbrot (1964/ 1997) evidenzia una riflessione di Hadamard il quale, studiando alcune equazioni della meccanica, constatò -sorpreso- che gli effetti di un piccolo cambiamento nella situazione di partenza non sono necessariamente limitati; al contrario, ci sono situazione in cui procedendo a valanga hanno effetti rilevanti; immaginate una matita in equilibrio sulla propria punta: anche un microintervento potrebbe alterare il sistema. Il fenomeno, sottolinea Ekeland (1992, p. 82), è conosciuto in matematica sotto il nome di instabilità potenziale. Le equazioni che governano la circolazione atmosferica sono tali che, in certe condizioni, il battito d'ali di una farfalla in Madagascar può scatenare un ciclone nelle Filippine. Se è così, la relazione causa-effetto può pure essere non aleatoria e conosciuta con la massima precisione, ma non potrà essere utilizzata a fini predittivi. Quindi non solo la conoscenza della situazione all'avvio dell'esperimento avviene sempre entro limiti approssimati, ma gli errori -anche quelli di entità più modesta- potrebbero avere effetti progressivi estremamente sensibili alle condizioni di partenza rendendo indeterminata ogni previsione.

Esempio:

a) J.H. Poincaré ci ricorda "Una causa così piccola da sfuggire alla nostra attenzione può determinare un effetto considerevole, ma siccome non possiamo vedere la causa, diciamo che è un effetto dovuto al caso.

b) A. Eistein non abbandonò mai la tesi che esistono delle variabili nascoste che ci tagliano fuori da una parte importante delle informazioni ed è questo che crea l'illusione della casualità. Altri esperimenti hanno dimostrato che la casualità non è riconducibile ad un sottostante determinismo (Ekeland, 1992, p. 36). Anche Cantelli (1921) parla di effetti complessivi che si comportano come se fossero "dovuti al caso".

c) Prigogine (1997, p. 122): *"Il caso puro è, non meno del determinismo, una negazione della realtà e della nostra esigenza di capire il mondo. Quella che noi abbiamo cercato di costruire è una stretta via tra queste due concezioni che conducono entrambe alla alienazione: quella di un mondo governato da leggi che non lasciano alcun posto alla novità e quella di un mondo assurdo in cui non si può prevedere né descrivere nulla in termini generali".*

d) A. France: *"Caso è lo pseudonimo che il Signore usa quando vuole agire in incognito".*

La casualità è un filo rosso che cuce astronomia, fisica, informatica, meccanica, genetica, farmacologia, biologia molecolare, sociologia, economia, diritto, la finanza e tante altre discipline ed è per questo che la sua trattazione è nata in contesti senza alcun aspetto in comune se non la casualità e la sua valutazione.

Esercizio_TP04: *un esperimento ripetuto un certo numero elevato, ma finito di volte nelle medesime circostanze, ha prodotto sempre lo stesso evento. Ne concludete che la prova è:*
1) Deterministica; 2) Casuale; 3) Sia casuale che deterministica; 4) Né l'uno né l'altro.

Esercizio_TP05: *il comportamento di ogni fenomeno di interesse scientifico sembra governato da due classi di fattori: quelli che da soli sono in grado di esercitare un influsso apprezzabile e quelli la cui influenza è molto piccola. I primi sono poco numerosi e si scoprono relativamente presto; gli altri sono tantissimi e vengono via via scoperti in ragione della loro rilevanza (o quasi). All'aumentare delle scoperte ed al miglioramento delle tecniche di osservazione e di analisi cresce il numero dei fattori individuati; si può quindi ritenere che il progresso scientifico porterà alla conoscenza completa dei fenomeni più rilevanti. Vero o falso?*

Gut (1991, p. 2) sostiene che la differenza tra esperimenti deterministici e casuali sia una questione di scala: i primi descriverebbero i fenomeni a livello macro guardando al loro comportamento da lontano e, a grande distanza, le fluttuazioni sono eliminate, le asperità limate, la percezione delle tendenze più nitida; a livello micro, per un effetto *zoom*, tutto diventa importante, anche le turbolenze più impercettibili possono dominare la scena. Questo è ciò che succede all'acqua dei fiumi: la tendenza è dettata dalla legge di gravità, ma a livello atomico le particelle si muovono in maniera caotica. Se guardate alla riga da disegno da molto distante vedrete una linea retta; se cominciate ad avvicinarla noterete delle irregolarità che aumentano man mano che la portate più vicino all'occhio. Se la guardate al microscopio sparirà ogni fattezza riconoscibile per mostrarvi contorni confusi e insospettabili.

Significato della casualità.

Il significato del termine casuale nel linguaggio comune (controllatene la definizione in un paio di dizionari) è diverso da quello usato nel corso di Statistica.

Esempio:
Bertrand Hansen, un grande esperto di analisi tempi e metodi ha suggerito il seguente esperimento: si è chiesto agli studenti in classe di scegliere -senza cooperare o farsi scorgere- un numero scelto a caso tra 1 e 4 e di scriverlo su di un foglio. Il docente ha chiamato i numeri e gli studenti hanno risposto per alzata di mano. Nella tabella è riportata le frequenza.

Numero	Scelte
1	Pochi
2	Molti
3	Moltissimi
4	Pochi

La "3" è fatta da troppe persone perché possa considerarsi "casuale". Tra l'altro, questo è l'ultimo numero chiamato ed è accolto sempre con esclamazioni di meraviglia e sorpresa dopo che la stragrande maggioranza dell'aula ha alzato la mano. Peraltro, questa tabella si è mostrata la stessa in più di una dozzina di esperimenti con corsi di varia numerosità per cui la si può considerare un risultato empirico consolidato.

Nel parlar comune i fatti casuali sono avvenimenti involontari, imprevedibili, accidentali, fortuiti, occasionali; nel corso di Statistica sono una nozione più complessa e sfuggente. Aleatorio è un sinonimo che evidenzia l'impossibilità di dare regole di accadimento certe in analogia alla impossibilità di stabilire condotte per la vincita sicura nei giochi di azzardo.

Esempi:
a) Galavotti e Costantini (1992, p. 50) avvertono: la nozione della casualità appare quanto mai recalcitrante a una definizione.

b) Ekeland (1992, p. 25). La nozione di casualità si decompone in effetti in una moltitudine di proprietà, talmente diverse fra loro da apparire a volte contraddittorie.

c) La casualità non ha nulla a che vedere con qualcosa fatto a caso (Knuth, 1981, p. 5).

La sorte

La sorte (o fato, moira, fortuna, caso, alea) è una forza autonoma, neutra e imprevedibile, esterna ed estranea che agisce con un meccanismo inaccessibile, cieco, cinico, smemorato, capriccioso, incorreggibile, inappellabile, impassibile, indifferente che opera così perché è così che opera: ogni direzione è uguale di fronte alla sorte e a tutte dà la stessa attenzione (o meglio, disattenzione) rispettando a pieno il principio della non discriminazione tra le scelte possibili. E' temuta perché impone l'abbandono di ogni discrezione, di ogni merito, di ogni esperienza e di ogni intervento correttivo, ma è invocata per la trasparenza e l'equità. Non manca nemmeno chi -a torto o a ragione- considera la sorte un'entità paranormale.

Esempi:
a) Il consiglio di amministrazione di un ente pubblico chiese all'ordine provinciale degli ingegneri di indicare dei nominativi qualificati da inserire in una commissione. L'ordine rispose inviando l'elenco completo degli iscritti all'albo invitando l'ente ad estrarre a sorte i nominativi.

b) Per l'assegnazione dei lavori in economia un assessore numerò da 1 a 90 le ditte inserite nell'elenco dei fornitori di fiducia e programmò di affidare gli incarichi in base agli estratti sulla ruota di Napoli.

c) Nel film "La banda degli onesti" i tre improvvisati falsari decidono chi debba spacciare la prima banconota per mezzo della conta: ognuno propone un numero scelto a caso tra 1 e 5 basando poi la scelta sul totale ottenuto.
d) Il comune di Sorrento ha avviato la verifica dell'evasione e dell'elusione della tassa sui rifiuti solidi urbani (TARSU) con estrazione a sorte per gruppi di strade e per singoli contribuenti.

e) L'alea contrattuale è un istituto italiano (legge 41/1986) che porta gli enti pubblici a ridurre del 10% le richieste -comunque argomentate e documentate- di revisione prezzi da parte delle imprese appaltatrici.

f)Ruelle (1992, p. 45). *Nella vita di tutti i giorni troviamo numerosi esempi in cui il nostro datore di lavoro, un nostro congiunto o il nostro governo tentano di manipolarci. Essi ci propongono un gioco sotto forma di una scelta fra varie possibilità, di cui una appare chiaramente preferibile. Noi la scegliamo, dopo di che ci viene proposto un nuovo gioco e così via di seguito. Abbastanza rapidamente, da una scelta razionale all'altra, ci troviamo in una situazione che non ci piace per niente: siamo in trappola. Per evitare questa conclusione, è bene ricordarsi che agire un po' a caso cioè in modo variabile ed imprevedibile è forse la migliore strategia.*

g) A Napoli è stata organizzata una lotteria per distribuire i loculi del cimitero sovraffollato. Utilizzando il sorteggio si è inserita trasparenza in un settore che è sempre al centro di polemiche e favoritismi.

h) Il sorteggio è un congegno operativo, il più oggettivo fra quelli ritenuti possibili, per soddisfare esigenze di obiettività ed imparzialità: alcune commissioni esaminatrici, dopo aver pubblicato le domande, le fanno sorteggiare ai candidati. Attenzione! Quando il sistema è stato applicato in un concorso pubblico e negli esami universitari il Commissario autore e fautore della proposta ha subito minacce, denunce e lettere anonime.

i) L'azzardo in alcuni esami universitari raggiunge livelli di sorte così violenta da sfuggire ad ogni irregimentazione e gli studenti hanno l'impressione di partecipare ad una riffa strapaesana piuttosto che ad un sistema ordinato di valutazione delle competenze raggiunte.

Esercizio_TP06: *ripreso da una sentenza della Cassazione. Nella ricerca del nesso di causalità tra la condotta dell'imputato e l'evento, al criterio della certezza degli effetti della condotta, si può sostituire quello della probabilità di questi effetti a produrre determinati eventi. Pertanto, il rapporto causale sussiste anche quando l'opera del medico, se correttamente e tempestivamente intervenuta, avrebbe avuto non già la certezza del successo, bensì serie ed apprezzabili possibilità di successo, tali che la vita del paziente sarebbe stata salvata. Commentate la sentenza dal punto di vista della contrapposizione causalità/casualità.*

La sorte nella vita quotidiana

La pervasività della sorte è tale che la ritroviamo in una grande varietà di fenomeni naturali e sociali, ma anche nei giochi d'azzardo. Questi, basati su schemi che tutti possono capire e discutere, consentono di studiare l'azione della sorte senza i coinvolgimenti emotivi, culturali e filosofici che inevitabilmente filtrerebbero esaminandola in altri contesti. E' per questo (e non per manie ludiche, peraltro legittime) che la trattazione didattica di argomenti come la casualità ha come riferimenti iniziali e/o esplicativi i lanci di dadi, di monete, la ruota della fortuna, la roulette, estrazioni di biglie da una o più urne, lotterie, gratta -e-vinci, etc.

Esercizio_TP07: *si consideri il gioco delle tre carte: un classico in molti film, romanzi e racconti consistente nell'indovinare una carta su tre rapidamente mischiate da chi tiene il banco e disposte coperte davanti al giocatore. Lo si può considerare un gioco d'azzardo (e quindi soggetto all'azione della sorte)?*

Il ricorso alla sorte è legittimo in situazioni che sfuggono all'applicazione di criteri più rispettosi della individualità o in cui la loro applicazione è impossibile. Eppure si sono levate molte proteste, condanne e insurrezioni dell'opinione pubblica quando si è tentato di applicare il sorteggio ai malati da avviare ad una cura, agli immigrati cui concedere il permesso di soggiorno, all'assegnazione di un alloggio di servizio, agli aspiranti ammessi ad un corso universitario, alla nomina per una carica. Le obiezioni rimanevano forti e corali anche quando si faceva notare la sproporzione di risorse da impiegare per una valutazione individualistica: assegnare un contratto di lavoro ad uno delle migliaia di iscritti nelle liste di collocamento richiederebbe tempi lunghi e costi di gran lunga superiori all'importo dei contratti. D'altra parte, scegliere "razionalmente" la presidente di una commissione di collaudo tra centinaia di esperte di pari competenza e prestigio può avere ben poco di oggettivo.

Esempio:

I partecipanti ad un concorso per infermieri tenutosi di recente a Catania furono 8900 e 300 di essi vennero designati come vincitori. Dall'elenco dei risultati si notò che tutti i cognomi dei vincitori iniziavano con la lettera C.

Dice Dacunha-Castelle (1998, p. 239): "*Le nostre società, avendo smesso di vedere nel caso la mano di Dio, hanno orrore del sorteggio, non capiscono più come questo possa introdurre una forma di equità*". Infatti, nonostante il caso sia diventato un prodotto di largo consumo grazie ai quiz televisivi ed alla enorme diffusione delle lotterie e del Superenalotto, la gente continua a diffidare e cerca di cautelarsi contro i freni messi alla sorte. In fondo truccare dei dadi non è difficile: basta riempire le incisioni dei punti che interessano di un composto a base di piombo e quelle che non interessano con un composto ferroso e magnetizzare la superficie su cui i dadi rotolano. Peraltro, non sono mancati scandali anche sulla selezione dei partecipanti a trasmissioni a premi sia sulle reti private che pubbliche tanto nella scelta delle domande che nella scelta per niente casuale dei numeri di telefono da chiamare.

Esempi:

a) Hanno destato molto scalpore le irregolarità scoperte a Milano relativamente alle estrazioni del lotto. Tra i 90 bussolotti, 10 erano riconoscibili perché più nuovi. Prima di ogni estrazione i numeri erano inseriti nei bussolotti nello stesso ordine. A questo punto era necessaria la complicità dei bambini istruiti a selezionare i bussolotti più lucenti e di chi bendava i bambini perché lasciasse uno spiraglio sufficiente. Non era un sistema infallibile (forse proprio questa era la sua forza), ma ha fatto vincere molto denaro fino a che la catena di complicità non si è spezzata. In proposito, una interrogazione parlamentare ha chiesto la sostituzione dei bambini con un robot meccanico che assicuri più trasparenza nel procedimento di sorteggio evitando qualsiasi intervento manipolativo tutelando la credibilità del gioco e la speranza dei più deboli e della povera gente. Non è sicuro che il rimedio non sia peggio del male: i robot debbono essere programmati e sono persone i programmatori.

b) DeGroot (1986, pp. 53-54) illustra la seguente strategia di vincita certa. La società Totomio s.r.l. ha come motto "Vincitori o rimborsati con guadagno"; vende infatti -su *Internet*- l'esito di scommesse semplici (tipo: esce/non esce con sostanziale equiprobabilità). Se il risultato non è quello previsto la Totomio rimborsa il costo della consulenza più la metà di tale costo a titolo di consolazione. Il cliente è indotto a giocare, almeno fino alla concorrenza del rimborso promesso visto che vince o recupera più della spesa; d'altra parte, per fare simili offerte, la società deve avere informazioni recenti e sicure e quindi è bene seguirne il consiglio. In realtà, la Totomio non ha alcuna informazione particolare. E' una società minuscola che opera al minimo di costi consentiti dalla via telematica. Essa distribuisce a caso i vincenti delle scommesse incassando il premio da coloro cui ha predetto l'esito corretto, restituisce il premio di coloro che non hanno vinto e con la metà degli incassi paga ai perdenti il premio di consolazione. Alla Totomio rimane il 25% dei premi complessivamente pagati. Certo, non tutto può andare liscio e ci sono sbilanciamenti temporanei tra entrate e uscite, ma il l'idea funziona e diverse imprese operano sul *Web* o in borsa con questo principio.

Nonostante i trucchi e le alterazioni che incrinano la fiducia nella sorte c'è il sospetto che il richiamo ai diritti individuali, agli interessi legittimi, da opporre alla disumanità dei sorteggi casuali, non sia, in fondo, la richiesta di una impraticabile selezione basata su principi etici e morali, ma il tentativo di mantenere le posizioni acquisite pericolosamente messe in discussione da un potere, quello della sorte, che nessuno è in grado di influenzare se non con artifici ed irregolarità penalmente rilevanti.

Esercizio_TP08: *Downton (1982). Un'autorithy statunitense esamina la condotta di gioco che un esperto invia a fronte di un esborso di 100 dollari. Lo schema opera sulle sestine: 1-6, 7-12,13-18,19-24,25-30, 31-36.*
1) Si scommetta sulla sestina appena uscita: se si vince si continua a scommettere sulla stessa, se si perde si puntano due poste: una sulla sestina uscita ed un'altra su quella dove si è perso.
2) Si continua a giocare puntando una posta sull'ultima sestina uscita e tante altre poste, una per ogni sestina, quante sono quelle su cui si è perso. Se esce lo zero si ripete la giocata.
Il rivenditore affermava che, per ragioni troppo complesse a spiegarsi, la vincita si otteneva in media due volte su cinque e non due volte su sei portando a guadagnare fino a 500 dollari la settimana. Qual'è la vostra opinione?

Tentativo di definizione della casualità

La differenza tra "casuale" e "deterministico" non si può applicare ad una singola manifestazione o a poche manifestazioni. Ipotizziamo di lanciare un dado equilibrato:

$$\{1,2,3,4,5,6,1,2,3,4,5,6,1,2,3,4,5,6,1,2,3,4,5,6,1,2,3,4,5,6,1,2,3,4,5,6,1,2,3,4,5,6,\ldots\}$$

la successione appare troppo regolare: è possibile prevedere ciò che succederà ad ogni successivo lancio e non è questo ciò che si intende per sorte. E' sospetta anche la serie:

$$\{1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,11,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,\ldots\}$$

visto che non c'è alcuna differenza nelle uscite. Più accettabile sembra:

$$\{1,5,3,4,2,3,1,4,3,5,2,6,1,4,3,6,1,5,5,6,2,3,2,5,1,5,3,4,2,3,1,4,3,5,2,6,1,4,3,6,1,5\ldots\}$$

che ha però un difetto: se si considerano coppie di lanci si vede che il primo è sempre minore del secondo e la sorte non dovrebbe avere alcuna propensione a creare periodicità. La sequenza:

$$\{3,1,4,5,2,3,1,6,4,3,5,6,4,1,5,6,3,1,5,4,3,6,1,4,1,6,4,3,5,6,4,1,5,6,3,1,5,4,3,6,1,4\ldots\}$$

è meglio conformata, ma ha anch'essa un difetto: il "2" c'è solo una volta su 42 lanci e ciò non convince se si ha presente l'equità della sorte rispetto ai 6 risultati possibili. Ecco ora una serie realizzata dall'autore:

$$\{2,1,5,3,4,5,1,4,3,6,2,6,2,5,5,1,3,5,2,3,4,3,6,2,4,2,1,6,5,1,2,6,4,3,5,4,1,2,1,3,5,6\ldots\}$$

in cui le uscite sembrano ragionevolmente "sorprendenti" e senza ritardi insoliti. Si potrebbe obiettare che 42 lanci siano pochi per ragionarci a dovere e che forse si cambierebbe parere se la serie fosse più lunga. Il problema però non è la lunghezza della serie; anzi, maggiore è il numero di lanci, maggiore è il numero di successioni che ci appariranno sospette senza esserlo: in 1000 lanci vi sono 5^{500} (un numero con 350 cifre) successioni formate

da 500 “1” nelle prime posizioni e cifre diverse nelle altre 500 e, per brutta che sia, non sarebbe affatto strano incontrarla. Se l’unico indizio per stabilire la casualità di una serie fosse la sua origine tutte sarebbero tutte casuali poiché la medesima procedura può generare ogni successione: e infatti, a rigore, non si dovrebbe parlare di un evento casuale o di una successione finita di eventi causali, ma di una sequenza infinita di eventi casuali.

Knuth (1981, pp. 146-148) cerca di superare le difficoltà considerando coppie di modalità che, nell’ipotesi di casualità, dovrebbero rispettare l’equilibrio delle frequenza relative e cioè le 36 possibili combinazioni: (1° lancio, 2° lancio) dovrebbero mostrarsi con la stessa frequenza. Questo, ad esempio basterebbe ad escludere l’ultima sequenza dato che non sono mai presenti coppie di elementi uguali. E’ possibile costruire delle sequenze “non casuali” che rispettano l’equifrequenza per coppie, terne, quaterne, k-tuple di modalità a partire dalla 1ª posizione aumentando di conseguenza la numerosità delle successioni in esame. Ad esempio, una serie costruita disponendo tutte le possibili coppie muovendo il secondo indice più rapidamente del primo:

$$\{1,1,1,2,1,3,1,4,1,5,1,6,2,1,2,2,2,3,2,4,2,5,2,6,3,1,3,2,3,3,3,4,3,5,3,6,4,1,4,2,4,3\ldots\}$$

mostrerebbe cifre singole con la stessa frequenza, coppie di cifre con la stessa frequenza, ma nessuno la giudicherebbe casuale. Anche se si fanno ruotare gruppi di, diciamo k-tuple le combinazioni di (k-1), (k-2), …, le cifre saranno presenti con la stessa frequenza, ma la trasparenza della regola di costituzione le farebbe escludere dalle serie casuali. Più difficile, secondo la congettura di Knuth, costruire sequenze non casuali che mostrino equifrequenza per k-tuple a partire dalla posizione m-esima per ogni “m” e per ogni “k” che quindi potrebbero superare l’esame di casualità. Restano però dubbi sulla operatività di tale definizione e non è dimostrato che esistano sequenze agevolmente reperibili che verificano la condizione di Knuth. Lo stesso autore, dopo aver impegnato 26 pagine alla definizione di casualità, si dichiara insoddisfatto del risultato.

La casualità, in effetti, è una nozione ostica e sfuggente, più facile da cogliere istintivamente che definire formalmente.

Esempi:

a) David e Barton (1962, p. 184) affermano: *“E’ un fatto pacifico che la casualità non possa essere definita con precisione, ma ciò è possibile per la non-casualità. Non si può affermare con sicurezza che una sequenza sia casuale, ma solo che lo risulta rispetto ad un particolare tipo di non-casualità”.*

b) De Finetti ritiene sufficiente un’ idea intuitiva e non si pone nemmeno il problema di darne una definizione rigorosa (Gavalotti e Costantini,1992, p. 52).

c) Bradley (1976, p.58): *“Una mole notevole di sperimentazioni ha accertato che è impossibile per un essere umano agire da selettore casuale semplicemente decidendo di esserlo”.*

Esercizio_TP09: *un pensiero di S. Agostino: “Nos eas causas quae dicuntur fortuitae ... non dicimus nullas, sed latentes; easque tribuimus vel veri Dei ...” a) C’è idea di casualità? b) Quale influenza ha potuto esercitare sullo sviluppo della probabilità?*

Esercizio_TP10: *la trattazione del calcolo delle probabilità danno a volte per scontati alcuni concetti e definizioni ritenendoli concetti primitivi che cioè non possono essere scomposti in nozioni più semplici e che tutti possono intendere, almeno in linea generale anche senza alcuna particolare spiegazione. Uno di questi è il termine “ugualmente probabili”. Provate a descrivere il seguente esperimento: “scelta di un oggetto tra un numero finito di oggetti ugualmente probabili”.*

Definizione algoritmica di casualità

Nel corso del tempo si sono affermate definizioni alternative di casualità: ad esempio quella di Lehmer e Franklin che desume la casualità a posteriori e cioè se la serie, o meglio, il meccanismo per produrre la serie supera certi test allora se ne ammette la casualità (Knuth però pensa che un generatore di numeri che passi tutti i test di casualità proponibili non possa avere niente di casuale). Un’altra nozione, non direttamente basata sulle frequenze relative, ha origine in alcuni lavori di A.N. Kolmogorov, A. Church, G. Chaitin, P. Martin-Löf che fanno leva sul concetto di algoritmo e cioè riconducono la casualità alle regole da usare per memorizzare e comunicare la successione.

Esempio:

Si supponga di dover trasmettere -per telegramma- il dominio dei valori possibili nel primo estratto di una ruota del gioco del lotto. A questo fine si può comunicare la frase: “conta da uno a novanta con incrementi unitari” quindi otto parole invece delle novanta necessarie a descrivere l’intero dominio. La semplificazione è abbastanza forte per poter considerare la successione: {1,2,…,90} non casuale. Inviare invece i risultati dei cinque estratti per ogni ruota richiede quasi sempre la copia integrale delle estrazioni.

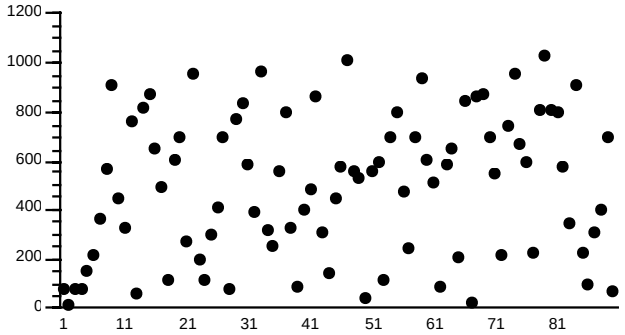
E' importante cogliere l'aspetto euristico della definizione che sembra sia stato trascurato da Chaitin e dagli altri, ma non da M.G. Kendall (1941-42) che afferma "*In Statistica per selezione casuale intendiamo una scelta che se continuata abbastanza a lungo fà comparire tutti i membri ugualmente spesso. Non è il suo carattere erratico che rende casuale una sequenza, ma la sua capacità di produrre limiti definiti"*. La mancanza di sequenze facili da memorizzare può essere solo apparente nel senso che, se nella successione non si sono trovate strutture questo è da attribuire o alla casualità della sequenza oppure al fatto che non si è avuto abbastanza intuito, tempo ed informazioni sufficienti per individuarla. Peraltro, in ogni sequenza finita di numeri è possibile individuare una struttura "non casuale" purché ci si possa ragionare con tempi e risorse sufficienti.

L'assenza di struttura non può infatti essere provata e l'invisibilità non ne implica l'assenza. Ordinamenti più subdoli potrebbero sfuggire.

Esempio:
La successione di numeri compresi tra 0 e 1024 qui riportata appare destrutturata e caotica e non si riesce ad intravedere nessuna regola di comportamento.

68	319	257	572	467	549	500	535	793
1	757	950	381	857	582	73	205	566
69	52	183	953	300	107	573	740	335
70	809	109	310	133	689	646	945	901
139	861	292	239	433	796	195	661	212
209	646	401	549	566	461	841	582	89
348	483	693	788	999	233	12	219	301
557	105	70	313	541	694	853	801	390
905	588	763	77	516	927	865	1020	691
438	693	833	390	33	597	694	797	57

$$Resto(a;b) = a - \left[\frac{a}{b}\right] b$$



In realtà è una sequenza pseudo-casuale ottenuta con il meccanismo dei resti (Knuth, 1981, vol. 2, cap. 3) delle serie nota in matematica come serie di Fibonacci. La sequenza perciò non è casuale, ma ne ha l'apparenza. Se si ignora il meccanismo, ovvero se il meccanismo che genera la sequenza non entra logicamente in contatto con in problema in cui è applicata, la sequenza pseudo-casuale simula egregiamente la sorte.

Peraltro, intravedere una struttura nota per una serie di valori non è affatto una garanzia che quella struttura si manterrà inalterata ovvero seguirà modificazioni prevedibili in base a quanto si è già osservato.

Esempi:
a) La serie dei numeri di Mersenne è espressa dalla formula: $m_k=2^k-1$; se k è un numero primo tale dovrebbe anche essere m_k. La congettura funziona per k=1, 2, 3, 5, 7 e funziona per k=13 e k=17, ma è smentita per k=11 in quanto $m_{11}=2047=89*23$.

b) Kendall (1941) sostiene che non esiste una casualità assoluta come non esiste la velocità assoluta ed entrambe hanno significato relativo: la seconda rispetto ad un sistema di coordinate, la prima rispetto ad un meccanismo di scelta. Kendall, consapevole che nella realtà si ha sempre a che fare con successioni finite di numeri, cerca anche di introdurre il concetto di casualità locale chiedendo alle serie di rispettare, almeno approssimativamente, le proprietà delle sequenze infinite, ma su questo il discorso è meno fluido. Ogni sequenza finita che porti ad una certa struttura di frequenza potrebbe far parte di una sequenza infinita che invece porta ad una struttura diversa senza che si possa escludere o ammettere alcun legame tra le due.

La conclusione è che non esiste una definizione di casualità soddisfacente in ogni occasione ovvero non esistono serie che non manchino di casualità sotto un qualche aspetto fallendo uno dei tanti test cui possono essere sottoposte. In via provvisoria una successione è considerata casuale -dato lo stato delle conoscenze sul meccanismo che la produce- se non è stato possibile stabilire un insieme finito e noto di regole che consenta di prevedere quale sia la modalità nella prossima manifestazione ovvero quelle regole possono rimanere opache nel contesto in cui la serie è adoperata. Non è molto e ci sono predicati vaghi, ma basterà per le nostre applicazioni.

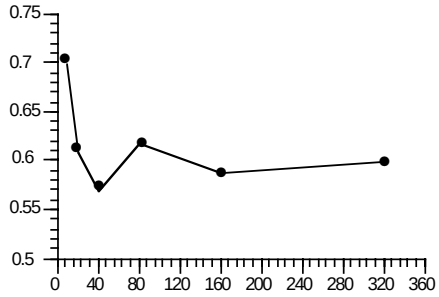
Esercizio_TP11: *Mandelbrot (1964/1997) distingue tre tipi di casualità: sorte benigna, sorte selvaggia e sorte lenta. Nei primi la sorte è addomesticabile: ha la sua autonomia, ma agisce per schemi conosciuti e può essere agevolmente rimossa. La si riscontra ad esempio i giochi d'azzardo, esperimenti di laboratorio, analisi socio-economiche circoscritte. Nei secondi si verificano scarti enormi, regolarità inattese e crolli sorprendenti. Ad esempio gli indici di borsa e i fenomeni meteorologici. La sorte lenta si riscontra in fenomeni con code spesse cioè situazioni in cui possono presentarsi valori grandi o piccoli con una probabilità elevata anche per posizioni molto estreme. Tali fenomeni perdono la selvatichezza solo con in numero enorme di repliche dell'esperimento). Quali situazioni concrete rientrano in quest'ultimo caso?*

6.1.2 Postulato empirico del caso

Una tipica espressione incontrata in Statistica, ma anche in altre discipline è “scegliere a caso” in un gruppo di oggetti cioè scegliere in modo che ogni oggetto possa comparire, ma nessuno in particolare sarà certamente incluso. Questo ha delle implicazioni giudicate positive come si è visto nel paragrafo precedente soprattutto ai fini della trasparenza. Una scelta casuale applicata ripetutamente ad un insieme di oggetti identici, garantisce che ognuno sarà scelto con una frequenza fissa purché l’azione della sorte possa esplicarsi indisturbata abbastanza a lungo. Il perché la frequenza relativa delle manifestazioni di un esperimento casuale tenda a stabilizzarsi su di un valore costante se la prova è ripetuta in circostanze omogenee non è ancora chiaro. E’ però stato verificato in innumerevoli occasioni e del resto possiamo contribuire anche noi a confermarlo con una semplice prova.

Esempio:
Prendete una puntina da disegno -di quelle piccole, tutte in metallo- con la testina non troppo larga. Tenetela per la punta tra il pollice e l’indice perpendicolare ad un tavolo dalla superficie liscia e piana. Mantenete la puntina sospesa a circa 25 cm e poi lasciatela cadere. L’esito della prova si può rilevare con una variabile dicotoma:

$$X_i = \begin{cases} 1 & se\ \ la\ \ punta\ \ non\ \ è\ \ rivolta\ \ verso\ \ l'alto \\ 0 & altrimenti \end{cases}$$



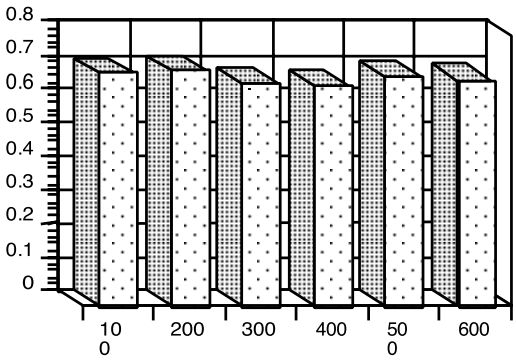
dove “i” è l’ordine della prova. Replichiamo 10 volte il lancio ed annotiamo l’esito in un sistema cartesiano riportando in ascissa il numero di prove e in ordinata la frequenza relativa di X_i=1. Effettuiamo quindi un ciclo di lanci -cercando di mantenere costante l’altezza da cui la puntina ricade, l’impulso dato con l’apertura delle dita, etc. per valori di n=20, 40, 80, 160,… (è noioso, ma ne vale la pena e d’altra parte era raccomandato da Yule come simpaticamente ci ricorda Keynes, 1994, p. 390). Noterete che la frequenza relativa della modalità X=1 differisce sempre meno (sia pure con delle oscillazioni) da 0.6 (o altra costante dipendente dal tipo di puntina) e se si potesse aumentare sempre più il numero di prove, gli scarti da 0.6 diventerebbero trascurabili. Peraltro, il risultato collima con la previsione di Walley (1991, p. 20). Il valore intorno a cui su raggruppano le frequenze potrebbe essere tanto p=0.599 che p=6.001 ed ogni altro valore compreso nell’intervallo. Si è scelto 0.6 perché semplifica i calcoli che, nella maggior parte delle applicazioni, non avrebbero alcuna variazione significativa usando un valore leggermente diverso.

Ogni lancio della puntina da disegno è una prova a sé che ha solo legami fisici impercettibili con le altre (l’usura del tavolo, del metallo, della pazienza di chi sperimenta, etc. cioè tutti aspetti sanabili e quindi singolarmente insignificanti ai fini del risultato dell’esperimento) per cui tutte le prove rientrano nella categoria “repliche indipendenti dello stesso esperimento”. Perché allora la frequenza dell’evento tende ad essere la stessa in prove diverse? Giacomo Bernoulli scriveva a Leibniz nel 1703: *“... anche la più stupida delle persone sa per non so quale istinto di natura -e senza nessun ammaestramento precedente- che più cresce il numero delle osservazioni e minore è il pericolo di allontanarsi dal vero. Tuttavia, darne accurata dimostrazione matematica è indagine tuttaltro che spregevole”*.

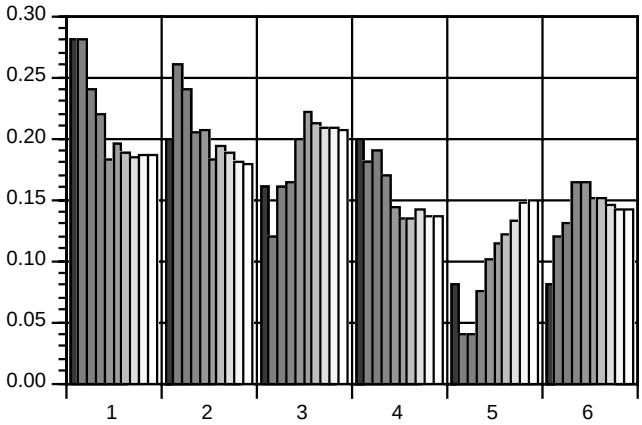
Esempi:
a)Parzen (1960, p. 3-4) conduce il seguente esperimento: in una scatola si introducono sei biglie: due di colore grigio e quattro bianche (indistinguibili per forma, materiale, dimensione, peso, temperatura, porosità, lucentezza). La scatola è agitata e ruotata tante volte da rendere inutile il ricordo dell’ordine e del lato in cui una singola biglia è stata deposta. Si sceglie -alla cieca- una biglia dalla scatola si annota il colore e la si rimette dentro per rifare una energica mischiata. La biglia estratta potrà essere bianca o grigia, oppure dar luogo ad una sequenza ininterrotta di biglie dello stesso colore senza che sia possibile congetturare qualcosa sull’esito di una singola estrazione. Questo però non vuol dire che non si possa dire qualcosa sull’intero esperimento: è stato constatato in molte occasioni che i fenomeni aleatori, considerati per numeri grandissimi, mostrano regolarità sorprendenti, ma significative. Parzen svolse n=600 repliche ottenendo:

Dalla	alla	Estr.	Estr. Tot.	Fr.	Fr. tot.
1	100	69	69	0.69	0.690
101	200ª	70	139	0.70	0.695
201	300	59	198	0.59	0.660
301	400	63	261	0.63	0.653
401ª	500	76	337	0.76	0.674
501	600	64	401	0.64	0.668

Come si vede dall’ortogramma, la frequenza della modalità “biglia bianca” è circa 2/3, sia nelle singole *tranches* di cento prove che nei blocchi cumulativi di 100, 200, etc.



b) Fraser (1958, pp. 6-8), esprime la convinzione che se un esperimento è replicato un numero sufficientemente elevato di volte in condizioni simili, porta all'emersione di una qualche struttura stabile di comportamento nelle frequenze relative. A questo fine effettua un esperimento con un pezzo di plastica avente una grossolana forma di cubo (nessun giocatore serio lo avrebbe considerato un dado da gioco). Dopo aver numerato le facce con gli interi da 1 a 6 lo lancia e ne annota la modalità che si trova sulla faccia rivolta verso l'alto. Fraser ha ripetuto la prova per n=12'800 volte riportando lo stato dei risultati per n=25, 50,100, 200, 400, 800, 1'600, 3'200, 6'400, 12'800 di cui diamo la rappresentazione grafica con un ortogramma multiplo.



Le frequenze relative sembrano inizialmente fluttuare, ma poi, all'aumentare di "n", le oscillazioni si smorzano e le frequenze convergono su dei valori fissi. Le cifre confermano la grossolanità dell'intaglio del dado (in caso di perfetta simmetria dovrebbero essere tutte intorno a 0.167). Fraser pubblica invece:

$$f_1 = 0.186,\ f_2 = 0.179,\ f_3 = 0.207,\ f_4 = 0.137,\ f_5 = 0.149,\ f_6 = 0.142$$

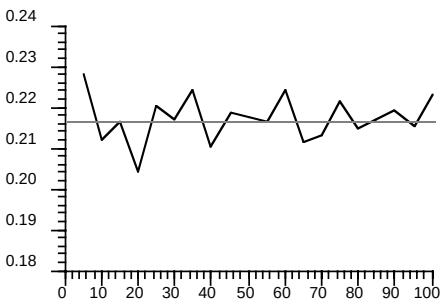
Tali valori possiedono i requisiti della non negatività e somma unitaria, ma non sono delle vere e proprie frequenze relative dato che non sono stati osservati in realtà.

Concezione frequentista alla probabilità

I valori prima ottenuti, opportunqmente rivisti, sono un modello di riferimento al quale le frequenze relative sembrano tendere. Questo modo di costruire le probabilità si chiama "frequentista" in quanto definisce la probabilità come limite delle frequenze relative allorché valgano le condizioni del postulato empirico del caso.

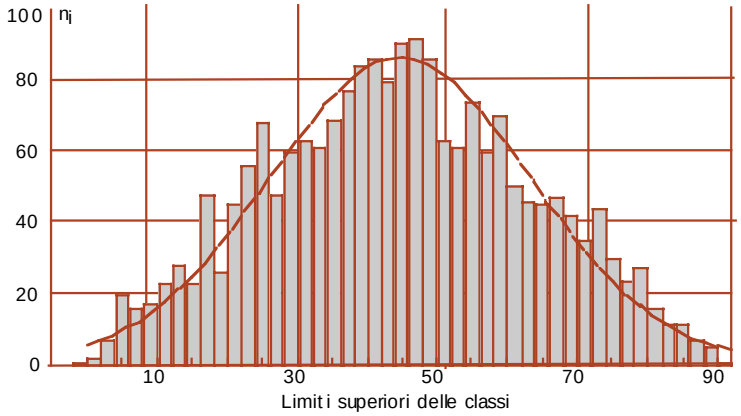
Esempi:

a) Gnedenko (1989, pp. 241-245) mischia un mazzo di carte, lo divide in due mucchi di 26 carte e conta le volte che il primo contiene lo stesso numero di carte rosse e nere. Nel grafico è riportato l'esito di una simulazione informatica di vari cicli di 100 mazzi.



La frequenza tende a fissarsi intorno al valore inferiore 0.218 che è quello che ci si aspetta di ottenere se le carte fossero mischiate in modo soddisfacente.

b) Ho realizzato il seguente esperimento: estrazione di due numeri casuali tra 1 e 90, X_1 e X_2 e calcolo del loro valore centrale: VC=(X_1 + X_2)/2. Estrazione di due altri numeri casuali tra X_1 e X_2 ricalcolo del loro valore centrale e così via per 2000 repliche dello stesso esperimento. Le frequenze (in questo caso assolute) dei valori tendono a seguire il modello detto Normale o gaussiano.



L'impressione che si ricava è che se l'esito di una singola sperimentazione non può essere previsto con certezza, siamo abbastanza sicuri di ciò che succede nel complesso purché si possa disporre di una serie considerevole di repliche e che le frequenze teoriche da stimare rimangano costanti nel frattempo che si conduce la sperimentazione.

Esempi:

a) La giocata sui ritardatari è un'occasione per gli appassionati. Forti dell'idea che un numero in ritardo prima o poi debba uscire, arrivano a sperperare ingenti patrimoni. Dietro questa convinzione c'è un fatto vero: la frequenza relativa di ogni numero dovrebbe essere 1/90 per cui alcuni giocatori, dopo 135 settimane attivano la sequenza delle scommesse al raddoppio. C'è però una falsa premessa: è vero che il numero in ritardo uscirà, ma non è detto che ciò avvenga nell'arco delle possibilità finanziarie dei giocatori e dei loro discendenti.

b) Gli scommettitori del Totocalcio sanno bene che la composizione più ricorrente della schedina vincente è la 6-5-2 cioè sei segni "1", cinque segni "x" e due segni "2" che nel corso dell'anno si presenta varie volte. Il problema non è solo indovinare la giusta combinazione dei segni, ma anche se si verifica nel particolare concorso in cui la si gioca.

Questo fatto è noto come postulato empirico del caso (sembra un ossimoro, ma non lo è). Il valore a cui tende la frequenza relativa ha un ruolo importante nel calcolo della probabilità al punto che è spesso confusa con questa. I valori di convergenza dei rapporti di frequenza possono essere un riscontro empirico di talune scelte di probabilità ed altre volte sono una base di partenza per costruire delle probabilità, ma non sono, almeno non lo sono da sole, le probabilità.

Esercizio_TP12: *scegliete una situazione di vita quotidiana (e quindi facilmente osservabile) per controllare che l'incertezza del suo accadere sia soggetta alla legge empirica del caso. Ad esempio, il numero di semafori verdi incontrati giornalmente recandovi all'università o il numero di clienti che vi precedono al banco mensa.*

L'uso del postulato empirico come base di determinazione della probabilità ha due debolezze:
1) le condizioni degli esperimenti non si possono mantenere costanti molto a lungo e ci sono esperimenti in cui l'ambito di determinazione può modificarsi da prova a prova (ad esempio in prove sequenziali).
2) Ci sono esperimenti che hanno una perfetta natura casuale, ma sono necessariamente finiti per cui la tendenza al limite deve prendere delle scorciatoie (accade ad esempio nelle scelte da popolazioni finite).

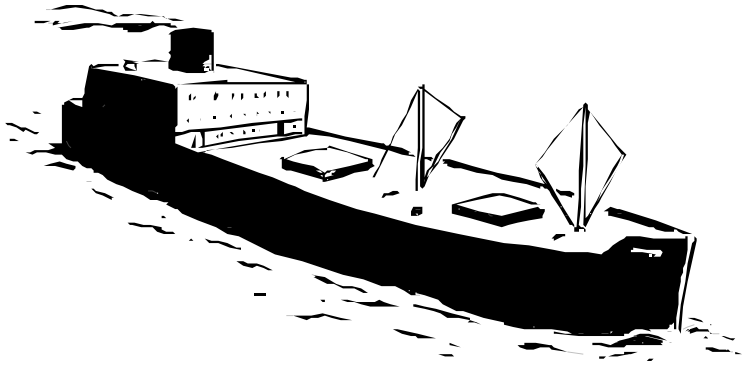
Esempio:

Qual'è la probabilità che il governo chiamato ad iniziare la legislatura sia lo stesso di quello che la conclude? Possiamo considerare come probabilità approssimata il rapporto avente al denominatore il numero di legislature dalla proclamazione della Repubblica ed al numeratore il conteggio (piuttosto esiguo) delle volte che la formazione ministeriale sia rimasta invariata cosicché la probabilità coincide con tale frazione. Naturalmente non c'è alcuna garanzia che il comportamento passato continui nel futuro ed è difficile pensare che un governo del periodo 1953-1957 sia assimilabile ad un governo del 1996-2000. D'altra parte, le legislature finora succedutesi sono troppo poche per poter fornire garanzie di attendibilità.

Esercizio_TP13: *secondo il postulato empirico del caso si può considerare probabilità di un evento la frazione di casi in cui l'evento è accaduto su di un totale moltp elevato di ripetizioni. Ad esempio, la percentuale di votanti alle ultime elezioni amministrative in Calabria è stata del 64.6%. In mancanza di altre informazioni, è possibile affermare che la probabilità che una persona residente in Calabria voti alle prossime amministrative è 64.6?*

6.2 Il modello di Kolmogorov

La costruzione di modelli è importante in tutte le scienze per poter lavorare su una realtà più grande, complessa e mutevole e quindi un modello della casualità è essenziale in contesti dove i cambiamenti avvengono anche sotto l'azione della sorte. I modelli danno però risposte in ragione della loro vicinanza a ciò che rappresentano. Per studiare il comportamento della nave:



non si userà una barchetta di carta, ma una serie di equazioni, disegni e modelli in scala. Il moto dei pianeti è rappresentato con precisione tale che si possono determinare anche l'ora ed i minuti di una congiunzione astrale. Il plastico di un edificio consente di valutarne l'impatto sulle strutture già esistenti, la galleria del vento permette lo studio della aerodinamicità delle auto; è poi notissima la struttura elicoidale dei geni del DNA. Noi siamo alla ricerca di un modello che ci consenta di studiare le componenti casuali dei fenomeni ovvero di una rappresentazione idealizzata delle osservazioni effettuate su di un fenomeno casuale.

Lo sviluppo moderno della teoria della probabilità si è attestato sulla terna (S, W, P) che ha avuto notevole successo nel trattamento di molti fenomeni naturali e sociali. Ogni componente della terna ha un suo ruolo: S elenca gli esiti possibili in una prova casuale, W è un insieme i cui elementi sono a loro volta degli insiemi costruiti combinando-con le regole dell'algebra degli insiemi- gli elementi di S e P(.) è una funzione che misura la casualità associata ad un elemento di W.

La teoria della probabilità è un modello matematico, semplice e potente, per descrivere il comportamento della sorte della quale riporta i tratti più semplici e più caratteristici per poi manovrarli ed arricchirli con il linguaggio matematico.

6.2.1 Insiemi ed eventi

Un esperimento o prova in senso statistico è una situazione di studio circoscritta ad uno o più aspetti di un fenomeno soggetto a variazioni almeno in parte dovute alla sorte. Per prima cosa si deve stabilire quali siano i suoi possibili esiti e cosa debba intendersi per esito. Indichiamo con e_i la descrizione di una delle manifestazioni alternative realizzabili nella prova. L'esito e_i è detto evento elementare perché, ai fini dell'esperimento, è considerato non ulteriormente frazionabile cioè e_i potrà entrare -in tutto e mai in parte- nella composizione di altri eventi, ma nessun altro evento può essere pensato, anche parzialmente, al suo interno. L'insieme dei risultati o eventi elementari della prova forma l'universo degli eventi (si dice anche spazio campionario) cioè un elenco di espressioni incompatibili ed esaustive: $S=\{e_1, e_2, \ldots, e_n\}$ tali che, qualunque sia l'esito dell'esperimento, esso sarà riconducibile in modo univoco ad uno (ed uno solo) elemento di S. La composizione di S riflette le condizioni materiali in cui avviene l'esperimento, il grado di conoscenza raggiunto sul problema, il livello di approfondimento con cui procedere e il punto di vista di chi l'effettua l'esperimento.

Esempi:
a) Per accertare la X= “situazione occupazionale” di 100 laureati nello scorso anno accademico in Economia dopo che siano trascorsi almeno tre anni (e non più di quattro). La sorte entra in gioco sia perché non potendo esaminare tutti i laureati ne scegliamo casualmente solo 100 e sia perché accertare la condizione professionale dei giovani in questi anni defluenti è una specie di avventura. Gli eventi elementari sarebbero moltissimi tanto è variegato il fenomeno, ma possiamo limitarci ai casi: S={occupazione stabile, occupazione precaria, a tempo determinato disoccupazione, non cerca lavoro}.

b) Si intende analizzare una barra radioattiva. Utilizzando un contatore Geiger-Müller si registra il “numero di particelle” che decadono in una data unità di tempo. L’evento elementare è un intero naturale, zero incluso: S={0, 1, 2, …, }. La sorte è da tempo affiancata a questo tipo di studi perché non è possibile stabilire con precisione quante particelle decadranno in un dato intervallo temporale.

c) Si vuole verificare il rendimento di un nuovo ibrido agrumicolo e, a tale scopo, è coltivato un giardino di dimensione, composizione, altitudine ed esposizione standard per la specie. L’evento elementare potrebbe essere “quintali di prodotto” con dominio -in chilogrammi- dato da S={0, 225}.

d) Su di un tavolo ci sono due scatole: una cilindrica ed una cubica. Nella prima si trovano due biglie: una di colore bianco ed una nera; nella seconda altre due biglie una è ancora di colore bianco, ma l’altra è rossa. La prova consiste nello scegliere casualmente una biglia dalla scatola cilindrica e collocarla in quella cubica; da questa poi si estraggono due biglie rimettendo la prima estratta nell’urna per poi estrarre la seconda. Come è formato l’universo degli eventi?

$$S = \{(B,B);\ (R,R);\ (R,B);\ (B,R);\ (N,N);\ (B,N);\ (N,B);\ (R,N);\ (N,R)\}$$

Esercizio_TP14: *un’urna contiene tre biglie di cui una bianca, una rossa ed una nera; una seconda urna contiene due biglie cave di uguale colore in cui sono inserite le cifre “1” e “2”. L’esperimento consiste nello scegliere a caso una biglia da ciascun urna. Qual’è l’universo degli eventi dell’esperimento?*

Esercizio_TP15: *nel poker giocato in Europa il numero di carte è proporzionato al numero di giocatori con la regola seguente: la carta più piccola con cui si gioca è determinata sottraendo da 11 il numero di giocatori. Ad un tavolo con 4 persone si gioca con i 7=11-4 togliendo i 6, i 5, i 4, i 3 e i 2 cosicché rimane un mazzo di 32 carte. Qual’è l’universo degli eventi dell’esperimento scelta casuale del numero di carte ad un tavolo in cui sia scelto a caso (fra 3 e 10) il numero di partecipanti?*

Universo degli eventi
L’universo degli eventi riporta le circostanze che possono succedere in una prova in modo sintetico ed operativo garantendo che, dopo lo svolgimento di ogni prova, non si abbiano dubbi su che cosa si sia o non si sia verificato e che quello che si riscontra è uno degli esiti previsti.

Esempi:
a) L’universo degli eventi non è univocamente individuato dalla descrizione dell’esperimento, ma deve essere chiaramente specificato. Nel caso del lancio di un dado si potrebbe essere tanto interessati alla faccia rivolta verso l’alto che quella poggiata sulla superficie su cui il dado è rotolato od anche al baricentro del dado o al numero di giri che ha compiuto con la forza impressa dal lancio (Ferrari, Leoni, Marliani, 1992, p. 12).

b) All’inizio di una partita di calcio, l’arbitro lancia in aria una moneta per decidere il campo e chi dà il calcio d’avvio. L’universo degli eventi potrebbe essere S={testa, croce, in bilico nel terreno, persa nell’erba, scomparsa in una pozzanghera, rubata da una gazza}. Le possibilità degli ultimi eventi elementari sono così remote che l’attenzione si concentrerà sui primi due e si opererà con S={testa, croce}.

c) I risultati dell’Auditel possono essere analizzati con un S molto dettagliato: per tutte le fasce orarie e per tutti canali televisivi, locali o nazionali e per il segmento di pubblico coinvolto. Spesso, è sufficiente una analisi per macrodati dei canali nazionali per una distinzione semplice: S=*{day time, prime time*}.

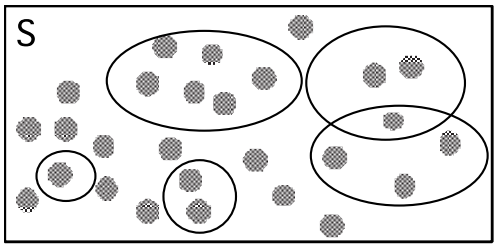
Esercizio_TP16: *Feller (1950, pp. 9-10) introduce uno schema interessante di descrizione dell’universo degli eventi. “k” biglie possono essere collocate in “h” buche e, come in una partita a bigliardo, possono finire tutte nella stesa buca oppure una per ogni buca oppure lasciare alcune buche vuote. Si abbia k=h=3.*
a) Formate l’universo degli eventi ipotizzando biglie distinte (“a”,”b”,”c”, oppure “1”, “2”,”3”);
b) Formate l’universo degli eventi ipotizzando biglie indistinguibili (ad esempio tutte dello stesso colore);
c) Lo schema si applica ad esempio ad un ascensore che collega “k” piani e porta “h” passeggeri che possono scendere o non scendere a qualsiasi piano. Proponete almeno un altro paio di circostanze in cui l’universo degli eventi può essere costruito sulla traccia biglie-e-buche.

Per ora il calcolo delle probabilità è sviluppato su degli universi semplificati, didattici, poco agganciati alle applicazioni pratiche. In questo capitolo privilegiamo la facilità di formalizzazione piuttosto che il realismo dei contesti affinché l’apprendimento dei concetti nuovi non sia disorientato e ostacolato dalle abbreviazioni e forzature cui si è fatto ricorso nell’analisi esplorativa subito presentata nelle sue applicazioni realistiche.

Operazioni con gli eventi

L'universo degli eventi S associato ad un esperimento è un insieme ed è naturale perciò parlare di eventi come insiemi importando uno strumento consolidato e flessibile per discutere degli eventi perché le regole sugli insiemi valgono indipendentemente dalla natura degli elementi in essi inseriti; possiamo inoltre illustrarne i risultati con i diagrammi di Venn, dal matematico britannico John Venn (1834-1923), cioè figure geometriche esplicative disegnate sul piano. L'uso dell'insiemistica non è senza costo perché per poter essere applicata agli eventi, questi devono essere formati da elementi univoci; non possono essere contemplati eventi ibridi, abbinati, sfocati, frazionari o riferibili a più di una prova (sono esclusi ad esempio interessanti esperimenti di meccanica quantistica). Tuttavia, gli aspetti negativi di tali forzature sono compensati dal vantaggio di una trattazione snella e rigorosa.

Una prima estensione sono gli eventi composti, cioè eventi formati da uno o più eventi elementari.



L'evento composto si configura come un sottoinsieme di S che si verifica se si verifica almeno uno degli eventi elementari che contiene. Nel diagramma di Venn, l'universo S è rappresentato con il rettangolo in cui ricadono gli eventi elementari (punti grigi). Gli eventi composti sono i cerchi e le ellissi, interne al rettangolo , disposte intorno a gruppi di eventi elementari. Uno stesso evento elementare può essere comune a più di un evento composto.

Esempi:

a) Per occupare la posizione di vicedirettore si esaminano cinque candidature: {A,Q,R,T,O}. I parametri dell'esperienza, della capacità, lo spirito aziendale, nonché l'età, gli studi, condizioni di salute e moralità sono tutti a livelli ottimi. La selezione deve essere ben ponderata perché ponendo la persona giusta al posto sbagliato (o peggio: la persona sbagliata al posto sbagliato) provocherà disastri all'impresa. Essa è però soggetta a casualità perché dovrà basarsi sulla capacità di interazione e personalità che sono fattori incerti e mutevoli. Prima di arrivare alla decisione si potranno considerare eventi composti del tipo: E_1={A, O}; E_2={Q,R,T}; E_3={Q,R,T,O}; E_4= {A} etc. L'evento E_2 si verifica se la scelta cade sulla candidatura "Q" oppure sulla R o sulla T. Se la scelta cadesse su Q allora si verificherebbe anche E_3 oltre il già citato E_2, l'evento E_4 si verifica solo se si verifica l'evento A.

b) Nella tris di Cesena corrono S={Golden Tango, Bernadette, Can Can, Mon Amour, King, Mambo, Jolly, D'Artagnan, Piripicchio, Butterfly, Soldatino, Antonello da Messina}. L'amico Mandrake, su richiesta della fidanzata Gabriella, dovrebbe giocare M={King, Soldatino, D'Artagnan}; M è un evento composto con gli eventi elementari di S.

c) L'insieme - e quindi l'evento- può essere specificato stabilendo la sua regola di composizione interna e cioè una proprietà che tutti e solo i suoi elementi verificano: A={x| x è una regione italiana}; se x=Molise allora la proprietà è soddisfatta; se x=Brezia la proprietà non è soddisfatta.

L'evento composto deriva da una asserzione logica relativa agli eventi elementari di una prova. Se e_1, e_2, e_3, e_4 sono degli eventi elementari allora E_1={e_1, e_2} e E_2={e_1, e_3, e_4} sono eventi composti. Per indicare lo stato di appartenenza di un evento elementare ad un particolare evento composto si utilizza la simbologia:

$$e_i \in E_j \quad se \ e_i \ è \ un \ esito \ incluso \ in \ E_j$$
$$e_i \notin E_j \quad se \ e_i \ non \ è \ un \ esito \ incluso \ in \ E_j$$

Esempi:

a) Capoluoghi di provincia calabresi: S={Catanzaro, Cosenza, Crotone, Reggio Calabria, Vibo Valentia}.

$$Crotone \in S; \ Castrovillari \notin S$$

b) Giorno della settimana: S={lunedì, martedì, mercoledì, giovedì, venerdì, sabato, domenica}.

$$mercoledì \in S; \ dicembre \notin S$$

c) La cardinalità di un evento, come la cardinalità di un insieme, denota il numero di esiti inclusi in un evento composto: card(Capoluoghi di provincia calabresi)=5, card(giorni della settimana)=7.

Esercizio_TP17: *un ufficio ha quattro sportelli aperti al pubblico. Ogni sportello può essere impegnato al servizio degli utenti oppure libero e per indicare la situazione dell'ufficio si usa il simbolo (a,b) dove "a" è il numero di sportelli in servizio e "b" il numero di quelli senza utenti. Da quali eventi elementari è costituito l'evento composto E= "Almeno due sportelli occupati"?*

Esercizio_TP18: *una sala studio ha 2 tavoli di cui uno con 10 posti e l'altro con 16. L'evento elementare è il numero di posti complessivamente occupato. Da quali eventi è composto l'evento "17 posti liberi"?*

Singoletti, insieme vuoto e universo

Un caso estremo di evento composto è il singoletto cioè l'evento descritto da un singolo evento elementare come l'evento E_4 del primo esempio $E_4=\{A\}$ che ha un solo esito a favore ovvero si verifica solo se la prova genera A. In verità, il termine "evento" dovrebbe essere attribuito solo a quello composto, anche in forma di singoletto, evitando la locuzione "evento elementare" anche se questa è ormai radicata nell'uso. Quando si afferma: "si è verificato e_1" si deve intendere: si è verificato il singoletto $E=\{e_1\}$. La nozione di singoletto consente di applicare le operazioni dell'insiemistica a tutti e solo eventi composti. All'altro estremo c'è l'evento formato da tutti gli elementi in S ovvero E=S. Inoltre, un insieme può risultare vuoto quando non esiste alcun oggetto che possa verificarne la legge di composizione interna: $\varnothing=\{x|\ x$ è un numero dispari multiplo di due$\}$, $\varnothing=\{x|\ x$ è una parola del dizionario italiano con più di 26 lettere$\}$. Per come è costruito l'insieme vuoto è unico, così come è unico l'insieme universo E=S.

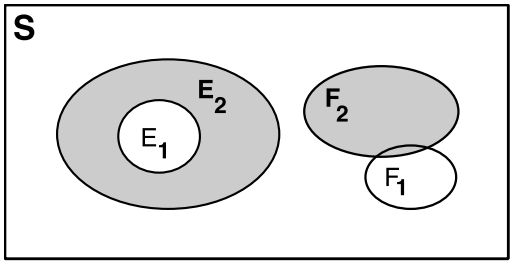
Esempi:

a) Se la nostra prova è il lancio di una moneta con due faccele possibili manifestazioni sono S={C,T} all'interno del quale riconosciamo i due singoletti {C) e {T}.

b) Estrazioni del lotto: $S=\{x|\ 1\leq x\leq 90;\ x$ è intero$\}$. Allora $E_1=\{80, 81,82, 83,84, 85\}$ ed $E_2=\{5, 25, 45, 65, 85\}$ sono eventi composti in S. Lo è pure $E_3=\{12\}$, ma nono lo è $E_4=\{13, 17, 5.5, 91\}$ in quanto due degli elementi di E_4 non sono in S.

c) Estrazione di una carta da un mazzo francese di cui si rileva il seme. L'universo degli eventi è S={C,Q,F,P} tra cui individuiamo gli eventi composti: {C}{Q}{F}{P} {C,Q} {C,F} {C,P} {Q,F} {Q,P} {F,P} {C,Q,F},{C,Q,P},{C,F,P},{Q,F,P},{C,Q,F,P}.

In generale, un insieme E è un sottoinsieme dell'insieme F, scritto $E\subset F$ (ovvero $F\supset E$), se ogni evento elementare inserito in E appartiene anche ad F e almeno un evento di F non è in E; quindi F implica E perché quest'ultimo si verifica ogni volta che si verifica F. L'opposto non è necessariamente vero.



E_1 è un sottoinsieme di E_2, cioè $E_2\subset E_2$, ma F_1 non lo è di F_2 dato che non vi è tutto incluso.

Esempi:

a) Mercati internazionali: S= {NYSE, AMEX, NASDAQ, LSE, SEAQ, LIFE, LTOM, FSE, XTRA, DTB, MONEP, MATIF, ALEX-E, ALEX-D, SWX, ASE, CED, DEKB, RSE, ISMA}. Mercati che hanno Londra come riferimento: E={LSE, SEAQ, LIFE, LTOM}; quindi $E\supset S$.

b) Editoria in Piazza Affari: S={Buffetti +3.4, Class Editori +2.65, Espresso +0.01, Mediaset 1.44, Mondadori -0.47, Poligrafici +0.84, Seat -0.68}. Titoli in calo: E={Mondadori -0.4, Seat -0.68}. Inoltre, card(S)=7, card(E)=2.

Esercizio_TP19: *un funzionario pubblico -esistente in vita- è presente sul posto di lavoro oppure è assente; in questo caso l'assenza può essere giustificata oppure ingiustificata. Nella prima ipotesi rientrano: malattia, indisposizione, congedo familiare, permesso sindacale, incarico istituzionale, permesso breve, missione fuori sede, riunione di lavoro, pausa pranzo. Nella seconda ipotesi si debbono includere: fuga prolungata, uscita breve senza permesso, pausa caffè, papariamento presso altri colleghi, sonno profondo. Costruite l'universo degli eventi ed individuate uno o più eventi composti di possibile interesse in uno studio sull'efficienza organizzativa.*

Uguaglianza di eventi

La nozione di evento composto aiuta a definire in modo rigoroso la scrittura E = F. Cioè due eventi si dicono uguali se e solo se, ogni volta che si verifica E si verifica anche F e viceversa. In altre parole, E è un evento in F e quest'ultimo è un evento di E. Se ciò non succede allora i due eventi sono diversi: E≠F.

$$E = F \text{ se } E \subset F \text{ e } F \subset E$$

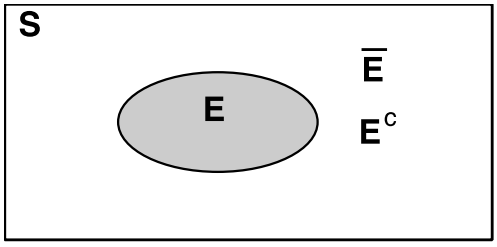
Esempio:
Estrarre da un'urna quattro biglie rosse e due biglie gialle è lo stesso che estrarre due biglie gialle e quattro rosse quando l'ordine di estrazione non è rilevante per definire l'evento.

Esercizio_TP20: *è noto che E≠F e che F≠G. Ne consegue che E≠G. Vero o falso?*

Esercizio_TP21: *sia E= "Giorno del prossimo compleanno di una persona"; F= "Due persone festeggiano il compleanno nello stesso giorno". Si può affermare che E=F?*

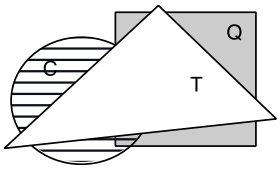
Negazione di un evento

Altri eventi si creano applicando le tipiche operazioni degli insiemi. Dato un evento E, l'evento "Non E" indicato con $\overline{E}$("E" negato") oppure con E^c (E complementare) si verifica se non si verifica E. Nota la costituzione dell'evento E, quella del suo complementare si ricava considerando tutti gli elementi di S non inseriti in E.



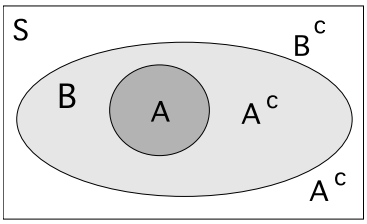
Esempi:
a) La riduzione dei costi di attività è un obiettivo sentito in modo preponderante dalle imprese che lavorano su commessa. Occorre decidere se intervenire su di uno o più sottoprocessi tra quelli che compongono gli standard dell'azienda. L'individuazione dell'area di intervento è lungi dall'essere una scienza esatta nonostante i proclami di alcune società di consulenza. Se i sottoprocessi sono indicati con SP_1, SP_2 , …, SP_8 l'evento: E= "Si interviene su SP_2" ha come negazione "Si interviene su SP_1, SP_3, SP_4,…,SP_8".

b) Immaginate una carta topografica sulla quale siano stati tracciati delle figure geometriche.



Le località che ricadono in una delle figure (è esclusa ogni incertezza di assegnazione) formano l'universo degli eventi. Nel disegno è riportato con tratteggio l'evento complementare a E={Località incluse in "T"}

c) Se B $\subset$ A allora -necessariamente- $A^c \supset B^c$.



Infatti, il complemento di A è sia il cerchio che il rimanente del quadrato; invece il complemento di B è solo la parte bianca esterna alla ellisse. E' chiaro che la definizione del complemento necessita sia della specificazione di S che dell'evento da negare.

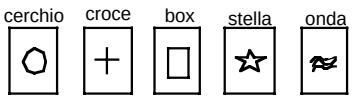
Esercizio_TP22: *un sondaggio intende accertare il canale acceso nei cinque minuti prima della chiamata. Sia S=(R1, R2, R3, C5, R4, I1, E7, MTV, TMC, TMC2, RM, Telecapri, Locali, Satellitari, Pay TV}.*
a) Descrivere l'evento "guardavo canali per i quali si paga un canone";
b) Descrivere l'evento "non guardavo un canale commerciale".

Evento certo ed evento impossibile

Definiamo l'universo degli eventi S come evento certo e cioè l'evento che si verifica in ogni replica dell'esperimento: noi, come chiunque altro, saremmo disposti a scommettere qualsiasi cifra avendolo a favore e nessuna cifra avendolo contro. Che interpretazione dare alla negazione di S? L'evento S^c accade quando non accade l'evento S cioè se si verifica un elemento che non è incluso in S; poiché S contiene già tutti gli eventi possibili, S^c sarà l'insieme vuoto ∅. Logicamente, l'evento ∅ si verificherebbe se l'esperimento desse luogo ad una modalità non in S, ma ciò è impedito ed è per questo che l'evento $\emptyset=S^c$ è detto evento impossibile: nessuna promessa di vincita, comunque grande rispetto alla posta da pagare per entrare in gioco, potrebbe indurci (e nessuna persona ragionevole potrebbe essere indotta) a scommettere in suo favore.

Esempi:
a) Un consiglio è composto da 18 membri. Si vota una mozione molto combattuta. L'esito è incerto: i voti a favore hanno come dominio S={0, 1,2,…,18} che include -attraverso lo zero- anche l'evento "nessuno partecipa alla votazione". L'evento impossibile è che i voti favorevoli siano 19 o più. Attenzione, in alcuni collegi, in caso di parità di voti, prevale la mozione votata dal presidente, il cui voto quindi vale più d'uno e questo non è un evento impossibile anche perché la mozione è combattuta. La contraddizione è solo apparente. I voti espressi sono un esperimento, il destino della mozione è un'altro e richiede un distinto universo degli eventi quale S={approvata, respinta, votazione non valida, votazione rinviata}.

b) Le carte di Zener, inventate dal Dott. Rhine per studiare (e soprattutto confutare) le percezioni extrasensoriali, è formato da 25 carte a 5 a 5 contrassegnate con lo stesso simbolo:



Un tipico esperimento è il seguente: dopo una energica e prolungata mescolatura chi conduce l'esperimento sceglie a caso una carta -senza guardarla- e chiede ad un soggetto di indovinarne il disegno. L'evento certo è che questo sia uno dei cinque disegni; l'evento impossibile è che non lo sia. Perché Rhine ha pensato a queste nuove carte e non le tradizionali carte francesi?

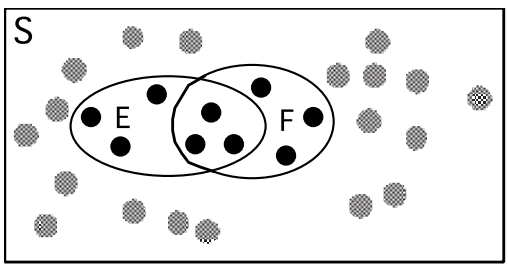
In ogni prova sono sempre presenti l'evento certo ed il suo complemento, l'evento impossibile che, d'ora in avanti, saranno sempre sottintesi in ogni esperimento. Tutti gli altri eventi sono da considerarsi incerti o casuali.

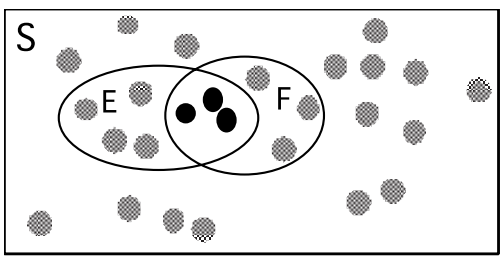
Esercizio_TP23: *sia S={1,2,3,…,15} l'universo degli eventi e siano E={1, 3, 5, 7, 9, 11, 12, 13, 14, 15) ed F={2, 4, 6, 8, 10, 12, 13, 14, 15} due sottoinsiemi di S.*
a) Cosa significa E≠F? b) Determinate E^c ed F^c;

Esercizio_TP24: *Dimostrate che se E non è un sottoinsieme di F allora F^c non può essere un sottoinsieme di E^c.*

Unione ed intersezione di eventi

Vediamo ora come le due operazione binarie degli insiemi: unione ed intersezione applicate agli eventi possano servire a costruire interessanti manifestazioni di una prova.





Dati due eventi E e F. L'evento unione si verifica se accade o l'uno o l'altro o entrambi:

$$(E \cup F) = \{x | x \in E \ \ oppure \ \ x \in F\}$$

Nell'unione rientrano gli esiti contenuti in E, quelli in F e quelli che contemporaneamente appartengono sia ad E che ad F, questi però conteggiati solo una volta.

Esempi:
a) Un rivenditore di computer intende lanciare una campagna promozionale con particolare attenzione alla "rottamazione" delle macchine obsolete. La tipologia trattata riguarda macchine con sistema operativo Windows, MacOs, Unix, Linux. Definiamo come prova casuale il cliente e come universo degli eventi: S={W, M, U, L}. Per il singolo cliente si verifica l'evento {M, L} se il cliente acquista una macchina con sistema operativo MacOS oppure una basata sul Linux o due o più macchine di cui almeno una con sistema operativo MacOS o Linux.

b) Se S={x|x ≤31, x intero}, E={x∈ S|x è multiplo esatto di 3}, F={x∈ S|12≤x≤24, x è pari} allora E∪F={3,6,9,12,14,15,16,18,20,21,22,24}

L'evento intersezione si verifica se accadono entrambi gli eventi:

$$(E \cap F) = \{x | x \in E \ \ e \ \ x \in F\}$$

L'intersezione si compone degli esiti di E che sono anche in F ovvero di quegli elementi di F presenti pure in E.

Esempi:
a) I punti vendita di una catena commerciale attiva nel settore abbigliamento hanno un numero di dipendenti *part-time* ricadenti in S={x intero|1≤x≤20}. Se quelli medio-piccoli sono considerati i punti vendita con *partimers* in E={x intero|1≤x≤14} e quelli medio-grandi i punti ricadenti in F={x intero|12≤x≤20}, quelli medi saranno E∩F={12,13,14}.

b) Se in un concorso per laureate si considerano gli eventi A={x| x è laureata in economia e commercio}, B={x|x è laureata in economia aziendale} allora A∪B={x|x possiede una laurea in economia e commercio o in economia aziendale}; A∩B={x|x è laureata sia in economia e commercio che in economia aziendale}.

Esercizio_TP25: *sia S={i| i=0, 1,2,…, 1'000} e J={iεS|i^2εS} e K={iεS|i(i+1)/2εS}. Come si compone (J∩K)?*

Esercizio_TP26: *dimostrare che E ⊂ F se e solo se E∩F=E.*

Esercizio_TP27: *E= "Carmela otterrà un aumento" ed F= "Carmela otterrà una promozione". Descrivere simbolicamente i seguenti eventi:*
a) "Non è promossa", *b) "Non ottiene un aumento , ma è promossa",*
c) "Non ottiene l'aumento e non è promossa", *d) "O ottiene l'aumento o è promossa".*

Se E ed F non hanno alcun evento in comune si dicono mutualmente incompatibili: E∩F=∅. E' impossibile che si verifichino insieme cioè se accade E non può accadere anche F o viceversa e quindi:

$$E \subset F^c \ \ oppure \ \ F \subset E^c \ \ \Rightarrow E \bigcap F = \varnothing$$

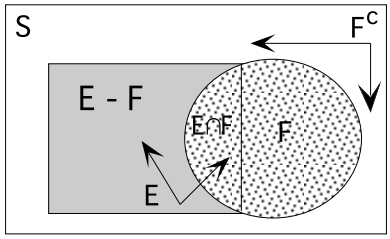
Esempi:
a) Il completamento di un progetto richiede un numero di settimane ricadenti nell'intervallo di interi: S={x|5≤x≤15). Se per un particolare tipo di progetto si impiegano E={x| x≥8} settimane è evidente che ciò è incompatibile con la scadenza F={x| x≤7}.

b) In una indagine campionaria sulle difficoltà di relazione in un campeggio si considerano gli eventi: A=(x|x ha meno di 6 anni}, B={x|x ha più di 77 anni}. I due eventi non hanno alcun elemento in comune e sono incompatibili.

Esercizio_TP28: *si consideri S={x|x è una regione italiana} dal quale si ricavano gli eventi composti: E={x|x è una regione del Mezzogiorno} e F={x|x è una regione con sbocco sul mare}.*
a) Quali regioni formano l'evento: $E \cap F^c$? *b) Quali regioni formano l'evento $(E \cap F)^c$?*
c) Quali regioni formano l'evento: $E^c \cap F^c$? *d) Quali regioni formano l'evento $(E \cup F)^c$?*

Evento differenza
Il passaggio dagli eventi agli insiemi non è automatico ed ogni operazioni insiemistica andrà sempre ben asseverata prima di proporla per degli eventi. Un insiemeinteressante, legato alla negazione, è la differenza tra due insiemi: E - F. Letto in chiave di eventi indica la determinaizione dell'esperimento che si verifica allorché si verificano gli esiti in E, ma non tutti perché ne sono esclusi quelli comuni ad F.

$$E - F = E \cap F^c$$



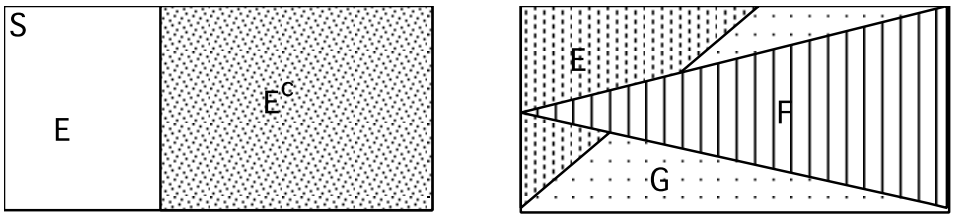
La differenza di due eventi (E - F) è uguale all'intersezione di E con il negato di F.

Esempio:
Lancio di un dado. L'universo degli eventi è formato dagli interi {1, 2, 3, 4, 5, 6}. Definiamo gli eventi composti E: "Punteggio dado > 2", F: "Punteggio dado <5". Poiché A={3, 4, 5, 6} e B={1, 2, 3, 4} ne consegue che A - B ={5, 6} e B - A ={1, 2}.

Esercizio_TP29: *dimostrare che A - B e B - A sono incompatibili e quindi A-B≠B-A;*

Esercizio_TP30: *un'urna contiene 10 biglie numerate da zero a nove. Se ne scelgono due con reimmissione e si rilevano i loro valori con X_1 ed X_2. Detto $E=\{(X_1,X_2)|X_1 + X_2=5\}$ e $F=\{(X_1*X_2>5\}$ determinate E - F.*

Eventi necessari, coperture e partizioni
Se due o più eventi sono tali che almeno uno deve verificarsi, si dicono necessari cioè, congiuntamente considerati, formano l'evento certo:



Gli eventi necessari possono sia essere incompatibili come nel caso dell'evento E e del suo complementare E^c del 1° grafico che sovrapponibili come gli eventi E, F, G del 2° grafico. In tal caso gli eventi formano una copertura finita dell'universo degli eventi. In generale, "k" eventi: $E_1, E_2, \ldots, E_k$ formano una copertura di S se:

$$\bigcup_{i=1}^{k} E_i = S \quad \text{(unione per i che va da 1 a k di E con i)}$$

cioè si tratta di eventi globalmente necessari perché riuniti formano l'evento certo, gli eventi componenti non sono però necessariamente distinti.

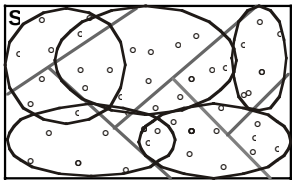
Gli eventi: $E_1, E_2, \ldots, E_k$ formano invece una partizione finita di S se:

$$\bigcup_{i=1}^{k} E_i = S \quad e \quad E_i \cap E_j = \varnothing \quad per \quad ogni \quad i \neq j$$

Esempi:
a) Indichiamo con S={x|1≤x≤k} il numero di interruzioni con cui gli interventi dei parlamentari di un partito sono stati disturbati nel corso delle varie sedute. Una copertura di S è data dalla unione degli E_i={interruzioni subite dal parlamentare i-esimo} dato che lo stesso parlamentare potrebbe aver subito interruzioni in più di una seduta. La partizione dell'insieme delle interruzioni avviene considerando l'unione delle interruzioni subite da ciascun parlamentare per ogni distinta seduta.

b) La suddivisione dell'Italia in grandi comparti territoriali: {Sud-Isole, Centro, Nord-Est, Nord-Ovest} è una partizione anche se in molti commenti sui media sembra intesa come copertura dato che parti del centro, finiscono al Nord (Emilia-Romagna) e parti del Sud si attribuiscono al Centro (Molise).

Esercizio_TP31: *l'universo degli eventi è costituito dalle suddivisioni di una zona in competenze amministrative (le linee) e aree di interesse commerciale (le ellissi). Quale costituisce una copertura e quale una partizione?*



Le operazioni di unione ed intersezione sugli insiemi (e quindi sugli eventi) hanno diverse proprietà algebriche in comune con le operazioni elementari sui numeri, rimanendone però concettualmente distinte ed è bene ricordarsene per non esserne confusi e commettere errori grossolani. Consideriamo tre eventi E, F, G:

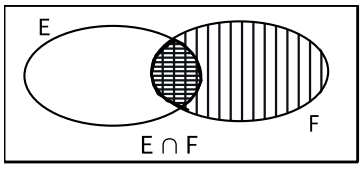
Legge commutativa :	$E \cup F = F \cup E$	$E \cap F = F \cap E$
Legge associativa	$E \cup (F \cup G) = (E \cup F) \cup G$	$E \cap (F \cap G) = (E \cap F) \cap G$
Legge distributiva	$E \cup (F \cap G) = (E \cup F) \cap (E \cup G)$	$E \cap (F \cup G) = (E \cap F) \cup (E \cap G)$
Idempotenza	$E \cup E = E$	$E \cap E = E$
Monotonia : $E \subset F$	$E \cup F = F$	$E \cap F = E$
Convoluzione	$\left(E^c\right)^c = E$	

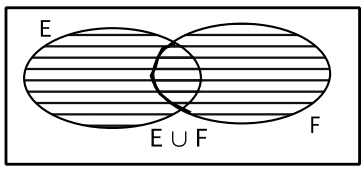
Esempi:

a) Ecco alcune relazioni notevoli che riguardano un generico evento E ed i due eventi estremi: quello certo e quello impossibile.

$$1)\ E \cup S = S; \qquad E \cap S = E; \qquad 2)\ E \cup \varnothing = E; \qquad E \cap \varnothing = \varnothing;$$

b) Verifichiamo che, qualunque siano E ed F, si ha E∪(E∩F) e E∩(E∪F)





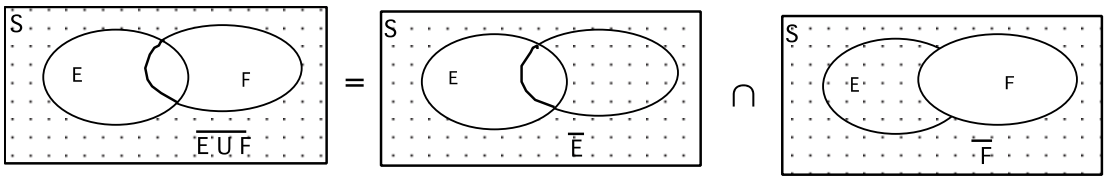
L'algebra degli eventi non richiede in realtà che due operazioni: la terza può essere ricavata dalle altre due in base alle cosiddette leggi di De Morgan:

$$1)\ \overline{E \cup F} = \overline{E}\ \cap \overline{F}; \qquad 2)\ \overline{E \cap F} = \overline{E}\ \cup \overline{F};$$

Il negato dell'unione è pari all'intersezione dei negati ed il negato dell'intersezione è pari all'unione dei negati.

Esempi:

a) Diagrammi di Venn per la prima regola.



b) La candidata ideale ha meno di 30 anni (E<30) ed è laureata da più di 4 anni (L>4) cioè sono escluse le candidate con almeno 30 anni $(E<30)^c$ oppure le candidate con massimo 4 anni già trascorsi dalla laurea $(L>4)^c$: $(E<30)\cap(L>4)=(E<30)^c\cup(L>4)^c=(E>30)\cup(L<4)$

c) La vigilanza di un villaggio turistico opera per quattro turni di sei ore. Indichiamo con T_i l'evento "un addetto è presente nell'i-esimo turno". Supponiamo che un addetto possa essere presente al 2°, oppure negli altri ad esclusione del 4°. Come si esprime in termini insiemistici?

$$T_2 \cup \overline{T}_4 = \{T_1, T_2, T_3\} = \overline{\overline{T}_2 \cap T_4} = \overline{\{T_3, T_2, T_4\} \cap \{T_4\}} = \overline{\{T_4\}} = \{T_1, T_2, T_3\}$$

Poiché le operazioni di unione ed intersezione sono commutative e associative non è difficile estendere le proprietà insiemistiche nonché le leggi di De Morgan ad un generico numero finito di eventi.

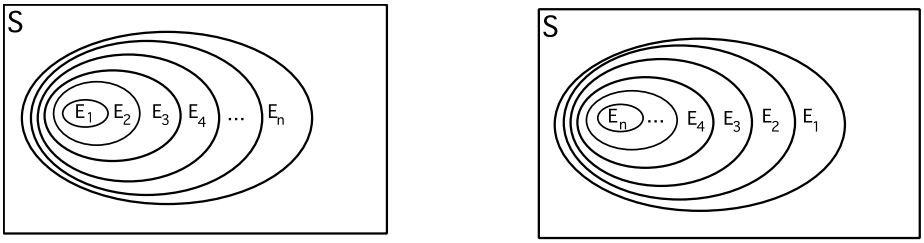
Esercizio_TP32: *dimostrare le seguenti relazioni:*

$$a)\ F=(F\cap E)\cup\left(F\cap E^c\right);\quad b)\ (E\cup F)=E\cup\left(F\cap E^c\right);\qquad c)\ E\cap\left(F\cap E^c\right)=\varnothing;$$
$$d)\ (E\cup F)=E\cup(F-E\cap F);\qquad e)\ (E\cup F)-F=E$$

Successioni di eventi

Il concetto di successione monotona di eventi è importante per alcune estensioni del modello di Kolmogorov che utilizzeremo nel prossimo capitolo. Una successione infinita di eventi (E_i, i=1,2,…,} si dice:

Crescente se $E_1\subseteq E_2\subseteq\ \ldots\ \subseteq E_i\subseteq\ \ldots$ *Decrescente* se $E_1\supseteq E_2\supseteq\ \ldots\ \supseteq E_i\supseteq\ \ldots$



Nella successione crescente il termine che segue implica quello che lo precede, in quella decrescente ne è implicato. Se gli eventi in S sono infiniti altrettanto infinite sono le operazioni che li coinvolgono. I puntini sospensivi significano infatti che la successione prosegue all'infinito.

Esempi:

a) G. Cantor definì enumerabile ogni insieme che potesse entrare in corrispondenza biunivoca con l'insieme degli interi positivi, i cui elementi perciò potessero essere contati uno ad uno. Tale corrispondenza esiste ad esempio per i numeri pari dato che l'abbinamento -un elemento per ciascuno degli insiemi- può continuare all'infinito senza mai esaurire nessuno dei due; La stessa corrispondenza esiste, anche se istintivamente non convince, per i numeri razionali. La corrispondenza non esiste con un intervallo di numeri reali in cui il tentativo di descriverlo elencando tutti i suoi elementi porterebbe ad una contraddizione . Ogni intervallo di numeri reali ha la cardinalità del continuo che è di ordine superiore rispetto alla cardinalità dell'enumerabile (Dauben, 1983).

b) Per gli insiemi infiniti si verifica il fenomeno controintuitivo che un sottoinsieme possa avere la stessa cardinalità -enumerabile o continua- dell'insieme di cui fa parte e cioè {1,2,3,…} e {$10^{1000000}$,$10^{1000000}$+1, $10^{1000000}$+2, …} contengono lo stessa infinità di elementi. L'insieme dei numeri pari ha stessa infinità dell'insieme dei pari e dei dispari. L'apparente illogicità si risolve considerando il meccanismo di abbinamento di elementi presi uno ad uno dai due insiemi e si vedrà che nessuno si esaurisce prima dell'altro.

c) Analizziamo la sequenza di intervalli:

$$E_i=\left\{x\left|\frac{1}{i+1}\le x\le\frac{i}{i+1}\right.\right\}\Rightarrow E_1=\left\{\frac{1}{2}\right\};\ E_2=\left\{x\left|\frac{1}{3}\le x\le\frac{2}{3}\right.\right\};\ E_3=\left\{x\left|\frac{1}{4}\le x\le\frac{3}{4}\right.\right\};\ \ldots E_n=\left\{x\left|\frac{1}{n+1}\le x\le\frac{n}{n+1}\right.\right\};\ E_\infty=\{x|0<x<1\}$$

che risulta monotona crescente ed ha un intervallo limite nell'intervallo unitario. Quella che segue è invece monotona decrescente ed ha come limite l'evento elementare x=1/2.

$$E_i=\left\{x\left|\frac{1}{2}-\left(\frac{1}{2}\right)^i\le x\le\left(\frac{1}{2}\right)^{\frac{i}{i+1}}\right.\right\}\Rightarrow E_1=\left\{x|0\le x\le\frac{1}{\sqrt{2}}\right\};\ E_2=\left\{x\left|\frac{1}{4}\le x\le\frac{1}{\sqrt[3]{4}}\right.\right\};\ E_3=\left\{x\left|\frac{3}{8}\le x\le\frac{2}{\sqrt[4]{8}}\right.\right\};\ \ldots$$
$$E_n=\left\{x\left|\frac{1}{2}\left[1-\left(\frac{1}{2}\right)^{n-1}\right]\le x\le\left(\frac{1}{2}\right)^{\frac{n}{n+1}}\right.\right\};\ E_\infty=\left\{\frac{1}{2}\right\}$$

Le successioni monotone tendono ad un limite definito se proseguite per un numero illimitato di termini:

$$\textit{Decrescente}:\ \lim_{i\to\infty}E_i=\bigcap_{i=1}^{\infty}E_i;\qquad \textit{Crescente}:\ \lim_{i\to\infty}E_i=\bigcup_{i=1}^{\infty}E_i;$$

e cioè la successione monotona decrescente converge all'evento più piccolo incluso in tutti gli altri: $E_i\downarrow E_1$, quella crescente tende all'evento più grande cioè quello che include tutti gli altri: $E_i\uparrow E_\infty$. E' evidente che se gli eventi (E_i} formano una partizione allora:

$$\bigcap_{i=1}^{\infty} E_i = \varnothing; \qquad \bigcup_{i=1}^{\infty} E_i = S$$

Esercizio_TP33: *date le successioni:* $A_i = \left\{ x \middle| -5 + \frac{1}{i+1} < x < 20 - \frac{1}{i+1} \right\};\ i = 0,1,\ldots,\ B_i = \left\{ x \middle| 1 < x < 5 - \frac{i}{i+1} \right\};\ i = 1,2,\ldots,$

a) Verificare che gli A_i *formino una successione monotona crescente e determinarne il limite;*
b) Verificare che i B_i *formino una successione monotona decrescente e determinarne il limite.*

Algebre

L'universo degli eventi descrive i risultati alternativi di un esperimento che confluiscono nei singoletti; ma questi non sono i soli eventi a cui si può essere interessati. In effetti, si è visto che, assimilando gli eventi ad insiemi ed utilizzando le regole stabilite per questi ultimi è possibile definire tanti altri eventi: unione, intersezione, negazione, differenza. Tutti gli eventi costruiti con gli elementi in S formano a loro volta un evento, detto classe o famiglia, i cui elementi sono degli eventi. Ci interessa un particolare tipo di classe: l'algebra, indicata con W che ha le seguenti proprietà:

$$1.\ S \in W; \qquad 2.\ Se\ \ E, F \in W \Rightarrow (E \cup F) \in W,\ (E \cap F) \in W,\ E^c, F^c \in W.$$

cioè l'algebra è "chiusa" sotto le operazioni di unione e negazione di un numero finito di eventi. In altre parole, ogni operazione insiemistica ed ogni loro sequenza finita effettuata sugli eventi nell'algebra W produce sempre e comunque eventi che ricadono in W. L'algebra non contiene eventi elementari ovvero li contiene solo nella forma di singoletti e si deve distinguere tra l'evento elementare "a" che fa parte dell'universo degli eventi S ed il singoletto E={a} che invece fa parte dell'algebra W. Si tratta di una sottigliezza che rende omogenea la composizione di W che così conterrà solo degli eventi composti (almeno uno deve essere presente in W perché si possa parlare di algebra). La terminologia, come si è già osservato, è poco felice dato che gli eventi elementari, in quanto tali, non fanno parte dell'algebra.

Esempi:
a) L'algebra più piccola che si può formare per un esperimento con universo S è W_0={∅,S} cioè l'algebra include solo l'evento impossibile e l'evento certo. Per controllare che si tratti di un'algebra occorre verificare la presenza dell'evento certo (affermativo) e che, considerati due eventi qualsiasi dell'algebra siano soddisfatte le condizioni di appartenenza indicate dalla seconda proprietà:

$$(\varnothing \cup S) = S \in W_0,\ (\varnothing \cap S) = \varnothing \in W_0,\ \varnothing^c = S \in W_0, S^c = \varnothing \in W_0$$

pertanto W_0 è un'algebra, magari troppo ristretta per poter affrontare compiutamente un esperimento, ma del tutto legittima dal punto di vista formale.

b) Controlliamo che anche W_1={E,E^c,∅,S} formato dando all'evento E un ruolo di primo piano, sia un'algebra. Poiché S∈ W_1 la prima proprietà è soddisfatta e tale risultano le condizioni che coinvolgono sia l'evento certo che l'evento impossibile. Per le altre si ha:

$$(\varnothing \cup E) = E \in W_1,\ (\varnothing \cap E) = \varnothing \in W_1,\ E^c, E \in W_1;\ \ \left(\varnothing \cup E^c\right) = E^c \in W_1,\ \left(\varnothing \cap E^c\right) = \varnothing \in W_1,\ \left(E^c\right)^c = E \in W_1$$
$$(S \cup E) = S \in W_1,\ (S \cap E) = E \in W_1; \left(S \cup E^c\right) = S \in W_1,\ \left(S \cap E^c\right) = E^c \in W_1;$$

Si tratta perciò di un'algebra a tutti gli effetti. W_1 è l'algebra più piccola contenente l'evento E.

c) Se W è un'algebra di S e E∈ W, la minima algebra di S contenente anche E è già contenuta in W (Parpinel e Provasi, 1999, p. 468).

Esercizio_TP34: *verificare che l'insieme formato da tutti i possibili eventi composti ottenibili dall'universo S:* $W=\{E|\ E \subseteq E\ \}$ *costituisce un'algebra.*

Esercizio_TP35: *sia S={C,Q,F,P} e si consideri la classe di eventi:* ϕ*, S, E={C,Q}, F={F,P}. Verificare che si tratta di un'algebra anche se non considera tutti i possibili sottoinsiemi di S.*

Se S contiene "n" esiti si può costruire un evento composto considerando o non considerando il primo esito considerando o non considerando il secondo e così via sino all'n-esimo. Le possibilità sono due per il 1° elemento che si combinano con le due del 2° che si combinano con le due del 3° e così via. Ogni composizione è quindi un numero binario ...01011010... Il totale degli eventi che confluisce in un'algebra è perciò: 2^n: se n=10 gli eventi possibili sono 1024. In realtà se ne trattano molti di meno, ma il modello di Kolmogorov si estende a tutto ciò che è coerente con i suoi presupposti e non solo a ciò che riveste interesse in una data applicazione.

Esempi:
a) Scelta di una direzione di marcia con S={M, N, E, O}. Siamo interessati a modellare le opzioni lungo le direttrici M-N e E-0. Le algebre che si possono costituire sono diverse:

$$W_0 = \{S;\varnothing\}; \qquad F_1 = \{M,N\};\ F_2 = \{E,O\} \ \Rightarrow\ W_1 = \{S;\varnothing;F_1;F_2\}$$
$$F_1 = \{M\};\ F_2 = \{N\};\ F_3 = \{E,O\} \Rightarrow\ W_2 = \{S;\varnothing,F_1;F_2;F_3;F_1\cup F_2;F_1\cup F_3;F_2\cup F_3\}$$
$$W_3 = \{S;\varnothing;M;N;E;O;(M,N);(M,O);(M,E);(N,E);(N,O);(E,O);(M,N,E);(M,N,O);(N,E,O);\}$$

W_3 è l'algebra più grande indotta da S. Pesarin (1989, p. 21) evidenzia come le partizioni W_1 e W_2 siano equivalenti, ma la seconda è più fine e quindi più ricca di possibilità operative: $W_1 \subset W_2$. E' evidente che molti elementi di W_3 possono essere sostituiti con espressioni più sintetiche.

b) Lancio del dado. S=(1,2,3,4,5,6}. Consideriamo la copertura E={1,2,4,6} ed F={1,2,4,5}. Per gestire i possibili eventi che possono scaturire dalla prova si propone la classe di insiemi: W={S; ∅; E;F;E^c;F^c;E∪F;E∩F}. E' un'algebra? E' cioè additiva? La risposta è negativa perché ad esempio manca il {3}=E^c∩F^c.

Nell'impostare un modello probabilistico occorrerà anche procedere alla scelta efficace ed efficiente dell'algebra da utilizzare per descrivere gli eventi di interesse nell'esperimento casuale. Nella teoria elementare si sceglie l'algebra più grande costruibile a partire da S.

Esercizio_TP36: *in un settore sono presenti tre aziende: E, F, G che hanno formato un cartello. L'accordo prevede che, in ogni gara d'appalto tranne la prima, una rinunci e solo chi perde abbia diritto a partecipare alla gara successiva senza ritirare l'offerta economica. Supponendo che in un anno si bandiscano 3 gare: a) Definire l'evento elementare per l'aggiudicazione delle gare; b) Definire l'universo degli eventi; c) Proporre un'algebra per l'esperimento.*

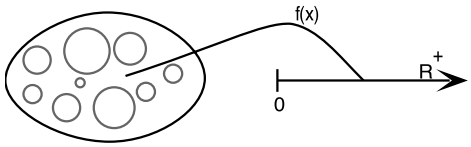
La funzione di insieme

Per misurare la casualità di un evento partiamo dal concetto di funzione di insieme cioè di una regola che associa agli insiemi di una classe un numero reale.

Esempi:
a) Consideriamo la classe W di insiemi E_n={x∈N$^+$|1≤i≤n} dove N$^+$ è l'insieme dei naturali positivi, allora è una funzione di insieme:

$$f(E_n) = \sum_{i=1}^{n} i \Rightarrow \qquad f(E_1) = 1; \qquad f(E_2) = 3; \ \ldots \qquad f(E_n) = f(E_{n-1}) + n$$

b) Insieme dei cerchi nel piano con associata la circonferenza:



La funzione di insieme è additiva se, data la coppia di insiemi E_1 ed E_2 rientranti in W e tali che $E_1 \cap E_2 = \varnothing$, allora $f(E_1 \cup E_2) = f(E_1) + f(E_2)$; è inoltre monotona se dati $E_1 \subset E_2$ ciò implica $f(E_1) \le f(E_2)$.

Esempi:
a) Esito del lancio di un dado: S={1,2,3,4,5,6}; l'algebra più grande W comprenderà 2^6=64 insiemi, singoletti ed universo degli eventi inclusi. Consideriamo la funzione di insieme che associa all'insieme il numero di valori. Abbiamo: f(∅)=0, f(S)=6. Se, ad esempio, E_1={6}, E_2={5}, E_3={4,5} allora $f(E_1 \cup E_2)=2$ che coincide con la somma $f(E_1)+f(E_2)=1+1=2$; anche $f(E_1 \cup E_3)=3=f(E_1)+f(E_3)$ e così via per ogni coppia, terna, etc. che rientrano in W. Quindi la funzione è additiva. E' anche monotona dato che $f(E_3) \ge f(E_2) \ge f(E_1)$. Non è additiva la funzione che associa ad ogni insieme il totale dei suoi valori: infatti $f(E_1 \cup E_2)=11 \ne f(E_1)+f(E_2)=6+11=17$. E' però monotona dato che in S non vi solo elementi negativi.

b) Altre funzioni di insieme non negative, monotone ed additive si ritrovano nelle applicazioni più comuni: la produzione di auto in un Paese determinata attraverso l'aggregazione dei prodotti degli stabilimenti installati. La superficie espropriata da un ente pubblico ottenuta per somma delle particelle catastali, il numero di reti segnate da una squadra di calcio determinato a partire da ciascun calciatore presente almeno una volta in una partita.

Esercizio_TP37: *un esperimento consiste nel considerare un periodo scelto a caso in un brano considerando come funzione di insieme il numero di parole. Ad esempio, nel primo capitolo dei Malavoglia si trova: A= "Diceva pure -Gli uomini sono fatti come le dita della mano: il dito grosso deve fare il dito grosso, e il dito piccolo deve fare da dito piccolo". A tale evento è associata f(A)=28. Verificate che la funzione goda della proprietà di non negatività, monotonicità e additività.*

6.2.2 Assiomi del calcolo delle probabilità

Agli inizi del 1900 si è consolidata l'analogia tra la misurazione di una grandezza fisica e la misurazione della casualità di un evento che è così assimilata alla determinazione di quante unità di misura sono in essa contenute. Ne è conseguita una teoria dell'incertezza in cui la casualità dell'evento E è espressa con un numero non negativo P(E) ad esso associato -con il meccanismo della funzione di insieme- detto probabilità dell'evento. Un'altra idea affermatisi nel corso del tempo è di imperniare la trattazione della casualità sulla legge di stabilizzazione delle frequenze relative nel senso che se, nell'esperimento ricorrono le condizioni per il postulate empirico del caso, allora P(E) deve avere caratteristiche analoghe alle frequenze relative. Alcune le ricordiamo:
1) La frequenza relativa è un numero dell'intervallo unitario; 2) Una modalità che non si verifica ha frequenza zero; 3) La somma delle frequenze relative è pari ad uno; 4) La frequenza relativa di due modalità distinte: (X_1 oppure X_2) è pari alla somma delle frequenze relative delle due modalità.

La funzione di insieme coinvolta nel calcolo delle probabilità presenterà le caratteristiche della non negatività, additività e monotonicità configurandosi -dal punto di vista matematico- come una funzione di misura o una misura. A.N. Kolmogorov (1933/1995, p. 11) sostiene: *la teoria della probabilità come disciplina matematica può e deve essere assiomatizzata esattamente nello steso senso della geometria e dell'algebra. Ciò significa che, dopo aver attribuito i nomi agli oggetti da studiare, le loro relazioni e gli assiomi che tali relazioni debbono soddisfare, tutti gli ulteriori sviluppi debbono poggiare su tali assiomi.* Gnedenko (1962, p. 20) aggiunge: *"la teoria della probabilità, al pari delle altre discipline matematiche, si è evoluta ignorando la necessità di pratiche applicazioni"*.

L'introduzione più limpida dell'approccio di Kolmogorov è quella contenuta nel primo paragrafo di un articolo che C.E. Bonferroni scrisse nel 1942.

> "... Come avviene in tutti i rami delle matematiche, anche nella statistica matematica è impossibile definire tutti i concetti in modo logico esplicito, cioè con una *definizione esplicita o nominale*. Tali definizioni, infatti, consistono nel ridurre un concetto ad altri precedentemente definiti, e quindi costituiscono una catena che necessariamente ha uno o più anelli di partenza: questi corrispondono ai *concetti primitivi*, che non si definiscono, ma dei quali si enunciano solo alcune proprietà, utilizzate nelle successive deduzioni. Si ha inoltre, una *definizione per postulati*, o *implicita* o *descrittiva*. Fissati i postulati, si può costruire attraverso a dimostrazioni e definizioni nominali, tutta la teoria, applicando le regole della logica generale e, ove sia possibile, i procedimenti della matematica, che di tale regole non sono che sviluppo ed affinamento. Ma come scegliere i postulati? Se si vuole che la teoria svolta non sia una semplice raccolta di concatenazioni e combinazioni logiche -com'è, in fondo, la teoria di un qualsiasi "giuoco"- ma abbia carattere di "scienza", occorre che i postulati siano aderenti, per dir così, al concetto cui si attribuiscono. Onde la necessità di chiarire la natura di questo concetto, non più con la pretesa di definirlo logicamente, ma con lo scopo di far comprendere di che cosa si parli quando di esso si parla: in altre parole, occorre quella che può chiamarsi, genericamente, *definizione fisica* del concetto. Essa è la sorgente, per così dire, alla quale debbono essere attinti i postulati."

Caratteristiche dei postulati

I postulati (cfr. ad esempio Piccolo e Vitale, 1984, p. 121) debbono risultare:

a) Coerenti e cioè non devono generare contraddizioni interne;
b) Utili e quindi devono essere subito operativi;
c) Non ridondanti ovvero non devono poter essere dedotti da altri postulati.

L'intuizione di Kolmogorov è di adoperare come postulati dei fatti inequivoci concernenti le frequenze relative perché grandezze vicine alle probabilità dando così una rappresentazione soddisfacente del mondo reale (cfr. Zenga, 1991, pp. 10-13); ma quali in particolare? I risultati menzionati non sono gli unici e ne esistono altri (ad esempio la frequenza con cui non si verifica una modalità) che potrebbero essere sfruttati ovvero dai quali ricavare quelli già citati ed altri. La scelta si è orientata sui postulati seguenti:

1. Le possibili manifestazioni di una prova S formano un'algebra W di eventi composti costituita da tutti i possibili sottoinsiemi di S.
2. La probabilità dell'evento E è una funzione di insieme -detta funzione di probabilità- che associa ad ogni evento in W un numero reale non negativo: $P(E)\geq 0$. La P(.) è definita esclusivamente per gli eventi composti e non per i punti elementari $e_1, e_2,\ldots,e_n$ inclusi in S. Gli e_i diventano visibili per la funzione di insieme P solo come $E_i=\{e_i\}$.
3. La probabilità dell'evento certo è pari ad uno: P(S)=1; cioè la funzione di probabilità P(.) è normalizzata.
4. La funzione di insieme è additiva. La probabilità dell'unione di "n" eventi mutualmente incompatibili E_i, i=1,2,…,n è pari alla somma delle probabilità dei singoli eventi:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) \quad se \quad E_i \cap E_j \quad per \quad i \neq j$$

La terna (S,W, P) che soddisfa i postulati 1-4 è detta spazio di probabilità perché in essa è contenuto tutto ciò che serve per trattare gli eventi e la casualità del loro verificarsi.

Esempi:
a) La normalizzazione è la caratteristica che distingue la misura della casualità dalla misura di distanze, aree, volumi che invece possono, almeno in teoria, tendere all'infinito.

b) Melsa e Sage (1973, p. 22) rilevano: la ragione per cui gli eventi considerati nel modello di Kolmogorov debbono formare un'algebra dovrebbe ora essere evidente. Se E_1 ed E_2 sono due eventi incompatibili ed $E_1 \cup E_2$ non fosse un evento dell'algebra, allora $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ non avrebbe significato dato che $E_1 \cup E_2$ non potrebbe essere probabilizzato. Peraltro, se P(E) è la probabilità di un evento $E \subset W$ allora, poiché $P(S)=1=P(E \cup E^c)=P(E)+P(E^c)$, anche 1-P(E) è la probabilità di un evento ed in particolare di E^c che quindi, per coerenza dovrebbe essere un evento dell'algebra.

c) Un esperimento tanto semplice quanto utile è la prova bernoulliana in cui gli eventi elementari alternativi sono solo due: successo (1), insuccesso (0). All'evento E={1} è assegnata probabilità "p" e quindi all'evento E^c={0} probabilità (1-p). Lo spazio di probabilità risulta così formato da: S={0,1}; W={∅,S,{0}, {1}}; P(1)=p, P(0)=1-p, P(1 e 0)=0, P(1 o 0)=1.
c) L'assegnazione in cui S={a,b,c,d} con P[{a}]=0.21, P[{b}]=0.58, P[{c}]=-0.14, P[{d}]=0.35 non è corretta per la presenza di una probabilità negativa P(c)=-0.14. E' anche sbagliata l'assegnazione: P[{a}]=0.08, P[{b}]=0.27, P[{c}]=0.36, P[{d}]=0.39 in quanto la somma è 1.1 che è superiore all'unità.

d) Abbiamo visto che, se card(S)=n allora è possibile formare 2^n eventi composti per i quali la funzione P(.) deve fornire la probabilità. Se n=26 e potessimo effettuare una assegnazione ogni miliardesimo di secondo sarebbe necessario almeno un anno per completare l'opera. La procedura seguita in pratica è di assegnare le probabilità ai singoletti e di procedere -per ogni evento nell'algebra W che interessi- secondo il quarto postulato. Sia S={a,b,c,d,e,f,g,h} con P(a)=0.1=P(b)=P(c)=P(d), P(e)=0.15=P(f)=P(g)=P(h) e si abbia inoltre M={{c},{d},{g},{h}); ne consegue che P(M)=p[{{c},{d},{g},{h}}]=p(c)+p(d)+p(g)+p(h)=0.1+0.1+0.15+0.15=0.5

***Esercizio_TP38**: ad alcune esponenti del mondo della finanza sono state chieste delle valutazioni probabilistiche rispetto all'andamento futuro del mercato mobiliare; In particolare, i singoletti su cui ragionare erano: E= "Forte guadagno", F= "Moderato guadagno", G= "Stabilità", H= "Moderata perdita", K= "forte perdita". Ecco le opinioni. Quali sono quelle coerenti con i postulati?*
1) P(E)=0.15, P(F)=0.15, P(G)=0.15, P(H)=0.15, P(I)=0.15; 2) P(E)=0.15, P(F)=0.20, P(G)=0.25, P(H)=0.30, P(I)=0.35; 3) P(E)=0.11, P(F)=0.29, P(G)=0010, P(H)=0.33, P(I)=0.17; 4) P(E)=-0.05, P(F)=-0.25, P(G)=1.00, P(H)=0.05, P(I)=0.25.

Teoremi sul calcolo delle probabilità

Per apprezzare la forza dei postulati esaminiamo alcuni corollari che torneranno poi utili in seguito.
1. La probabilità dell'evento impossibile è zero. Tenuto conto che: $S \cap \varnothing = \varnothing$ ciò implica che:

$$P(S \cup \varnothing) = P(S) + P(\varnothing) = 1 + P(\varnothing); \quad S \cup \varnothing = S \Rightarrow P(S \cup \varnothing) = P(S) = 1 \ \ e \ \ P(\varnothing) = 0$$

2. La probabilità dell'evento negato è il complemento ad uno della probabilità dell'evento negato.

$$E \cap E^c = \varnothing \Rightarrow P(E \cup E^c) = P(E) + P(E^c); \quad E \cup E^c = S \Rightarrow P(E \cup E^c) = P(S) = 1$$

Quindi: $1 = P(E) + P(E^c) \Rightarrow P(E^c) = 1 - P(E)$

Esempio:
Le due scommesse: "10,000 che esce il 27 sulla ruota di Napoli" al 10%" e "Non esce il 27 sulla ruota Napoli al 90%" dovrebbero risultare indifferenti. Si usa il condizionale in quanto la razionalità nelle scommesse è compromessa dall'avversione al rischio, del fascino del "9" o da un sentimento contrario all'azzardo ovvero da sfiducia e diffidenza sulla regolarità delle condizioni della scommessa o su chi la propone.

Le espressioni della probabilità sono diverse. Come decimale: 0.25, in percentuale: 25%, come frazione: 1/4, come casi contro e a favore: 3:1 (tre a uno) cioè su quattro *chances* una è a favore e tre contro.

3. La funzione P(.) dà valori compresi tra zero ed uno. Poiché $P(E \cup E^c)=P(E)+P(E^c)=1$, per avere $P(E)>1$ sarebbe necessario che $P(E^c)<0$, ma ciò contraddirebbe il 2° postulato che impone $P(E) \geq 0$ e quindi $0 \leq P(E) \leq 1$.

Esempio:
Lei chiede a Lui: mi ami? Lui risponde: al 101%. Cosa vuol dire? Che ricambia totalmente il suo amore e lo squillo del cellulare a cui risponde -scattando- è di sicuro la madre che vuol sapere come è andato l'esame. Comunque, è empiricamente dimostrato che è meglio non fidarsi dei paradossi.

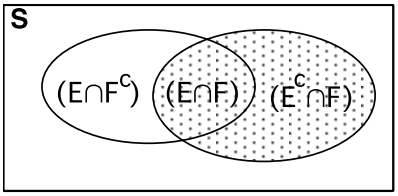
Esercizio_TP39: *dimostrare che se E ed F sono due eventi equivalenti allora il sistema dei postulati assegna ad entrambi la stessa probabilità;*

Esercizio_TP40: *è vero che se P(A)=P(B) allora A=B?*

4. La probabilità dell'unione di eventi compatibili è pari alla somma della probabilità degli eventi meno la probabilità della loro intersezione (probabilità totale):

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Spesso, ci ricorda Dall'Aglio (1987, p. 19), è necessario semplificare eventi complicati e saper riconoscere se due eventi sono uguali. Altre volte occorre riscrivere gli eventi in forma più complessa per evidenziare proprietà non immediate. Un esempio è la seguente riformulazione dell'evento unione in termini di eventi incompatibili:



$$(E \cup F) = \left(E \cap F^c\right) \cup (E \cap F) \cup \left(E^c \cap F\right)$$
$$\left(E \cap F^c\right) \cap (E \cap F) = E \cap E \cap F \cap F^c = E \cap \varnothing = \varnothing$$
$$\left(E \cap F^c\right) \cap \left(E^c \cap F\right) = E \cap E^c \cap F \cap F^c = \varnothing \cap \varnothing = \varnothing$$
$$(E \cap F) \cap \left(E^c \cap F\right) = E \cap E^c \cap F \cap F = \varnothing \cap F = \varnothing$$

Discende dai postulati che se due eventi sono uguali, la probabilità loro assegnata è uguale; quindi, possiamo ottenere la probabilità dell'unione considerando l'espressione alternativa:

$$\begin{aligned} P(E \cup F) &= P\left(E \cap F^c\right) + P(E \cap F) + P\left(E^c \cap F\right) \pm P(E \cap F) = P\left(E \cap F^c\right) + P(E \cap F) + P\left(E^c \cap F\right) + P(E \cap F) - P(E \cap F) \\ &= P\left[\left(E \cap F^c\right) \cup (E \cap F)\right] + P\left[\left(E^c \cap F\right) \cup (E \cap F)\right] - P(E \cap F) = P[E \cap S] + P[F \cap S] - P[E \cap F] \\ &= P(E) + P(F) - P[E \cap F] \end{aligned}$$

Esempi:

a) In un processo di produzione si sceglie a caso un *item*. Sia: E= "difettoso sul peso" e F= "difettoso nella forma". Per varie ragioni che qui non interessa chiarire la funzione di probabilità assegna: P(E)=0.38, P(F)=0.33, P(E∩F)=0.26. Ne consegue che, la stessa funzione, per coerenza con i postulati, deve assegnare: P(E U F)= 0.38 + 0.33 - 0.26 =0.45. Ciò conferma quanto già l'intuito aveva suggerito: nel valutare la probabilità dell'unione conteggiamo gli eventi elementari in E e poi quelli in F, ma così facendo quelli contenuti nell'intersezione di E con F sarebbero contati due volte ed ecco quindi la necessità di sottrarre una volta il conteggio degli elementi comuni.

b) Ciccillo ha mezzora per navigare in Internet. Con probabilità del 44% si collegherà ad un sito di viaggi e con probabilità del 53% si collegherà sia ad un sito di viaggi che ad uno di cinema; invece, la probabilità che non si colleghi ad un sito di cinema è del 65%. Si può concludere che si collegherà ad un sito di viaggi o ad un sito di cinema con probabilità del 25%. Vero o falso?

$$P(V) = 0.44,\ \ P(V \cap W) = 0.53,\ \ P\left(W^c\right) = 0.65 \Rightarrow P(V \cup W) = 0.44 + 0.35 - 0.53 = 0.26$$

c) Consideriamo due eventi E ed F con P(E)=0.40, P(F)=0.30, P(E∩F)=0.10. Calcolare la probabilità che si verifichi E o F, ma non entrambi. L'evento che interessa è A=E∪F-E∩F con probabilità: P(A)=P(E∪F)-2P(E∩F)=0.40+0.30-0.20=0.50.

Esercizio_TP41: *sia P(E)=0.3, P(F)=0.2, P(G)=0.6, P(E∪F)=0.5, P(E∪G)=0.8, P(F∪G)=0.7. Quale di queste coppie è formata da eventi incompatibili: (E,F); (E,G); (F,G)?*

Esercizio_TP42: *ipotizzando che E, F, G siano mutualmente incompatibili e che P(E)=0.25, P(F)=0.65, P(G)=0.15 determinare: a) P(E^c); b) P(F∪G); c) Cosa si può dire su P(E∪F∪G)?*

Esercizio_TP43: *dati due eventi F e G per i quali P(F)=0.54, P(G)=0.29 e P(F∩G)=0.17 determinare: a) P(F∩G^c); b) P(F^c∩G^c), c) P(F^c∪G^c)*

5. Monotonicità della funzione di probabilità. Se F⊃E allora la probabilità dell'evento contenitore è non minore della probabilità dell'evento contenuto: P(F)≥P(E). Sfruttando le regole dell'insiemistica possiamo scrivere:

$$F = E \cup \left(E^c \cap F\right) \;\; con \;\; E \cap \left(E^c \cap F\right) = \varnothing$$

cioè l'evento è espresso come unione di altri eventi incompatibili. La sua probabilità è: $P(F)=P[E»(E^c \cap F)]=P(E)+P(E^c \cap F)$ e poiché le probabilità sono non negative, si ha P(F)≥P(E).

Esempio:
Se E ⊂ F allora P(F - E)=P(F) - P(E). Ricordiamo che F - E = F ∩ E^c e che P(F ∪ E^c) = P(F) + P(E^c) - P(F∩E^c). Quindi:

$$P(F - E)=P(F \cup E^c) - P(E^c) - P(F)=P(S) - P(E^c) - P(F)=1 - P(E^c) - P(F)=P(E) - P(F)$$

All'evento che si realizza se accade uno degli eventi in F, ma non in E deve essere assegnata la probabilità di F defalcata dalla probabilità di E cioè quelle parti di F che non possono più verificarsi.

Esercizio_TP44: *siano E, F∈W. Dimostrare che la probabilità della differenza E - F è pari alla probabilità di E meno la probabilità dell'intersezione di E con F: P(E-F)=P(E)-P(E∩F).*

6. Disuguaglianza di Boole:

$$P\left(E \cap F\right) \le Min\left\{P(E), P(F)\right\} \le Max\left\{P(E), P(F)\right\} \le P\left(E \cup F\right) \le P(E) + P(F)$$

La probabilità dell'unione è sempre maggiore o uguale della probabilità massima tra quelle a confronto e che la probabilità dell'intersezione è sempre minore o uguale della probabilità minima.

Esempio:
Apprendimento cumulativo. Disse una volta Bearzot, allenatore della nazionale di calcio italiana che vinse i mondiali del 1982. *"Se addestri un cane ad attraversare la strada e gli fai ripetere l'esperienza per 50 volte ti puoi aspettare che alla fine impari qualcosa. Ma se per 50 volte cambi il cane ti troverai sempre al punto di partenza"*. Le probabilità di un evento composto aumentano man mano che si aggiungono eventi elementari che non siano eventi impossibili.

Esercizio_TP45: *data la seguente configurazione della probabilità di alcuni eventi: P(A)=0.52, P(B)=0.48, P(C)=0.53, C⊂A, P(A∩B)=0.64 verificatene la coerenza con i postulati.*

7. La probabilità totale può essere estesa a più di due eventi. Partiamo da E, F, G. Sia A=(F ∪ G) ed applichiamo la regola agli eventi E ed A: P(E∪A)=P(E)+P(A)-P(E∩A). Sostituendo ad A la sua nuova formulazione si ha:

$$\begin{aligned} P\left(E \cup F \cup G\right) &= P(E) + P\left(F \cup G\right) - P\left[E \cap \left(F \cup G\right)\right] = P(E) + P(F) + P(G) - P\left(F \cap G\right) - P\left[\left(E \cap F\right) \cup \left(E \cap G\right)\right] \\ &= P(E) + P(F) + P(G) - P\left(F \cap G\right) - P\left(E \cap F\right) - P\left(E \cap G\right) + P\left(E \cap F \cap G\right) \end{aligned}$$

Sommando la probabilità dei tre eventi si sommano due volte le intersezioni delle coppie di eventi ed ognuna di queste deve essere sottratta. Così facendo però si toglie troppo perché le parti comuni a tutti e tre gli eventi vengono sottratte una volta in più del necessario e l'equilibrio si ripristina sommando la probabilità congiunta dei tre eventi.

Esempi:
a) La Teseia s.r.l. ha formulato le probabilità per i punti di aumento in percentuale del prodotto interno lordo:

Aumento (S)	*0.00*	*0.25*	*0.50*	*0.75*	*1.00*	*1.25*	*1.50*	*1.75*	*2.00*
Probabilità	0.05	0.05	0.10	0.10	0.10	0.15	0.25	0.15	0.05

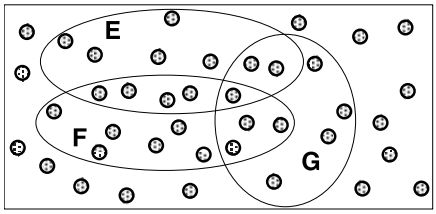
La *leader* di un movimento politico, prima di avviare una forte opposizione, valuta i seguenti eventi: E={xεS|0.25< x ≤0.75}, F{xεS|0.50≤ x ≤1.25}, G={xεS|0.75≤ x≤ 1.75}. Per calcolare P(E U F U G) bisogna ricostruire le probabilità degli eventi coinvolti:

$$P(E) = 0.20, \;\; P(F) = 0.45, \;\; P(G) = 0.75, P\left(E \cap F\right) = 0.20, \;\; P\left(E \cap G\right) = 0.10, \;\; P\left(F \cap G\right) = 0.35$$

$$P\left(E \cap F \cap G\right) = 0.10; \qquad P\left(E \cup F \cup G\right) = 0.20 + 0.45 + 0.75 - 0.20 - 0.10 - 0.35 + 0.15 = 0.85$$

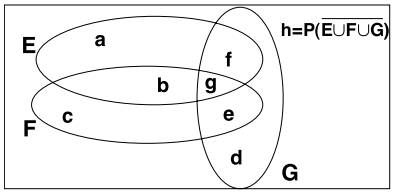
b) Ad ogni punto del diagramma sia associata la probabilità p=1/40. Proviamo a calcolare P(E∪F∪G).

$$P(E\cup F\cup G)=\frac{12}{40}+\frac{14}{40}+\frac{13}{40}-\frac{5}{40}-\frac{3}{40}-\frac{4}{40}+\frac{1}{40}$$
$$=\frac{(12+14+13+1)-(5+3+4)}{40}=\frac{40-12}{40}=\frac{28}{40}$$



Esercizio_TP46: *la responsabile degli acquisti rinvia -con il rischio di essere licenziata in tronco se sbaglia- ha sospeso un importante ordine nella speranza che uno dei tre fornitori F_1, F_2 o F_3 diminuisca i prezzi. La probabilità di riduzione sono: $P(F_1)=0.93$, $P(F_2)=0.88$, $P(F_3)=0.91$. Gli accordi di cartello sospettati sono: $P(F_1\cap F_2)=0.85$, $P(F_1\cap F_3)=0.90$, $P(F_2\cap F_3)=0.87$, $P(F_1\cap F_2\cap F_3)=0.79$.*
1) Qual'è la probabilità che almeno uno dei fornitori abbassi il prezzo di vendita?
2) Qual'è la probabilità che più di uno abbassi il prezzo di vendita?

Esercizio_TP47: *l'universo degli eventi è stato diviso in otto regioni ed ai corrispondenti singoletti è stata assegnata la probabilità indicata con la lettera interna alla regione. Supponete che f=g=e=k, a=c=d=mk e che h=0.5 Per quali valori di "m" e "k" si ottiene un sistema di probabilità coerente con i postulati?*



8. Una disuguaglianza fondamentale è quella di Bonferroni. Se $\{E_i\}$, i=1,2,…, n è una classe di eventi, allora:

$$1-\sum_{i=1}^{n}P\left(E_i^c\right)\le P\left(\bigcap_{i=1}^{n}E_i\right)\qquad ovvero\qquad 1-\sum_{i=1}^{n}P\left(E_i\right)\le P\left(\bigcap_{i=1}^{n}E_i^c\right)$$

La probabilità dell'intersezione deve essere non minore del complemento ad uno della somma delle probabilità degli eventi negati. Il teorema può essere provato per induzione. Per n=2 la relazione è certamente valida:

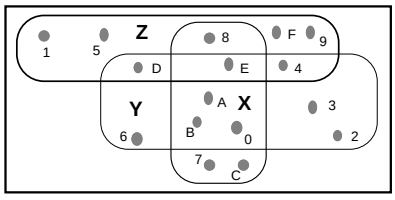
$$P\left(E_1\cap E_2\right)\ge 1-\left[P\left(E_1^c\right)+P\left(E_2^c\right)\right]=1-\left[P\left(E_1^c\cup E_2^c\right)+P\left(E_1^c\cap E_2^c\right)\right]=1-P\left(E_1^c\cup E_2^c\right)-P\left(E_1^c\cap E_2^c\right)$$
$$P\left(E_1\cap E_2\right)\ge 1-P\left(E_1\cap E_2\right)^c-P\left(E_1\cup E_2\right)^c\ge P\left(E_1\cap E_2\right)-P\left(E_1\cup E_2\right)^c$$

e poiché il lato sinistro è uguale al lato destro meno una quantità non negativa, la disuguaglianza è soddisfatta. Lo stesso ragionamento può essere esteso ad un "n" qualsiasi (cfr. Rohatgi, 1976, p. 27).

Esempio:
Verifichiamo la disuguaglianza di Bonferroni per i dati del punto a) dell'esercizio precedente:

$$P\left(F_1\cap F_2\cap F_3\right)\ge 1-\left[P\left(F_1^c\right)+P\left(F_2^c\right)+P\left(F_3^c\right)\right];\qquad 0.79\ \ge\ 1-[0.07+0.12+0.09]=0.72$$

Esercizio_TP48: *la "Sperticato Carmela S.p.A." ha allo studio 4 nuovi prodotti combinabili con due tipi di marketing e due zone di collocazione. Le strategie sono nel diagramma di Venn insieme alle caratteristiche prevalenti: X=minori costi di produzione, Y= minori costi di promozione, Z= migliori caratteristiche. Costruire i seguenti eventi ed assegnare loro una probabilità proporzionale al numero di eventi elementari in essi ricadenti.*



a) $X\cap Y$; *b)* $Z\cup X$; *c)* $(X\cup Y)^c$; *d)* $(X\cap Y\cap Z)$; *e)* $(X-Y)$; *f)* $(Z-Y)$

Esercizio_TP49: *la nuova politica di acquisto prevede che il 75% delle forniture venga affidato a ditte locali ed il 50% a ditte guidate da una donna (proprietaria o amministratore delegato); all'interno di questa categoria però le ditte locali dovranno essere pari al 40%.*
a) Qual'è la probabilità che una fornitura non venga affidata ad una ditta locale non guidata da una donna?
b) Qual'è la probabilità che la fornitura sia affidata ad una ditta non locale guidata da una donna?

6.2.3 Che cos'è la probabilità?

Per comprendere la flessibilità introdotta dalla assiomatizzazione di Kolmogorov consideriamo un universo degli eventi che contenga un numero finito di esiti: $S=\{e_1, e_2, \ldots, e_n\}$. L'algebra W costruita su S con le usuali operazioni insiemistiche contiene 2^n eventi composti per cui sarebbe necessario abbinare ognuno di questi eventi con una probabilità; tali probabilità, inoltre, dovrebbero essere coerenti con i postulati e con i corollari da essi derivati. Questo però, come si è visto, non è necessario perché la funzione di probabilità può essere specificata per i singoletti ottenendo per mera via di calcolo le probabilità di tutti gli altri eventi in W.

Esempi:

a) Wilks (1962, p. 11) osserva: "... *Il punto cruciale della formulazione di Kolmogorov è che la teoria matematica interviene dopo l'assegnazione delle probabilità. Potremmo ovviamente dubitare circa la corretta scelta delle probabilità, ma questo è un problema di verifica di ipotesi che è discusso altrove".*

b) Shirayev (1996, p.14) afferma: *"La questione cruciale non è il come assegnare la probabilità degli eventi elementari, ma di come calcolare la probabilità di eventi complessi a partire da quelle attribuite agli eventi elementari."*

c) DeGroot (986, p.6): *"L'impegno maggiore nella trattazione matematica della probabilità, sia nei testi introduttivi che in quelli avanzati, si concentra su due questioni fondamentali:*
1) Come determinare la probabilità di un evento qualsiasi a partire dalle probabilità già assegnate ai risultati elementari di un prova;
2) Come aggiornare tali probabilità allorché si rendono disponibili nuove informazioni rilevanti sulla prova".

Il sistema dei postulati di Kolmogorov è incompleto: non perché la scelta dei postulati sia poco felice o inadeguatamente sviluppata, ma perché lascia indeterminata la definizione di probabilità scegliendo una posizione apparentemente *super partes*. Le regole proposte, infatti, evitano di pronunciarsi su che cosa sia la probabilità. Sappiamo che misura qualcosa presente negli esperimenti casuali, ma si sono stabilite solo le regole per combinare dei valori non per scegliere quei valori. E' possibile formulare un meccanismo che misuri il grado di incertezza di un evento così come i chilometri esprimono le distanze sulla superficie terrestre, i litri la capacità per i contenitori di liquidi? Occorre chiarire che, sebbene il modello di Kolmogorov abbia avuto come riferimento le frequenze relative (la cosiddetta concezione frequentista della probabilità, discussa nel paragrafo iniziale), questo non implica la piena adeguatezza di tale concezione a dare una spiegazione esauriente della probabilità e nel corso del tempo si sono consolidate altre due linee interpretative.

Esempi:

a) L'agenzia spaziale europea ha progettato un vettore per portare una navicella su Titano. L'esperimento può essere schematizzato con un dominio semplicissimo: S={successo, insuccesso} ed assegnando P(successo)=p, P(insuccesso)=1-p. Il valore da dare a "p" non è ottenibile dai postulati di Kolmogorov. Ci si aspetta che "p" sia elevato per non mettere troppo a rischio l'equipaggio, ma quanto? 0.75, 0.90, 0.99? La sua determinazione è esterna alla assiomatizzazione. Saranno le conoscenze tecniche, le esperienze passate di alcune o tutte le persone coinvolte nel progetto, a proporre il valore (ammesso che ve ne sia uno solo) più plausibile per "p".

b) La commissione esaminatrice di un concorso ha predisposto 12 argomenti distinti ed autonomi su cui interrogare i concorrenti. E' noto che chi risponde bene al primo quesito vince il concorso, anche rispondendo poco alle altre domande. Cinzia, con il poco tempo che ha, prepara un solo argomento e per il resto mette in preventivo degli insinceri, ma efficaci: "l'ho studiato, ma ora non ricordo", "non mi sento bene", "mi sento confusa, ma sono preparata". Qual'è la probabilità che Cinzia vinca il concorso?

c) L'ISTAT (1998, p. 179) pubblica la ripartizione per comparto geografico e per zona altimetrica dell'unità "ettaro di suolo italiano":

	Montagna	Collina	Pianura	Totale
Nord	5'531'815	2'272'878	4'187'322	11'992'015
Centro	1'576'048	3'723'862	535'516	5'835'426
Sud	3'503'136	6'548'277	2'555'249	12'606'662
Totale	10'610'999	12'545'017	7'278'087	30'434'103

Se non conosco la collocazione e la fascia altimetrica di un ettaro, ma debbo pronunciarmi su queste due caratteristiche, in mancanza di altre informazioni dovrò assegnarlo a "Collina" e "Mezzogiorno" in quanto interpretando gli ettari censiti come equiprobabili, il rapporto di casi favorevoli su casi possibili comporterebbe una probabilità del 21.5% che è la più alta fra le nove combinazioni.

Solo il terzo esempio è inquadrabile nell'approccio frequentista. Per i primi due non esiste una casistica alla quale rifarsi per determinare delle frequenze relative da trasformare poi in probabilità

Concezione classica o matematica

Una situazione in cui si può ricavare la probabilità -all'interno del solo sistema dei postulati- è quella di esperimenti con simmetrie cioè uscite casuali in cui gli eventi elementari sono considerati equiprobabili: o per ragioni fisiche legate alla uniformità delle loro *chances* di uscita, o per ignoranza, o per le semplificazioni che ciò comporta o solo per la mancanza di indicazioni in senso contrario.

Esempi:
a)Il principio di ragione non sufficiente (o di indifferenza) nel calcolo delle probabilità afferma che se non esiste alcuna ragione conosciuta per sostenere un modello particolare, occorre utilizzare il modello di probabilità uniforme. Hodges e Lehmann (1971, p. 39) ritengono però che spetti a chi costruisce il modello avanzare le ragioni per l'uso di un modello piuttosto che di un altro.

b) Se fate scegliere a caso le facce di una moneta ad un gruppo di persone vi accorgerete che la maggior parte opterà per testa. La prevalenza si può speigare con il fatto che i disegni sulle due facce sono in rilievo e che quello relativo alla "testa" sia sempre stato più elaborato cioè più pesante rispetto alla croce spostando il baricentro della moneta verso questa uscita.

Il porre gli eventi di S sullo stesso piano implica una specifica funzione di probabilità. Se "p" con $0 \leq p \leq 1$ è la comune probabilità da assegnare ai singoletti $E_1=\{e_1\}$, $E_2=\{e_2\},\ldots,E_n=\{e_n\}$ si ha:

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} p = np = 1 \Rightarrow p = \frac{1}{n}$$

Se i singoletti sono equiprobabili, la probabilità da assegnare a ciascun E_i è pari al reciproco del numero degli eventi in S: $P(E_i)=1/n$ per $i=1,2,\ldots,n$ che è nota come funzione di probabilità uniforme perché ripartisce in modo paritario la dote di probabilità (uno) fra gli "n" singoletti. In questo caso la probabilità è la conseguenza automatica delle simmetrie presenti (ma più spesso solo ipotizzate) nella prova.

Esempi:
a) Le lotterie sono un tipico esperimento casuale in cui ogni biglietto ha la stessa probabilità di essere estratto rispetto a tutti gli altri (a meno di disfunzioni o di imbrogli). Due persone che comprano lo stesso numero di biglietti hanno la stessa probabilità di vincere il premio purché la scelta sia interamente demandata alla sorte. Il fatto che voi non avete mai vinto niente e che quel vostro amico o parente vince una settimana sì e l'altra pure è una questione non spiegabile con le probabilità.

b) Se i possibili numeri da estrarre da un'urna sono 45 non è affatto detto che dopo 45 estrazioni i numeri comincino a ripetersi ovvero che dopo la 45ª estrazione la probabilità di un numero non estratto sia maggiore di quella di un numero già estratto. Pensare questo significa ritenere più probabili i numeri meno frequenti contraddicendo il significato frequentistico e, soprattutto, esponendosi a cattivi pensieri sulla regolarità delle estrazioni.

Esercizio_TP50: *è in corso una pesca per beneficenza. La bambina bendata inserisce la mano in un'urna non trasparente per estrarre una biglia. Due spettatori si pronunciano. Il signor A afferma: "la biglia è di colore rosso oppure non lo è per cui la probabilità che sia rossa è del 50%. Il signor B afferma: "la biglia può essere rossa, verde, gialla, blù, arancione, viola, nera" per cui la probabilità che sia rossa è del 14% (1/7). Chi ha ragione?*

L'uniformità rende semplice determinare la probabilità da assegnare ad un evento composto diverso dai singoletti. Ipotizziamo che l'evento E contenga "h" singoletti:

$$E = \bigcup_{i=1}^{h} E_i, \quad E_i = \{e_i \in S\}$$

su un totale di "k" eventi equiprobabili. Allora la probabilità di E sarà:

$$P(E) = \sum_{i=1}^{k} P(E_i) = \overbrace{\frac{1}{k} + \frac{1}{k} + \ldots + \frac{1}{k}}^{h \text{ volte}} = h\left(\frac{1}{k}\right) = \frac{h}{k} = \frac{card(E)}{card(S)}$$

che è pari al rapporto tra il numero di risultati contenuti in E (casi favorevoli: E si verifica solo se si verifica uno di essi) e quelli in S (casi possibili). Un evento è "probabile" se i casi a favore sono più numerosi dei casi contro:

$$\frac{h}{k} > \frac{k-h}{k} \Rightarrow \frac{h}{(k-h)} > 1$$

maggiore è (h/k) più grande è la probabilità dell'evento finché questa non arrivi alla certezza (probabilità uno) che però non può essere raggiunta a meno che l'opposto non sia un evento impossibile. E' questo il significato che Laplace dava alla probabilità che però è meno generale di quanto egli non intendesse trattandosi solo di una conseguenza del postulato di simmetria nell'assegnazione della probabilità e non una accettabile definizione di probabilità, nemmeno nella forma di concetto primitivo.

Esempi:
a) L'approccio classico viene seguito nelle scommesse esprimendo le probabilità come rapporto *(odds)* tra interi positivi ridotti ai minimi termini e pronunciando prima il numero più grande (dopo aver opportunamente chiarito se si parla di *odds* contro o a favore). Ad esempio, scegliendo a caso una carta da un mazzo francese, la scommessa sull'uscita dell'asso sarebbe espressa come 12:1 (12 contro uno) visto che la probabilità di estrarlo è 4/52=1/13. La scommessa è dunque: per ogni unità di conto che punti sull'uscita dell'asso devi ottenerne 12 in caso di uscita effettiva ovvero contentarti di un dodicesimo della posta nel caso in cui sei tu il banco e non esce l'asso. Allo stesso modo, l'uscita del 7 nel lancio di due dadi è espressa con il rapporto 5:1 dato che i casi a favore sono 6 i casi possibili sono 3 e quindi P(7)=6/36=1/6 e 6=5+1.

b) Una SpA ha emesso un milione di obbligazioni numerate sequenzialmente da "000000" a "999999". Il programma di rientro prevede che ogni anno si rimborsino tutte le obbligazioni che contengono - le ultime due cifre dell'anno in posizioni adiacenti a partire dalla prima. Qual'è la probabilità che una data obbligazione sia rimborsata nell'anno in corso? I casi favorevoli sono 3x104=30'000 (fissata la coppia di posizioni le altre possono combinarsi liberamente e le coppie adiacenti sono tre). La probabilità è del 3%.

Esercizio_TP51: *una confezione di lattine per una bibita analcolica contiene 84 pezzi di cui 4 sono difettate rispetto alla linguetta di apertura. Si scelgono a caso (si applica quindi il modello di probabilità uniforme) due lattine distinte. Qual'è la probabilità che entrambe presentino difetti?*

Esercizio_TP52: *il controllo antidoping colpisce un giocatore scelto a caso nell'insieme delle due liste che le squadre consegnano all'arbitro prima dell'inizio della partita e che include 11 giocatori, 6 in panchina ed il portiere di riserva.*
a) Qual'è la probabilità che sia scelto un giocatore partente dalla panchina?
b) Qual'è la probabilità che il controllo colpisca uno dei giocatori della formazione iniziale?
c) Esprimete come odds contro la possibilità che tocchi ad un portiere.

Il successo della funzione di probabilità uniforme è dovuto al suo carattere "spontaneo" perché nel valutare un evento eseguiamo a mente il rapporto tra circostanze a favore e contro. Non è quindi strano ritrovarlo nella probabilità. Questo però non è del tutto accettabile perché:
1) Include una tautologia: "ugualmente possibili" è già una definizione di probabilità e quindi dovrebbe essere inclusa nel sistema dei postulati.
2) Non può essere richiamato se non si hanno conoscenze sulla struttura fisica della prova e/o si ignora come questa ne influenzi le manifestazioni.
3) Non è applicabile se l'universo degli eventi è di tipo continuo o enumerabile.

Esempio:
Landenna e Marasini (1986, pp.26-27) propongono il seguente studio: da due mazzi di carte francesi si sceglie una carta per ogni mazzo. Una di esse è di colore nero. Qual'è la probabilità che l'altra sia pure di colore nero? Poisson ragionò così: i casi possibili sono: (N1,N2), (N1,R2), (R1,N2) e (R1,R2). Se la prima è nera, restano solo 3 casi di cui uno a favore. Perciò la probabilità è 1/3. Von Kries invece partì dal fatto che la scelta della 1ª carta non ha alcuna influenza sulla costituzione dell'universo degli eventi nella scelta della 2ª per cui la probabilità è 26/52=1/2. La scelta tra le due formulazioni -ugualmente valide- dell'universo degli eventi è arbitraria come concludono sia E. Poincaré che J.M. Keynes.

Tali obiezioni riducono la portata della probabilità uniforme che rimane confinata ai casi in cui l'esperimento sia descrivibile con un numeri finito di eventi simmetrici dal punto di vista dell'occorrenza; essa è, infatti, l'ideale riferimento per tutti i giochi d'azzardo e di alcuni esperimenti della fisica delle particelle. In generale, occorrerebbero sistemi diversi di proporre le probabilità, modelli più ricchi e di validità provata; solo che è difficile trovarne di altrettanto semplici e potenti come quello di probabilità uniforme che continua a vivere ed influenzare teoria ed applicazioni anche oltre la sua reale validità.

Esercizio_TP53: *la congiunzione venerdì 17 gode di pessima fama (almeno in Italia). Qual'è la probabilità che un 17 qualsiasi sia venerdì? Una risposta immediata potrebbe essere 1/7 dato che sette sono i casi possibili ed uno solo quello favorevole. Si ragioni con un calendario gregoriano ipotizzando di partire dal 1.1.1601 lunedì. E' meritata la sinistra tradizione di questo giorno?*

Probabilità soggettiva

Sono emersi due modi di proporre la probabilità: quello classico derivato dalle conoscenze teoriche sull'esperimento (ad esempio le simmetrie) e l'altro basato sul postulato empirico del caso (approccio frequentista). Resta da spiegare come esprimere le probabilità in situazioni in cui c'è casualità, ma le condizioni fisiche dell'esperimento non sono note o siano conoscibili solo in parte e- contemporaneamente- non possano essere replicate o replicate un numero adeguato oppure, se replicate, non mostrino tendenza alla stabilizzazione delle frequenze relative.

Esempi:
a) Chi compila la schedina del totocalcio e deve decidere quale segno proporre per il primo incontro nella storia del calcio Juventus-Schiavonea ha poche possibilità di sfruttare l'esperienza dato che non ci sono casi assimilabili. Peraltro, le simmetrie dell'esperimento portano a favorire la squadra che gioca in casa e/o che gode di una migliore posizione in classifica, ma difficilmente può andare oltre l'affermazione: P("1")≥P("x")≥P("2").

b) L'impianto di una attività produttiva permanente in una zona sismica richiede la valutazione del rischio terremoto (massima magnitudo) nell'area che circonda la sede dei futuri stabilimenti. Le esperienze passate sono limitate e non ci sono simmetrie geologiche che possano aiutare. La forza probatoria degli argomenti diventa quindi determinante.

c) La compagnia dei Lloyd's è pronta ad assicurare chiunque contro qualsiasi rischio, in qualsiasi luogo ed in qualsiasi epoca, ma per fissare il premio deve stabilire il rischio probabile. Qual'è l'alea di una soprano che perde la voce o di un tennista che perde irreversibilmente la capacità di giocare?

Siccome c'è casualità anche in queste situazioni si rende necessario estendere il significato di probabilità integrandolo con elementi qualitativi. In genere, la probabilità soggettiva è illustrata con le scommesse: se una persona è disposta a partecipare ad una scommessa che attribuisce all'evento scelto 2/3 vuol dire che è disposta a pagare 2 euri in cambio di 3 in caso di vincita. E' pero facile tradurla nel solito intervallo unitario.

Una prima debolezza può subito essere colta nel quadro delle scommesse in cui è collocabile. Senza troppo cercare troverete persone che, sull'esito di Bologna-Juventus trovano sensate frasi del tipo: al 50% vince la Juve, al 50% vince il Bologna ed al 75% pareggiano. Non avrete neanche difficoltà a provare scommettitori che considerano accettabile una strategia di gioco che li vede puntare contemporaneamente 2:1 per un evento del tipo "Can Can vince la tris di Agnano" e 3:2 su "Mambo non vince la tris di Agnano". Sulla coerenza il discorso è molto complesso.

Esempio:
Un giocatore ha osservato abbastanza a lungo le uscite di un tavolo di roulette e si è formato il seguente quadro di opinioni: il rosso esce con probabilità del 36%, il nero con probabilità del 45% e lo zero (che annulla sia il rosso che il nero) esce con probabilità del 9%. Se gli si fa notare che la probabilità dell'evento certo non è il 100% magari è disposto a revisionare le sue aspettative spalmando il 10% che manca sui tre eventi del suo universo. Potrebbe anche asserire che il 10% che manca è riferito ad una possibile catastrofe: il lampadario che crolla sul tavolo rendendolo inutilizzabile, un *black-out* elettrico prolungato, il *crupier* che perde il controllo e regala *fiches* a tutti. In questo caso gli eventi sono riferiti ad un diverso esperimento con un universo di quattro elementi. E' anche possibile riscalare proporzionalmente le probabilità:

$$P(R)=\frac{0.36}{0.90}=0.4;\quad P(N)=\frac{0.45}{0.90}=0.5;\quad P(0)=\frac{0.09}{0.90}=0.1$$

Non c'è tuttavia alcuna garanzia che questo rispecchi il pensiero del giocatore che preferisce un mondo imperfetto in cui l'evento certo ha probabilità variabile da esperimento ad esperimento.

Esercizio_TP54: *supponiamo che Luciana dia all'evento E probabilità zero e che le venga proposta la scommessa cui è tentata di partecipare per non sembrare scortese. Se E non esce non vince niente; se E esce deve pagare diecimila euri. Secondo voi, cosa deve fare Luciana, per essere coerente ?*
1. Accettare; 2. Non accettare; 3. E' in una posizione di indifferenza.

La probabilità è dunque anche un'espressione numerica del grado di convinzione o fiducia personale, auspicabilmente fondata, sulla verità di una certa asserzione in base ad un *corpus* di conoscenze, razionali ed anche istintive. Nella formulazione logico-soggettivista la probabilità è il giudizio che un osservatore (persona, organizzazione, sistema esperto) "j" esprime sulle possibilità di verificarsi dell'evento elementare "e_i" ricadente in un certo universo degli eventi S, dato il quadro di evidenze, intuizioni ed emozioni F_k. Questa definizione ci porta a navigare in acque molto profonde: fascinose per le opportunità di nuove scoperte, ma con seri rischi di naufragio.

$$P_j\left(E_i|F_k\right),\ \ j=1,2,\ldots,N;\ \ k=1,2,\ldots,h,\ \ E_i \in W$$

Se cambia il quadro di evidenza F_k cambia la probabilità dell'evento. Questo non crea troppi problemi se il quadro di riferimento è definito al punto da poterne seguire i cambiamenti.

Esempi:

a) Un giudice per le indagini preliminari che deve decidere se confermare la custodia cautelare o liberare un accusato si baserà sulle evidenze raccolte dall'autorità giudiziaria e sulla personalità del soggetto. La casistica di riferimento è vaga e se l'accusato non ha precedenti la decisione si baserà sul confronto delle prove contro e a favore così come sono percepite dal giudice. Se però il giudice cambia, cambierà anche la percezione dei fatti. Così si spiega la presenza di più gradi di giudizio.

b) Supponiamo che le conoscenze su di un esperimento porti ad assegnare le probabilità secondo la funzione:

$$P\left(E_i\right)=\left(\frac{1-q}{1-q^n}\right)q^{i-1};\quad i=1,2,\ldots,n$$

Se le modifiche nel quadro informativo si concretano in una variazione nei parametri "n" o "q" il modello può essere conservato fino a che non ci siano esigenze di cambiamento. Se per qualche ragione le opinioni rimangono inespresse o espresse nella direzione contraria alle evidenze consolidate, le probabilità diventano uno strumento poco utile.

Supponiamo quindi che il quadro informativo esterno all'esperimento sia unico. Questo però risolve solo una parte dell'indeterminatezza dello spazio di probabilità di Kolmogorov. Il cambiamento potrebbe anche essere ascritto ad un mutamento dello stato d'animo dell'osservatore "j" ovvero ad una modifica sostanziale nel suo modo di vedere gli stessi fatti oppure alla scelta di un altro osservatore. Peraltro, le informazioni F possono essere talmente vaghe e generiche che è difficile farle confluire in un unico valore ed il soggetto riesce solo a proporre un limite inferiore ed un limite superiore alla probabilità. Niente impedisce un passo di generalità superiore:

$$\underline{P}_j\left(E_i|F\right) \leq\ \ P_j\left(E_i|F\right) \leq \overline{P}_j\left(E_i|F\right)$$

può essere data una probabilità che a sua volta è inclusa in un ulteriore intervallo di probabilità (cfr. Medolaghi, 1920) aumentando la difficoltà di mantenere la coerenza tra le varie asserzioni. Anche senza queste sofisticazioni (che però diversi autori considerano tuttaltro che gratuite o prive di senso) il concetto rimane inestricabile.

Scala di misurazone della probabilità

Una questione da approfondire (cfr. Monari, 1992) è su quale scala misurare la probabilità . Secondo Feller (1950, p.19) la scala deve essere quella proporzionale ritenendo la valutazione della casualità di un evento riconducibile alla valutazione di una distanza di cui sia impossibile la misurazione diretta (l'imprecisione con pregiudica la proporzionalità della scala come abbiamo imparato nel paragrafo dedicato alle tecniche di misurazione). Landenna e Marasini (1986, pp. 98-101) illustrano la concezione comparativa della probabilità che misura l'incertezza su di una scala ordinale. Keynes (1994, p.35) dubita perfino della scala ordinale perché ci sono eventi che hanno probabilità diversa, ma sui quali non si può essere conclusivi su quale sia il più probabile. Gnedenko (1962, p.25) è del parere che se la probabilità fosse solo giudizio personale di credibilità a larga componente emotiva bisognerebbe rivolgersi alla psicologia e non alla matematica per trattarla. Sono tanti gli elementi di cui bisogna tenere conto, forse troppi perché gli studenti di un corso di base possano formarsi delle opinioni sulla complessità di questo approccio. Una bussola è l'opera monumentale di P. Walley (1991). Anche la lettura del 1° capitolo di Scozzafava (1996) è illuminante.

La probabilità è un fatto ed è anche la conoscenza di un fatto, un'intuizione ed un ragionamento; in parte è intrinseca all'esperimento ed in parte deriva dall'osservatore dell'esperimento. Può riguardare un episodio singolo mai verificatosi fino al momento della valutazione e può riguardare un numero sterminato di repliche della stessa prova. E' conoscenza teorica, è percezione soggettiva, è esperienza ed è comune sentire. Tutto concorre a formulare un giudizio su quale sia la probabilità da dare ad un certo evento. Molti autori sono consapevoli della vaghezza di formulazione del concetto di probabilità, come del resto sono imprecise le misurazioni fisiche e soprattutto la misura delle attitudini psicofisiche in cui però la difficoltà pratica di misurare un concetto non ha precluso -grazie ad opportuni accorgimenti e semplificazioni- soddisfacenti sviluppi teorici.

Esercizio_TP55: *una interessante lettura per comprendere l'evoluzione storica del calcolo delle probabilità ed il suo intreccio con le altre radici della Statistica è l'eccellente capitolo che Boldrini (1968) dedica alla storia della Statistica. Cercate il testo in biblioteca e scoprite: a) Quale sia stato il ruolo degli studiosi italiani; b) Quale siano stati gli apparentamenti religiosi con questa disciplina.*

6.3. Probabilità e calcolo combinatorio

La legge empirica del caso è il dato di fatto che in certe circostanze le frequenze relative si stabilizzano se valutate su di una successione molto lunga di prove omogenee. In verità, quando statistici di professione ed appassionati si sono sobbarcati fatiche e sbadigli per controllare la rispondenza tra simmetrie fisiche dell'esperimento e frequenze relative, hanno trovato sì accordo sostanziale, ma con scarti più consistenti di quanto non ci si aspettasse. Tra l'altro, scarti oscillanti che si mantengono significativi anche per successioni enormi, almeno nella scala del ragionevole tempo che si può dedicare a tali sperimentazioni. Certo, ciò è spiegabile con l'usura degli strumenti di rilevazione e della caduta di attenzione in chi rileva; anche le insopprimibili imperfezioni ed irregolarità delle condizioni sperimentali potrebbero avervi un ruolo. Sono escluse, almeno fino a prova contraria, altre cause. Questo ci porta a due considerazioni:
1) E' irragionevole negare la stabilizzazione delle frequenze relative intorno ai valori prestabiliti dalle condizioni fisiche a premessa dell'esperimento;
2) E' superfluo esplicare una nuova verifica sperimentale ogni volta che ricorrano le medesime simmetrie.

Applichiamo queste considerazioni ad una prova semplice e fondamentale: la scelta casuale di un gruppo di "n" oggetti tra un numero finito di N. Per semplificare ricorriamo al modello dell'urna che contiene biglie di vario colore in proporzione nota. Si può parlare effettivamente di scelta casuale solo dopo aver eliminato ogni differenza a proposito di forma, temperatura, peso, superficie esterna e posizione iniziale delle biglie ovvero siano state ridotte a distinzioni inutilizzabili o irrilevanti per poter scegliere una determinata biglia e questo per tutte le biglie (è implicito lo scuotimento prolungato e deciso dell'urna nonché la sua rotazione ripetuta dal basso verso l'alto e viceversa). I gestori delle case da gioco (o il ministero delle finanze) non temono che qualcuno inventi un "sistema" per vincere basato sui sogni, sui fondi di caffè o su qualche formula tipo il famoso "passo del capitano" per giocare sicuro alla roulette di cui mena vanto Alberto Sordi nel film "Crimen" (mirabile esempio di commedia italiana). Il vero incubo è che nei loro strumenti -per difetti di costruzione, deterioramenti, mero errore materiale o per alterazioni provocate da malintenzionati- si realizzi una qualche regolarità, anche minima, che possa essere sfruttata da giocatori accorti (fatti del genere sono successi e certamente sono nella cronaca dei quotidiani al momento in cui state leggendo queste pagine).

6.3.1 Formazione dell'universo degli eventi

Uno stesso esperimento può essere affrontato con un universo degli eventi diverso secondo le finalità dell'indagine ed è perciò necessario specificare a quale insieme si fa riferimento ovvero elencare -senza alcuna omissione- le possibili manifestazioni della prova (o perlomeno di stabilire quante siano qualora l'elencazione fosse resa impossibile dalla enormità dl loro numero). Molte contraddizioni dell'approccio classico alla probabilità hanno come denominatore comune la costituzione errata o confusa dell'universo degli eventi.

Il calcolo combinatorio, basato sulle idee primitive di distinzione e di classificazione, stabilisce in quanti modi diversi si possono combinare degli oggetti e torna utile nell'enumerazione delle alternative in un esperimento semplificando una operazione che può rivelarsi lunga e noiosa ed in cui è facile omettere o duplicare degli eventi. Un diretto beneficio è la facilità di assegnare le probabilità ai singoletti (e quindi a tutti gli altri) qualora ricorressero le condizioni del modello di probabilità uniforme.

Esempio:
Ripreso da E. Lombardo (1984, p. 497). Trattando il problema di quale fosse la probabilità di ottenere testa per due volte nel lancio di due monete, D'Alembert enumerò i tre casi seguenti:
1) Croce al primo lancio. E' inutile continuare perché ora non possono più verificarsi due teste;
2) Testa al primo e croce al secondo;
3) Due teste in entrambi i lanci.
Le tre possibilità non sono però equivalenti; manca, infatti, la distinzione del primo caso in due sotto-casi: (croce, croce) e (croce, testa) che sono stati erroneamente accomunati nel primo lancio.

Le nozioni necessarie per evitare banali mancanze (si veda comunque Thomasian 1969, pp. 24-25, per una giustificazione del ragionamento di D'Alembert) dovrebbero essere già note dalle scuole secondarie o dai corsi di matematica. Tuttavia, per la loro importanza propedeutica, è bene riprendere alcune nozioni di calcolo combinatorio seguendo la costruttiva impostazione di E. Lombardo (1984, cap. 9).

Moltiplicazione combinatoria

In primo luogo dobbiamo perfezionare il modo in cui si costituisce S estendendolo a prove in cui l'evento elementare risulti dal combinato di diverse classificazioni ovvero ai casi in cui l'esperimento si componga di sottoprove -parallele o in sequenza - ognuna dotata di una propria descrizione.

Esempi:

a) Una calcolatrice tascabile ha 33 tasti con funzioni che si attivano in tre modi: pressione diretta, freccia blu + tasto e freccia rossa + tasto. Quante funzioni esistono in tutto? 33x3=99.

b) Un taxi deve andare da Piazza "A" a Piazza "D" passando per Piazza "B" e piazza "C". Per il primo tratto può imboccare tre vie, per il secondo quattro e cinque per il terzo. Quanti percorsi può seguire? Ogni scelta per A→B si combina con le scelte della tratta B→C che a loro volta si combinano con quelle C→D e dunque: 3x4x5=60.

b) Nel pianificare un' indagine sul consumo di caffè si preparano tre miscele: arabica, colombiana, mista; con due diverse confezioni: busta o scatola e quattro diversi formati: singola, doppia, famiglia, bar. Qual'è l'universo degli eventi?

S	B	A	S	B	C	S	B	M
D	B	A	D	B	C	D	B	M
F	B	A	F	B	C	F	B	M
B	B	A	B	B	C	B	B	M
S	S	A	S	S	C	S	S	M
D	S	A	D	S	C	D	S	M
F	S	A	F	S	C	F	S	M
B	S	A	B	S	C	B	S	M

Ogni scelta della miscela si combina con le due scelte della confezione formando 3x2=6 coppie di scelte (miscela, confezione). A loro volta, ciascuna di queste 6 coppie si combina con le quattro scelte del formato dando luogo a 6x4=24 terne di scelte (miscela, confezione, formato).

L'universo degli eventi si configura come un prodotto cartesiano $S=\{C_1 \otimes C_2, \otimes \ldots \otimes C_n\}$ dove "n" è il numero di sottoprove in cui si articola l'esperimento. Le sottoprove possono sia essere ripetizioni di una stessa operazione che operazioni diverse. Se fosse valido lo schema di probabilità uniforme sarebbe sufficiente stabilire il numero degli eventi elementari per definire tutto ciò che è necessario a gestire l'esperimento. In questo senso lo strumento più semplice è il principio della moltiplicazione combinatoria cioè di procedere al conteggio dei casi suddividendo l'operazione in sub-operazioni più semplici: invece di contare gli elementi di un insieme si contano gli elementi di vari sottoinsiemi componenti e si moltiplicano le numerosità.

Esempi:

a) Nel lancio di tre monete l'universo degli eventi è formato dalla successione dei tre risultati {CCC, CCT, CTC, TCC, TTC, CTT, TCT, TTT} cioè 2x2x2=8 elementi. L'evento elementare è formato con tre informazioni di stato (testa o croce), ma costituisce un oggetto unico, non frazionabile, almeno in questo esperimento.

b) Un portafoglio contiene tre azioni (Giat, Mirelli, Nocetti) che alla chiusura possono risultare: (in salita, in discesa, stabile). L'universo degli eventi è : S={(Giat,in salita), (Giat, in discesa), (Giat, stabile), (Mirelli,in salita), (Mirelli, in discesa), (Mirelli, stabile), (Nocetti,in salita), (Nocetti, in discesa), (Nocetti, stabile)}. Invece di elencare gli eventi elementari si può stabilire il loro numero tenendo conto che ognuna delle tre azioni può trovarsi in ciascuno dei tre stati e quindi S comprende 3x3=9 esiti diversi.

c) La Società Generale Servizi Turistici ha organizzato una lotteria in due sue divisioni. Chi vince avrà diritto ad un viaggio premio di un mese per 4 persone con copertura integrale delle spese. Per vincere bisogna essere in possesso del numero vincente scelto tra i 10'000 possibili. Le estrazioni delle due lotterie sono separate. I dipendenti possono comprare solo un biglietto che è valido per entrambe le estrazioni. Quanti sono i casi possibili? Nella 1ª estrazione ci sono 10x10x10x10=10^4 possibilità che si combinano con altrettante possibilità della 2ª per cui chi gioca ha una possibilità su cento milioni (10^8) di vincere entrambi i premi.

d) Un settimanale riporta i libri più venduti di visi in tre categorie: narrativa (7 titoli), saggistica (8 titoli), divulgazione (5 titoli). Se qualcuno ne volesse un *set* di tre, uno per categoria, quante scelte avrebbe a disposizione? 7*8*5=280.

e) Una agenzia di viaggi ha organizzato un *tour* per 9 città d'arte per l'anno in corso ed un *tour* di 8 per l'anno venturo (in entrambi è prevista la visita a Firenze che costituisce l'unico duplicato). Una proposta *last minute* offre a prezzo stracciato la visita ad una città d'arte quest'anno ed una il prossimo anno, scegliendo però a caso le due città. Stefania non vuole passare per Firenze perché è troppo cara. Quante possibili scelte le sono favorevoli? 8*7=56.

Supponiamo che l'evento elementare sia identificato dall'accostamento di "n" tipologie C_i ognuna contenente n_i con i=1,2,…, n categorie. Il totale dei casi possibili si ottiene dalla moltiplicazione combinatoria:

$$card(S) = k = \prod_{i=1}^{n} n_i = n_1 n_2 \ldots n_n$$

k sarà pari a zero quando uno dei domini componenti è vuoto.

Esempi:
a) Il codice a barre è una tecnica identificativa in uso per prodotti, libri, componenti. La codifica EAN prevede 13 caratteri ognuno rappresentato da due barre: 2 caratteri rappresentano il paese di origine, 5 l'azienda produttrice, 5 il prodotto; l'ultimo è un carattere fisso di controllo. Quante possibili configurazioni di codici sono possibili? $2^2x2^5x2^5$=4'096.

b) La targa automobilistica ha ora tre blocchi di caratteri: due lettere, tre cifre e altre due lettere. Le diverse targhe possibili sono: 26*26*10*10*10*26*26=456'976'000.

c) Un menu prevede due soli tipi di primi, quattro tipi di secondo, tre sole scelte per il contorno e due dessert. Quanti menu diversi è possibile richiedere? 2x4x3x2=48.

Esercizio_TP56:
a) Ciccillo si è iscritto all'Università e, sebbene sia quasi finito il primo semestre, non ha ancora imparato a muoversi nel campus. Dal centro residenziale alle aule esistono quattro percorsi; dalle aule alla biblioteca tre percorsi; dalla biblioteca al laboratorio linguistico due percorsi e dal questo alle sale informatiche tre percorsi. Quanti sono i possibili percorsi andata e ritorno?
b) "O.K. il prezzo è giusto!". Un concorrente sa che il prezzo è di otto cifre e che la 1ª è più grande di sette e l'ultima è pari (zero incluso). Quante prezzi rimangono da scegliere?
c) Per gli anni 1996 e del 1997 si vuole valutare come le aziende hanno modificato il ricorso alle fonti di finanziamento: soci, aziende ed istituti di credito, credito agevolato, autofinanziamento, obbligazioni, altri intermediari finanziari. Quante scelte occorrerà considerare se si include la distinzione tra credito a breve, a medio e lungo termine?
d) Per molto tempo sono state in uso le schede perforate come input per i computer. Erano cartoncini organizzati in 80 colonne e 12 righe. Su ogni colonna trovava posto un carattere rappresentato da perforazioni in una o più righe. La corrispondenza tra carattere e perforazioni era espressa da codici. Sapendo che un codice impiegava 3 fori e che usava solo 6 righe, quanti caratteri si potevano rappresentare sulle 80 colonne?
e) Una terapia prevede l'uso di tre principi attivi: A, B, C. Il primo ha 3 livelli (basso, medio, alto), il secondo ne ha 2 (presente, assente), ed il terzo deve essere dosato su 5 livelli. Quante diverse terapie sono possibili?

6.3.2 Enumerazioni

Un esperimento molto semplice, ma di grande utilità per la formulazione di molti problemi statistici, è basato su di un' urna opaca contenente N bussolotti di vario colore (però indistinguibili per ogni altro aspetto) al cui interno siano poste delle indicazioni ad esempio le N=26 lettere dell'alfabeto oppure le N=10 cifre arabe o gli N=90 numeri del lotto e che l'esperimento consista di "n" estrazioni di un bussolotto. L'evento elementare è una n-tupla di elementi. L'esperimento include diverse varianti che si possono classificare in base a tre aspetti:
1) L'ordine delle estrazioni all'interno della n-tupla è rilevante oppure no;

$$Ordinata: \left(x_1 \in C_1,\ x_2 \in C_2,\ldots,\ x_n \in C_n\right);\quad Non\ ordinata: \left\{x_1 \in C_1,\ x_2 \in C_2,\ldots,\ x_n \in C_n\right\}$$

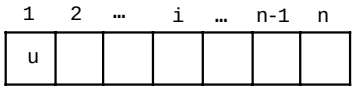
Nella prima, ogni alterazione dell'ordine genera un evento distinto; ciò non succede nella seconda dove, per avere un evento diverso, è necessario modificare almeno un elemento.
2) Gli elementi all'interno dei bussolotti possono essere tutti diversi oppure ripetuti. Ad esempio se si inseriscono lettere e cifre scritte con un carattere come il *times* che non distingue la cifra uno dalla lettera elle minuscola: "1" e "l" questo simbolo deve essere considerato ripetuto.
3) L'estrazione avviene con o senza reimmissione. Occorre cioè precisare se -dopo l'estrazione- la biglia è rimessa nell'urna- oppure ne resta fuori. Nel primo caso si parla si estrazione con reimmissione; nel secondo caso si parla di estrazione senza reimmissione perché dopo ogni estrazione la scelta si riduce di una unità.

Disposizioni senza reimmissione

Si tratta di scelte ordinate senza ripetizione e senza reimmissione: l'evento elementare è una n-tupla costituita con i risultati delle "n" estrazioni tra le N possibili. Si ammette che l'ordine sia importante e cioè che {A, B} è diverso da {B,A} anche se contengono gli stessi elementi. Peraltro, gli elementi nei bussolotti sono tutti distinti. Se n=1, le scelte possibili sono N dato che ogni elemento è legittimato ad entrare nella scelta. Come varia il numero delle n-tuple al variare dell'ampiezza?

Esempio:
Consideriamo l'insieme dei punti cardinali P={E,N,O,S}. Le possibili coppie, senza ripetizione, sono 12 in quanto ciascuno dei 4 può trovarsi appaiato ad uno dei rimanenti 3 e quindi 4*3=12. Per scelte di ampiezza n=3 le possibilità aumentano in quanto ciascuna delle 12 coppie già ottenute può combinarsi con uno dei due elementi rimasti per cui il numero di scelte diventa 12*2=24. Se l'ampiezza passa ad n=4 le opportunità di scelta non aumentano visto che ora si può solo completare la terna con l'elemento mancante.

L'evento elementare in questo esperimento è una disposizione di N oggetti presi "n" alla volta con un prefissato ordine: $D_{SR}(N,n)$ dove n≤N. Il pedice SR indica che lo stesso oggetto non può ricomparire nella stessa disposizione. Per stabilire il loro numero usiamo un procedimento induttivo: definiamo la disposizione per n=1 e poi cerchiamo una regola per costruire la disposizione di ordine "i+1" a partire da quella di ordine "i". Pensiamo perciò alla n-tupla come formata da "n" caselle distinte tali che ognuna possa essere occupata da un solo elemento:



Nella prima casella trova posto uno qualsiasi degli N elementi, diciamo quello etichettato "u"; nella seconda uno dei restanti (N-1) dato che non possiamo ripetere "u"; nella terza possiamo scegliere tra (N-2) elementi ancora rimasti e così via fino a che non si siano occupate le "n" caselle; per l'n-esima posizione rimangono liberi [N-(n-1)]=(N-n+1) elementi. La regola di moltiplicazione combinatoria comporta:

$$D_{SR}(N,n) = N*(N-1)*(N-2)*\ldots*(N-n+1)$$

Esempi:
a) Un'insegnante deve scegliere quattro studenti rappresentativi dello stato di preparazione di una classe composta da 25 alunni. Supponendo che gli alunni siano ordinati secondo una graduatoria di profitto, quante sono le possibili scelte? $D_{SR}(25,4)=25*24*23*22=303'600$.

b) Tra i 7 clienti che aspettano di essere richiamati si prevede che solo 3 confermino la prenotazione. In quanti modi diversi possono dislocarsi le tre telefonate di conferma? 7*6*5=210.

c) Il codice di un certo prodotto è formato con lo schema ordinato: {Lettera, Lettera, Numero, Numero, Numero} dove "Lettera" è uno dei 21 caratteri dell'alfabeto italiano e "Numero" è una delle dieci cifre arabe. Supponendo che né le lettere né i numeri possano ripetersi più di una volta, quanti prodotti è possibile etichettare? Ci sono $D_{SR}(21,2)$= 21*20=420 possibilità per le lettere e $D_{SR}(10,3)$= 10*9*8=720 per le cifre. In tutto sono: $D_{SR}(21,2)$* $D_{SR}(10,3)$=302'400.

Esercizio_TP57:
a) Un'associazione che persegue l'obiettivo delle pari opportunità tra i sessi è formata da 25 persone. Si devono nominare presidente, segretario e tesoriere e si decide di occupare a turno le cariche. Tutte possono ruotare su qualsiasi posizione purché prima di diventare segretario si sia stati tesoriere e prima di presidente si sia tenuta la carica di segretario. Quante sono le possibili nomine?
b) Le lettere hanno -come i numeri- un valore posizionale: cambia la parola se le lettere sono scambiate di posizione. Quante parole diverse di 5 lettere si possono formare con le vocali e le consonanti: "s", "c" ed "r"?
c) Gli abilitati alla libera professione di ragioniere sono stati 150. Volendo intervistarne 15 in un dato ordine (per controllare la comunicazione tra gli intervistati) quante sequenze si dovrebbero controllare?
d) Nella corsa tris si scommette sui cavalli che arrivano nei primi tre posti. Ipotizzando un numero di partecipanti tra 15 e 25 quante sono le giocate alternative?

Lo schema di calcolo del numero di disposizioni può essere impostato in modo ricorsivo:

$$D_{SR}(N,n) = N \quad per \;\; n=1; \quad D_{SR}(N,n-1) = (N-n)*D_{SR}(N,n) \quad per\; n>1$$

cioè le disposizioni di N oggetti in blocchi di "n" si ottengono combinando le disposizioni di "n" oggetti in blocchi di (n-1) con ognuno degli (N-n) non compresi nel blocco.

Esempi:
a) $D_{SR}(30,4) = 30*29*28*27 = 657'720 = (30-3)*D_{SR}(30,3) = 27*30*29*28$

b) La caposquadra coordina 12 tecnici e deve inviare una squadra di 3 persone in cui l'ordine di scelta determina il tipo di responsabilità: attrezzature, comunicazioni, conduzione automezzo. Quante squadre può formare? D(12,3)=12*11*10=1'320.

Esercizio_TP58:
a) Nell'assegnazione di codici con accostamento di linee colorate sono disponibili N=7 colori ed i codici sono formati da n=3 linee; ordinamenti di colori diversi sono codici diversi. Quanti sono i codici?
b) In una scommessa ippica si vince se nella corsa di "n" concorrenti si individuano correttamente il 1°, il 2° ed il 3° in ordine di arrivo. Quante alternative esistono?
c) Il giudice deve affidare una perizia contabile ad un collegio di quattro periti. Tra il personale di fiducia vi sono 20 commercialisti di cui 8 nella fascia B e 12 nella fascia più specializzata A.
1. Quante scelte sono possibili se la perizia non richiede competenze specialistiche (possibili sia B che A)?
2. Quante scelte sono possibili in caso siano necessarie competenze specialistiche (solo fascia A)?

Disposizioni con reimmissione

Le biglie estratte sono ora rimesse garantendo nel contempo il ripristino della situazione di scelta antecedente l'estrazione. Le biglie , come in precedenza, sono tutte diverse; quindi, in ogni posizione sono date N possibilità e le alternative di selezione -ordinata o no- con reimmissione di "n" bussolotti da un'urna che ne contiene N è:

$$D_{CR}(N,n) = \overset{n\ volte}{N * N * \ldots * N} = N^n$$

Esempi:
a) Riprendiamo l'insieme: P={E,N,O,S}. Se n=1 niente è cambiato rispetto al caso senza ripetizione e le possibilità sono sempre N. Se invece n≥2 qualcosa cambia. Le possibilità per n=2 sono infatti 16 e non più 12 come prima in quanto bisogna aggiungere i 4 casi: "EE", "NN","OO", "SS"; Per n=3 i possibili eventi elementari sono 4*4*4=64.

b) Ad un compito sono stati dati esercizi con difficoltà: "*", "**", "***". Il compito è composto con l'aiuto del computer che sceglie a caso 5 esercizi. Nessun controllo selettivo è fatto cosicché può capitare un compito di esercizi di uguale difficoltà. Quante possibili configurazioni di livelli di difficoltà si possono generare? 5^3=625.

c) Nel concorso pronostici del TOTIP si ottiene la vincita massima indovinando n=14 risultati legati alle corse ippiche. Per ogni risultato le possibilità sono N=3 e cioè ("1", "X", "2"); ne consegue che le diverse disposizioni con ripetizione sono: 3^{14}=4'782'969.

d) Un quotidiano elenca il nome del Presidente del Consiglio degli ultimi undici governi con a fianco il nome del Ministro dell'Interno; tali nominativi sono però disposti a caso. Quanti "ticket" alternativi sarebbero possibili? 2^{11}=2'048.

Esercizio_TP59:
a) Nel codice ASCII un carattere è rappresentato da un byte cioè 8 bit ed ogni bit può assumere il valore "0" oppure l'1 tranne l'ultimo bit che assume un valore determinato dagli altri sette. Dato che i valori possono ripetersi, quanti sono i possibili simboli?
b) All'uscita di un cinema si presentano 80 spettatori. Dieci intervistatori devono interrogarne dieci (uno per ciascuno) per creare un indice di gradimento del film. Tenuto conto che gli intervistatori non sono in contatto tra di loro, quanti sono i possibili gruppi di intervistati che si possono determinare?
c) Un nuovo tipo di collirio è somministrato a 5 volontarie scelte fra 30 coinvolte nella sperimentazione. Poiché nello studio è indifferente effettuare le applicazioni alla stessa persona o a persone diverse, quante sono le scelte a disposizione degli sperimentatori?

Permutazioni semplici

Supponiamo che da un'urna sia stata scelta una n-tupla e che sia utile considerare l'ordinamento dei suoi elementi ovvero, quante sono le scelte ordinate possibili se dall'urna si estraggono tutti i bussolotti?

Esempio:
Una organizzazione di volontari ha sorteggiato quattro nomi: I_1, I_2, I_3, I_4 per le cariche di portavoce, vicario, segretario e tesoriere: P,V,S,T. Supponendo che ogni nominativo possa essere destinato ad una qualsiasi delle cariche abbiamo le possibilità seguenti:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
P	I_1	I_1	I_1	I_1	I_1	I_1	I_2	I_2	I_2	I_2	I_2	I_2	I_3	I_3	I_3	I_3	I_3	I_3	I_4	I_4	I_4	I_4	I_4	I_4
V	I_2	I_2	I_3	I_3	I_4	I_4	I_1	I_1	I_3	I_3	I_4	I_4	I_2	I_2	I_1	I_1	I_4	I_4	I_2	I_2	I_3	I_3	I_1	I_1
S	I_3	I_4	I_4	I_2	I_2	I_3	I_3	I_4	I_3	I_1	I_1	I_3	I_4	I_1	I_4	I_2	I_1	I_1	I_1	I_3	I_2	I_1	I_2	I_3
T	I_4	I_3	I_2	I_4	I_3	I_2	I_4	I_3	I_3	I_4	I_3	I_1	I_1	I_4	I_2	I_4	I_2	I_2	I_3	I_1	I_1	I_2	I_3	I_2

Ciascuna colonna costituisce una permutazione cioè un evento elementare di questo esperimento. Come si arriva a 24? Ogni nominativo si combina con gli altri tre e ne risultano 12 coppie; ognuna di queste si combina con i due elementi rimasti ed ecco il 24.

La permutazione è la disposizione di tutti gli elementi in cui nessuno è presente più di una volta (in parecchi testi, disposizioni e permutazioni sono accomunati in una unica definizione e simbologia). Il loro numero è pertanto:

$$P_{SR}(N,N) = N*(N-1)*(N-2)*\ldots*2*1 \;=\; N!$$

Esempi:
a) In un esame universitario sono stati prescelti quattro argomenti da esporre in un ordine qualsiasi. Quante possibilità ha la persona interrogata? 4*3*2*1=4!!= 24.

b) Un ordine del giorno -per ragioni di tempo- è stato circoscritto a 7 punti. Quante sono le sequenze? 7*6*5*4*3*2*1= 7!=5'040.

c) Una psicologa deve colloquiare con 6 pazienti. Quante sequenze di visite sono possibili? 6!=720.

d) E' la festa della donna ed il ristorante è affollato da signore e signorine in vena di goliardate. Ciccillo serve ai tavoli e prende -su foglietti distinti- le ordinazioni di 5 tavoli. Qualcuna, accorgendosi che non ha scritto il numero di tavolo sulle ordinazioni gli mescola i foglietti. Fra quanti possibili ordinamenti è finito quello giusto? 5!=120. Il 1° abbinamento tavolo/foglietto ha cinque possibilità di cui una giusta. Il 2° ne ha quattro se la prima è giusta, il 3° ne ha tre perché due sono già state individuate, il 4° due e il 5° una.

Esercizio_TP60:
a) Una cura consiste in una sequenza di 7 trattamenti. Non è però ancora definito quale sia l'ordine di somministrazione più efficace. Tenuto conto che uno stesso trattamento non può essere ripetuto all'interno della terapia, quante di queste sono possibili?
b) Un prodotto richiede 4 fasi ed ognuna può essere effettuata da una diversa macchina: A, B, C, D. Ogni macchina può svolgere una qualsiasi delle fasi, ma non più di una fase. Quanti diversi processi produttivi si possono organizzare?
c) Le sei ragazze dell'appartamento al 5° piano-interno 2- hanno ricevuto la proposta di un appuntamento alla cieca con sei bravi ragazzi istruttori di nuoto. Quante sono le possibili coppie?
d) Si consideri la frase "Tutti i giovani amano le canzoni dei Beatles". Quante frasi si possono formare senza mai ripetere una stessa parola?
e) Una catena di ristorazione sta considerando 9 sedi in cui realizzare delle nuove filiali. L'investimento è proporzionale al numero di residenti nella città. Quante politiche di investimento sono possibili?

Il fattoriale di un numero
Il fattoriale di un numero cresce velocemente come si può vedere dalla tabella e già 15 fattoriale che non sembrerebbe preoccupante ha un valore superiore a mille miliardi.

n	$n!$	n	$n!$	n	n
1	1	6	720	11	39' 916' 800
2	2	7	5' 040	12	479' 001' 600
3	6	8	40' 320	13	6' 227' 020' 800
4	24	9	362' 880	14	871' 782' 291' 200
5	120	10	3' 628' 800	15	1' 307' 674' 368' 000

Il numero 10000! è un intero con 2'500 cifre e per scriverlo non basterebbe una pagina di questo libro. Per il fattoriale è possibile dare una definizione ricorsiva: n!=n*(n-1)! cioè definiamo un qualcosa attraverso il qualcosa da definire. La tautologia è però solo apparente perché (n-1)! è "più semplice" di n!

$$(n-1)! = (n-1)*(n-2)! \;; \qquad (n-2)! = (n-2)*(n-3)! \;; \qquad \ldots$$

Si arriva a 1!=1*0! e qui -convenzionalmente- si pone 0!=1 in modo che si regga la definizione per ogni intero. Uno strumento utile per la valutazione dell'ordine di grandezza dei fattoriali è la formula di Stirling:

$$n! \cong \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

(cfr. ad esempio, Longo, 1962, pp. 91-94) dove il simbolo $\cong$ significa "approssimativamente uguale". La formula di Stirling è soprattutto utile nelle applicazioni teoriche, ma è in anche grado di fornire delle accettabili approssimazioni: $10! \cong 3{,}598{,}695.619$ che non è troppo lontano dal valore esatto (l'errore è inferiore allo 0.2%).

Esercizio_TP61: *una migliore approssimazione del fattoriale può essere ottenuta dalla formula di Stirling modificata. Verificatene l'efficacia su 15!*

$$n! \cong \sqrt{2\pi n}\left(\frac{n}{e}\right)^n\left(1+\frac{1}{12n}\right)$$

Tra permutazioni semplici e disposizioni (cioè entrambe senza ripetizione) esiste una comoda relazione basata sui fattoriali. Ogni disposizione di N unità prese ad "n" la volta può essere abbinata alle permutazioni delle restanti (N-n)! unità:

$$N! = \left[N*(N-1)*(N-2)*\ldots*(N-n+1)\right]*\left[(N-n)*(N-n-1)*\ldots*2*1\right] = D_{SR}(N,n)*(N-n)!$$

Ne consegue che: $D_{SR}(N,n)=N!/(N-n)!$ e ciò permette di ricavare il numero di disposizioni dalle permutazioni avvantaggiandosi dell'uso dei fattoriali.

Esempi:

a) Calcoliamo le disposizioni di N=7 biglie prese a n=3 ed n=5 alla volta.

$$D_{SR}(7,3) = \frac{7!}{(7-3)!} = \frac{7*6*5*4*3*2*1}{4*3*2*1} = \frac{7*6*5}{1} = 210; \quad D_{SR}(7,5) = \frac{7!}{(7-5)!} = \frac{7*6*5*4*3*2*1}{2*1} = \frac{7*6*5*4*3}{1} = 2'\,520$$

b) Un'urna contiene N=4000 bussolotti dalla quale si deve ricavare una disposizione di n=200 biglie. Quante sono le possibilità?

$$k = \frac{4000!}{200!3800!} \cong \frac{\sqrt{2\pi}\sqrt{4000}}{2\pi\sqrt{200}\sqrt{3800}}\left(\frac{4000}{e}\right)^{4000}\left(\frac{3800}{e}\right)^{-3800}\left(\frac{200}{e}\right)^{-200}$$

$$Ln(k) \cong 0.5\left[Ln(4000)-Ln(3800)-Ln(200)-Ln(2\pi)\right] + 4000Ln(4000)-3800Ln(3800)-200Ln(200)$$

$$= -3.5422 + 794.061 = 790.5188 \Rightarrow k \cong 10^{343}$$

c) Una scatola contiene 100 numeri: da 00 a 99 inseriti in altrettanti bussolotti identici. Ogni numero corrisponde ad una persona in una lista di candidate per un colloquio di lavoro. Peraltro, le candidate sono tutte sullo stesso piano e perciò si decide di sceglierne 10 estraendo senza reimmissione dei bussolotti. Poiché il numero associato stabilisce anche l'ordine di presentazione della domanda di assunzione si vuole considerare la scelta ordinata. Quante possibilità esistono?

$$D_{SR}(100,10) = \frac{100!}{90!} = 100*99*98*97*96*95*94*93*92*91 = 62'\,815x10^{15}$$

Esercizio_TP62:

a) Alle corse dei cavalli o dei cani, una scommessa sulla exacta significa scegliere due dei concorrenti che arriveranno -nell'ordine- al primo e al secondo posto. Si supponga che la corsa preveda 12 partecipanti. Quante alternative di exacta esistono?

b) La giuria di un film-festival deve scegliere i primi tre classificati tra 18 opere concorrenti. Quante sono le possibili terne di finalisti?

c) Ad una selezione pubblica partecipano 45 concorrenti. Le prime quattro classificate frequenteranno un corso-concorso per l'assunzione. In quanti modi diversi possono essere occupate le posizioni vincitrici?

d) Presso il CATI (computer aided telephonic interviewing) è in azione un dispositivo che chiama automaticamente i numeri di un distretto telefonico basato su sei cifre. Quanti sono i numeri formati da cifre diverse?

e) Un commesso viaggiatore deve recarsi una e una sola volta in ciascuna delle 9 province della Sicilia. Supponendo che da ciascuna possa recarsi in una qualsiasi delle altre quanti sono i potenziali itinerari?

Combinazioni semplici

In questo caso l'ordine con cui gli oggetti compaiono nella n-tupla non è rilevante: è come se fossero presi in un unico blocco fermo restando che gli oggetti trattati sono tutti distinti. Ad esempio, nella scelta dall'insieme dei punti cardinali: P={E,N,O,S}, le coppie che rispondono a tali requisiti sono: (E,N); (E,O); (E,S); (N,O); (N,S); (O,S) quindi sei in tutto. Questa è una combinazioni di 4 oggetti presi a due alla volta. Rispetto alle disposizioni non sono più considerati alternativi gli allineamenti degli stessi elementi cioè: (O,S) ed (S,O) coincidono e contano per una. La scelta è ancora senza reimmissione. Per stabilire il numero delle combinazioni partiamo proprio dalle disposizioni. Ognuna di queste è formata da "n" delle N unità, ma esattamente n! (cioè le loro permutazioni) sono da considerarsi identiche e debbono essere conteggiate una sola volta. Quindi:

$$C(N,n)=\frac{D_{SR}(N,n)}{n!}=\frac{N!}{(N-n)!*n!}$$

Esempi:
a) Un gruppo di 10 volontari è pronto a sottoporsi ad una nuova terapia. Solo tre di loro potranno fruire della nuova cura sperimentale. Quante combinazioni, senza ripetizione, di volontari si possono avere?

$$\mathrm{C}(10,3)=\frac{10!}{3!*7!}=120$$

b) Da una lista di N persone si deve formare un comitato rappresentativo di "n" membri all'interno del quale sarà poi nominato un/una presidente. Quante sono le opportunità di scelta se N=15 e n=6? Occorre prima formare, senza ripetizioni, il comitato e poi, per ogni comitato, abbinare le possibilità di nomina alla carica di presidente:

$$C(N,n)*C(n,1)=\frac{N!}{(N-n)!*n!}*\frac{n!}{1!*(n-1)!}=\frac{N!}{(N-n)!*(n-1)!}$$

nel caso in esempio si hanno 30'030 risultati diversi.

c) Nel totogol si devono indovinare 8 risultati legati alle partite di calcio su 32 possibilità indicate dalla schedina. Le combinazioni alternative sono C(32,8)=10'518'300.

d) Un sistema informativo è composto da 10 CPU e rimane operativo purché almeno 6 dei processori sono in linea. Tenuto conto che la posizione dei processori in funzione non è rilevante, in quante combinazioni il sistema sarà operativo? C(10,6)=210

Esercizio_TP63:
a) In un banco sono presenti 11 diverse marche di detersivo. Se l'acquirente intende comprarne 3, quante scelte ha a disposizione, tenuto conto che l'ordine non ha alcuna importanza?
b) Il maresciallo della finanza ha una lista di 20 aziende da sottoporre a controllo dettagliato. Se, per il primo giorno decide di controllarne solo 5 tra cui certamente quella più grande, quante solo le scelte possibili?
c) Assunta ha 9 amiche con cui si tiene molto in contatto e 7 amici a cui è affezionata. Al party della sua amica Stefania può portarne però solo 5 ed i maschi non possono essere più di due. Di quante scelte dispone?

La formula per calcolare il numero di combinazioni può essere semplificata:

$$\frac{N!}{(N-n)!n!}=\frac{N*(N-1)*(N-2)*\ldots*(N-n+1)*(N-n)!}{(N-n)!*n!}=\frac{N*(N-1)*(N-2)*\ldots*(N-n+1)}{n!}$$

Esempi:
a) Al Superenalotto si vince indovinando 6 numeri sui 90 possibili in qualsiasi ordine si presentino. Le possibilità di uscita sono:

$$C(90,6)=\frac{90!}{6!84!}=\frac{90*89*88*87*86*85}{6*5*4*3*2}=622'\,614'\,630$$

b) Carmela Morandi deve scegliere 5 fornitori tra 20 per sondare la disponibilità a ridurre del 10% il prezzo di contratto in cambio della conferma degli ordini. Quante sono le opportunità potenzialmente esaminabili?

$$C(20,5)=\frac{20*19*18*17*16}{5!}=15'\,504$$

c)Il nuovo allenatore della nazionale ha idee strane su come formare la squadra. Considera i giocatori di nazionalità italiana delle prime otto in classifica -portieri esclusi- che hanno giocato la domenica precedente: 56 in tutto. Fra questi ne convoca 20 che poi disporrà a piacimento nei vari ruoli. Quante convocazioni distinte può effettuare? C(56,20)=785613.56x10^6.

Coefficiente binomiale

Per le combinazioni si usa anche il simbolo noto come coefficiente binomiale:

$$\frac{N!}{(N-n)!*n!}=\binom{N}{n}\quad da\ \ leggere\text{: "}N\ \ su\ \ n\text{"}$$

Affinché la formula sia definita per ogni valore di "N" ed "n" si conviene che:

$$a)\binom{N}{n}=0\quad se\quad n>N;\qquad b)\binom{N}{N}=1;\qquad c)\binom{N}{0}=1$$

La a) stabilisce che non c'è modo di scegliere più elementi di quelli contenuti; b) afferma che c'è un solo modo di scegliere tutti gli elementi e c'è pure un solo modo di non sceglierne alcuno.

Esempi:

a) Una sequenza è costituita da "p" successi e "q" insuccessi per cui vi sono (p+q)! sequenze possibili. Fra queste però p! e q! corrispondono allo stesso ordinamento per cui le possibili combinazioni sono:

$$\binom{p+q}{p}=\binom{p+q}{q}=\frac{(p+q)!}{p!q!}$$

b) In quanti modi si possono allineare 8 segni "+" e 3 segni "-" cosicché i segni "-" non siano mai contigui?
I segni "-" possono essere collocati nelle posizioni indicate con lo "0":

0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0

a questo punto basta sostituire tre "0" con tre "-" e questo si può fare in C(8,3) modi diversi.

Esercizio_TP64:

a) Un revisore ha scelto 30 transazioni che presentano un saldo negativo di almeno 50 milioni. Se decidesse di esaminare un campione di 7, quante scelte potrebbe fare?
b) Un supermercato ha 9 uscite di cui 4 debbono essere provviste di videocamera. In quanti modi si possono collocare le videocamere?
c) Una associazione per la difesa dei consumatori ha ricevuto 5000 reclami. Per verificarne tenore e portata si decide di sceglierne 100 per un colloquio più approfondito. Quante sono le possibili scelte? (Sugg. Usate l'approssimazione di Stirling);
d) Che significato dare e che valore attribuire al coefficiente binomiale C(0,0)?
e) "n" colli di peso diverso debbono essere collocati in un bagagliaio. Tenuto conto che ciascun collo può essere caricato oppure lasciato a terra, quanti sono i possibili carichi?

Vediamo alcune a le numerose relazioni notevoli per i coefficienti binomiali.

1. Condizione di simmetria:

$$\binom{N}{n}=\binom{N}{N-n}$$

Questa identità nasce dalla considerazione che:

$$\binom{N}{n}=\frac{N!}{n!*(N-n)!}=\binom{N}{N-n}=\frac{N!}{(N-n)!*n!}$$

Per ogni scelta di "n" elementi c'è una "non scelta" di (N-n) elementi che sebbene più contorta della prima prevede lo sesso numero di possibilità. Inoltre, conferma le assegnazioni logico-convenzionali uguali a uno per i casi C(N,N) e C(N,0).

*2.*Formula dell'addizione:

$$\binom{N}{n}=\binom{N-1}{n}+\binom{N-1}{n-1}$$

Con questa proprietà si instaura una relazione ricorsiva per il calcolo dei coefficienti binomiali per un fissato "n" e per N crescente.

$$\binom{N-1}{n}+\binom{N-1}{n-1}=\frac{(N-1)!}{n!*(N-n-1)!}+\frac{(N-1)!}{(n-1)!*(N-n)!}=\frac{(N-1)!}{n!*(N-n-1)!}+\frac{n}{(N-n)}*\frac{(N-1)!}{n!*(N-n-1)!}$$

$$=\left[1+\frac{n}{(N-n)}\right]*\frac{(N-1)!}{n!*(N-n-1)!}=\frac{N}{(N-n)}*\frac{(N-1)!}{n!*(N-n-1)!}=\frac{N!}{n!*(N-n)!}=\binom{N}{n}$$

L'idea è semplice: se la n-tupla non contiene un elemento le alternative sono date dal primo addendo; se lo contiene le scelte saranno quelle del secondo dato che una posizione è impegnata e le possibilità si sono ridotte di una unità.

3. Fattorizzazione:

$$\binom{N}{n+1}=\left[\frac{(N-n)}{(n+1)}\right]*\binom{N}{n}$$

Questa è un'altra relazione ricorsiva per calcolare il numero di combinazioni di ordine (n+1) a partire dal numero di combinazioni di ordine "n" per un valore di N fissato.

$$\binom{N}{n+1}=\frac{N!}{(n+1)!*(N-n-1)!}=\frac{N!}{(n+1)*n!*(N-n-1)!}=\frac{(N-n)*N!}{(n+1)*n!*(N-n-1)!*(N-n)}$$
$$=\frac{(N-n)N!}{(n+1)n!(N-n)!}=\left[\frac{N-n}{n+1}\right]*\frac{N!}{n!*(N-n)!}=\left[\frac{N-n}{n+1}\right]*\binom{N}{n}$$

4. Formula del binomio:

$$(a+b)^n=\sum_{i=0}^{n}\binom{n}{i}a^{n-i}b^i=\binom{n}{0}a^n+\binom{n}{1}a^{n-1}b+\binom{n}{2}a^{n-2}b^2+\ldots+\binom{n}{n-1}a^1b^{n-1}+\binom{n}{n}b^n$$

Poiché $(a+b)^n=(a+b)*\overset{n\ volte}{(a+b)}*\ldots*(a+b)$ il termine a^n non può che essere ottenuto prendendo "a" da ogni binomio e c'è solo una possibilità: C(n,n); il termine $a^{n-1}b$ si ottiene prendendo "a" da (n-1) binomi e "b" da uno solo. Le scelte sono ora: C(n,n-1). L'addendo "i" della sommatoria: $a^{n-i}b^i$ si ottiene scegliendo "b" in "i" binomi e "a" in (n-i) e per questo esistono C(n,n-i) modi distinti.

Esempi:

a) Sviluppo binomiale per n=2:

$$(a+b)^2=a*(a+b)+b*(a+b)=a^2+ab+ba+b^2=a^2+2ab+b^2=\binom{2}{0}a^2+\binom{2}{1}ab+\binom{2}{2}b^2$$

b) Un caso interessante si ha per b=1. La formula del binomio diventa:

$$\sum_{i=1}^{n}\binom{n}{i}a^i=(1+a)^n$$

c) per a=-1 e b=1 si ha:

$$\sum_{i=0}^{n}\binom{n}{i}(-1)^i=\binom{n}{0}-\binom{n}{1}+\binom{n}{2}-\ldots+(-1)^n\binom{n}{n}=0$$

d) Coefficienti binomiali generalizzati.

La capacità di sintesi dell'espressione binomiale è tale da poter includere anche il caso di un ambito di scelta non intero: ad esempio nella scelta tra "α" unità convenzionali espresse con frazioni (cfr. paragrafo 1.3).

$$\binom{\alpha}{n}=\frac{\alpha(\alpha-1)(\alpha-2)\ldots(\alpha-n+1)}{n!}=\prod_{i=1}^{n}\left(\frac{\alpha+i-1}{i}\right)$$

dove "n" è un intero. Vale la convenzione già adottata per α intero: C(α,0)=1. Nulla impedisce di utilizzare la formula con un α negativo:

$$\binom{-\alpha}{n}=\prod_{i=1}^{n}\left(\frac{-\alpha+i-1}{i}\right)=(-1)^n\frac{\alpha(\alpha+1)(\alpha+2)\ldots(\alpha+n-1)}{n!}=(-1)^n\binom{n+\alpha-1}{n}$$

A parte il fattore $(-1)^n$ vedremo tra poco che è ancora possibile interpretare l'espressione come il conteggio delle alternative di scelta tra unità frazionarie.

Esercizio_TP65: *dimostrare che:*

$$\sum_{i=0}^{r}\binom{n}{i}\binom{N-n}{r-i}=\binom{n}{0}\binom{N-n}{r}+\binom{n}{1}\binom{N-n}{r-1}+\ldots+\binom{n}{r}\binom{N-n}{0}=\binom{N}{r}$$

Partizioni

Diamo adesso alle entità dell'evento elementare la possibilità di essere presenti più di una volta nell'urna. A questo fine i bussolotti cavi del nostro esperimento diventano portatori di due informazioni: una relativa alla categoria di appartenenza ed una distintiva all'interno del gruppo. Supponiamo che i gruppi diversi siano "m" e che di ogni gruppo siano presenti N_i, i=1,2,…,m esemplari con $\sum N_i=N$. Procediamo ad estrarre senza reimmissione, "n" bussolotti. Tale scelta includerà n_i biglie del gruppo i-esimo per i=1,2,…,m con i vincoli:

$$0 \le n_i \le min\{n, N_i\} \quad e \quad \sum_{i=1}^{m} n_i = n;$$

La scelta ordinata delle biglie del 1° gruppo può avvenire in $D_{SR}(N_1,n_1)$ modi diversi, le biglie del 2° gruppo sono selezionabili in $D_{SR}(N_2,n_2)$ modi che si combinano con i casi provenienti dalla 1ª scelta per formare $D_{SR}(N,n_1)D_{SR}(N\text{-}n_1,n_2)$ possibilità; le scelte nel 3° gruppo sono $D_{SR}(N_3, n_3)$ che si affiancano a gli abbinamenti dei due precedenti per formare $D_{SR}(N_1,n_1)D_{SR}(N_2,n_2)D_{SR}(N_3,n_3)$ e così via. Il numero di partizioni ordinate è:

$$D_{sr}(N_1,n_1) * D_{sr}(N_2,n_2) * D_{sr}(N_3,n_3) * \ldots * D_{sr}(N_m,n_m) = \frac{N_1!}{(N_1-n_1)!} * \frac{N_2!}{(N_2-n_2)!} * \ldots * \frac{N_m!}{(N_m-n_m)!}$$

Se i bussolotti non fossero distinguibili aldilà del colore ogni coefficiente $D_{SR}(N_i,n_i)$ dovrà essere diviso per il fattore $n_i!$ dato che tante sarebbero quelle identiche. In definitiva, il numero di partizioni non ordinate è:

$$\frac{N_1!N_2\ldots N_m!}{n_1!n_2\ldots n_m!}\left[\frac{1}{(N_1-n_1)!(N_2-n_2)!\ldots(N_m-N_m)!}\right] = \binom{N_1}{n_1}\binom{N_2}{n_2}\ldots\binom{N_m}{n_m} = \prod_{i=1}^{m}\binom{N_i}{n_i}$$

Esempi:

a) Un lotto contiene 20 prodotti di cui 14 buoni, 4 mediocri e 2 difettosi. Se si scelgono 6 prodotti -senza reimmissione e senza ripetizione- quante sono le scelte alternative in cui compaiono 3 mediocri e due buoni?

$$\binom{14}{2}\binom{4}{3}\binom{2}{1} = \frac{14!\,4!\,2!}{2!3!1!}\frac{1}{12!1!1!} = 728$$

b) Una comitiva di 100 turisti è diretta all'ufficio cambi per ottenere degli euro: 29 sono americani, 31 giapponesi, 18 australiani, 15 russi, 7 argentini. Si decide che il cambio sia effettuato solo da due per nazionalità. Quante sequenze di cambiavalute sono possibili?

$$\binom{29}{2}\binom{31}{2}\binom{18}{2}\binom{15}{2}\binom{7}{2} = \frac{29*28*31*30*18*17*15*14*7*6}{(2!)^5} = 63'\,691'\,138'\,350$$

c) In una smazzata di tressette la possibilità di una mano con 3 spade, 4 coppe, 2 bastoni e 1 denari è:

$$\binom{10}{3}\binom{10}{4}\binom{10}{2}\binom{10}{1} = \frac{10^4 9^3 8^2 7}{4!3!2!1!} = 11'\,340'\,000$$

Esercizio_TP66:

a) Un campione di 26 comuni è così suddiviso: 3 in sviluppo, 4 stabili e 19 in transizione. Tenuto conto che sui 100 comuni della popolazione le categorie erano (20, 30, 50). Quante erano le partizioni non ordinate che si potevano ottenere con la composizione (3,4,19)?

b) Nove biglie scelte tra 23 raggruppate per colore nella composizione (8,6,9) debbono essere collocate in 3 urne distinte nella composizione (4,2,3). Quante sono le scelte ordinate? Quante sono quelle non ordinate?

c) Un partito politico ha quattro componenti: progressista, conservatrice, ecologista, cattolica. La composizione del consiglio nazionale di 42 membri deve rispettare le proporzioni riscontrate nelle ultime elezioni: (5,4,2,3). Quante composizioni diverse risultano (l'ordine non è importante)?

d) Un organismo di N=20 Paesi di cui (7 ricchi, 5 in via di sviluppo e 8 poveri) deve nominare una commissione di 10 membri di cui 4 Paesi ricchi, 3 in via di sviluppo e 3 poveri. Quante commissioni alternative si possono formare?

e) Ad una gara automobilistica partecipano 5 diverse automobili per 12 squadre. Al traguardo arrivano solo 12 macchine, due per ogni squadra. Quante possibilità aveva questa partizione in modo ordinato e non ordinato?

Permutazioni con ripetizione

Manteniamo la composizione dell'urna come nel caso delle partizioni e consideriamo i possibili ordinamenti considerando indistinguibili quelle dello stesso gruppo. La scelta delle biglie del 1° gruppo può avvenire in $C(N,N_1)$ modi diversi dato che le biglie del gruppo "1" potrebbero trovarsi in qualsiasi posizione delle N scelte; le biglie del 2° gruppo sono selezionabili in $C(N\text{-}N_1,N_2)$ modi che si combinano con i casi provenienti dalla 1ª scelta per formare $C(N,N_1)C(N\text{-}N_1,N_2)$ possibilità; le scelte nel 3° gruppo sono $C(N\text{-}N_1\text{-}N_2, N_3)$ che si affiancano a gli abbinamenti dei due precedenti per formare $C(N,N_1)C(N\text{-}N_1,N_2)C(N\text{-}N_1\text{-}N_2,N_3)$ e così via. Il numero dei casi distinti è perciò:

$$\binom{N}{N_1}\binom{N-N_1}{N_2}\binom{N-N_1-N_2}{N_3}\cdots\binom{N-N_1-N_2-\ldots-N_{m-1}}{N_m}$$

Sviluppando i coefficienti binomiali si realizzano delle semplificazioni:

$$\binom{N}{N_1}\binom{N-N_1}{N_2}\binom{N-N_1-N_2}{N_3}\cdots=\frac{N!}{N_1!(N-N_1)!}\frac{(N-N_1)!}{N_2!(N-N_1-N_2)!}\frac{(N-N_1-N_2)!}{N_3!(N-N_1-N_2-N_3)!}\cdots$$

$$=\frac{N!}{N_1!N_2!N_3!\ldots N_m!}=\binom{N}{N_1,N_2,\ldots,N_m}=P(N_1,N_2,\ldots,N_m)$$

L'ultimo simbolo è il coefficiente multinomiale e generalizza quello binomiale. Il risultato, come è agevole controllare, è invariante rispetto all'ordine di considerazione dei gruppi.

Esempi:

a) Una nave deve inviare dei messaggi con bandiere di tre colori diversi: 3 rosse, 4 gialle, 5 verdi. Ogni messaggio è costituito da un allineamento di 12 bandiere; ad esempio: RRR, GGGG, VVVVV potrebbe significare seria perdita di carburante. Quanti sono in tutto i messaggi diversi che si possono inviare?

$$P(3,4,5)=\frac{12!}{3!*4!*5!}=\frac{12*11*10*9*8*7*6}{3!*4!}=27'720$$

b) Sei amiche vanno al cinema e vogliono sedere nella medesima fila. La sola fila che ha sei posti vuoti (in tutto i posti sono 10) ha già occupate le sedie "1", "5", "9" e "10". In quanti modi possono accomodarsi le amiche?

$$P(6,3)=\frac{6!}{3!*3!}=\frac{6*5*4*3*2*1}{3*2*1*3*2*1}=20$$

c) Nei primi undici posti di una gara automobilistica si sono classificate: 4 Lancia, 3 Tojota, 2 Peugeot, 2 Subaru. Fra quante classifiche alternative si può indovinare quella effettivamente formatasi?

$$P(4,3,2,2)=\frac{11!}{4!*3!*2!*2!}=69'300$$

d) N=30 delegati sindacali debbono formare un comitato. Fra i delegati N_1=15 sono per una trattativa ad oltranza e N_2=15 sono già pronti allo sciopero. Se la delegazione è formata da 4 persone quanti sono i casi in cui prevalgono i trattativisti?

$$\binom{15}{0}\binom{15}{4}+\binom{15}{1}\binom{15}{3}=1*\frac{15!}{11!4!}+15*\frac{15!}{3!12!}=27'360$$

e) Consideriamo gli anagrammi di ABBA (l'intramontabile quartetto *rock*). Se fossero tutte lettere diverse avremmo 4!=24 permutazioni; ma ora A B B A è identica a A B B A in cui si siano scambiate di posto la prima e la quarta A e/o la seconda e la terza B. Gli elementi uguali sono due rispetto alla B e due rispetto alla A, quindi gli anagrammi diversi sono 24/2/2=6:

f) Un noto *data set* per l'analisi multivariata consta di 12 tribù e caste indiane da suddividere in gruppi. La partizione ritenuta più efficace è in cinque gruppi ovviamente non vuoti e incompatibili. Quante sono le possibili composizioni dei gruppi? Il problema è diverso dai precedenti in quanto non è prestabilita la numerosità dei gruppi. Si tratta in breve di risolvere l'equazione:

$$x_1+x_2+x_3+x_4+x_5=12$$

limitatamente ai valori interi delle incognite. Bose e Manvel (1984, p. 48) propongono una soluzione di grande semplicità. Si parte dall'equazione: 1+1+1+1+1+1+1+1+1+1+1+1=12. Poiché gli addendi sono cinque alcuni degli "1" debbono essere accorpati: 1+(1+1+1)+(1+1)+(1+1+1+1)+(1+1)=12 per la composizione: (1,3,2,4,2). Tale scelta può essere fatta in C(12-1,5-1)=C(11,4)=330 modi diversi. In generale, se "m" è il numero da suddividere ed "n" gli addendi interi allora il numero di modi è C(m-1,n-1) se poi la soluzione deve essere data per interi non negativi allora i modi distinti sono c(m+n-1,n-1). Supponiamo ora che nessun gruppo possa avere meno di due unità. Quanti sono ora le possibili partizioni? Poniamo $y_i=x_i\text{-}2 \Rightarrow y_1+y_2+y_3+y_4+y_5=12 \Rightarrow x_1\text{-}2+x_2\text{-}2+x_3\text{-}2+x_4\text{-}2+x_5\text{-}2=12 \Rightarrow x_1+x_2+x_3+x_4+x_5=2$ con soluzioni intere non negative: C(2+5-1,5-1)=C(6,4)=15.

Esercizio_TP67:
a) Si conduce uno studio sugli elettrodomestici posseduti dalle famiglie. Una di queste ha tre radio, tre televisori, due videoregistratori e uno stereo. In quanti modi distinti si possono considerare gli oggetti?
b) Le etichette di un prodotto possono essere formate con i simboli {2,2,2,2,2, 6,6,6, 7, 8,8}. Ogni codice contiene 11 di questi simboli. Quante sono le possibili etichette?
c) Una portafoglio di 20 azioni comprende 10 titoli già quotati, 6 non quotati e 4 che hanno fatto richiesta di essere quotati. Quante permutazioni distinte si possono avere?
d) Per: {a,a,a, b,b, c, d,d,d,d}. Le permutazioni potenziali sono 10! =3'628'800. In realtà, quelle distinte (e cioè formate dagli stessi elementi, ma con almeno un elemento in una posizione diversa) sono molto meno. Quante?
e) In magazzino c'è una partita di 24 pneumatici usati di tre marche diverse presenti in egual numero; la marca si è cancellata su tutti. Dovendone scegliere 4 quante possibilità ci sono che siano tutte della stessa marca?

Combinazioni con ripetizione

In questo esperimento si deve costituire una sequenza di "n" biglie. C'è un'urna che ne contiene N diverse alla quale si affianca un'altra urna che contiene "n" biglie numerate da 1 ad "n". Si estrae senza reimmissione una biglia dalla prima urna e poi si estrae una biglia dalla seconda; il numero qui selezionato indica quante volte si deve ripetere la prima biglia nella sequenza. Se il numero uscito dalla seconda urna è "n" l'esperimento si interrompe perché la sequenza è stata completata. Nella seconda estrazione la composizione delle urne cambia: la prima è ridotta di una unità dato che la scelta è senza reimmissione e dalla seconda si tolgono i numeri che sommati a quello già uscito danno un risultato superiore ad "n". Si continua così fino a che non si siano occupate le "n" posizioni della sequenza. Per calcolare il numero delle combinazioni con ripetizione si aumenta idealmente di (n-1) l'insieme base da cui si intende scegliere gli oggetti e si considerano le combinazioni semplici prese a blocchi di "n". Il numero complessivo di tali combinazioni è:

$$\binom{N+n-1}{n} = \frac{(N+n-1)!}{n!(N-1)!} = \sum_{i=0}^{n-1}\binom{N-1+i}{n-1}$$

Esempi:
a) Consideriamo l'insieme {a, b, c, d, e} e supponiamo di sceglierne n=3 come combinazione semplice. Le possibilità sono dieci:

(a,b,c) (a,b,d) (a,b,e) (a,c,d) (a,c,e) (a,d,e) (b,c,d) (b,c,e) (b,d,e) (c,d,e)

se ora le lettere possono ripetersi le possibilità aumentano. Per ogni unità occorre aggiungere le terne in cui ne compaiono due e quella in cui ne compaiono tre: (a,a,a) (a,a,b) (a,a,c) (a,a,d) (a,a,e) … In tutto, bisogna sommarne 25 che unite alle dieci di prima porta il numero complessivo a 35=C(7,3).

b) Una moltiplicazione prevede 10 fattori non nulli. Quanti sono i possibili allineamenti se due segni negativi non possono essere contigui? Senza segni negativi esisterebbero C(10,0)=1 allineamento:

$$+0 \quad + \quad 0 \quad + \quad 0 \quad + \quad 0 \quad + \quad 0 \quad + \quad 0 \quad + \quad 0 \quad + \quad 0 \quad + \quad 0 \quad + \quad 0$$

Se ci fosse un solo segno "-" potrebbe essere disposto in C(10,1) modi diversi. Se i segni "+" sono 8 i modi sono C(9,2) e così via fino a cinque segni "-" perché se fossero di più allora la loro contiguità sarebbe forzata. Quindi:

$$1+\sum_{i=1}^{5}\binom{10-i+1}{i} = \binom{10}{1}+\binom{9}{2}+ \ \dots \ +\binom{6}{5} = 144$$

c) Ipotizziamo che gli "n" oggetti siano rappresentati con degli asterischi e che l'evento elementare sia una n-tupla ottenuta collocando delle linee tra gli asterischi più due all'esterno. Per suddividere gli oggetti in "n" gruppi sono necessarie n+1 linee di demarcazione; due di queste però sono vincolate a rimanere all'esterno per cui sono libere di muoversi solo (n-1) linee che racchiudono -a coppie- un numero n_i di asterischi. Tale numero però non può essere superiore ad "n"; ne consegue che le entità libere di muoversi sono le (n-1) linee di demarcazione più le N scelte degli asterischi. Fra queste dobbiamo scegliere senza reimmissione gli "n" posti in cui inserire gli oggetti e quindi:

$$\binom{N+n-1}{n} = \frac{N*(N+1)*(N+2)*\dots*(N+n-1)}{n!}$$

che è simile alle combinazioni senza ripetizione, solo che ora i fattori crescono e non diminuiscono in progressione aritmetica.

Nonostante l'apparente semplicità la formula non ha una derivazione facile (si vedano comunque Lombardo, 1984, pp. 524-527; Knuth, 1981, pp. 488-489; Feller 1950, p. 38).

Esempi:
a) Sia S={a_1, a_2, a_3} ed N=4. Esistono perciò 4+3-1=6 celle da riempire con un asterisco o con una barra. Le barre sono due (contano quelle interne dato che le esterne sono fisse) e gli asterischi quattro. Ecco alcune delle possibili combinazioni:

| * | | * * * | | * | * | * * | | | * * * * | |
$n_1=1,\ n_2=0,\ n_3=3$ $(a_1a_3a_3a_3)$ $n_1=1,n_2=1,n_3=2$ $(a_1a_2a_3a_3)$ $n_1=0,n_2=4,n_3=0$ $(a_2a_2a_3a_2)$

b) La *password* di una rete è formata da 5 caratteri scelti tra le 26 lettere dell'alfabeto e le dieci cifre arabe. Non c'è limite al numero di presenze di uno stesso carattere. L'ordine con cui i caratteri entrano nella chiave è però ininfluente. Qual'è il numero massimo di tentativi che un *hacher* dovrebbe fare per entrare nella rete?

$$\binom{36+5-1}{5}=\frac{36*37*38*39*40}{5!}=658'\,008$$

c) Un gruppo di n=5 amiche cerca posto al concerto. C'è una fila di N=12 posti. In quanti modi possono sedersi se vogliono stare vicine?

$$\binom{12+5-1}{5}=\frac{16!}{5!11!}=4368$$

Esercizio_TP68:
a) Lungo un tratto di strada sono state poste 5 stazioni di segnalazione di modo che se un'auto percorre l'intero tratto sarà conteggiata 5 volte. Supponendo che ne passino 10 quante registrazioni si possono verificare?
b) Un messaggio si compone di 200 caratteri scelti tra le 21 lettere dell'alfabeto italiano più lo spazio per separare le parole ed il punto. Un programma di decodifica quante possibilità dovrebbe considerare?
c) Se conoscete il calcolo differenziale questo esercizio proposto da Feller (1950, p. 39) è per voi. Quante derivate parziali di ordine "r" possiede una funzione di "n" variabili?

Partizioni con ripetizione
Supponiamo che la scelta avvenga con reimmissione e che pertanto siano possibili le ripetizioni della stessa biglia fermo restando la non rilevanza dell'ordine di estrazione. Con il ragionamento e simbologia adottati per la scelta senza reimmissione si ottiene:

$$\prod_{i=1}^{m}\binom{N_i+n_i-1}{n_i}=\binom{N_1+n_1-1}{n_1}\binom{N_2+n_2-1}{n_2}\cdots\binom{N_m+n_m-1}{n_m}$$

Esempio:
I donatori abituali di sangue sono: 5 "AB", 10 "B", 10 "A", 15 "0". Nel mese cinque di loro possono essere convocati più volte per una trasfusione. Quante sono le possibili scelte in cui "AB"=2, "A"=1, "O"=2, "B"=0?

$$\binom{5+2-1}{2}\binom{10+1-1}{1}\binom{15+2-1}{2}\binom{10+0-1}{0}=\frac{6!}{2!\,4!}10*1*\frac{16!}{2!14!}=18'\,000$$

Esercizio_TP69: *un gruppo parlamentare ha tre correnti: 8 "storici", 9 "riformisti" e 6 "liberal". Il gruppo deve designare un comitato di 9 saggi tra cui individuare i titolari delle varie cariche. Calcolate le scelte possibili di una commissione paritaria (3,3,3) nell'ipotesi che i nominativi siano scelti: a) Senza reimmissione; b) Con reimmissione.*

Posizioni vincolate
Alcuni problemi di calcolo combinatorio richiedono una maggiore attenzione. In particolare, le permutazioni (e conseguentemente le disposizioni) su di un ordinamento circolare o con delle posizioni vincolate (Bose e Manvel,1984). La peculiarità di queste situazioni è che per il primo o i primi elementi della scelta non tutte le posizioni sono disponibili o non sempre esiste la "prima" posizione e/o alcune posizioni vanno saltate.

Esempi:
a) Le permutazioni di "n" oggetti disposti in circolo sono dette permutazioni circolari (Freund e Walpole, 1980, p. 7) ed il loro numero è (n-1)! Si consideri un tavolo circolare con otto sedie. La 1ª persona può occupare solo una posizione in quanto il cerchio non ne ha una di riferimento. La 2ª, usando come riferimento l'altra, può occupare 7 posizioni (una è già occupata); la 3ª, in riferimento agli altri due, ne potrà occupare 7*6 e così via per un totale di 7!=5'040 permutazioni.

b) Determiniamo il numero di allocazioni di N oggetti in "n" gruppi in cui nessuna rimane vuoto. Questo significa che gli N oggetti lasciano (N-1) spazi tra i quali debbono trovare posto (n-1) linee (l'n-esima è preclusa per il fatto che due linee non possono essere adiacenti). Quindi, le possibilità sono:

$$\binom{N-1}{n-1}=\frac{(N-1)!}{(n-1)!(N-n)!}=\frac{n}{N}\binom{N}{n}$$

c) L'estrazione del biglietto vincente di una lotteria avviene scegliendo -senza reimmissione- le cifre inserite in dieci bussolotti contenuti in un'urna. Il numero si compone di cinque cifre disposte nell'ordine di estrazione. Ciccillo ha un biglietto che inizia con il "2" che è appena stato estratto. Che possibilità ha di vincere? Poiché una posizione è già impegnata occorre considerare le rimanenti quattro. Su queste possono ruotare in modo ordinato i numeri da tre a nove e quindi le possibilità sono $D_{SR}(7,4)=840$.

d) Le permutazioni senza ripetizione delle lettere A-B-C-D-E-F sono numerate progressivamente a partire da 1 secondo l'ordine alfabetico. Quale permutazione si troverà in posizione 314ª? Con l'A in 1ª posizione ne esistono (6-1)!=5!=120; con il B in 1ª posizione altre 120 e si arriva a 240. Con il C in 1ª l'A in 2ª ne esistono (6-2)!=4!=24 ed altre 24 sono quelle con il C in 1ª ed il B in 2ª e così via fino al C in 1ª ed D in 2ª che ci porta alla permutazione n. 312: CDABEF Quindi, si ha: n. 313=CDABFE, 314=CDAEBF

e) I numeri di Stirling del secondo tipo: S(N,n), tra gli altri usi, indicano il numero di modi alternativi di collocare N biglie in "n" urne nessuna delle quali deve rimanere vuota. Riordan (1958, p.33) ottiene le relazione ricorsiva seguente: S(N+1,n)=S(N,n-1)+nS(N,n) con S(N,1)=N e S(N,N)=1. Poiché S(1,1)=1 si ha S(3,2)=S(2,1)+2S(2,2)=2+2=4; S(4,2)=S(3,2)+2S(3,3)=4+2=6.

Esercizio_TP70:
a) In quanti modi possono essere disposti i numeri da 0 a 36 sui 37 tasselli di una roulette? Come si modifica la risposta aggiungendo lo"00"?
b) In quanti modi quattro coppie possono sedersi ad un tavolo circolare alternando persone di sesso diverso?
c) Durante una selezione di personale ci si è accorti che tra i compiti consegnati dai candidati in una fila di 13 ben 4 risultavano copiati. Si ritiene che la "copia" sia passata per persone sedute in posti adiacenti ci si domanda quante possibilità ci siano.
d) La rosa di una squadra di calcio è formata da 22 giocatori. Tutti tranne i tre portieri possono giocare in qualsiasi ruolo. Quante formazioni sono possibili?
e) Il controllo delle acque di una sorgente può avvenire con prelievi in 16 punti strategici tra i quali se ne scelgono 8 distinti per ogni controllo. Tuttavia, 2 punti debbono essere forzatamente inseriti tra gli 8. Quanti sono i possibili controlli?

Se è difficoltoso contare le entità che godono di una proprietà si può ragionare contando le entità che non godono della proprietà per poi sottrarle dal totale (un segnale importante in questa direzione è la presenza dell'avverbio "almeno" o "alpiù" nella formulazione del problema).

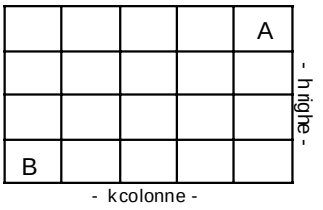
Esempi:
a) Quante parole di cinque lettere contenenti almeno una vocale si possono formare con l'alfabeto italiano? In questo caso la risposta è 21^5 permutazioni con ripetizione in totale meno 16^5 permutazioni con ripetizione che NON contengono alcuna vocale: 3,035,525.

b) vediamo in quanti modi si possono scegliere -senza ripetizione- i numeri {1, 2,…,9} facendo però in modo da avere almeno due numeri consecutivi cioè (4,5,6) oppure (2,3,6), ma non (1,3,5). Immaginiamo le 9 possibili scelte come una sequenza binaria con un bit per ogni numero che si pone nello stato "1" se il numero è stato scelto ed è "0" altrimenti: le terne prima citate darebbero le configurazioni: (0001110000), (0110010000), (1010100000). Scriviamo i sei "0" della stringa alternati con delle "x":x0x0x0x0x0x0. Scegliere i tre numeri significa sostituire tre "x" con tre "1" e cancellare le restanti "x". Tale scelta può essere fatta in c(7,3)=35 modi diversi per cui le scelte con almeno due numeri consecutivi sono: $D_{SR}(9,3) - 35 = 469$.

c) Un mazzo di carte francesi è stato diviso per i quattro semi ed ogni mazzetto adeguatamente mischiato. Da ogni mazzetto si sceglie a caso una carta. Quante possibilità ci sono che non sia un asso? Esistono 13^4 alternative di scelta per la quaterna di carte. Quelle che non contengono nessun asso sono 12^4 e quindi le possibilità sono $(13^4 - 12^4) = 7825$.

Esercizio_TP71:
a) In quanti modi si possono porre in fila 6 ragazze e 2 ragazzi se i ragazzi non si debbono mai trovare vicini?
b) In quanti modi si possono scegliere quattro cifre tra "0" e "9" fermo restando che nessuna quaterna sia formata da numeri consecutivi? c) In quante permutazioni dell'alfabeto italiano non compaiono mai "a" e "t"?
d) Quante coppie di carte del mazzo italiano contengono il sette ed una figura?

Esercizio_TP72: *David e Barton (1962, p.17) suggeriscono il seguente problema: la principessa è in posizione "A" e deve raggiungere il suo principe in "B". Può però muoversi -di uno o più passi- solo come una torre: in verticale o in orizzontale, ma non in diagonale. Quanti percorsi esistono?*



Esercizio_TP73: *4 coppie hanno 8 biglietti contigui per lo stadio. In quanti diversi modi si debbono sedere se:*
a) I coniugi debbono rimanere seduti accanto; b) Ogni marito deve avere sulla destra la propria moglie;
c) I mariti siedono su di un lato e le mogli all'altro estremo;
d) Se non debbono sedere vicino persone dello stesso sesso.

Sequenze di inclusioni/esclusioni

Gli eventi elementari di un esperimento sono talvolta costruiti considerando in successione la verifica (o la non verifica) di alcuni criteri o proprietà.

Esempi:

a) In un incubatore di imprese sono presenti N=76 ditte di cui 20 esportano nel mercato extracomunitario, 16 in quello comunitario e 18 in entrambi. Quante ditte non esportano in nessuno dei due mercati? Indichiamo con n(A) il numero di elementi che verificano la proprietà A. Dal totale delle N aziende dobbiamo sottrarre quelle che esportano nel mercato comunitario MC e quelle che esportano nel mercato extracomunitario (MEC). Così operando però si sono sottratte due volte le 18 aziende che esportano in entrambi i mercati: una volta come MC ed una volta come MEC. Per ripristinare la correttezza del conteggio dobbiamo sommare la numerosità dell'intersezione:

$$n\left(\overline{MC}\cap\overline{MEC}\right)=N-n(MC)-n(MEC)+n(MC\cap MEC)=76-20-16+18=54$$

b) Una società di consulenza ha 120 clienti importanti: 40 operano nel ramo finanziario, 30 in quello industriale e 20 in entrambi; inoltre, 10 sono attivi nelle comunicazioni, 3 nelle comunicazioni e nel ramo finanziario, 7 nelle comunicazioni e nel ramo industriale; solo 5 clienti operano simultaneamente nei tre settori. Quanti clienti importanti sono esclusi dai tre settori? Indichiamo con n(F), n(I) ed n(C) il numero di clienti attivi in ognuno dei tre rami; quelli che non vi operano sarebbero: $n(F^c\cap I^c\cap C^c)=n(S)-[n(F)+n(I)+n(C)]$ con $S=F\cup I\cup C$. A questi bisogna aggiungere chi opera in due rami perché defalcati due volte: $n(F^c\cap I^c\cap C^c)=n(S)-[n(F)+n(I)+n(C)]+n(F\cap C)+n(F\cap I)+n(I\cap C)$; la correzione però è andata oltre perché ha riportato per intero i clienti che operano contemporaneamente nei tre rami ed occorre scorporarli dal conteggio. In definitiva:

$$n(F^c\cap I^c\cap C^c)= n(S) - n(F) - n(I) - n(C) + n(F\cap C) + n(F\cap I) + n(I\cap C) - n(I\cap C\cap F) =120-40-30-20+3+7+20-5=55$$

Cerchiamo ora una soluzione più generale adottando la simbologia adeguata. In particolare, definiamo:

$$S_0=N;\ S_1=\sum_{i=1}^{m}n(A_i);\ S_2=\sum_{i_1=1}^{m-1}\sum_{i_2=i_1+1}^{m}n\left(A_{i_1}\cap A_{i_2}\right);\ S_j=\sum_{i_1=1}^{m-j+1}\sum_{i_2=i_1+1}^{m-j}\cdots\sum_{i_j=i_{j-1}+1}^{m}n\left(A_{i_1}\cap A_{i_2}\cap\ldots\cap A_{i_j}\right);\ S_n=n\left(\bigcap_{i=1}^{m}A_i\right)$$

La formula per calcolare le cardinalità di B_k={Si verificano insieme "k" proprietà $A_1, A_2, \ldots, A_m$} è:

$$n(B_k)=\sum_{i=k}^{m}(-1)^{i-k}\binom{i}{k}S_i$$

Esempi:

a) Nel caso della società di consulenza le proprietà sono m=3 e riguardano N=120 soggetti. L'evento che interessa è B_0 cioè si vogliono considerare gli eventi elementari (i clienti) che non sono attivi in nessuno dei tre rami: finanza, industria, comunicazioni. Quindi:

$$n(B_0)=\sum_{i=0}^{3}(-1)^iS_i=S_0-S_1+S_2-S_3=120-(40+30+20)\ +\ (3+7+20)\ -5=55$$

b) Il problema degli abbinamenti. Consideriamo una successione di N interi consecutivi da 1 ad N. Occorre conteggiare il numero di permutazioni, tra le N! possibili, in cui si verificano uno o più abbinamenti e cioè sequenze in cui il numero "i" si collochi nella posizione i-esima, il numero "j" nella posizione j-esima, il "k" nella k-esima e così via. Per chiarire: 321 ha un solo abbinamento dato che il "2" occupa la 2ª posizione laddove "3" e "1" sono fuori posto; 312 non ne ha nessuno e 123 ne ha tre. L'abbinamento in i-esima posizione può avvenire ponendo il numero "i" in nella posizione che gli corrisponde e lasciando le altre libere di variare per cui ci sono (N-i)! possibilità. Poiché siamo interessati al numero complessivo di abbinamenti e non su quale posizione avvengono dobbiamo considerare sullo stesso piano tutte le combinazioni degli N interi prese a gruppi di "i": C(N,i). Ne consegue che le cardinalità delle combinazioni delle proprietà A_i={posizione i-esima occupata dal numero "i") sono date da:

$$\binom{N}{j}(N-j)!=\frac{N!(N-j)!}{J!(N-j)!}=\frac{N!}{j!}$$

cosicché la cardinalità dell'evento "nessun abbinamento" è pari a:

$$n(B_0)=N!-\frac{N!}{1!}+\frac{N!}{2!}-\frac{N!}{3!}+\ldots+(-1)^m\frac{N!}{m!}=N!\left[1-\frac{1}{1!}+\frac{1}{2!}-\frac{1!}{3!}+\ldots+(-1)^m\frac{1}{m!}\right]$$

L'espressione in parentesi, all'aumentare di "m", converge rapidamente al numero $e^{-1}\cong 0.36788$ per cui la cardinalità cercata è ben approssimata da [N*0.36788]. In particolare, se disponiamo in fila le m=10 carte di un seme del mazzo napoletano, tra le 3.6 milioni di permutazioni ve ne sono poco più di 1.3 milioni prive di ogni abbinamento. Inoltre si ha:

$$n(almeno\ \ un\ \ abbinamento)=N!-n(B_0)=N!\left[1-\frac{1}{2!}+\frac{1!}{3!}+\ldots+(-1)^{m+1}\frac{1!}{m!}\right];$$

Esercizio_TP74: *un'imprenditrice ha un numero di telefono è formato da 11 cifre che è lo stesso numero di cifre della partita IVA. a) In quanti casi si ha almeno un abbinamento? b) In quanti casi se ne trovano 11?*

6.3.3 Applicazioni dello schema di equiprobabilità

Le tecniche di conteggio discusse nel precedente paragrafo hanno molte applicazioni nel calcolo delle probabilità soprattutto se affiancate dal modello di probabilità uniforme. Sia S l'universo degli eventi finito e discreto e sia E un evento compreso in W, l'algebra di S. Secondo la concezione classica la probabilità di E è data da:

$$P(E) = \frac{card(E)}{card(S)}$$

Esempi:

a) Un test si compone di 8 domande di cui 6 di teoria e 2 di applicazioni. Per superare il test è necessario rispondere correttamente a 6 domande su otto di cui un minimo 4 di tipo teorico. Se si risponde a caso qual'è la probabilità di superare il test?

$$\frac{\binom{6}{4}\binom{2}{2}+\binom{6}{5}\binom{2}{1}+\binom{6}{6}\binom{0}{0}}{\binom{8}{6}} = \frac{15+12+1}{168} = \frac{1}{6}$$

b) Se si uniscono tutte le coppie di vertici di un poligono di "n" lati, qual'è la probabilità che, scelti a caso due vertici, essi formino una diagonale? Le combinazioni di vertici presi a due a due sono C(n,2). Da queste si devono escludere quelle che formano dei lati cioè "n" e quindi P(diagonale)=[C(n,2)-n)]/C(n,2)= 1-2/(n-1)

c) Ad una gara d'appalto hanno partecipato "n" ditte. La commissione procede per confronti a coppie: le ditte sono abbinate casualmente: l'offerta migliore tra le due passa alla fase successiva e la peggiore è scartata (se "n" è dispari una delle ditte scelta a caso passa direttamente al turno seguente). Le ditte A e B hanno concertato le offerte in modo che se si incontrano una delle due guadagna automaticamente un passaggio. Qual'è la probabilità che le due ditte siano abbinate in uno dei turni di confronto? Il numero dei possibili abbinamenti è C(n,2). I turni di accoppiamenti sono "m" dove m=min{m|2^m>n) e le due ditte possono accoppiarsi in uno qualsiasi dei turni con la stessa probabilità cosicché:

$$P(A \ \ confronta \ \ B) = \frac{m}{c(n,2)} = \frac{2m}{n(n-1)}$$

d) Il famoso caso del Cavalier De Méré. Quale evento è più probabile: ottenere almeno un "1" nel lancio di 4 dadi oppure un doppio "1" in 24 lanci di 2 dadi? Nel primo esperimento i casi possibili sono 6^4=1'296. Per determinare i casi favorevoli definiamo prima l'evento A_i={esce un "1" in "i" dadi e non negli altri} la cui cardinalità è:

$$card(A_i) = \binom{4}{i}5^i; \quad i = 0,1,2,3$$

L'evento di interesse è E=(A_1UA_2UA_3UA_4) che ha probabilità:

$$P(E) = \frac{500+150+20+1}{6^4} = \frac{671}{1296} = 0.5177$$

Nel secondo esperimento i casi possibili sono 36^{24}. Qui conviene conteggiare i casi sfavorevoli cioè le serie di 24 lanci di due dadi in cui non si verifica un doppio "1" che ha cardinalità 35^{24} poiché in ogni prova abbiamo escluso l'esito (1,1). Possiamo perciò concludere che se A_i={escono due "1" nel lancio i-esimo} la probabilità di E=(A_1UA_2U...UA_{24}) è:

$$P(E) = 1 - P(\overline{E}) = 1 - \left(\frac{35}{36}\right)^{24} = 0.4914$$

che è leggermente più bassa della prima.

e) Le foto di "n" personaggi dello sport, dello spettacolo e della politica sono state abbinate ad altrettante loro foto da neonati. Il quiz consiste nell'abbinarli correttamente. A questo fine ci si regola soprattutto con la conformazione degli occhi che è la parte del volto meno soggetta a cambiamenti. Se si rispondesse a caso, la probabilità di almeno un abbinamento sarebbe:

$$P(almeno \ \ un \ \ abbinamento) = \left[1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots + (-1)^{n+1}\frac{1}{n!}\right]$$

che tende ad attestarsi, con oscillazioni smorzate, al valore di 0.36788.

f) In una città operano 4 medici specialisti di malattie infettive: Rossi, Bianchi, Verdi, Neri. Durante un'epidemia 12 ammalati cercano uno specialista. Se ognuno di loro scegliesse a caso da una guida telefonica, qual'è la probabilità che i pazienti si distribuiscano equamente tra i medici? Indichiamo con A_j la scelta del paziente cioè $A_j \varepsilon$ {R, B, V, N}; l'evento elementare di questo esperimento è la dozzina delle scelte {A_1, A_2, … ,A_{12}} in cui ognuna ha quattro possibili modalità. I casi possibili sono pertanto 4^{12} laddove i casi favorevoli sono dati dal numero di partizioni di 12 elementi in quattro gruppi di 3. La probabilità cercata è:

$$\frac{\binom{12}{3,3,3,3}}{4^{12}} = \frac{12!}{(3!)^4 4^{12}} = 0.022$$

Quindi è molto probabile (98%) che almeno uno dei medici abbia più pazienti della media e qualcunaltro ne abbia di meno.

g) Le lettere che formano la parola ABRACADABRA sono state mischiate e riprese una ad una. Qual'è la probabilità che le cinque "A" si ritrovino vicine? Le posizioni sono 11. Se le prime cinque sono occupate dalle "A" ne rimangono libere sei (il posizionamento contiguo delle "A" può avvenire in 7 modi diversi). Le 6 posizioni libere possono essere occupate da due "B", due "R", una "C" e una "D". Il numero di permutazioni complessivo sarebbe 11! ma è chiaro che il loro numero è minore dato che alcune scelte sono identiche perché frutto di lettere uguali scambiate di posto. Qualcosa di analogo accade le permutazioni delle consonanti nelle 6 posizioni non vincolate. In definitiva il rapporto tra casi favorevoli e possibili è:

$$P("AAAAA") = \frac{7*\binom{6}{2,2,1,1}}{\binom{11}{5,2,2,1,1}} = 0.015$$

h) Un gruppo di 2N persone è in fila per acquistare il biglietto del teatro che costa 5 euri. Metà delle persone ha in tasca solo un biglietto da 5 euri e l'altra metà solo un biglietto da 10 euri. All'apertura il botteghino non ha un fondo cassa per cui se il primo cliente ha un biglietto di 10 euri non è in grado di dare il resto ed il cliente sarebbe costretto ad aspettare. Qual'è la probabilità che si formi una fila in cui nessuno dei clienti sia costretto ad aspettare? Si tratta di una partizione in due gruppi: di tipo "5" con N elementi e di tipo "10" con N elementi per cui i casi possibili sono P(N,N). I casi sfavorevoli sono le partizioni in due gruppi in cui nelle prime (N-1) posizioni siano presenti "i" di tipo "5" e nelle restanti (N+1) posizioni si trovino (N+1- i) di tipo "10" e questo per ogni scelta di "i" tra zero e (N+1). I casi sfavorevoli sono perciò:

$$\sum_{i=0}^{N+1} C(N-1,i)C(N+1,N+1-i) = C(2N,N+1) = \binom{2N}{N,N}$$

Quindi, la probabilità di una coda senza interruzioni è:

$$\frac{\binom{2N}{N,N} - \binom{2N}{N+1,N-1}}{\binom{2N}{N,N}} = \frac{\frac{(2N)!}{N!N!} - \frac{(2N)!}{(N+1)!(N-1)!}}{\frac{(2N)!}{N!N!}} = 1 - \frac{N!N!}{(N+1)!(N-1)!} = \frac{1}{N+1}$$

che tende a zero all'aumentare del numero delle persone in coda ossia l'evento "nessun cliente aspetta" tende a sovrapporsi all'evento impossibile per N crescente.

Esercizio_TP75:

a) Questo problema coinvolse anche I. Newton la cui soluzione non risultò convincente. Due scommettitori sono all'opera: A vince se lanciando sei dadi (o per sei volte un dado) ottiene almeno un "1"; B vince se lanciando dodici dadi realizza un doppio "1". Qual'è la vera tra le seguenti: 1) P(A)=P(B), P(A)<P(B), P(A)>P(B)? (Sugg. Lavorate sulla probabilità degli eventi complementari).

b) Galileo e il Granduca di Toscana. Su quale evento ritenete sia razionale scommettere nel lancio di tre dadi: il 10 o il 9?

c) Un gruppo di 8 giovani si reca in discoteca. Nel prezzo di ingresso è prevista una consumazione da scegliere tra le 8 disponibili ed ognuno ne sceglie una diversa. Tra le disponibilità c'è un cocktail a base di alcool. Anna e Marco debbono guidare per cui tale opzione è sgradita. Se chi serve al banco desse a caso le consumazioni qual'è la probabilità che Anna o Marco ricevano il cocktail?

d) Una scimmia digita a caso i tasti di una tastiera con 22 simboli: le lettere dell'alfabeto e lo spazio separatore. Ogni minuto batte 50 tasti. Un milione di sue compagne la imita. Quanto tempo è necessario perché si abbia probabilità uno che si formi la frase: GLI UMANI SONO PRIMATI POCO EVOLUTI RISPETTO A NOI.

Esercizio_TP76:

a)N biglie debbono essere inserite in "n" buche. Qual'è la probabilità che esattamente una buca rimanga vuota? N. B. Sono possibili le ripetizioni cioè ogni buca può ricevere più di una biglia; inoltre, l'ordine non conta.

b) Si scommette sugli arrivi ad un gran premio automobilistico. Se i concorrenti sono 8 qual'è la probabilità di indovinare, nell'ordine i primi 4? Se riuscite a sapere che le condizioni pilota/mezzo sono tali che uno dei concorrenti arriverà certamente ultimo, quali sono ora le probabilità?

c) L'identificazione di un prodotto è formata da 5 lettere distinte scelte con equiprobabilità tra le 16 consonanti dell'alfabeto italiano. Il software che genera il codice ha avuto un guasto e non è più in grado di distinguere tra lettere uguali. Qual'è la probabilità che l'esito del programma formi un codice?

d) Una commissione formata da 3 conservatori, 3 progressisti ed un ambientalista deve designare una delegazione per una delegazione di tre membri. Qual'è la probabilità che le tre parti politiche vi siano rappresentate se i membri sono scelti a caso?

Schema ipergeometrico

Un particolare tipo di partizione è la divisione degli N oggetti in due gruppi di cui uno comprendente N_1 elementi è indicato come "speciale" perché i suoi elementi verificano una certa proprietà ed un altro di $(N\text{-}N_1)$ elementi "comuni" per i quali la proprietà non è soddisfatta. L'esperimento consiste nella scelta casuale di un numero fissato di "n" di elementi di cui n_1 speciali ed i restanti $(n\text{-}n_1)$ comuni. Qual'è la probabilità che la scelta -senza reimmissione- degli "n" elementi contenga n_1 elementi speciali?

La scelta di questi può avvenire in $C(N_1,n_1)$ modi diversi. Ognuno può abbinarsi con le combinazioni di $(N\text{-}N_1)$ elementi comuni presi a blocchi di $(n\text{-}n_1)$ e quindi i casi favorevoli, grazie alla moltiplicazione combinatoria, sono: $C(N_1,n_1)*C(N\text{-}N_1,n\text{-}n_1)$ con C(N,n) casi possibili. Ne consegue:

$$P(n_1)=\frac{\binom{N_1}{n_1}\binom{N-N_1}{n-n_1}}{\binom{N}{n}};\ n_1=1,2,\ \ldots,\ n$$

Esempi:

a) Una lotteria ha venduto k^2 biglietti e ha messo in palio "k" premi. Un gruppo di scommettitori decide di comprare "k" biglietti: qual'è la probabilità di vincere almeno uno dei premi? Nell'ambito dello schema ipergeometrico i biglietti vincitori diventano le unità speciali scelte senza reimmissione dall'insieme delle unità. La probabilità che si sta cercando è allora:

$$P(0)=\frac{\binom{k}{0}\binom{k(k-1)}{k-0}}{\binom{k^2}{k}}=\frac{\frac{[k(k-1)]!}{k!(k^2-2k)!}}{\frac{(k^2)!}{k![k(k-1)]!}}=\frac{\{[k(k-1)]!\}^2}{(k^2)!(k^2-2k)!}$$

b) Lo *staff* è composto da 10 dirigenti. La presidente firma 5 lettere di promozione di colore verde e 5 censure gravi di colore rosso. Le missive sono affidate ad un segretario con preghiera di inserirle in buste dello stesso colore. Il segretario, daltonico e incosciente, imbusta distrattamente le lettera con il rischio di errori e confusione. Determiniamo la probabilità che esattamente "x" lettere siano imbustate con lo stesso colore. Poiché gli abbinamenti vanno a due a due (se una lettera di colore verde finisce in una busta rossa, una lettera rossa sarà finita in una busta verde) si crea lo schema ipergeometrico:

$$P(X=x)=\frac{\binom{5}{k}\binom{5}{5-k}}{\binom{10}{5}},\ k=\frac{x}{2},\ x=0,2,4,6,8,10$$

c) E' in corso il gioco delle coppie. I nomi di 6 ragazzi e di 6 ragazze sono scritti su dei bigliettini ben piegati e riposti in un cappello. Dopo una energica mescolatura si scelgono a caso 4 biglietti ed i nomi di coloro che sono estratti dovranno organizzarsi un coppie, anche di membri dello stesso genere.

1. Qual'è la probabilità che siano scelti due ragazze e due ragazzi? 2. Qual'è la probabilità che siano scelte più ragazze che ragazzi?

1. $P(D=2,U=2)=\frac{\binom{6}{2}\binom{6}{2}}{\binom{12}{4}}=0.4546$; 2. $P(D>U)=P(D=3,U=1)+P(D=4,U=0)=\frac{\binom{6}{3}\binom{6}{1}}{\binom{12}{4}}+\frac{\binom{6}{4}\binom{6}{0}}{\binom{12}{4}}=02727$

Esercizio_TP77:

a) Nel consiglio direttivo di un consorzio intercomunale (cui aderisce anche Roccasecca) ognuno dei 25 comuni nomina due propri rappresentanti cosicché il consiglio generale è costituito da 50 membri. In questo si deve formare per scelta casuale e senza reimmissione un comitato di 25 persone.

1) Calcolare la probabilità che Roccasecca vi sia rappresentato; 2) Calcolare la probabilità che tutti i 25 comuni vi siano rappresentati.

b) Qual'è il numero più probabile di carte di denari in una mano di dieci carte del mazzo napoletano?

c) Un dado regolare è lanciato per 3 volte. Si ignora l'esito, ma da indiscrezioni si apprende che le facce sono tutte diverse. Qual'è la probabilità che sia uscito il "6"?

d) Un improvvisato archeologo propone alla casa d'aste di bandire una raccolta di 18 rarissime monete brezie (popolo preromanico della Calabria citra). In realtà le monete autentiche sono 9 e solo buone imitazioni le altre. Il banditore ne può controllare solo 4. Qual'è la probabilità che le quattro esaminate siano tutte buone? Qual'è la probabilità che due siano false?

Tentativi ripetuti

Feller (1968, pp. 47-50) descrive due schemi di tentativi ripetuti che possono rivelarsi la chiave interpretativa di molti problemi. Gli elementi base sono ancora urne e biglie.

Nel primo schema le biglie sono collocate casualmente in N urne finché non si tenti di inserire una biglia in un'urna già occupata; a questo punto l'esperimento si interrompe. Il numero delle biglie è indeterminato, ma non può essere inferiore a 2 (solo dalla 2ª in poi è possibile un duplicato) e non può essere superiore ad (N+1) perché a questo punto il duplicato è sicuro. Qual'è la probabilità che si collochino "m" biglie prima di provocare una duplicazione? I casi possibili sono N^m dato che, per ogni biglia, sono possibili N scelte. Se il doppione si verifica all'm-esima biglia vuol dire che ci sono (m-1) urne che contengono già la biglia; le scelte di tali urne sono disposizioni senza ripetizioni: $D_{SR}(N,m-1)$. Peraltro, alla m-esima collocazione, ciascuna delle (m-1) urne già impegnate può essere la candidata a generare il doppione cosicché i casi favorevoli sono: $(m-1)D_{SR}(N,m-1)$ e la probabilità cercata è perciò:

$$P(\textit{successo alla } m^a \textit{ prova}) = q_m = \frac{(m-1)D_{SR}(N,m-1)}{N^m} = \frac{N(N-1)(N-2)\ldots(N-m+2)(m-1)}{N^m}$$
$$= \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\ldots\left(1-\frac{m-2}{N}\right)\left(\frac{m-1}{N}\right)$$

in cui si pone convenzionalmente $q_1=0$.

Esempio:
Il problema del compleanno. In un'aula ci sono N studenti sotto esame. I colloqui continuano finché non si trovi uno studente nato nello stesso giorno e mese di uno di quelli già chiamati. Se "m" è il numero di quelli interrogati, che valore raggiungerà?

N	$1-q_N$	N	$1-q_N$
5	0.0271	35	0.8144
10	0.1169	40	0.8912
15	0.2529	45	0.9410
20	0.4114	50	0.9704
25	0.5687	55	0.9863
30	0.7063	60	0.9941

Ipotizziamo l'anno di 365 giorni e consideriamo queste le nostre urne: N=365; gli studenti sono le biglie. In tabella è dato il riassunto dei calcoli per vari valori di "m". La probabilità di dover interrompere i colloqui è già molto alta per N=30 ed è quasi certezza dopo N=60

Esercizio_TP78: *un solitario consiste nel disporre in linea le 13 carte di un seme. Dalle restanti carte, adeguatamente mischiate, si scelgono ad una ad una delle carte coprendo con queste le corrispondenti scoperte. Il gioco si interrompe non appena si presenta una delle carte già coperte. Qual'è la probabilità di riuscire a coprire tutte le 13 carte?*

Esercizio_TP79: *Filomena si reca ad un party a cui partecipano anche N ragazzi. La giovane donna è convinta che se incontra un ragazzo del suo segno (vergine) la serata sarà piacevole. Quanti ragazzi debbono essere presenti affinché la probabilità che Filomena debba ballare con tre partner prima di incontrare il suo cavaliere ideale sia superiore al 95%?*

Nel secondo schema si collocano, una alla volta, biglie nelle varie urne (che ora possono contenere più di una biglia) finché un'urna prefissata, diciamo la 1ª, rimane vuota. L'universo degli eventi non è finito perché non vi è ragione di attendersi che una biglia finisca certamente nella 1ª urna. Se l'interruzione avviene alla m-esima biglia i casi possibili sono N^m. Per determinare i casi favorevoli si tiene conto che, per le precedenti (m-1) biglie, le urne disponibili erano (N-1) poiché la 1ª era preclusa a pena dell'interruzione dell'esperimento:

$$q_m^* = \frac{(N-1)^{m-1}}{N^m} = \left(\frac{N-1}{N}\right)^{m-1}\left(\frac{1}{N}\right) = \left(1-\frac{1}{N}\right)^{m-1}\left(\frac{1}{N}\right)$$

Inoltre, la probabilità che l'esperimento richieda più di "m" prove è:$[1-(1/n)]^m$ per m=1,2,…,

Esempi:
a) Il prof. Paletta, geniale, ma distratto collega gira con un mazzo di N=12 chiavi tutte dello stesso tipo; non solo, ma essendo inanellate in un portachiavi a forma di cerchio, non c'è verso di ricordare quella che apre l'ufficio ed ogni volta è un'impresa trovare quella giusta. Qual'è la probabilità che la ricerca termini alla terza chiave? Qual'è la probabilità che sia necessario provare più di 6 chiavi?

$$q_3^* = \left(1-\frac{1}{12}\right)^2 \frac{1}{12} = 0.07; \qquad p_6^* = \left(1-\frac{1}{12}\right)^6 = 0.5933$$

b) Il segreto di Pulcinella. In un villaggio di 100 abitanti una persona racconta in gran segreto un pettegolezzo ad un'altra persona che, a sua volta e sempre in gran segreto, lo racconta ad un'altra che prosegue allo stesso modo. Calcoliamo la probabilità che il segreto venga raccontato a tutti senza tornare al progenitore. La risposta è facile perché ricorre lo schema appena tracciato con l'urna bloccata corrispondente al progenitore e gli ascoltatori interpretati come biglie. Quindi la probabilità è:

$$q_{100}^* = \left(1-\frac{1}{100}\right)^{99} \frac{1}{100} = 0.0037;$$

Se invece di raccontare il segreto ad una sola persona per volta ad ogni incontro si forma un crocchio di "k" persone, la formula si modifica:

$$q_m^* = \left(1-\frac{k}{N}\right)^{m-1}\left(\frac{k}{N}\right) \qquad p_m^* = \left(1-\frac{k}{N}\right)^m$$

e per k=5 ed N=100 la probabilità che tutti ne vengano a conoscenza all'insaputa del progenitore scende a 0.0003.

c) La roulette russa. Due amiche discutono su chi debba uscire con il ragazzo che piace ad entrambe. Si affidano alla sorte formando un mazzo di 6 carte in cui c'è un solo asso. Le ragazze mischiano ogni volta il mazzo. La carta prescelta non è rimessa nel mazzo. La prima che trova l'asso vince. Liberata, ritiene di essere favorita se sceglie per prima. E' vero? Le carte sono equiprobabili e ricorre lo schema delle urne con posizione vincolata: la scoperta dell'asso alla i-esima carta è pertanto:

$$p_i = \left(1-\frac{1}{6}\right)^{i-1}\frac{1}{6};\ i = 1,2,\ldots,$$

$$P(Liberata\ \ scopre\ \ l'\ asso) = p_1 + p_3 + p_5\ +\ \ldots\ = \frac{1}{6}\left[1+\left(\frac{5}{6}\right)^2+\left(\frac{5}{6}\right)^4+\ldots+\right] = \frac{6}{11}$$

e quindi Liberata ha ragione.

d) In una sala dove sono in voga i balli di coppia è in programma lo scambio casuale del cavaliere. Osvalda si ritiene così sfortunata che quasi certamente a lei toccherà quello che ha già. Supponiamo che nella sala ci siano N coppie qual'è la probabilità che scambiando a caso i membri delle coppie ad Osvalda tocchi proprio suo marito? Identifichiamo le dame con un numero da 1 a N così pure i cavalieri. Il problema di Osvalda è allora un problema di abbinamento analogo a quello discusso nel paragrafo precedente. In particolare la probabilità che Osvalda debba ballare con il consorte è:

$$\frac{n(B_1)}{N!} = \frac{\sum_{i=1}^{N}(-1)^{i-1}\binom{i}{1}S_i}{N!} \quad con\ \ S_i = \binom{N}{i}(N-i)^N$$

in cui la S_i è ottenuta tenendo conto che nelle permutazioni in cui si sono vincolate una o più posizioni, sono possibili delle ripetizioni:

$$\frac{n(B_1)}{N!} = \sum_{i=1}^{N}\frac{(-1)^{i-1}}{(i-1)!} = \sum_{i=0}^{N-1}\frac{(-1)^i}{i!} \cong e^{-1} = 0.36788$$

L'approssimazione migliora con l'aumentare del numero delle coppie. E' sorprendente come la probabilità non muti di fatto dopo un N moderatamente piccolo. In definitiva, Osvalda non si deve meravigliare troppo se si ritrova a ballare con il marito.

Esercizio_TP80: *la catena di Sant'Antonio. Ciccillo decide di giocare con la posta elettronica ed invia un messaggio a due corrispondenti scelti a caso con richiesta di fare altrettanto pena l'installazione automatica di un virus mortale per il sistema. I due contattati da Ciccillo (la 1ª generazione) obbediscono così come i loro corrispondenti (2ª generazione) allungando sempre più la catena. Ipotizzate che la popolazione sia di (N+1) utenti e determinate la probabilità che il processo si replichi per 1,2,…,m generazioni senza che Ciccillo riceva indietro il suo messaggio.*

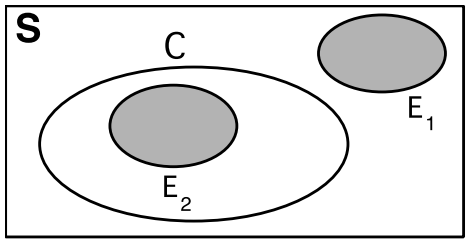
Esercizio_TP81: *si consideri l'esperimento consistente nella allocazione casuale (con probabilità uniforme) delle biglie in "n" urne. E'possibile che ogni urna riceva più di una biglia. Le biglie continuano ad essere inserite fintanto che la prima urna non arrivi a contenere esattamente "k" biglie. Con quale probabilità questo succede alla r-esima biglia con (r≥k)?*

6.4 L'indipendenza

La probabilità è un concetto complesso e difficile se considerato ad una scala sufficientemente piccola per percepirne i dettagli. Fra le sue tante articolazioni abbiamo visto il legame con il quadro di evidenze empiriche-teoriche-percettive con cui si affronta un problema. La natura di questa relazione non è per ora in questione, siamo piuttosto alla ricerca di uno schema per valutare le modifiche indotte nell'universo degli eventi da una informazione aggiuntiva che pervenga su uno o più degli eventi elementari. Questo ci porta al problema della causa più probabile ed al concetto di indipendenza tra due eventi che è centrale nel calcolo delle probabilità.

6.4.1 Probabilità condizionata

In uno spazio di probabilità isoliamo un evento $C \subset W$ che giudichiamo di un qualche interesse per il nostro esperimento (pertanto: $P(C)>0$) e vediamo come modificare lo spazio di probabilità nell'ipotesi che C si verifichi.



Che ciò sia possibile è evidente dato che se si è interessati ad E_1 si ha: $C \cap E_1 = \varnothing \rightarrow P(C \cap E_1)=0$; se invece l'evento di interesse è E_2 allora il fatto che $C \supset E_2 \rightarrow P(E_2)=1$. Per comodità manteniamo lo stesso universo S anche se qualche suo evento potrebbe essersi tramutato in un evento impossibile alla luce di ciò che si è verificato; (dal rettangolo S si è passati all'ellisse C e l'evento E_1 non può più verificarsi). Lo stesso vale per l'algebra W anche se in si opera solo con la classe di eventi compatibili con il verificarsi di C e cioè $W \cap C$.

Esempio:
Supponiamo che le facce di un dado siano equiprobabili e che il dado debba essere lanciato su di una superficie piana e rigida. La funzione di probabilità è :

E	1	2	3	4	5	6	
$p(E)$	1/6	1/6	1/6	1/6	1/6	1/6	1

E	1	2	3	4	5	6	
$p(E)$	1/6	0	1/6	0	1/6	0	3/6

Limitiamoci all'evento C= "esce un dispari". Ne consegue che alcuni eventi sono ancora possibili, altri no. Poiché non è alterata l'equiprobabilità si ha la tabella a destra. Le nuove probabilità debbono essere riscalate per sommare ad uno (probabilità dell'evento certo)

E	1	2	3	4	5	6	
$p(E)$	$\frac{1/6}{3/6}$	0	$\frac{1/6}{3/6}$	0	$\frac{1/6}{3/6}$	0	$\frac{3/6}{3/6}$

=

E	1	2	3	4	5	6	
$p(E)$	1/3	0	1/3	0	1/3	0	1

Il riscalamento delle probabilità è un fatto ovvio. Poiché la massa di probabilità da distribuire tra gli eventi non è più l'unità, ma P(C), le vecchie probabilità devono essere aggiornate dividendole per P(C). Inoltre, poiché l'evento certo adesso è C e non più S, è giusto che, nella nuova funzione di probabilità, a questo tocchi l'unità. Rimane da spiegare il fattore di riparto. L'incidenza del possibile verificarsi di C su di un qualsiasi altro evento E non può che essere misurata dalle parti che i due eventi hanno in comune perché ora si può verificare solo ciò che è in $E \cap C$. Il numeratore della nuova probabilità è: $P(E \cap C)$. Infatti, nel lancio del dado: P("1"$\cap$ Dispari)=1/6, P("2"$\cap$ Dispari)=0/6. Questa è una procedura intuitiva, ma del tutto generale: alla luce della restrizione C le probabilità vanno riscritte con la formula:

$$P(E \subset W|C) = \frac{P(E \cap C)}{P(C)}; \quad con \;\; P(C) > 0$$

La probabilità di E sotto la condizione C (è questo ciò che indica il simbolo "|") è determinata dalla probabilità che i due eventi si presentino insieme (nello spazio di probabilità originario) rapportato alla probabilità assegnata (sempre nello spazio originario) all'evento condizionante. Per comodità espositiva abbiamo mantenuto lo stesso simbolo "P" per indicare la funzione di probabilità condizionata, ma è chiaro che, una volta riscalata, la funzione di probabilità non è più la stessa di quella originaria anche se a questa strettamente connessa.

Esempi:

a)I potenziali clienti di una *data warehousing* sono classificati secondo la disponibilità all'acquisto: alta, media, bassa e alla possibilità di acquisto: immediata, dilazionata, nulla. Il modello di probabilità che guida il *management* è dato in tabella.

Poss./Disp.	Alta	Media	Bassa	
Immediata	0.20	0.09	0.01	0.30
Dilazionata	0.30	0.15	0.05	0.50
Nulla	0.05	0.05	0.10	0.20
	0.55	0.29	0.16	1

Le probabilità interne sono dette congiunte e quelle sull'ultima riga o sull'ultima colonna sono le marginali. Supponiamo che, a giudicare da segni esteriori, un cliente sia classificato nella possibilità dilazionata, Qual'è la probabilità che abbia disponibilità alta?

$$p(alta|dilazionata) = \frac{p(alta \cap dilazionata)}{p(dilazionata)} = \frac{0.30}{0.50} = 0.60$$

b) Le frasi ambigue sono le trappole del ragionamento condizionale. L'evento che la figlia di operai frequenti l'università non è la stesso che una studentessa universitaria sia figlia di operai. Scelta casualmente una famiglia ci dobbiamo chiedere in che modo sapere che sia operaia incide sull'avere una figlia all'università; nell'altro caso, scelta una ragazza dobbiamo chiederci in che modo sapere che studia all'università influenzi l'aspettativa che provenga da una famiglia operaia.

c) Un classico (Falks, 1996). Si lanciano tre monete. Qual'è la probabilità che presentino la stessa faccia? Prima soluzione. I casi possibili sono 8: (CCC, CCT, CTC, TCC, TTC, CTT, TCT, TTT); i casi favorevoli sono 2 e quindi la probabilità cercata è 2/8=1/4. 2ª soluzione. Due monete sono sicuramente uguali; quindi il risultato è determinato dalla 3ª; questa può ricadere tanto come testa che come croce quindi la probabilità richiesta è 1/2. La seconda soluzione è sbagliata perché parte da una falsa premessa. La conoscenza dell'evento "almeno due monete uguali" non è rilevante dato che non modifica l'universo degli eventi originario. Se E={tre facce uguali} e F={almeno due facce uguali} allora P(E|F)=P(E∩F)/P(F)=P(E∩F)/1=P(E) dato che E è già incluso in F.

d) L'agente di viaggio ha ricevuto due fax di prenotazione poco leggibili, ma con destinazioni possibili solo per Cipro e Baleari. L'universo degli eventi è S={(C_1,C_2); (C_1,B_2); (B_1,C_2); (B_1,B_2)}. Si presuppone l'equiprobabilità. Leggendo meglio il primo fax si riesce a stabilire che era per Cipro, qual'è la probabilità che lo sia anche il secondo? E' diversa che per Baleari?

$$P(C_1 \cap C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)} = \frac{1/4}{1/2} = \frac{1}{2}; \;\; P(C_1 \cap B_2|C_1) = \frac{P(C_1 \cap B_2)}{P(C_1)} = \frac{1/4}{1/2} = \frac{1}{2}$$

Esercizio_TP82: *siano E, F ⊂ W con P(F)>0. Verificare che:*

a) $Se \;\; E \cap F = \varnothing \Rightarrow P(E|F) = 0;$ *b)* $Se \;\; E \subset F \Rightarrow P(E|F) \geq P(E);$

c) $Se \;\; E \subset F \Rightarrow P(F|E) = 1;$ *d)* $P(E|F) + P(E|\overline{F}) \neq 1;$ *e)* $P(E|F) + P(\overline{E}|\overline{F}) \neq 1$

Esercizio_TP83: *il successo V o l'insuccesso V^c di un programma di incentivi ai dipendenti da parte della compagnia "Alfa" dipende in gran parte dal fatto che il suo maggiore concorrente, la compagnia "Beta", cambi (C) o non cambi (C^c) la propria politica di incentivi. Alcune probabilità sono note.*

	V	V^c	
C	$P(C \cap V)$	$P(C \cap V^c)$	$P(C)$
C^c	$P(C^c \cap V)$	$P(C^c \cap V^c)$	$P(C^c)$
	$P(V)$	$P(V^c)$	1

$\Rightarrow$

	V	V^c	
C			
C^c	0.5		0.7
	0.7		1

1) Completare la seconda tabella;

2) Quale regola si applica nella seconda riga?

Esercizio_TP84: *un'inchiesta sul fumo tra ha prodotto le seguenti frequenze relative cioè probabilità di fatto.*

Sesso/atteg.	Fuma	Non fuma	Ha smesso	
M	0.10	0.35	0.05	0.50
F	0.15	0.25	0.10	0.50
	0.25	0.60	0.15	1.00

Calcolate la probabilità che, scelta a caso una persona in quella fascia d'età, si abbia:
a) P(M|NF); b) P(ha smesso|F).

La probabilità condizionata non introduce alcun concetto nuovo e non c'è bisogno di un assioma *ad hoc* per definirla. Di parere opposto sono Pompilj (1984, pp. 59-61), Piccolo e Vitale (1984, pp. 133-134) e Pieraccini (1991, p.25) che considerano il principio della probabilità composta un ulteriore postulato ed in questo differenziandosi da Kolmogorov che ottiene la probabilità condizionata come definizione (cfr. Piccolo, 1999, p. 244; Monfort, 1980, p. 71).

La probabilità assoluta può essere espressa come una probabilità condizionata (rispetto ad S):

$$P(E) = \frac{P(E \cap S)}{P(S)} = \frac{P(E)}{P(S)} = \frac{P(E)}{1} = P(E)$$

Esercizio_TP85: *dato l'universo S, la funzione di probabilità P(.) ed un evento possibile $E \subset W$ cioè con $P(E)>0$, dimostrare che P(./E) è una legittima funzione di probabilità.*

La probabilità condizionata è un concetto semplice e fecondo di cui appropriarsi subito cercando però di non perdere mai di vista la premessa essenziale che l'evento di interesse C sia isolabile dagli altri eventi dell'algebra e che per P(.|E) valgano le stesse proprietà della funzione originaria di probabilità.

Esempi:

a) Ekeland (1992, p.110) riflette: *"... Se oggi un demone spostasse di qualche centimetro la Terra dalla sua orbita, ad una scadenza abbastanza lontana ne risentirebbero tutte le orbite planetarie e questo effetto non potrebbe essere calcolato e neppure esaminato se non considerando il sistema solare nel suo complesso".*

b) Ruelle (1992, p. 31) spiega con quella che chiama "mescolanza" l'apparente paradosso che il tempo di oggi pomeriggio da un lato dipende in modo sensibile dalla posizione in cui si trovava qualche settimana fa Venere e dall'altro sia statisticamente indipendente da tale posizione. La mescolanza è una proprietà di un universo degli eventi che si modifica ad ogni prova ampliandosi, scompigliandosi, ripiegandosi su stesso finché si perde l'effetto delle condizioni iniziali.

Principio della probabilità composta

Un modo diverso, ma equivalente di descrivere la probabilità condizionata è:

$$P(E \cap C) = P(C)P(E|C)$$

nota come formula della probabilità composta (o regola della moltiplicazione). Anzi, alcuni autori preferiscono definire così la probabilità condizionata in quanto rimane valida anche quando C è un evento con probabilità zero (ma non necessariamente impossibile come vedremo nel prossimo capitolo).

Esempi:

a) L'esperimento consiste nel lanciare due dadi uguali ed equilibrati. Si apprende che la somma dei due punteggi è un numero pari. Qual'è la probabilità che il punteggio più alto sia il 4? Poniamo E= "la somma è pari", F= "il punteggio massimo è 4". Dobbiamo calcolare P(E∩F) e tale calcolo può avvenire con P(E)*P(F|E)=(18/36)*(3/18)=(3/36).

b) L'ente che gestisce un titolo finanziario vende, compra o rimane fermo con varie probabilità ed ha una analoga strategia, sia pure con diverse probabilità, se il titolo è in ribasso. Nell'ipotesi che il titolo sia in ribasso l'ente compra con probabilità del 35% e la probabilità che il titolo ribassi è del 40%. Qual'è la probabilità di "titolo in ribasso, acquista il titolo"? P(T-)=0.40, P(C|T-)=0.35, P(T-∩C)=0.35*0.40=0.14

c) La signora è in ritardo per prendere il treno. Il viaggio è lungo e vorrebbe comunque comprare qualcosa da leggere. Su di una pila espositiva vi sono 12 gialli di cui però 5 li ha già letti. Nella fretta ne prende due qualsiasi. Qual'è la probabilità che nessuno dei due comprati sia tra quelli che ha già letto? Poniamo L_1="1° libro già letto" e L_2="2° libro già letto". $P(L_1)$=5/12. Se il primo lo ha già letto, quando sceglie il secondo rimangono 11 libri di cui solo 4 sono già letti: $P(L_2|L_1)$=4/11 e quindi $P(L_1 \cap L_2)$=20/132=15.2%.

Spesso, la definizione del lato sinistro dell'equazione di probabilità composta è più difficile di quella del lato destro e quindi la formula diventa una utile scorciatoia.

Esempio:
Un'urna contiene 4 biglie bianche e 2 nere. La prova consiste nell'estrazione, senza reimmissione, di due biglie. Sia E_1={1ª bianca} e sia E2={ 2ª bianca}. Interessa calcolare la probabilità che entrambe siano bianche. Abbiamo due strategie: enumerazione e probabilità composta. Nel primo caso rapportiamo casi favorevoli e casi possibili.

(B1,N4)	(B2,N4)	(B3,N4)	(B4,N4)	(N4,B1)	(N5,B1)
(B1,N5)	(B2,N5)	(B3,N5)	(B4,N5)	(N4,B2)	(N5,B2)
(B1,B2)	(B2,B1)	(B3,B1)	(B4,B1)	(N4,B3)	(N5,B3)
(B1,B3)	(B2,B3)	(B3,B2)	(B4,B2)	(N4,B4)	(N5,B4)
(B1,B4)	(B2,B4)	(B3,B4)	(B4,B3)	(N4,N5)	(N5,N4)

Supponiamo che le biglie numerate da uno a quattro siano bianche mentre la "5" e la "6" nere. Su 30 casi, la doppia bianca compare 12 volte e quindi $P(E_1 \cap E_2)=12/30$. In alternativa si può partire dal fatto che $P(E_1)=4/6$ e $P(E_2|E_1)=3/5$ poiché l'estrazione di una biglia bianca priva l'urna di una biglia e riduce di uno le bianche. Pertanto, $P(E_1 \cap E_2)=P(E_1)*P(E_2|E_1)=(4/6)*(3/5)=12/30$.

Esercizio_TP86 : *con riferimento alle condizioni dell'esempio precedente, detti F_1={nera alla 1ª} e F2={nera alla 2ª}, calcolare: 1) $P(F_2|F_1)$; 2) $P(E_2|F_1)$; c)$P(F_2|E_1)$;*

Esercizio_TP87: *l'esperienza con l'insegnamento di Statistica è che solo 5% degli studenti che teme di essere respinto è respinto: $P(R_1^c \cap R_2^c)=0.05$ dove R_1="ritiene di superare" e R_2="supera l'esame"; Il 45% di chi teme di essere respinto è approvato $P(R_1^c \cap R_2)=0.45$; il 10% di chi riteneva di essere approvato è invece respinto: $P(R_1 \cap R_2^c)=0.10$; il 40% di coloro che ritenevano di essere approvati è poi approvato in effetti: $P(R_1 \cap R_2)=0.40$ Calcolare: 1. $P(R_1|R_2^c)$; 2. $P(R_1|R_2)$; 3. $P(R_1^c|R_2^c)$; 4. $P(R_1^c|R_2)$;*

Tutti i risultati della teoria della probabilità sono validi nel caso della probabilità condizionata, almeno nella teoria elementare. In particolare, valgono le relazioni:

$$1.\ P(C|C)=1; \qquad 2.\ P(E|C)=\frac{P(E)}{P(C)}P(C|E);\ 3.\ P(E\cap\overline{C})=P(\overline{C})P(E|\overline{C})$$

$$4.\ P(\overline{E}|C)=1-P(E|C); \quad 5.\ P(E)=P(C)P(E|C)+P(\overline{C})P(E|\overline{C})$$

che discendono pianamente da analoghi risultati con le probabilità assolute. La "5" è interessante perché mostra la relazione tra probabilità condizionata ed assoluta.

Impostazione degli esercizi
Dixon (1969, pp. 8-9) segnala che la difficoltà maggiore nei problemi di calcolo delle probabilità è la traduzione in simboli delle informazioni fornite e dal quesito posto spesso in termini vaghi e senza un esplicito riferimento ad un ben definito universo degli eventi.

I passi da seguire sono:

1) Individuare le parti che forniscono informazioni;
2) Tradurre le informazioni in simboli chiari ed univoci;
3) Circoscrivere le richieste del problema esprimendole con i simboli del punto 2;
4) Applicare le regole del calcolo delle probabilità.

Esempi:
a) La direzione vendite di accessori per telecomunicazioni ritiene che i rilievi dei clienti abbiano la seguente distribuzione di probabilità:

Status	parte elettrica	parte meccanica	esterno
in garanzia	0.12	0.13	0.05
post garanzia	0.18	0.35	0.17

Indichiamo con E={rilievo per la parte elettrica}, M={rilievo per la parte meccanica}, A={rilievo per l'aspetto esterno}, G={rilievo in garanzia}. Dalla tabella risulta che P(G)=P(G∩E)+P(G∩M)+P(G∩A) e quindi P(G)=0.12+0.13+0.05=0.30. La probabilità di lamentele sulla parte elettrica o meccanica fuori garanzia è: $P(E \cup M|G^c)=P(E|G^c)+P(M|G^c)-P(E \cap M|G^c)=(0.18+0.35+0.00)/0.70=0.76$. La probabilità di lamentele per la parte meccanica prescindendo dallo status è $P(M)=P(M\cap G)+P(M\cap G^c)= 0.13+0.35=0.48$ che, in base alla "5", può anche essere espressa come: $P(M)=P(G)P(M|G)+P(G^c)P(M|G^c)= 0.30(0.13/0.30)+0.70*(0.35/0.70)=0.48$.

b) Le possibilità che l'arbitro assegni un rigore (A) sono 4 a 1 e che, una volta assegnato, venga poi trasformato in goal (B) è del 90%. Qual'è la probabilità che la squadra segni su rigore? P(A)=1/(4+1)=0.20; P(B|A)=0.20*0.90=18%.

c) In una classe di 30 frequentanti, 9 non hanno mai fruito dell'assistenza del *tutor*. Fra questi, 8 sono stati bocciati ed 1 è stato promosso. Tra i 21 che hanno consultato il *tutor* 16 hanno superato l'esame. Qual'è dunque la probabilità che uno studente superi l'esame? Come si modifica tale probabilità richiedendo l'aiuto del *tutor*?

	T	*NT*	
S	16	1	17
B	5	8	13
	21	9	30

$$\Rightarrow P(S)=\frac{17}{30}=0.57, \qquad P(S|T)=\frac{16}{21}=0.76$$

c) Un'azienda si è accorta che le ragioni di rifiuto del suo prodotto sono attribuibili a difetti in una particolare componente. Un test di qualità rivela: il 20% delle componenti è difettoso; il 90% delle componenti passa il test di qualità; i prodotti privi di difetti passano il test nel 95% dei casi. Qual'è la probabilità che la componente non sia difettosa dopo aver passato il test? Poniamo E= "La componente è difettosa", F= "La componente passa il test". Il problema fornisce le seguenti indicazioni: P(E)=0.20; P(F)=0.90; P(F|non E)=0.95. Si vuole conoscere P(Non E |F). Solo a questo punto si applicano le regole di calcolo:

$$P(\overline{E}|F)=\frac{P(\overline{E})}{P(F)}P(F|\overline{E})=\frac{[1-P(E)]}{P(F)}P(F|\overline{E})=\frac{0.80*0.95}{0.90}=0.84$$

Esercizio_TP88: *l'urna di Polya. Un'urna contiene "b" biglie bianche e "r" biglie rosse. L'esperimento consiste nello scegliere a caso una biglia, annotarne il colore e rimetterla nell'urna; inoltre, si tolgono "c" biglie dello stesso colore (senza eliminarle tutte) e se ne aggiungono "d" dell'altro. Dopo una adeguata rimescolatura se ne estrae una seconda. Posto* E_1*={colore della 1ª} ed* E_2*={colore della 2ª}*
1) Calcolare: $P(E_2=B|E_1=R)$, $P(E_2=R|E_1=R)$, $P(E_2=B|E_1=B)$, $P(E_2=R|E_1=B)$; *2) Immaginate un'indagine campionaria che possa esser ricondotta a questo schema.*

Esercizio_TP89:*una prova consiste nell'estrarre a caso e con equiprobabilità una biglia dall'urna A che ne contiene 4 di cui 2 rosse e 2 nere; la biglia estratta è collocata nell'urna B che già ne conteneva 4 rosse e 2 Nere. Dopo l'inserimento della nuova biglia l'urna B è mischiata e da essa si estrae casualmente una ulteriore biglia che risulta essere rossa. Qual'è la probabilità che sia quella proveniente dall'urna A?*

Altre regole utili di calcolo delle probabilità condizionata sono le seguenti:

6. $P(E\cup F|C)=P(E|C)+P(F|C)-P(E\cap F|C)$;
7. $P(E\cap F|C)=P(E|C\cap F)*P(F|C)=P(E|C\cap E)*P(E|C)$;
8. $P(C|E\cup F)=\frac{P(C|E)*P(E)+P(C|F)*P(F)}{P(E\cup F)}$;
9. $P(C|E\cap F)=\frac{P(C\cap E|F)}{P(E|F)}$

Esempi:
a) In una scatola ci sono nove monete: 4 tipo C, 3 M e 2 D. Si scelgono -senza reimmissione- due monete. Si sa che alla prima estrazione non è stata ottenuta una moneta C. Qual'è la probabilità che si ottenga una moneta D alla seconda? Adottiamo la simbologia: C1, M1, D1, C2, M2, D2. Le informazioni date sono: $P(C1^c)=1$; è richiesta la probabilità di D2 dato C1 negato. Quindi:

$$P(D2|\overline{C1})=P(D2|M1\cup D1)=\frac{P(D2|M1)*P(M1)+P(D2|D1)*P(D1)}{P(M1\cup D1)}=\frac{\frac{2}{8}\frac{3}{9}+\frac{1}{8}\frac{2}{9}}{\frac{5}{9}}=\frac{1}{5}$$

b) Una concorrente ad un quiz televisivo deve scegliere una di tre buste identiche: A , B e C ognuna contenente due buste più piccole: nella A1 si vince una crociera come pure nella A2; nella B1 si vince di nuovo la crociera e nella B2 un CD di canzoni popolari bergamasche; lo stesso CD è il premio indicato nelle due buste piccole: C1 e C2 contenute nella busta C. La concorrente sceglie la busta grande e poi quella piccola. In questa ha vinto una crociera; qual'è la probabilità che anche l'altra vinca una crociera? E="esce la crociera" con P(E)=3/6 (tre casi favorevoli su sei possibili). La seconda busta ha come premio una crociera solo se la scelta iniziale è sta la A e, pertanto, P(A|E)=[P(A)/P(E)]P(E|A)=2/3.

c) Il ladro Fantomas ha individuato il mobile dove sono conservati i preziosi. Il mobile ha quattro cassetti chiusi a chiave ed ognuno contiene due scomparti interni, chiusi anche questi. Dall'inventario che il ladro si è procurato risulta la seguente composizione per cassetti e scomparti A=(gioielli, gioielli), B=(gioielli, vetri), C=(vetri, gioielli), D=(vetri, vetri). Fantomas ignora quale siano i cassetti e può aprirne uno solo. Aperto uno degli scomparti si scopre che contiene vetri colorati; qual'è la probabilità che l'altro contenga invece dei preziosi? L'esito è favorevole (per Fantomas) se la scelta è caduta sui cassetti B o C e quindi:

$$P(B\cup C|V)=P(B|V)+P(C|V)-P(B\cap C|V)=\frac{P(B\cap V)}{P(V)}+\frac{P(C\cap V)}{P(V)}-\frac{P(B\cap C\cap V)}{P(V)}$$
$$=\frac{1}{P(V)}\left[P(B)P(V|B)+P(C)P(V|C)-P(B\cap C)P(V|B\cap C)\right]=\frac{1}{4/8}\left[\left(\frac{1}{4}\right)\left(\frac{1}{2}\right)+\left(\frac{1}{4}\right)\left(\frac{1}{2}\right)-0\right]=\frac{1}{2}$$

Esercizio_TP90: *il meccanismo che previene gli ingressi abusivi di un sistema informatico conta su due allarmi: E ed F. Il primo è inattivo con una probabilità del 5%, il secondo del 2.5%. Se il primo non scatta, il secondo si attiva con una probabilità del 99%.*
1) Se il sistema richiedesse entrambi gli allarmi attivi quale sarebbe la probabilità di rimanere senza difesa?
2) Se bastasse che almeno uno fosse in funzione, quale sarebbe il grado di copertura?
3) Qual'è la probabilità che ne sia attivo uno e solo uno?

Esercizio_TP91: *in uno staff di 50 dipendenti ne sono presenti 20 con contratto di ingresso agevolato al lavoro. Tre dipendenti scelti a caso debbono costituire la delegazione di fabbrica. Qual'è la probabilità che siano tutti contrattisti? Adoperate sia il calcolo combinatorio che la probabilità condizionata.*

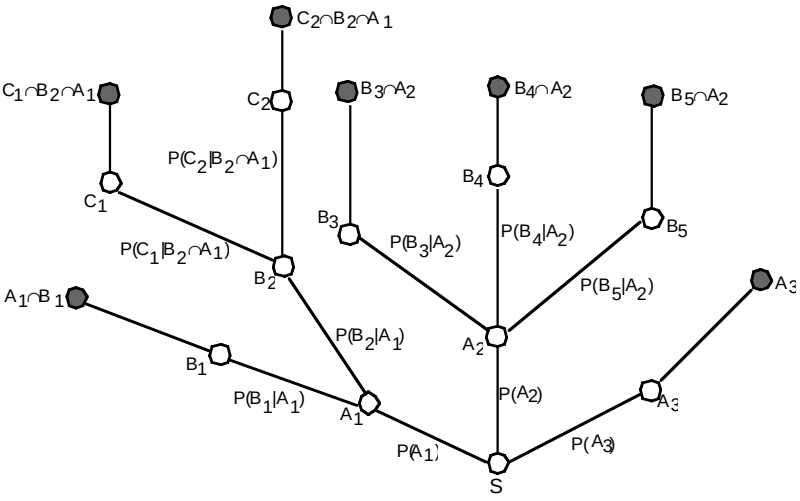
Esercizio_TP92: *nel blackjack il punto che dà il nome al gioco si ottiene con un asso ed una figura o un asso ed un dieci. Scegliendo a caso e senza reimmissione due carte da un mazzo francese, qual'è la probabilità di fare il punto?*

Esercizio_TP93: *il tetraetro presenta quattro facce regolari. Una volta lanciato si considera come risultato la faccia rivolta verso il basso. Un gioco consiste nel lanciarne tre e vince chi per primo ottiene tre facce uguali. Se uno dei giocatori ha già scoperto due facce con il due, qual'è la probabilità che anche la terza mostri il due?*

Esercizio_TP94: *una biglia è scelta a caso da un'urna che contiene 3 biglie bianche e 5 biglie rosse. La biglia estratta è rimessa nell'urna insieme ad un'altra biglia del suo stesso colore. A questo punto si estrae una seconda biglia. Qual'è la probabilità che*
a) Nessuna biglia estratta è bianca; b) Solo una è bianca; c) Entrambe sono bianche.

Albero delle decisioni

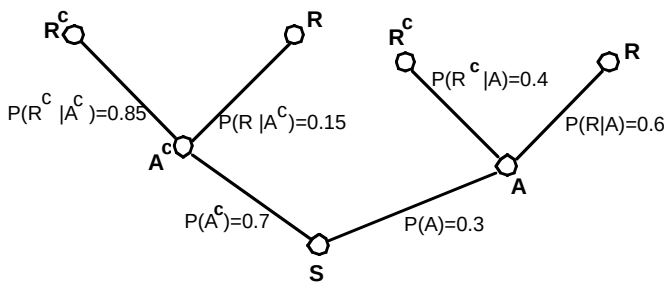
Un modo efficace di rappresentare le conoscenze probabilistiche sull'algebra di un esperimento casuale è il diagramma ad albero (cfr. Roberts, 1992, pp.46-53). I nodi dell'albero rappresentano gli eventi, la sequenza dei rami che li connette esprime l'ordine con cui gli eventi sono considerati. Ad ogni ramo è abbinata la probabilità dell'evento sul suo nodo terminale condizionata dall'evento sul nodo iniziale.



Esempio:

Alcuni clienti sono classificati in base all'acquisto o non acquisto di un prodotto e se ricordano o no uno *spot* che lo pubblicizza. Il passaggio dal 1° livello al 2° avviene con il meccanismo della probabilità condizionata. Il prodotto delle probabilità assegnate ai sue rami è pari alla probabilità congiunta dei due nodi. Questo tipo di diagramma è il nucleo iniziale della teoria statistica delle decisioni.

Spesa / Memoria	A	A^c	
R	120	60	180
R^c	80	340	420
	200	400	600



Esercizio_TP95: *nella tabella sono riportate le probabilità congiunte di coloro che completano gli studi universitari per il titolo di studio e per gruppi di discipline. Rappresentare le informazioni con un diagramma ad albero.*

	$Scientifico - A_1$	$Umanistico - A_2$	
$Dipl.\ Unin. - D_1$	0.15	0.05	0.20
$Laurea - D_2$	0.25	0.40	0.65
$Dott.\ Spec.\ D_3$	0.05	0.10	0.15
	0.45	0.55	1.00

La probabilità condizionata si estende senza troppe difficoltà al caso di tre o più eventi:

$$P(E \cap F \cap G) = P(F \cap G) * P(E|F \cap G) = P(G) * P(F|G) * P(E|F \cap G)$$

Esempi:

a) L'idea della difesa con barriere successive singolarmente insicure, ma nel complesso ben solide, è molto antica, ma sempre in voga. Berry e Lindgren (1990, p.58) riferiscono del principio alla base dello scudo stellare dell'amministrazione Reagan negli USA. Il sistema è concepito come strati successivi, in cui ciascuno strato ha probabilità, diciamo dell'80%, di fermare il missile nemico (F) e del 20% di mancarlo (M). Tale modello di probabilità si mantiene costante negli strati: $P(F_2|M_1)=80\%$ e $P(M_2|M_1)=20\%$. La probabilità che un missile sorpassi quattro livelli così predisposti è data da:

$$P(M_4 \cap M_3 \cap M_2 \cap M_1) = P(M_1)P(M_2|M_1)P(M_3|M_1 \cap M_2)P(M_4|M_1 \cap M_2 \cap M_3) = 0.2 * 0.2 * 0.2 * 0.2 = 0.2^4 = 0.0016$$

b) Le imprese più significative (almeno 50 dipendenti) di una provincia sono state classificate per attività prevalente: 200 nel settore alimentare, 40 in quello meccanico, 100 nell'edilizia e 60 negli innovativi. Qual'è la probabilità che in un scelta casuale -senza reimmissione- di n=4 imprese ne capiti una per ogni settore:

$$P(A_1 \cap M_2 \cap E_3 \cap I_4) = P(A_1) * P(M_2|A_1) * P(E_3|A_1 \cap M_2) * P(I_4|A_1 \cap M_2 \cap E_3) = \frac{200}{400}\frac{40}{399}\frac{100}{398}\frac{60}{397} = 0.0019$$

Nel modello delle urne si è visto come sia rilevante reimmissione/non reimmissione delle biglie per assegnare le probabilità. Possiamo aggiungere alcune riflessioni che nascono dal concetto di probabilità condizionata.

Esempi:

a) In un mazzo di carte francesi si scelgono a caso e senza reimmissione tre carte. Calcoliamo la probabilità che siano tre carte di cuori. Sia C_i l'evento "cuori alla carta i-esima". La probabilità di $C_1 \cap C_2 \cap C_3$ può essere espressa come:

$$P(C_1 \cap C_2 \cap C_3) = P(C_1)P(C_2|C_1)P(C_3|C_1 \cap C_2) = \frac{13}{52}\frac{12}{51}\frac{11}{50} = 0.0129$$

b) Un'indagine è mirata alle famiglie con tre figli ed in particolare al sesso della prole. I casi possibili sono 2^3 =8 che si ritengono equiprobabili, almeno ad uno stadio iniziale dell'indagine. Si viene a sapere che in un quartiere non c'è famiglia con tre figli che non abbia una figlia femmina. Qual'è la probabilità che siano tutte femmine? Una soluzione istintiva potrebbe essere: la prima c'è sicuro; la coppia di femmine ha probabilità 1/4 dato che ora i casi sono quattro: (F_2F_3, M_2M_3, M_2F_3, F_2M_3). Tale soluzione sarebbe ammissibile se l'evento di interesse fosse: "scelta a caso una famiglia con due figli, qual'è la probabilità che siano entrambe femmine?" La domanda era invece un'altra. L'informazione che su tre figli una è certamente femmina riduce gli eventi ai da 8 a 7 (è escluso solo $M_1M_2M_3$) e quindi $P(F_1F_2F_3|F_1)=1/7$.

c) I messi dell'imperatrice sono al villaggio per arruolare soldati. L'anziana che detiene il comando scrive il nome dei 24 giovani abili alla guerra. E' noto che 8 di questi sarebbero disposti a partire volontari, ma si preferisce sottomettere la scelta all'alea del sorteggio e così si estraggono -senza reimmissione- tre nominativi. Qual'è la probabilità che uno dei coscritti non sia un volontario? I non volontari sono 16; ciascuno di questi si può combinare con la coppia di volontari; di tali coppie ve ne sono C(8,2)=28. Quindi:

$$\frac{16 * 28}{C(24,3)} = \frac{448}{2024} = 0.2213$$

Sapendo che è stato sorteggiato un non volontario, qual'è la probabilità che gli altri due siano dei volontari?

$$P(V_2 \cap V_3|V_1^c) = P(V_1^c)P(V_2|V_1^c)P(V_3|V_1^c \cap V_2) = \frac{16}{24}\frac{8}{23}\frac{7}{22} = 0.0738$$

Non c'è incongruenza: le due probabilità sono associate a due eventi diversi.

Esercizio_TP96: *un lotto di 1000 prodotti contiene: 980 di qualità alta, 15 media e 5 bassa. Il cliente sceglie a caso e senza rimessa 5 item. Calcolare la probabilità che siano tutti difettosi (nel qualcaso l'ordine è rescisso).*

Esercizio_TP97: *il verificarsi di un evento improbabile è considerato manifestazione di volontà ultraterrene o di imbrogli o di errori e malfunzionamenti (l'affondamento del Titanic, l'esplosione di Chernobyl). Esprimete una vostra considerazione sul verificarsi di tali eventi.*

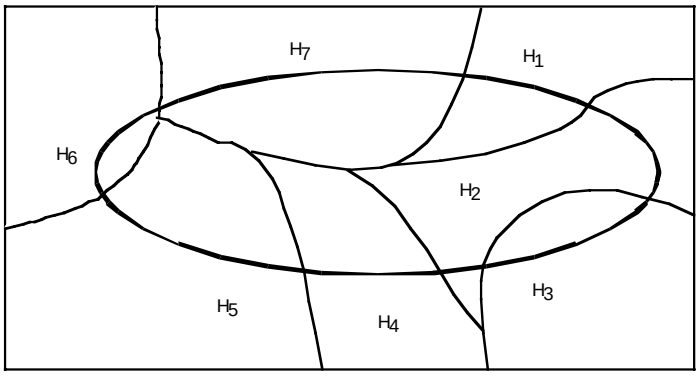
Teorema di Bayes

Le regole sugli insiemi consentono di esprimere un evento rispetto ad una partizione dell'universo in "k" parti esclusive ed esaustive:

$$H_i \cap H_j = \varnothing \;\; per \;\; i \neq j \quad e \quad \bigcup_{i=1}^{k} H_i = S$$

La stesse caratteristiche si estendono alle parti che ciascun elemento ha in comune con E che può essere scritto:

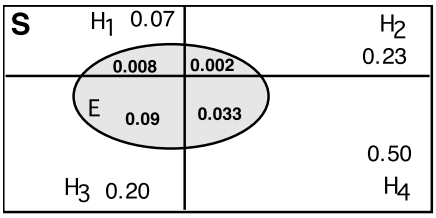
$$E = \bigcup_{i=1}^{k} (H_i \cap E) \;\; con \quad (E \cap H_i) \cap (E \cap H_j) = \varnothing \;\; per \;\; i \neq j$$



La probabilità del verificarsi di E è frazionabile nella probabilità del verificarsi dello stesso E in congiunzione con ciascun evento della partizione di S.

$$P(E) = \sum_{i=1}^{k} P(H_i \cap E) = \sum_{i=1}^{k} P(H_i) * P(E|H_i)$$

Esempio:
In un esperimento sono espressi i giudizi di probabilità per la partizione in figura. L'evento E può essere rappresentato come unione di quattro eventi incompatibili e la sua probabilità data come somma delle probabilità degli eventi componenti ottenuta in base alla formula della probabilità condizionata:



$$P(E) = 0.07 * \frac{0.008}{0.07} + 0.23 * \frac{0.002}{0.23} + 0.20 * \frac{0.09}{0.20} + 0.50 * \frac{0.033}{0.50} = 0.133$$

Con la spalmatura della probabilità dell'evento sulla partizione non si è fatto un vero e proprio passo avanti dato che molte altre partizioni potrebbero servire a riesprimere l'evento E senza che per questo si modifichi lo stato informativo del problema. La trovata semplice e geniale è di ragionare all'inverso e cioè non consideriamo più E l'evento condizionato, bensì il condizionante e poniamoci la domanda: nell'ipotesi che si verifichi E, qual'è la probabilità -sotto E- di ciascuno degli eventi della partizione? In altre parole, se l'effetto è il verificarsi dell'evento E, qual'è la sua causa più probabile fra le H_i i=1,2,…,k che costituiscono la partizione? Premettiamo innanzitutto che uno dei "k" eventi -necessariamente- si verificherà e mai due insieme dato che:

$$P\left(\bigcup_{i=1}^{k} H_i\right) = 1; \qquad P(H_i \cap H_j) = 0 \;\; se \;\; i \neq j$$

Principio della probabilità inversa
Si supponga che l'evento E abbia probabilità positiva: P(E)>0. Allora:

$$P(H_j|E) = \frac{P(H_j \cap E)}{P(E)} = \frac{P(H_j)P(E|H_j)}{P(E)} = \frac{P(H_j)P(E|H_j)}{\sum_{i=1}^{k} P(H_i)P(E|H_i)}$$

Tale risultato è noto come "principio della probabilità inversa" perché scambia il ruolo degli eventi causa e dell'evento effetto. Esso è dovuto a Thomas Bayes (1763/1958), ma fu pubblicato postumo per merito dell'amico di questi, Richard Price, al quale parvero irrilevanti certe reticenze di Bayes.

Esempi:
a) Ritorniamo al caso illustrato con il diagramma di Venn e determiniamo la causa più probabile di E. I calcoli mostrano che l'ipotesi più credibile -relativamente al verificarsi di E- è H_3:

$$P(H_1|E) = \frac{0.008}{0.133} = 0.0602; P(H_2|E) = \frac{0.002}{0.133} = 0.0150; \qquad P(H_3|E) = \frac{0.090}{0.133} = 0.6767; P(H_4|E) = \frac{0.033}{0.133} = 0.2481$$

Se, in una scommessa, le H_i dessero luogo alla stessa vincita, la logica ci imporrebbe di scegliere H_3.

b) Un test, basato su di un solo quiz a scelta multipla (4 risposte di cui solo una esatta), assegna 1 punto per la risposta esatta e non dà penalizzazioni per quella sbagliata. Adele è rimasta chiusa in stanza tutto il giorno prima dell'esame (si ignora se ha studiato oppure dormito). Indichiamo con "p" la probabilità che Adele abbia studiato. Il test è tale che se Adele ha studiato supera certamente l'esame. Poniamo gli eventi E= "Adele ha studiato", F= "Risponde correttamente". Nel caso superi l'esame, la probabilità che abbia in effetti studiato è:

$$P(E|F) = \frac{P(E)P(F|E)}{P(E)P(F|E) + P(E^c)P(F|E^c)} = \frac{p*1}{p*1 + (1-p)\frac{1}{4}} = \frac{4p}{3p+1}$$

che è pari a P(E) solo se p=0 oppure se p=1. Inoltre, se supera l'esame si ritiene "probabile" che abbia studiato solo se p>0.20 perché in questo caso P(E|F)>0.50.

c) Ciccillo ha due monete: una con due facce (M2) ed un'altra (M1) con la stessa faccia (croce) da entrambi i lati. Sceglie a caso (equiprobabilità) una moneta e la lancia. Sapendo che è uscito "croce", qual'è la probabilità che sia stata lanciata la moneta truccata?

$$P(M1|C) = \frac{P(M1)P(C|M1)}{P(M1)P(C|M1) + P(M2)P(C|M2)} = \frac{\frac{1}{2}*1}{\frac{1}{2}*1 + \frac{1}{2}*\frac{1}{2}} = \frac{2}{3}$$

Esercizio_TP98: *n uno stabilimento esistono cinque linee di produzione per uno stesso prodotto che però finiscono in un unico collettore per il confezionamento. Le linee producono lo stesso ammontare di pezzi. Nel complesso la probabilità che un prodotto sia imperfetto è P(I)=0.03; per le singole linee le probabilità di difetto sono $P(I/L_1)$=0.004, $P(I/L_2)$=0.003, $P(I/L_3)$=0.006, $P(I/L_4)$=$P(I/L_5)$. Qual'è la probabilità che, scelto a caso un prodotto e trovatolo difettoso, provenga da ciascuna delle linee?*

Esercizio_TP99: *la consulenza telematica di uno studio legale è distribuita a quattro team nelle proporzioni/ probabilità: 15%, 40%, 25%, 20%. I team hanno probabilità a priori di errore: 4%, 7%, 5%, 6%. Qual'è la probabilità di errore per lo studio nel suo complesso?*

Esercizio_TP100: *in una impresa per la pescicoltura vi sono tre vasche V_1,V_2,V_3 che contengono, rispettivamente: 200, 300, 500 trote iridate. Tali trote costituiscono le seguenti quote di pesci dell'impresa: 60%, 50%, 40%. L'acquirente sceglie la vasca in cui pescare con probabilità dettate dalla percentuale di trote iridate e ne prende una (racconterà poi di una lunga e difficile cattura). In quale vasca è più probabile l'abbia pescata?*

Probabilità a priori e a posteriori
La formula di Bayes non è un risultato eclatante di per sé e potrebbe apparire solo un modo diverso di esporre la probabilità condizionata. Bayes stesso la propone come risultato preliminare in un articolo in cui sembra interessato ad altro. Laplace ne comprese la portata e da allora essa riveste un ruolo centrale in Statistica. L'interpretazione moderna è la seguente: un esperimento casuale dà luogo ad una certa manifestazione E. Le cause possibili (ipotesi) formano una partizione $\{H_i, i=1,2,\ldots,k\}$ dell'evento certo. Le probabilità $P(H_i)$ valutate prima che l'evento E si verifichi sono dette probabilità a priori delle $\{H_i\}$.

Supponiamo che l'evento E abbia dei punti di contatto con uno o più eventi della partizione di modo che P(E)>0; ciò implica che il giudizio di credibilità sulle ipotesi H_i dovrà modificarsi dato che alcune diventeranno più verosimili ed altre meno a valle del verificarsi di E (almeno una delle probablità deve alterarsi a causa di E altrimenti la considerazione di tale evento sarebbe inutile per accertare la causa più porbabile). La $P(H_i|E)$ è detta probabilità a posteriori della causa o ipotesi H_i. Il fattore $P(E|H_i)$ che trasforma, al netto del fattore di scala: P(E), la probabilità a priori in probabilità a posteriori è detto verosimiglianza dell'evento H_i.

$$P\left(H_j\middle|E\right) \propto P\left(H_j\right)P\left(E|H_j\right)$$

dove "$\propto$" indica che il valore a sinistra è proporzionale a ciò che sta a destra.

Esempi:

a) Gli scioperanti premono per essere ricevuti dall'autorità amministrativa. Le aziende in crisi sono tre: A_1 con 200 dipendenti (25 donne), A_2 150 dipendenti (40 donne) e A_3 con 350 dipendenti (15 donne). L'autorità sceglie di ricevere una sola delegata. Ipotizzando l'equiprobabilità delle scelte, qual'è l'azienda da cui è più probabile sia dipendente?

$$D = \left(A_1 \cap D\right) \cup \left(A_2 \cap D\right) \cup \left(A_3 \cap D\right) \Rightarrow P(D) = P\left(A_1\right)P\left(D|A_1\right) + P\left(A_2\right)P\left(D|A_2\right) + P\left(A_3\right)P\left(D|A_3\right)$$
$$= \frac{200}{700}\frac{25}{200} + \frac{150}{700}\frac{40}{150} + \frac{350}{700}\frac{15}{350} = \frac{80}{700} = 0.1143$$

A questo punto la probabilità che la delegata provenga da ciascuna delle tre aziende è:

$$P\left(A_1|D\right) = \frac{25}{80} = 0.3125;\ \ P\left(A_2|D\right) = \frac{40}{80} = 0.5000;\ \ P\left(A_3|D\right) = \frac{15}{80} = 0.1875;$$

L'azienda favorita è la A_2 che ha più donne, anche se i dipendenti sono meno che nelle altre due.

b) Un esperimento consiste nell'estrarre -con reimmissione- "n" biglie da un'urna U_i scelta a caso fra (N+1) possibili urne equiprobabili. Ogni urna contiene lo stesso numero N di biglie di cui "i" sono rosse ed (N-i) sono bianche per i=0,1,2,…,N. All'atto dell'estrazione si scopre che tutte "n" biglie sono rosse. Qual'è la probabilità che sia rossa anche la (n+1)-esima? Scomponiamo E= "n biglie rosse" rispetto alle urne:

$$E = \bigcup_{i=0}^{N}\left(E \cap U_i\right) \Rightarrow P(E) = \sum_{i=0}^{N} P\left(U_i\right)P\left(E|U_i\right) = \sum_{i=0}^{N}\left(\frac{1}{N+1}\right)\left(\frac{i}{N}\right)^n = \left(\frac{1}{N+1}\right)\sum_{i=0}^{N}\left(\frac{i}{N}\right)^n$$

L'evento F= "n+1 biglie rosse" ha probabilità analoga con (n+1) al posto di "n". Quindi,P(F|E)=P(E∩F)/P(E)=P(F)/P(E) dato che F=(E∩G) con G= "la biglia estratta alla (n+1)-esima prova è rossa" e P(E∩F)=P(F).

$$P\left(F|E\right) = \frac{P(F)}{P(E)} = \frac{\left(\frac{1}{N+1}\right)\sum_{i=0}^{N}\left(\frac{i}{N}\right)^n}{\left(\frac{1}{N+1}\right)\sum_{i=0}^{N}\left(\frac{i}{N}\right)^{n+1}} = \frac{\sum_{i=0}^{N}\left(\frac{i}{N}\right)^n}{\sum_{i=0}^{N}\left(\frac{i}{N}\right)^{n+1}} N \frac{\sum_{i=0}^{N}(i)^n}{\sum_{i=0}^{N}(i)^{n+1}} \cong N \frac{\left(\frac{N}{n+1}\right)^{n+1}}{\left(\frac{N}{n+2}\right)^{n+2}} = \frac{n+1}{n+2}$$

Questo risultato è conosciuto come regola di successione. Una prima conclusione è che una teoria suffragata da "n" fatti sarà suffragata anche dal fatto (n+1)-esimo e si arriva alla certezza se gli eventi "n" sono numerosi. Da Laplace in poi la regola di successione è sempre stata fonte di applicazioni controverse perché ha due debolezze: equiprobabilità delle ipotesi a priori e validità solo in caso di un numero infinito di ipotesi (questo spiega l'ultima relazione nella formula) che nessuno sarà mai in grado di verificare o falsificare.

Esercizio_TP101: *la popolazione attiva di un comune è N=10'000 unità e fra questi si conta il 28% di disoccupati. Per accertare la diffusione del lavoro nero si scelgono casualmente n=200 persone attive; fra queste si selezionatno a caso e senza reimmissione m=20 per un'intervista più approfondita.*
a) Calcolare la probabilità che siano tutti disoccupati;
b) La 1ª persona scelta è disoccupata. Qual'è la probabilità che non sia la seconda?

Esercizio_TP102: *il design di un nuovo prodotto incontra i gusti del 96% dei consumatori. La società di marketing che lo assevera accetta il 97% dei design poi graditi dal pubblico e scarta il 95% di quelli che i consumatori rifiutano.*
a) Qual'è la probabilità che un design scartato sia in realtà gradito?
b) Qual'è la probabilità che un design accettato risulti successivamente sgradito?

La formula di Bayes si semplifica se le probabilità a priori delle "k" ipotesi sono equiprobabili: $P(H_j)=1/k$ che è lo schema preferito (ma non sempre giustificato) per valutare la probabilità delle cause in condizioni di totale ignoranza:

$$P(H_j|E)=\frac{P(H_j)P(E|H_j)}{\sum_{i=1}^{k}P(H_i)P(E|H_i)}=\frac{\frac{1}{k}P(E|H_j)}{\sum_{i=1}^{k}\frac{1}{k}P(E|H_i)}=\frac{P(E|H_j)}{\sum_{i=1}^{k}P(E|H_i)}$$

Dalla formula sono scomparse le probabilità a priori per lasciare tutta la scena alle realizzazioni sperimentali.

Esempi:

a) Una prova consiste nella scelta di biglie di colore diverso: bianche e rosse da tre urne con le seguenti composizioni: U_1={20B, 8R}, U_2={2B, 5R}, U_3={7B, 7R}. Si sceglie a caso l'urna e dall'urna prescelta si seleziona -sempre casualmente- una biglia. Poniamo E="la biglia è rossa". Qual'è l'urna da cui è più probabile che sia stata estratta? Ipotizziamo che le urne siano scelte con equiprobabilità: $P(U_1)=P(U_2)=P(U_3)=1/3$. La probabilità dell'evento E diviene:

$$P(E)=P\left[E\cap(U_1\cup U_2\cup U_3)\right]=P\left[(E\cap U_1)\cup(E\cap U_2)\cup(E\cap U_1)\right]=P(E\cap U_1)+P(E\cap U_2)+P(E\cap U_3)$$
$$=P(U_1)P(E|U_1)+P(U_2)P(E|U_2)+P(U_3)P(E|U_3)=\frac{1}{3}\left(\frac{8}{28}+\frac{5}{7}+\frac{7}{14}\right)=0.5$$

La probabilità a posteriori delle urne è: $P(U_1|E)=\frac{\frac{1}{3}}{\frac{1}{2}}\frac{8}{28}=\frac{8}{28}\frac{2}{3}=0.190$; $P(U_2|E)=\frac{5}{7}\frac{2}{3}=0.476$; $P(U_3|E)=\frac{7}{14}\frac{2}{3}=0.333$;

cioè l'urna U_2 è la provenienza più verosimile alla luce dell'equiprobabilità, ma è proprio l'equiprobabilità che contraddice lo stato fisico dell'esperimento ignorando che la U_1 contiene il quadruplo di biglie della U_2 e il doppio di quelle della U_3.

b) Ci si trova di fronte un parto trigemino. La partoriente ha già dato alla luce due maschi e si attende il terzo nato. Qual'è la probabilità che sia ancora di sesso maschile? Empiricamente si ha P(3M)=24/100 e P(2M, 1F)=27/100. Inoltre, la probabilità che i primi due siano maschi dato che il terzo è maschio è uno dato che in questo caso si forma l'evento certo. Ne consegue:

$$P(M3|M1\cap M2)=\frac{P(M1\cap M2\cap M3)}{P(M2\cap M1)}=\frac{P(M1\cap M2\cap M3)}{P(M2\cap M1|F3)+P(M2\cap M1|M3)}=\frac{\frac{24}{100}}{\frac{27}{100}\frac{1}{3}+\frac{24}{100}1}=\frac{8}{11}$$

Esercizio_TP103: *due società: "Turisud s.r.l. " e "Meridionale Tour s.a.s." hanno ciascuna due pacchetti leader A_1 e A_2 e B_1, B_2. La prima società riduce i prezzi di uno dei suoi pacchetti o di entrambi se l'altra riduce uno o entrambi i suoi prezzi. La strategia della Meridionale Tour prevede: $P(B_1^+\cap B_2)=0.5$, $P(B_1\cap B_2^+)=0.4$, $P(B_1^+\cap B_2^+)=0.1$. Le strategie della Turisud data quella della Meridionale Tour, sono:*

	$P(A_1^+\cap A_2)$	$P(A_1\cap A_2^+)$	$P(A_1^+\cap A_2^+)$
$B_1^+\cap B_2$	0.2	0.6	0.2
$B_1\cap B_2^+$	0.6	0.1	0.3
$B_1^+\cap B_2^+$	0.1	0.2	0.7

1) Qual'è la probabilità che aumentino tutti i prezzi? 2) Se la Meridionale Tour aumenta entrambi i suoi prezzi e la Turisud ne può aumentarne solo uno quale sarà quello che aumenterà e perché?

Esercizio_TP104: *un esperimento consiste nel lanciare due dadi regolari dal punto di vista del materiale e della distribuzione del peso, ma numerati in modo bizzarro: il 1° ha due facce con "1", due facce con "5" e due facce con "6"; il 2° ha due facce col "2", due con il "3" e due con il "4". Se si apprende che l'esito del lancio ha dato come somma otto, qual'è la probabilità che sia uscita la coppia (5,3)?*

Esercizio_TP105: *tre candidati di pari forza elettorale: Caruso, Ferrari, Spadafora alla presidenza regionale hanno come punto di forza la riduzione dei residui passivi nel bilancio dell'ente. Le rispettive maggioranze permetteranno di raggiungere l'obiettivo con probabilità: 0.35, 0.40, 0.45. Qual'è la probabilità che si riducano i residui passivi? Qual'è la probabilità che il merito sia di Caruso?*

Rapporto di verosimiglianza

Le probabilità a posteriori sono spesso contrapposte a due a due per confrontare la credibilità delle ipotesi:

$$\frac{P(H_i|E)}{P(H_j|E)}=\frac{P(E|H_i)P(H_i)}{P(E|H_j)P(H_j)}=\left[\frac{P(E|H_i)}{P(E|H_j)}\right]\left[\frac{P(H_i)}{P(H_j)}\right];\quad per\ \ i\neq j$$

tale quoziente, detto rapporto di verosimiglianza (*likelihood ratio*), esprime la probabilità del verificarsi di E sotto H_i in termini della probabilità sotto l'alternativa H_j. In caso di probabilità uniforme, la credibilità delle ipotesi è limitata ai soli rapporti di verosimiglianza. Supponiamo che dopo il verificarsi dell'evento E_1 si ipotizzi la verifica dell'evento E_2 e poi dell'evento E_3. Il rapporto diventa:

$$\frac{P(H_i|E_1\cap E_2\cap E_3)}{P(H_j|E_1\cap E_2\cap E_3)}=\left[\frac{P(E_3|H_i\cap E_1\cap E_2)P(E_2|H_i\cap E_1)P(E_1|H_i)}{P(E_3|H_j\cap E_1\cap E_2)P(E_2|H_j\cap E_1)P(E_1|H_j)}\right]\left[\frac{P(H_i)}{P(H_j)}\right]$$

All'aumentare dei dati si ridimensiona il ruolo delle probabilità a priori (il cui rapporto rimane costante) per dare sempre più spazio all'accumulo di fatti sperimentali. Questo spiegherebbe anche il progressivo ridursi dell'influenza delle condizioni iniziali dell'esperimento. Molte questioni della Statistica bayesiana e non bayesiana sono incentrate sulla reale portata dell'accumulo di esperienza.

Esempi:

a) Il 5% della popolazione residente in un comune è affetto da una malattia. Posto A= "Una persona scelta a caso fra i residenti del comune è ammalata" abbiamo: P(A)=0.05. Supponiamo di disporre di un test clinico che abbia sensitività, cioè la probabilità di essere positivo (T^+) dato che la persona è ammalata, $P(T^+|A)=0.90$; Ipotizziamo, che la probabilità di falso positivo (la persona è sana, ma il test indica il contrario) sia $P(T^+|A^c)=0.15$. Scelta a caso una persona si effettua il test e questo risulta positivo, qual'è la probabilità *a posteriori* che la persona sia effettivamente ammalata (sensitività del test)?

$$P(A|T^+)=\frac{P(T^+|A)P(A)}{P(T^+|A)P(A)+P(T^+|A^c)P(A^c)}=\frac{0.90*0.05}{0.90*0.05+0.15*0.95}=0.24$$

L'esito è sorprendente: nonostante il test abbia un buon grado di affidabilità (è un sintomo o un *marker* presente sui nove decimi delle persone ammalate), basarsi sulla sola presenza del sintomo o *marker* è rischioso dato che solo una volta su quattro il test positivo indica la presenza di malattia.

b) La specificità di un test è la probabilità che esso sia negativo dato che la persona è sana. Ipotizziamo che: $P(T^-|A^c)=0.80$. Se il test è negativo, qual'è la probabilità che la persona sia sana?

$$P(A^c|T^-)=\frac{P(T^-|A^c)P(A^c)}{P(T^-|A^c)P(A^c)+P(T^-|A)P(A)}=\frac{0.80*0.95}{0.80*0.95+0.10*0.05}=0.9935$$

Sotto questo aspetto il test è molto più soddisfacente. Pur presentando una specificità non elevata (*marker* presente una volta su cinque sani) nega la malattia con un errore inferiore al due per mille.

c) Un noto personaggio Y è stato coinvolto in un caso di riconoscimento di paternità. Trascurando i fattori legati al DNA analizziamo il problema dal punto di vista del gruppo sanguigno. La signora X è di gruppo A e il signor Y è di gruppo AB, il bambino è di gruppo B. Sia "a" la probabilità a priori di E= "Y è il padre" con P(E)=a. Le leggi di Mendel stabiliscono che P(B|E)=0.25 e $P(B|E^c)=0.08$. Ne consegue:

$$P(E|B)=\frac{P(B|E)P(E)}{P(B|E)P(E)+P(B|E^c)P(E^c)}=\frac{0.25a}{0.25a+0.08(1-a)}=\frac{a}{0.32+0.68a}$$

A questo punto è il giudice che, in base agli elementi in suo possesso, fissa "a" ed innesca il ragionamento probabilistico bayesiano (Dall'Aglio, 1982, pp. 66-68).

P(E)=a	P(E|B)
0.05	14.1%
0.25	51.0%
0.50	75.6%
0.75	90.4%
0.95	98.3%

Se il giudice fissa a=0.5 la probabilità a posteriori è dell'86% e se, per altri fatti noti (ad esempio, l'ammissione di un incontro), la ritiene ancora più alta: a=0.75 allora, unita all'evidenza dei dati sulle leggi di Mendel, si ha P(E|B)=94.9% che comincia ad essere alta.

d) L'8% degli operai risiede in una provincia diversa da quella in cui si trova la fabbrica. Il 10% di questi ha un livello salariale elevato (E), il 30% salario medio ed il 60% salario base. Per i residenti (R) le probabilità di quei livelli salariali sono 5%, 15%, 80%. Se, scelto a caso un operaio, si trova che ha un salario elevato, qual'è la probabilità che risieda fuori provincia (F)?

$$P(F|E)=\frac{P(E|F)P(F)}{P(E|F)P(F)+P(E|R)P(R)}=\frac{0.1*0.08}{0.1*0.08+0.05*0.92}=0.148$$

Esercizio_TP106: *un collegio giudicante emette sentenze giuste nel 95% dei casi (cioè il 95% di quelli giudicati colpevoli ed il 95% di quelli giudicati innocenti sono realmente tali). Se il 99% dei rinviati a giudizio è colpevole calcolare la probabilità che:*
a) La persona sia innocente dato che si è avuta una sentenza di assoluzione;
b) La persona innocente riceva un verdetto di colpevolezza;
c) La persona innocente sia giudicata innocente.

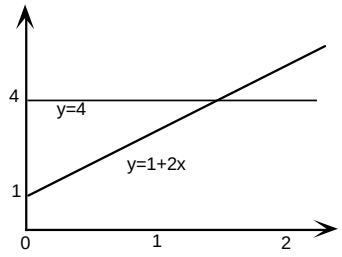
Esercizio_TP107: *Sono state lanciate due monete regolari. Sapendo che è uscita almeno una testa quale evento è più probabile: A= "una è croce" oppure B= "entrambe teste"?*

Esercizio_TP108: *la prigioniera è fuggita e si è nascosta in un bosco scelto a caso fra i tre che crescono nella zona del penitenziario. Se si trova nel bosco B_i l'elicottero la troverà con probabilità $(1-p_i)$. E' stata sorvolato il bosco B_1 e non è stata trovata traccia della prigioniera. Qual'è la probabilità che sia nel bosco B_j ,j=1,2,3?*

6.4.2 Indipendenza in probabilità

Nel concetto di indipendenza Kolmogorov (1933/1995, p. 19) vede il primo embrione della problematica specifica del calcolo delle probabilità. Tutte le nozioni legate alla casualità sono difficili e presentano aspetti non arrivabili per via della comune esperienza; a questo non si è sottratta la probabilità e non sfugge l'indipendenza. Procediamo per gradi e partiamo dal concetto opposto: la dipendenza.

Esempi:
a) Una variabile Y è dipendente da un'altra X se fissato il valore della seconda è univocamente determinato il valore della prima.



Y=1+2X implica che se si pone X=3, il valore della Y è quello ottenuto raddoppiando la X e aggiungendo l'unità: Y=7; questa è la dipendenza deterministica. Se la Y è costante al variare della X, ad esempio $y=4X^0$, la X non esercita alcuna influenza sulla Y (ovvero la Y non mostra variazioni al variare della X) e le due variabili sono indipendenti. Non solo, ma il ragionamento può essere svolto scambiando gli assi per cui in se la Y è funzionalmente indipendente dalla X questa lo è dalla Y. Pur importante come riferimento, la dipendenza o la indipendenza deterministica non si può estendere automaticamente ai fenomeni casuali.

b) Ekeland (1992, pp. 112-113) dubita persino della possibilità che esista l'indipendenza tra due fatti qualsiasi aldilà della astratta vita delle relazioni matematiche: "*Nell'universo non ci sono, né possono esserci eventi indipendenti. Il passante esercita dalla strada una forza di attrazione sulla tegola che si trova sul tetto dell'edificio e il colpo di vento che la fa cadere è inseparabile da tutto un contesto meteorologico in cui l'attività passata della vittima ha avuto la sua parte. Parlare di indipendenza non è altro che un'approssimazione comoda, una visione miope degli eventi che si deve necessariamente abbandonare se si ricerca una analisi più fine o un orizzonte più lontano".* Se nessun fatto può essere isolato da un altro per quanto diversi siano nell'ordine di grandezza e remoti sulla scala dello spazio-tempo perché ragionare di indipendenza?

In verità abbiamo già incontrato, senza rilevarli, questi problemi quando si è definito il dominio delle variabili dell'indagine statistica ed anche nel costruire l'universo degli eventi. L'esclusività delle manifestazioni della prova presuppone che ci si possa muovere in una realtà circoscritta nel cui perimetro solo alcuni eventi hanno rilevanza ed il resto è privo di interesse. Se non perdiamo di vista i limiti di applicabilità e la relatività delle nostre formulazioni, i risultati potranno ancora essere utili nonostante la consapevolezza del vortice di cause ed effetti in cui è immersa la piccola fetta di realtà che stiamo esaminando. In questo senso l'indipendenza fra due eventi è definita solo sul piano conoscitivo e cioè se l'apprendimento di un fatto offra o no un qualche fondamento razionale per aspettarsi il verificarsi dell'altro. In termini probabilistici, due eventi E ed F sono indipendenti se:

$$P(E|F) = P(E) \quad se \quad P(F) > 0$$

cioè se il verificarsi di F non altera o -allo stato delle nostre conoscenze- altera troppo poco per potersene rendere conto, la probabilità del verificarsi di E (cfr. De Cristofaro, 1992, p. 29 sul senso previsivo e non causale della indipendenza stocastica). Questa è l'indipendenza stocastica o in probabilità ("stocastico" significa sia congetturare, ma indica anche dei colpi tirati verso un bersaglio). In breve, non si afferma che l'evento E sia indipendente dall'evento F perché questo non influenza E (si tratterebbe di una affermazione più ampia di quanto non serva), ma solo che non si può verificare alcun evento che riguarda E che sia incluso anche in F, almeno in termini della funzione di probabilità adoperata nello spazio connesso all'esperimento.

Esempi:

a) Una società di marketing ha intenzione di condurre una ricerca sulla possibilità di migliorare il sistema a strappo dell'apertura delle lattine. Alcuni sondaggi preliminari hanno portato alla compilazione della seguente tabella relativa alle fasce d'età ed alla probabilità di acquistare una lattina per il costo indicato.

	<34	35-44	45-54	>55
Stessa cifra	22	23	18	15
Un po' di più	64	65	65	60
Molto di più	7	8	6	6
Rifiuto lattine	7	4	11	19
	100	100	100	100

Le prime due classi d'età presentano probabilità simili per cui sono da considerarsi stocasticamente indipendenti. Uno scostamento si realizza per le classi maggiori d'età evidenziando differenze di comportamento (e quindi dipendenza) rispetto ai più giovani.

b) In una ASL sono operative due apparecchiature per la TAC dislocate in edifici diversi e gestite da diverso personale. La prima ha probabilità 0.02 di disfunzione e l'altra ha probabilità 0.03 cosicché la probabilità che entrambe siano ferme è 0.0006 ritenendo vigente l'indipendenza. In realtà non è proprio così perché se una si ferma l'altra sarà sottoposta, almeno per un certo periodo, ad un superlavoro che potrebbe far lievitare la probabilità di guasto.

Esercizio_TP109: *nell'alfabeto italiano vi sono 10 lettere con tratti curvilinei {B, C, D, G, O, P, Q, R, S, U}. Verificare che tale informazione altera la probabilità a priori di ottenere una vocale;*

Esercizio_TP110: *da un sacchetto contenente 10 biglie di cui 5 bianche, se ne scelgono casualmente due. Sia E="Sono entrambe bianche" , F="Solo una è bianca". Gli eventi E ed F si possono considerare indipendenti? E' importante specificare che la scelta sia avvenuta con o senza reimmissione?*

Esercizio_TP111: *un'esperta giocatrice alla roulette consiglia:"seguite le uscite delle scommesse semplici (passe e mancque, ad esempio) e se per 3 volte ne esce una alla quarta volta giocate l'altra. E' una strategia razionale?*

Esercizio_TP112: *una persona temendo di arrivare in ritardo compra un biglietto dell'autobus perché così ha probabilità del 40% di essere in orario, chiama anche un tassì che gli dà probabilità del 70% di non fare tardi ed affitta una bicicletta che gli garantisce l'80% di probabilità di essere puntuale. Qual'è la probabilità che non faccia tardi? Si tratta di eventi indipendenti?*

Fattorizzazione della probabilità congiunta

La definizione dell'indipendenza porta alla formula moltiplicativa della probabilità (fattorizzazione):

$$P(E|F) = P(E) \;\Rightarrow\; \frac{P(E \cap F)}{P(F)} = P(E) \;\Rightarrow\; P(E \cap F) = P(E) * P(F)$$

L'indipendenza implica perciò che il verificarsi congiunto di due eventi indipendenti sia pari al prodotto delle rispettive probabilità. Detto in un altro modo, le tre probabilità seguenti:

$$P(F|E) = P(F); \qquad P(E|F) = P(E); \qquad P(E \cap F) = P(E)P(F)$$

sono tutte vere o tutte false contemporaneamente. L'ultima relazione, in particolare, mostra l'importanza del concetto di indipendenza: il fatto di poter calcolare la probabilità congiunta di due eventi dalla sola conoscenza della probabilità dei singoli eventi è, infatti, uno strumento teorico di grandissima rilevanza.

Esempi:

a) Un'impresa da intervistare per un sondaggio ha due proprietari. Poniamo E= "sono presenti entrambi i sessi" cioè E={mm,mf,fm,ff} ed F= "quello più anziano è di sesso maschile" per cui F={m,m,f,m}. Se le probabilità sono: P(mm)=0.20, P(mf)=0.22, p(f,m)=0.26, p(ff)=0.32, la conoscenza di "E" aiuta a conoscere "F"? P(F|E)=P(E∩F)/P(E)=P(F)/1=P(F) ovvero l'informazione non aggiunge nulla.

b) In uno *screening program* sull'ipertensione (misurata con la pressione sanguigna diastolica: DBP) si accerta che gli eventi A={moglie con DBP≥95} e B={marito con DBP≥95} hanno probabilità P(A)=0.1, P(B)=0.2 e P(A∩B)=0.02. Rosner (1990, p. 47) propone la seguente lettura: lo stato di ipertensione della moglie non dipende da quello del marito dato che nel 10% delle famiglie in cui la moglie è ipertensiva il marito non lo è, ma in un altro 10% lo è anche il marito. Se la causa fosse genetica questo risultato di indipendenza sarebbe quello atteso e smentirebbe un'eventuale ipotesi di causa ambientale.

c) Avete indetto una riunione per discutere una nuova strategia di vendita. L'addetta al marketing verrà con probabilità dell'80% e quella all'assistenza clienti al 95%. Ritenete che le decisioni di recarsi in riunione siano indipendenti. Con quale probabilità ne incontrerete almeno una?

$$P(M \cup C) = P(M) + P(C) - P(M \cap C) = 0.80 + 0.95 - 0.80 * 0.95 = 0.99$$

d) Vi viene suggerito un test per diagnosticare la solvibilità S o la volatilità V di un cliente per il credito al consumo. Si sa che i clienti solvibili sono il 60%, che il test risulta positivo nel 10% dei casi ed è basato su informazioni separate dalla condizione di solvibilità. Calcoliamo la specificità del test:

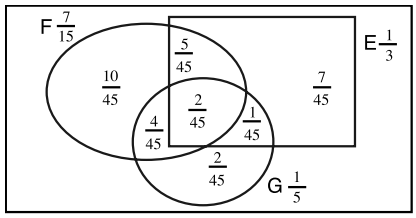
$$P(-|V) = \frac{P(- \cap V)}{P(V)} = \frac{P(\overline{+ \cup S})}{P(V)} = \frac{1 - P(+ \cup S)}{P(V)} = \frac{1 - P(+) - P(S) + P(+)P(S)}{P(V)} = 0.9$$

e) Fra i membri di una commissione il 40% eleggerebbe presidente Lojacono ed il 60% voterebbe per Cupiello se la votazione fosse fatta il giorno dell'intervista. Per studiare la stabilità del voto si scelgono a caso due membri e si indica: L1= "Una sola tra le persone prescelte vota per Lojacono" e L2= "Entrambe le persone prescelte votano per Lojacono". Gli eventi L1 ed L2 sono incompatibili dato che solo uno dei due si può verificare. Fra di essi c'è però dipendenza sia che la scelta avvenga con reimmissione che senza reimmissione.

Esercizio_TP113: *Parzen (1960, p. 89) propone il seguente quesito. Domenica giocano Bologna ed Inter. Gli eventi E="Bologna vince" e F="Inter vince" sono indipendenti, ma compatibili se le due squadre giocano contro squadre diverse. Sono dipendenti, ma incompatibili se si gioca Bologna-Inter. Come si spiega tale paradosso?*

Esercizio_TP114: *il treno da Cosenza a Salerno parte in orario con probabilità dell'80%; peraltro, parte e arriva in orario con probabilità del 70%.*
1. Qual'è la probabilità che se parte in orario arriva anche in orario;
2. Sapendo che la probabilità che se il treno parte in orario arriva in orario con probabilità del 75% qual'è la probabilità che il treno arrivato in orario sia anche partito in orario?

Esercizio_TP115: *verificare che, per gli eventi in figura, E è indipendente da G e da F, ma F e G sono dipendenti:*



Bilateralità della relazione di indipendenza

La relazione di indipendenza tra eventi è bilaterale:

$$P(E|F) = P(E) \;\Rightarrow\; P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E) * P(F)}{P(E)} = P(F) \quad per \quad P(E) > 0$$

Inoltre, la formula del prodotto, oltre a rendere evidente la simmetria P(E∩F)=P(F∩E), rende possibile definire l'indipendenza anche quando siano coinvolti eventi impossibili:

$$P(F \cap E) = P(E) * P(F)$$

Se P(E)=0 o P(F)+0 allora è nulla anche la probabilità dell'intersezione P(E∩F)=0.

Ciò è conforme all'idea che un evento a probabilità nulla non abbia elementi in comune con altri e che all'intersezione sia comunque assegnata probabilità zero. Se P(E)=0 allora ogni altro evento dell'algebra è indipendente da E (lo stesso si verifica se P(E)=1). E' per questa maggiore generalità che la formula del prodotto è preferibile al coinvolgimento esplicito delle probabilità condizionate.

Esempi:

a) La condizione di indipendenza, come si è detto, non è necessariamente ancorata alla natura degli eventi considerati o ad interrelazioni fisiche e logiche riscontrate nel fenomeno sotto analisi. Piuttosto è una conseguenza della funzione di probabilità e può cambiare se questa si modifica. Immaginiamo un esperimento in cui si lanciano un tetraedro ed un dado e sui 24 possibili esiti del lancio S={(1,1) (1,2) (1,3) (1,4) (1,5) (1,6) (2,1) (2,2) (2,3) (2,4) (2,5) (2,6) (3,1) (3,2) (3,3) (3,4) (3,5) (3,6) (4,1) (4,2) (4,3) (4,4) (4,5) (4,6)} si definiscono gli eventi: E= "esito del tetraedro maggiore o uguale dell'esito del dado" e F="somma dei due esiti= 7":

$$E=\{(2,1)\ (3,1)\ (3,2)\ (4,1)\ (4,2)\ (4,3)\ (1,1)\ (2,2)\ (3,3)\ (4,4)\}$$
$$F=\{(1,6)\ (2,5)\ (3,4)\ (4,3)\};\ E\cap F=\{(4,3)\} \Rightarrow P(E\cap F)=\frac{1}{24}\neq\left(\frac{10}{24}\right)\left(\frac{4}{24}\right)$$

Gli eventi sono quindi dipendenti. Si viene a sapere che C=" il risultato dei due poliedri non è uguale e l'1 del tetraedro non può abbinarsi con le uscite del dado superiori a 4"; si devono pertanto revisionare le probabilità:

$$E|C=\{(2,1)\ (3,1)\ (3,2)\ (4,1)\ (4,2)\ (4,3)\}$$
$$F|C=\{(2,5)\ (3,4)\ (4,3)\};\ E\cap F=\{(4,3)\} \Rightarrow P(E\cap F|C)=\frac{1}{18}=\left(\frac{6}{18}\right)\left(\frac{3}{18}\right)$$

Sotto la condizione C i due eventi sono indipendenti.

b) Si lanciano due tetraedri e sui possibili esiti S={(1,1) (1,2),(1,3) (1,4) (2,1) (2,2)(2,3) (2,4) (3,1) (3,2)(3,3) (3,4) (4,1) (4,2)(4,3)(4,4)} si definiscono gli eventi:

$$E=\{(1,1)\ (1,2)\ (1,3)\ (1,4)\ (2,1)\ (2,2)\ (2,3)\ (2,4)\}$$
$$F=\{(1,3)\ (2,3)\ (3,3)\ (4,3)\ (1,4)\ (2,4)\ (3,4)\ (4,4)\} \Rightarrow E\cap F=\{(1,3)\ (1,4)\ (2,3)\ (2,4)\}$$
$$P(E\cap F)=\frac{4}{16}=\left(\frac{8}{16}\right)\left(\frac{8}{16}\right)=P(E)P(F)$$

e quindi i due eventi sono indipendenti. Se si accerta che C= "La somma è maggiore o uguale a 6" e si riaggiustano le probabilità

$$E|C=\{(2,4)\}$$
$$F|C=\{(3,3)\ (4,3)\ (2,4)\ (3,4)\ (4,4)\};\ E\cap F=\{(2,4)\} \Rightarrow P(E\cap F|C)=\frac{1}{6}\neq\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)$$

i due eventi risultano ora dipendenti.

Esercizio_TP116:

a) Dimostrate che se $E,F\subset W$ e $P(E)>0$ e $P(F)>0$ allora E ed F non possono essere incompatibili e indipendenti;
b) Dimostrare che l'evento impossibile e l'evento certo sono indipendente da qualsiasi altro evento;
c) Dimostrate che se i due eventi sono indipendenti allora è vera una delle due asserzioni:
1) Almeno uno tra P(E) e P(F) è zero; 2) P(E\|F)=P(E) e P(F\|E)=P(F).

Indipendenza dei complementari

E' logico che la condizione: $P(E\cap F)=P(E)P(F)$ sussista anche tra i due eventi negati e tra un evento ed il negato dell'altro. Questo vuol dire che il verificarsi, diciamo di F, non solo non modifica la probabilità di E, ma anche quella di non E. D'altra parte, se così non fosse, alterando P(non E) si modificherebbe P(E) che ne è il complemento ad uno.

$$P(\bar{E}\cap\bar{F})=P(\overline{E\cup F})=1-P(E\cup F)=1-P(E)-P(F)+P(E\cap F)$$
$$=1-P(E)-P(F)+P(E)P(F)=[1-P(E)][1-P(F)]=P(\bar{E})P(\bar{F})$$

Analoga dimostrazione può essere data per le intersezioni tra eventi e loro complementari.

Esempi:

a) Suddivisione dei simpatizzanti del movimento contadino "Terra, subito!". Si vede subito che l'indipendenza è preclusa poiché le probabilità per i maschi aumentano con l'aumentare dell'età e diminuiscono invece per le femmine. Un controllo facile è:

$$P(M\cap 25-35)=\frac{203}{1320}=15.4\%\ \neq\ \left(\frac{700}{1320}\right)*\left(\frac{390}{1320}\right)=15.7\%$$

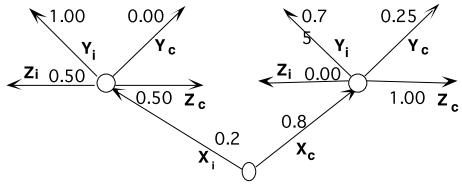
Età	Maschi	Femmine	Totale
< 25 anni	147	323	470
25 - 35	203	187	390
>35	350	110	460
Totale	700	620	1320

E' sufficiente che l'indipendenza sia sconfermata in una sola cella per definire dipendenti "Sesso" ed "Età", almeno secondo la tabella.

b) Ripreso da Drake (1967,p.21). Il presidente della commissione ritiene la signorina X colpevole di aver lanciato la copia del compito verso un altro candidato e attribuisce probabilità dell'80% all'evento. Le due colleghe Y e Z che siedono vicino ad X sono chiamate a testimoniare: Y è un'amica di X e se questa fosse colpevole mentirebbe per salvarla con probabilità del 25%; Z che non sopporta X mentirebbe se X fosse innocente con probabilità del 50%. Qual'è la probabilità che Y ed X diano un giudizio discorde?

$$P\left[(Y_c \cap Z_i) \cup (Y_i \cap Z_c) \middle| X_c\right] + P\left[(Y_c \cap Z_i) \cup (Y_i \cap Z_c) \middle| X_i\right] =$$

$$0.8 * 0.00 * 0.25 + 0.8 * 0.75 * 1 + 0.2 * 0.5 * 0.50 + 0.2 * 0.5 * 1 = 0.7$$



E' bene fare attenzione nel trattare con eventi incompatibili e con eventi indipendenti. I primi sono quelli che non si verificano mai insieme, i secondi sono quelli il cui verificarsi non altera le probabilità degli altri. La compatibilità/incompatibilità riguarda gli eventi e la dipendenza/indipendenza la probabilità. In ogni caso due eventi incompatibili sono sicuramente dipendenti, ma due eventi dipendenti non sono sempre incompatibili.

Esercizio_TP117:
a) dimostrare che la proprietà dell'indipendenza non è transitiva e cioè se A e B sono indipendenti e B e C sono indipendenti non necessariamente lo sono A e C. b) Dimostrare che A e B^c sono indipendenti se lo sono A e B.

Esercizio_TP118: *considerate la tabella del movimento "Terra, subito!" del precedente esempio e determinate quale debbano essere le entrate in caso di indipendenza.*

Domande sensibili

Per l'acquisizione di notizie molto delicate dalle persone su se stesse, sull'organizzazione cui appartengono o su altre persone si può ricorrere a varie tecniche. La più semplice è l'espressione in terza persona o comunque in forma indiretta dando a chi risponde la possibilità di non schierarsi apertamente (domande proiettive).

Esempi:
a) Questo prodotto è stato molto criticato. Su quali difetti ritiene si siano appuntate le lamentele dei clienti?

b) C'è chi sostiene che copiare durante un concorso non sia troppo scorretto se si ha veramente bisogno di ottenere il posto. Conoscete qualcuno cui è capitato di doverlo fare?

c) La pulizia personale è un segno evidente di civiltà, ma non occorre esagerare. Quali sono le attività che ritenete utili affinché ci si possa considerare una persona pulita?

La via più sicura per avere risposte su questioni riservate è la garanzia dell'anonimato più assoluto sia sulla persona che risponde che sulla risposta fornita. In questa direzione si colloca una tecnica molto interessante: le risposte casualizzate che mostrano un chiaro esempio in cui la Statistica aiuta a risolvere i problemi. L'idea è semplicissima e consiste nel porre ad ogni unità un quesito sull'argomento sgradevole (o su cui è sgradevole la domanda) che interessa sondare e, per non suscitare diffidenza eccessiva ed ottenere un certo numero di risposte valide, un'altra domanda più tranquilla che abbia la stessa modalità della prima, ma le cui percentuali di scelta tra le varie opzioni siano note e stabili nella popolazione indagata. Alla persona intervistata si chiede di rispondere casualmente ad una delle due domande di modo che chi intervista non sia in grado di conoscere la risposta data.

Esempio:
Un'indagine in un quartiere degradato richiede una domanda sull'uso di stupefacenti. Non necessariamente l'unità deve rispondere alla domanda sensibile. In base ad un meccanismo di sorteggio (ad esempio la scelta di una cifra a caso) può rispondere sul consumo di droga leggera oppure a quella più pacifica sulla lettura di quotidiani sportivi.

Prima di rispondere alle domande di questo riquadro è necessario lanciare (riservatamente) un dado: se il risultato è tre o meno di tre si deve rispondere alla domanda "A" , se è superiore a tre si risponderà alla domanda "B"				
A.	Ha fatto uso di spinelli?	☐ No	☐ Raramente	☐ Spesso
B.	Legge un quotidiano sportivo?			

Da precedenti indagini si sa che tra le persone intervistate il quotidiano sportivo è letto con le seguenti percentuali: No=20%, Raramente=35%, Spesso=45%. Le risposte sulla domanda casualizzata sono: No=45%, Raramente=30%, Spesso=25%, come si possono utilizzare queste informazioni?

Supponiamo che il meccanismo di sorteggio dia alla domanda più sensibile una *chance* "p" di essere scelta (nel caso in esempio p=0.5) e siano: λ la frazione di risposte complessive date ad una modalità di risposta, π_2 la frazione nota di chi sceglie quella modalità nella domanda tranquilla e π_1 la frazione incognita chi sceglie quella modalità per la domanda sensibile. Tra queste proporzioni -tenuto conto dell'indipendenza delle due domande esiste la relazione: $\lambda=p\pi_1 + (1-p)\pi_2$ dalla quale è facile ricavare l'unica vera incognita:

$$\pi_1 = \frac{\lambda - (1-p)\pi_2}{p}$$

Esempio:
Se dall'indagine risulta il 45% di "Si", allora la percentuale di favorevoli all'aborto libero e gratuito è

$$\pi_1 = \frac{0.45 - (1-0.5)0.5}{0.5} = 40\%$$

Prima di rispondere a questa domanda, lanciate in aria una moneta: se è testa rispondete alla "1" altrimenti alla "2"

1) Ritenete giusto che l'aborto sia libero è gratuito?
2) La somma dei vostri numeri di matricola è pari?

NO	SI

Esercizio_TP119: *dovete fornire una consulenza per una indagine sui furti nei supermercati effettuati da persone "perbene". Che tipo di domande proporreste per equilibrare l'esigenza di risposte accurate e salvaguardare la privacy di chi risponde?*

6.4.3 Indipendenza di "n" eventi

Per evitare le difficoltà del concetto di indipendenza e per esaltarne -con Kolmogorov- la mera natura concettuale, diremo che una m-tupla è costituita da eventi indipendenti se:

$$P\left(E_{k_1} \cap E_{k_2} \cap \ldots \cap E_{k_m}\right) = \prod_{i=1}^{m} P\left(E_{k_i}\right)$$

per ogni permutazione degli indici distinti ($2 \le k_1 < k_2 < \ldots < k_m \le m$). Questo significa che tutte le possibili coppie di eventi sono indipendenti: $P(E_i \cap E_j)=P(E_i)P(E_j)$ per $i \neq j$ e sono indipendenti anche tutte le combinazioni di tre eventi:$P(E_i \cap E_j \cap E_k)=P(E_i)P(E_j)P(E_k)$ per $i \neq j \neq k$ e così via fino ad arrivare alla indipendenza della m-tupla.

Esempio:
La condizione sulla terna e sulla coppia sono entrambe necessarie per assicurare l'indipendenza di ogni evento da tutti gli altri.

$$P\left[\left(E_i \cup E_j\right) \cap E_k\right] = P\left[\left(E_i \cap E_k\right) \cup \left(E_j \cup E_k\right)\right] = P\left(E_i \cap E_k\right) + P\left(E_j \cap E_k\right) - P\left(E_i \cap E_j \cap E_k\right)$$
$$= P\left(E_i\right)P\left(E_k\right) + P\left(E_j\right)P\left(E_k\right) - P\left(E_i\right)P\left(E_j\right)P\left(E_k\right) = \left[P\left(E_i\right) + P\left(E_j\right) - P\left(E_i\right)P\left(E_j\right)\right]P\left(E_k\right) = P\left(E_i \cup E_j\right)P\left(E_k\right)$$

Peraltro, la sola condizione sulla terna potrebbe non bastare per l'indipendenza delle coppie così come l'indipendenza a due a due (*pairwise independence)* non garantisce l'indipendenza delle terne.

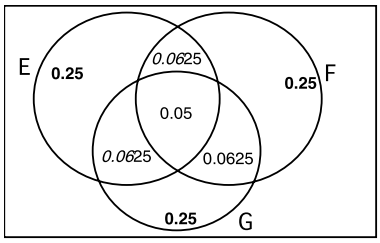
Esempi:
a) Ripreso da Cifarelli (1997, q,®69). Nel lancio di un dado si applica l'equiprobabilità e si considerano gli eventi: E={2,4,6), F={3,4,5,6}, G={3,5,6} con P(E)=1/2, P(F)=2/3, P(G)=1/2. Inoltre $P(E \cap F \cap G)=P(\text{"6"})=1/6$, ma anche $P(E)P(F)P(G)=1/6$ e quindi è soddisfatta la condizione sulla terna; non lo è per le coppie: $P(F \cap G)=1/2 \neq P(F)P(G)=1/3$, $P(E \cap G)=1/6 \neq P(E)P(G)=1/4$; solo $P(E \cap F)=P(E)P(F)=1/3$.

b) Paradosso di Bernstein. Ad un tetraedro si applica il modello di probabilità uniforme. Si considerano gli eventi E={1,2}, F={1,3}, G={1,4} con P(E)=P(F)=P(G)=1/2. Si può subito accertare che $P(E \cap F)=P(E)P(F)=1/4$, $P(E \cap G)=P(E)P(G)=1/4$, $P(F \cap G)=P(E)P(G)=1/4$ e che quindi la condizione di indipendenza è valida per coppie, ma non per la terna: $P(E \cap F \cap G)=1/4 \neq P(E)P(F)P(G)=1/8$.

c) Fisz (1963, p. 25). Un'urna contiene 4 bussolotti con all'interno i numeri: 110, 101, 011, 000. Indichiamo con E_i= " La cifra "1"compare in posizione i-esima". Si estrae a caso un bussolotto dall'urna e si considera il numero estratto sul quale si valutano gli eventi E_i. Si tratta di eventi indipendenti? Per ognuno i casi favorevoli sono due per cui $P(E_1)= P(E_2)=P(E_3)=2/4=0.5$. L'evento $(E_1 \cap E_2 \cap E_3)$ ha probabilità zero poiché è impossibile (nessun bussolotto contiene 111); se fossero indipendenti tale evento dovrebbe avere probabilità 1/8. Comunque, i tre eventi sono mutualmente indipendenti: $P(E_2|E_1)=1/2=P(E_2)$, $P(E_3|E_2)=1/2=P(E_3)$, $P(E_3|E_1)=1/2=P(E_3)$.

Esercizio_TP120: *l'indipendenza di "m" eventi implica ipotizzare la validità di un certo numero di equazioni. Ad esempio, l'indipendenza a due a due ne comporta C(m,2) combinazioni.*
a) Quante relazioni implica l'indipendenza di "m" eventi? b) Cambia qualcosa nel conteggio se nella relazione: di indipendenza alcuni eventi sono sostituititi dal loro complemento?

Esercizio_TP121: *in base agli elementi probabilistici riportati in figura verificate che l'evento E sia indipendente dall'evento F e dall' evento G presi separatamente, ma non dall'evento F∩G.*



Se $\{E_1, E_2 \ldots E_m\}$ sono indipendenti lo è qualsiasi insieme di "n" eventi con n≤m.

$$P\left(E_{k_1} \cap E_{k_2} \cap \ldots \cap E_{k_{n-1}}\right) = \sum_{k_n} P\left(E_{k_1} \cap E_{k_2} \cap \ldots \cap E_{k_n}\right) = \sum_{k_n} \prod_{i=1}^{n} P\left(E_{k_i}\right) = \prod_{i=1}^{n-1} P\left(E_{k_i}\right) \sum_{k_n} P\left(E_{k_n}\right) = \prod_{i=1}^{n-1} P\left(E_{k_i}\right)(1) = \prod_{i=1}^{n-1} P\left(E_{k_i}\right)$$

Si può accertare che ciò che è vero per (n-1) in rapporto ad "n" è vero anche per (n-2) in rapporto ad (n-1) fino ad arrivare al punto desiderato.

L'indipendenza è una condizione forte che talvolta sembra porsi contro il senso comune.

Esempi:
a) Ciccillo è un affezionato del 12 sulla ruota di Napoli. Indichiamo con E_i l'evento "Esce il 12 nella estrazione i-esima". Non si ha motivo di dubitare della indipendenza tra un'estrazione e l'altra. Ciccillo ha notato che il 12 non è uscito per 150 estrazioni. Che probabilità ha di uscire alla 151ª?

$$P\left(E_{151} \middle| E_1^c \cap E_2^c \cap \ldots \cap E_{150}^c\right) = \frac{P\left(E_1^c \cap E_2^c \cap \ldots \cap E_{150}^c \cap E_{151}^c\right)}{P\left(E_1^c \cap E_2^c \cap \ldots \cap E_{150}^c\right)} = P\left(E_{151}\right)$$

L'indipendenza tra le varie estrazioni impedisce, almeno in via teorica, il formarsi di una memoria nel congegno. Con un ragionamento simile si dimostra che la probabilità è la stessa non solo dopo 10, 100, 1000 estrazioni, ma che non c'è sequenza di ritardi, per quanto grande, che potrà mai modificare la probabilità di uscita del "12". Attenzione! Questo non significa che il "12" non uscirà, ma solo l'assenza di raziocinio nell'idea che la propensione ad uscire aumenti con il ritardo.

b) Blom (1989, p. 29) rileva come l'idea del ritardo che favorisce le uscite conviva col suo opposto allorché la ricevitoria in cui è appena avvenuta una forte vincita riscontri un aumento delle giocate soprattutto di clienti non abituali. Blom attribuisce tale attenzione alla convinzione -nei giocatori- che la sorte abbia preso a benvolere il locale. L'ipotesi è condivisibile, ma c'è un fattore che accomuna i due atteggiamenti. La fiducia nella sorte, o meglio nei meccanismi che la simulano, non è piena e gli scommettitori ritengono che un qualche difetto (colposo o doloso) nel meccanismo riduca o aumenti le *chances* di qualche evento e la vera abilità è di scoprirlo per sfruttarlo a proprio vantaggio. La sorte non teme critiche, non ha bisogno di compensare subito gli squilibri che creano i suoi capricci. Certamente lo farà, ma nei tempi che vuole in cui i 15 miliardi di anni serviti a formare l'universo contano meno del frullio d'ali del colibrì.

c) Il mago Sibillinus ha adottato questa strategia: incontra regolarmente i propri clienti costringendoli a ridurre le loro questioni ad una domanda con due sole risposte, diciamo Sì/No. Ad una metà -scelta casualmente- consiglia il "Sì" ed all'altra il "No". Nella consultazione successiva gli rimane la metà dei clienti cui ha dato risposta corretta. Tra questi ripartisce, sempre casualmente, il "Sì" ed il "No" cosicché nel nuovo turno di consultazioni ne ritrova solo la metà. Nuova suddivisione casuale dei consigli e perdita di un'altra metà. Sibillinus ha solo un cliente su 8 di quelli originali, ma sono ora clienti qualificati dato che ha loro fornito tre pronostici corretti consecutivi. Dopo altri cinque turni di consigli gli rimane solo un cliente su 256 di quelli che si erano rivolti a lui inizialmente, ma questi cui ha predetto il vero per ben 8 consulti: un evento del genere ha probabilità $(0.5)^8$=0.004. Se il mago avesse inizialmente abbindolato 50'000 persone ne avrebbe ora circa duecento che sono pronte a versargli l'intero patrimonio e seguirlo ovunque.

d) Thomasian (1969, p. 101) parte dalla relazione $1-x \leq e^{-x}$ per ogni numero reale "x". Riconsideriamo la disuguaglianza di Bonferroni:

$$P\left(\bigcup_{i=1}^{m} E_i\right) \geq P\left(\bigcap_{i=1}^{n} E_i\right) \geq 1 - \sum_{i=1}^{n} P\left(E_i^c\right) \geq 1 - \sum_{i=1}^{n} \left[1 - P\left(E_i\right)\right] \qquad P\left(\bigcup_{i=1}^{m} E_i\right) \geq 1 - e^{-\sum_{i=1}^{m} P\left(E_i\right)}$$

Poiché $1-P(E_i) \leq e^{-P(Ei)}$ si ottiene un limite inferiore alla probabilità dell'unione.

Esercizio_TP122: *un'esperimento consiste nel lanciare per due volte una moneta regolare. Si considerino gli eventi: A={(T,T), (T,C}, B={(T,T), (C,T)}, C={(T,T), (C,C)}. Verificare che gli eventi A, B, C sono mutualmente indipendenti, ma non lo è la terna.*

Esercizio_TP123: *l'affidabilità di un sistema è misurata dalla probabilità che continui a funzionare in condizioni di stress. Se un motore ha 3 cilindri che, separatamente e indipendentemente, sono operativi con probabilità del 99%, qual'è l'affidabilità del motore se questo è in grado di funzionare anche con due soli cilindri?*

Esercizio_TP124: *In un'urna sono stati inseriti 9 bussolotti contenenti 5 consonanti e 4 vocali. Scegliendo a caso e senza reimmissione un bussolotto alla volta, qual'è la probabilità della sequenza: CVCVCVCVC?*

Eventi curiosi

Grazie all' indipendenza si determina la probabilità di eventi singolari del tipo: numero di matricola coincidente con la data di nascita, biglietto di lotteria con serie uguale al numero di telefono o della targa della macchina. La natura "sorprendente" di tali eventi è dovuta al solo fatto di prestare loro attenzione perché ci colpisce la straordinarietà di una coincidenza e non che la coincidenza sia in effetti straordinaria. L'attenzione selettiva, infatti, trascura tutte le volte in cui la circostanza non si è verificata e che potrebbero essere numerosissime.

Esempi:

a) E' noto il caso della signora Adams che, nell'ottobre 1985, ha vinto la lotteria statale del New Jersey incassando circa otto miliardi e nel febbraio 1986 ha vinto la stessa lotteria incassando altri tre miliardi. Nel valutare coincidenze e accadimenti rari non bisogna perdere di vista il numero reale di tentativi effettuati Quindi non solo quello -riuscito- della signora Adams, ma tutte le giocate di tutti i giocatori nelle lotterie in cui l'insorgere dell'evento avrebbe destato -senza motivo- meraviglia.

b) Nella "Tammurriata nera" si tenta di spiegare la nascita di Ciro, bambino mulatto, da una donna bianca ricorrendo ai grandi numeri: "*Chisti fatti nun so' rari, se ne vegono a migliaia"* ed è solo la curiosità morbosa e impicciona del vicinato che fa apparire fuori dal comune un evento normalissimo. Purché si guardi a tutto il mondo e a tutte le epoche, non solo alla microrealtà del vicolo.

c) Siano $\{E_1, E_2 \ldots E_n\}$ indipendenti. Calcoliamo la probabilità che almeno uno si verifichi:

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = 1 - P(\overline{E_1 \cup E_2 \cup \ldots \cup E_n}) = 1 - P(E_1^c)P(E_2^c)\ldots P(E_n^c) = 1 - \prod_{i=1}^{n}(1 - p_i)$$

Se gli eventi hanno la stessa probabilità "p", posto q=1-p si ha: P(almeno uno si verifica)=1-q^n. Immaginiamo un automobilista che ha probabilità di incorrere in un sinistro pari a 1:500'000 e che guidi per 60 minuti al giorno. I percorsi su cui si muove sono tali che ogni 15 secondi ci sia un rischio (ipotizziamo l'indipendenza). La probabilità di un sinistro in 5 anni è:

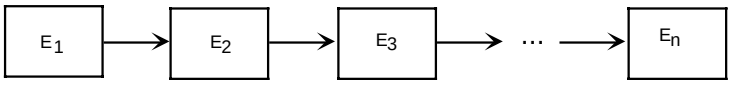
$$P(almeno\ \ un\ \ incidente) = 1 - \left(1 - \frac{1}{500'\,000}\right)^{438'\,000} = 58.35\%$$

Anche eventi con probabilità irrisorie possono verificarsi se la prova è replicata un numero elevato di volte. Quindi la probabilità piccola di incidente non deve, di per sé, fornire sicurezza all'automobilista.

Esercizio_TP125: *un pubblico ministero nell'accusare un imputato sostenne: se un evento si verifica solo una volta è un incidente; se si verifica due volte è una coincidenza; se si verifica tre volte è una prova. Esprimete una vostra opinione.*

Esercizio_TP126: *un sistema di compone di n elementi che possono funzionare con probabilità "p" e non funzionare con probabilità (1-p).*

1. Se le disfunzioni siano indipendenti qual'è la probabilità che il sistema non funzioni se a questo fine è sufficiente che anche un solo elemento smetta di funzionare?



2) A quale valore la probabilità se "n" tende all'infinito?

Indipendenza nelle inclusioni/esclusioni sequenziali

Dati "n" eventi qualsiasi $\{E_1, E_2, \ldots, E_n\}$ la probabilità della loro unione, come si è visto, è:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} P(E_i \cap E_j) + \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} P(E_i \cap E_j \cap E_k) - \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} E_i\right)$$

che costituisce la notevole semplificazione ottenibile in caso di indipendenza e che spiega (ma non giustifica) il ricorso frequente a tale ipotesi nella formulazione di modelli sperimentali e teorici.

Esempi:

a) Gli "n" creditori di Paolino Paperino decidono di incontrarlo percorrendo ognuno strade diverse. Se ciascuna successione di "n" creditori ha probabilità 1/n! di costituirsi, qual'è la probabilità che Paperino sfugga all'assedio? Ragioniamo sul suo complemento e cioè calcoliamo la probabilità che almeno uno lo incontri. Numeriamo i creditori da uno ad "n" ed indichiamo con E_i l'evento che Paperino sia affrontato dal creditore i-esimo. I casi favorevoli sono quelli in cui l'i-esimo creditore lo incontri per i-esimo a prescindere da quello che fanno gli altri: (n-1)! e $P(E_i)$=(n-1)!/n!=1/n. L'evento che i creditori (i,j) centrino l'obiettivo all'incontro i-esimo e j-esimo ha (n-2)! casi favorevoli poiché due posizioni sono fisse mentre le altre permutano e quindi $P(E_i \cap E_j)$=(n-2)!/n!=1/[n(n-1)]. Un gruppo di r-creditori ha probabilità $P(E_i \cap E_j \cap \ldots \cap E_r)$=(n-r)!/n! di incontrare Paperino nella sequenza prescritta dai loro indici. A questo punto la probabilità che almeno uno acchiappi il papero è:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \binom{n}{n}\frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} + \ldots (-1)^{n-1}\binom{n}{n}\frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + \frac{(-1)^{n-1}}{n!} \cong 1 - e^{-1} = 0.6321$$

L'approssimazione è già buona per n=6 (errore inferiore a due decimillesimi). Poiché i creditori di Paperino sono molti di più ogni volta che esce ha circa tre *chances* contro due di incontrarne uno. Da notare che la probabilità è quasi la stessa con n=10 o con n=100 e qui Paperino ha ragione ad indebitarsi sempre più.

b) Un bambino colleziona i premi inclusi nelle merendine. I premi sono numerati da 1 ad "m". Il bambino tenta di convince il padre a comprare n>m confezioni. Prima di sostenere la spesa si calcola qual'è la probabilità di ottenerne almeno uno di ogni tipo. Ipotizziamo che ogni uscita sia equiprobabile cioè abbia probabilità n^{-m} ed indichiamo con E_i l'evento "regalo i-esimo mancante nella confezione". I casi a favore sono $(n-1)^m$ dato che i premi sono allocati sulle restanti (n-1) confezioni in forma di permutazione con ripetizione. Allo stesso modo i casi favorevoli alla mancanza di due tipi sono $(n-2)^m$ e così via:

$$P(E_i) = \frac{(n-1)^m}{n^m} = \left(1-\frac{1}{n}\right)^m ;\ P(E_i \cap E_j) = \frac{(n-2)^m}{n^m} = \left(1-\frac{2}{n}\right)^m ;\ldots;\ P(E_i \cap E_j \cap \ldots E_r) = \left(1-\frac{r}{n}\right)^m ;$$

La probabilità di finire la collezione è il complemento ad uno della probabilità che ne manchi almeno uno:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \binom{n}{1}\left(1-\frac{1}{n}\right)^m - \binom{n}{2}\left(1-\frac{2}{n}\right)^m + \ldots + (-1)^{n-1}\binom{n}{1}\left(1-\frac{n}{n}\right)^m$$

se m=4 e n=5 la probabilità è del 30.4%; se n=10 la probabilità passa al 40.1% e se n=20 la probabilità è del 45% quindi non sembra conveniente comprare troppe confezioni.

Esercizio_TP127: *il mago Sibillinus si presenta ad un centro ricerche sulle percezioni extrasensoriali affermando di poter indovinare la sequenza di uscite di n=13 carte da gioco francesi. Le carte numerate da "1" ad "13"sono mischiate con cura e disposte in fila. Sibillinus dà la sua sequenza. Se ne indovina 11 o più firmerà un contratto di collaborazione molto ricco. Qual'è la probabilità -se rispondesse a caso- di tale evento?*

Modello moltiplicativo

L'analisi di una prova può essere spesso condotta con una articolazione in "n" sottoprove definendo inizialmente uno spazio di probabilità $(S_i, W_i, P_i\}$ specifico per ogni sottoprova. L'evento elementare dell'esperimento complessivo sarà una n-tupla ottenuta scegliendo in sequenza ordinata un elemento da ciascun S_i:

$$(e_1, e_2, \ldots, e_m) \in \left\{(E_1, E_2, \ldots, E_n) \middle| E_1 \in S_1,\ E_2 \in S_2,\ \ldots, E_n \in S_n\right\} = S_1 \otimes S_2 \otimes \ldots \otimes S_n = S$$

cioè l'ambito dell'esperimento multiplo sarà il prodotto cartesiano degli S_i e $W=W_1 \otimes W_2 \otimes \ldots \otimes W_n$ è l'algebra indotta da S (Loève, 1977, pp.104-105, 155-156).

Esempio:

Un processo di fabbricazione produce *item* eccellenti (E con p(E)=40%), buoni (B con p(B)=30%), tollerabili (T, con p(T)=25%) e difettati (D con p(D)=5%). Consideriamo come sottoprova la qualità di un singolo *item.* L'universo degli eventi alla prova i-esima è S_i={B, D, E, T}. L'universo degli eventi per tre prove consecutive include 64 eventi da (B,B,B) a (T,T,T). L'evento (B,D,B,T,B,T,D,D,B,E} è un evento elementare dello spazio prodotto costruito su n=10 sottoprove.

Per l'assegnazione della probabilità all'evento prodotto $E=(e_1, e_2, \ldots, e_n)$ a partire dalle funzioni di probabilità dei sottospazi è di aiuto la nozione di esperimenti indipendenti (in verità è proprio questa la convenienza di frazionare l'esperimento in subesperimenti). Se le parti di un esperimento multiplo sono mutualmente indipendenti (basta questo tipo di indipendenza) la probabilità può essere assegnata in base alla formula moltiplicativa della probabilità:

$$P(E) = P\left[\left(e_1 \in S_1, e_2 \in S_2, \ldots e_n \in S_n\right)\right] = \prod_{i=1}^{n} P_i\left(e_i\right)$$

Un meccanismo di questo genere fornisce probabilità non negative. Per la probabilità dell'evento certo si ha:

$$\sum_{e_1 \in S_1} \sum_{e_1 \in S_2} \cdots \sum_{e_1 \in S_n} P\left[\left(e_1, e_2, \ldots e_n\right)\right] = \sum_{e_1 \in S_1} \sum_{e_1 \in S_2} \cdots \sum_{e_1 \in S_n} \prod_{i=1}^{n} P_i\left(e_i\right) = \prod_{i=1}^{n}\left[\sum_{e_i \in S_i} P_i\left(e_i\right)\right] = \prod_{i=1}^{n}[1] \; = 1$$

L'additività segue dalla additività delle singole funzioni che compongono quella definita per lo spazio prodotto.

Esempi:

a) Uno scaffale contiene 30 prodotti di tipo A e 10 di tipo B. Un altro scaffale ne contiene 50 di tipo A e 25 di tipo B. Un cliente frettoloso sceglie a caso un prodotto da ciascuno dei due scaffali. Se le due scelte sono indipendenti, qual'è la probabilità che entrambi i prodotti siano A?

$$P\left(S_A \;\; x \;\; S_B\right) = P\left(S_A\right)P\left(S_B\right) = \frac{30}{40}\frac{25}{50} = \frac{750}{2000} = 0.375$$

b) Le variazioni percentuali di un indice di borsa sono meglio seguite pensando ad ogni chiusura come una sottoprova il cui esito sia descritto da S_i={"+,"-","0"} c con P(+)=p, P(-)=q e P(0)=1-p-q. L'algebra W_i conterrà gli eventi: {ϕ ,S,["+"],["-"],["0"],["+", "-"],["+", "0"]["-", "0"]} e pertanto l'algebra del prodotto sarà: $W=W_1 \otimes W_2 \otimes \ldots \otimes W_n$ Se si ipotizza l'indipendenza è possibile calcolare, ad esempio, la probabilità che dopo 4 variazioni negative l'indice chiuda la settimana con un variazione positiva o che non ci siano variazioni per l'intera settimana:

$$P(-,-,-,-,+) = P(-)P(-)P(-)P(-)P(+) = q^4 p; P(0,0,0,0,0) = P(0)P(0)P(0)P(0)P(0) = (1-p-q)^5;$$

c) Il modello moltiplicativo può descrivere anche l'estrazione senza reimmissione. Infatti, immaginiamo un'urna che contenga N biglie numerate e di doverne estrarre "n"; le scelte sono considerate indistinte purché contengano le stesse biglie a prescindere dall'ordine in cui si presentano. Ognuna di tali estrazioni è una sottoprova il cui dominio è ridotto di una possibilità ad ogni estrazione cioè nella prova i-esima i casi possibili sono (N-i+1). Applichiamo il modello di probabilità uniforme ad ogni sottoprova: $P(e_i)=1/(N-i+1)$. All'evento $(e_1, e_2, \ldots, e_n)$ si deve assegnare la probabilità prodotto perché le sottoprove sono indipendenti. Infatti, avendo abolito l'ordinamento, ogni singola biglia può manifestarsi in una qualsiasi delle sottoprove (mai in più di una a causa della mancata reimmissione). Ne consegue:

$$P\left[e_1, e_2, \ldots, e_n\right] = \frac{1}{D_{SR}\,(N,n)} = \frac{1}{N(N-1)(N-2)\ldots(N-n+1)}$$

Esercizio_TP128: *un investitore forma il suo portafoglio titoli affidandosi alla sorte con un esperimento a più stadi: al primo stadio sceglie una ed una sola tipologia in S={Bot, Azioni, Obbligazioni} con P(B)=0.25, P(A)=0.40, P(O)=35. Fatta la scelta, decide di investire da uno a 10 milioni secondo la funzione di probabilità:*

$$P(i) = \frac{i^3}{3025}; \;\; i = 1, 2, \ldots, 10$$

Calcolate le probabilità dei vari eventi dell'esperimento.

Parzen (1960, p. 96) osserva che non tutti gli elementi dello spazio prodotto S possono considerarsi degli eventi prodotto. Tuttavia, è possibile dimostrare che esiste un modo univoco di definire la funzione di probabilità P(.) in riferimento allo spazio prodotto.

Esercizio_TP129*: verificare che negli esperimenti indipendenti siano valide le seguenti condizioni:*

$$Se \;\; E_1 \subset S_A \;\; e \;\; E_2 \subset S_B \Rightarrow \begin{cases} P_{ab}\left(E_1 \in S_A \;\; qualunque \;\; sia \;\; E_2\right) = P_A\left(E_1\right) \\ P_{ab}\left(E_2 \in S_B \;\; qualunque \;\; sia \;\; E_1\right) = P_B\left(E_1\right) \end{cases}$$

Il modello moltiplicativo è un concetto semplice e analiticamente potente tanto da potersi considerare il punto più sviluppato nella teoria della probabilità (vi ritorneremo in altre parti del corso). Feller (1950, p. 132) invita a fare ogni sforzo per esprimere gli esperimenti complessi come prodotto di prove indipendenti. A questo però fa da freno il monito di Hodges e Lehmann (1970, p. 98) che avvertono di non indulgere troppo nell'uso di questo modello a causa della sua facilità d'uso perché ha una validità limitata dal presupposto che l'esito di un esperimento non influenzi l'esito dell'altro e non dovrebbe essere impiegato nei casi in cui questa condizione non sia verificata o almeno non sia verificata in modo sostanziale.

6.5 Selezione delle unità

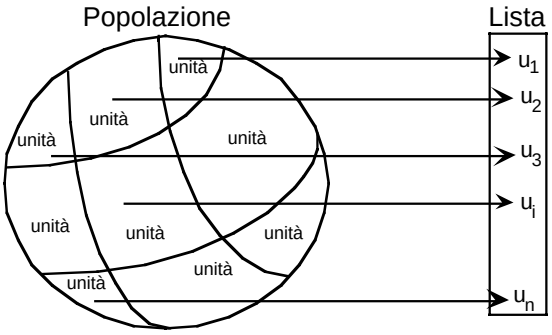
La crescente mole di informazioni che è necessario trattare in tante occasioni rende spesso impraticabile, ammesso che sia possibile ed opportuno, l'analisi di tutti i dati disponibili ed occorre procedere a qualche semplificazione. In questo paragrafo discuteremo il problema di come circoscrivere -grazie al calcolo delle probabilità- il numero di unità e dei modi in cui realizzare la loro scelta: il campionamento. Questa è una procedura fondamentale della Statistica e per la sua complessità è in genere presentata in cori più avanzati. In essa però, come osserva Kish (1965, p.4) c'è un duplice aspetto: la procedura di selezione (ovvero l'insieme di regole ed operazioni con cui si realizza la scelta delle unità) e la procedura di stima che riguarda i calcoli delle statistiche, il loro uso come valori presunti dei corrispondenti parametri della popolazione, l'accertamento del perché la popolazione presenti determinate caratteristiche. Il presente paragrafo è dedicato al primo aspetto (una scelta simile è fatta ad esempio in Becker ed Harnett, 1987) perché così si può dare la prima giustificazione della messa in opera di un impianto teorico così sofisticato -il calcolo della probabilità- della cui utilità aggiuntiva rispetto alle nozioni già impartite di statistica descrittiva qualcuno avrà delle perplessità.

Ad ogni unità soggetta ad indagine è attribuita una certa probabilità di essere effettivamente coinvolta: se si considerano tutte le unità, la probabilità non svolge alcun ruolo. Se non è possibile effettuare un'indagine completa (cfr. il paragrafo 1.2) ed occorre esaminare un campione ci saranno unità effettivamente esaminate ed altre no. Avremo modo di mostrare che se la selezione delle unità avviene in base alla sorte (campione casuale) le statistiche calcolate in base a esso tendono ad essere più attendibili; tuttavia, con l'inclusione o l'esclusione probabilistica delle unità, ci si trova di fronte a dei dati che sono quelli, ma avrebbero potuto essere altri, cosa si può dire allora sui risultati ottenuti?

6.5.1 Popolazione teorica ed effettiva

L'identificazione delle unità è una esigenza evidente. A prima vista non appare un compito improbo: se ci interessa conoscere l'atteggiamento sul contratto aziendale cercheremo sui ruolini paga i nominativi ed i recapiti dei dipendenti ai quali domanderemo -a tutti o a qualcuno- un'opinione in proposito. Se si deve notificare un atto giudiziario o contestare la violazione di qualche norma del codice della strada si cercherà all'anagrafe o al pubblico registro automobilistico i riferimenti del cittadino destinatario. Non sempre però la soluzione è così ovvia. Le popolazioni cui si rivolge la Statistica sono spesso formate da unità mai registrate in nessun elenco, repertorio, rubrica, albo, annuario, guida. Altre sono elencate in schedari vecchi e/o incompleti, altre ancora sono unità che vogliono rimanere celate.

Sia U la popolazione teorica e rappresentiamo con u_1, u_2, …, u_N, le sue unità (N indica l'ampiezza, finita o infinita, nota o incognita, della popolazione).



Il criterio organizzativo si traduce in una corrispondenza biunivoca tra le unità e l'insieme dei numeri naturali consecutivi che esprime l'ordine, arbitrario o precostituito, di considerazione delle unità. Questo però basta solo per ragionamenti teorici e negli esercizi di molti corsi di Statistica, ma non è così che riusciamo a gettare le basi di una corretta scelta delle unità.

Esempi:
a) L'iscrizione all'Università si materializza anche nell'assegnazione della matricola che semplifica la ricerca ed il controllo dei dati amministrativi dello studente. Di solito è assegnato in base all'ordine di presentazione agli sportelli anche se, sapendo che certi corsi si sdoppiano per numeri pari e dispari, qualcuno interessato potrebbe farsi assegnare un numero di matricola della parità desiderata.

b) Il codice fiscale contiene importanti dati identificativi del soggetto e permette di rintracciare buona parte delle transazioni legali in cui sono coinvolti i cittadini. L'elenco dei numeri già richiesti e lo schema di codifica sono gestiti dal Ministero delle Finanze.

c) Le titolarità dell'abbonamento alla televisione sono identificate attraverso il numero dell'abbonamento. L'elenco degli abbonati e la generazione dei nuovi numeri sono controllati dalla RAI.

d) Gli autori di software possono far valere i loro diritti esclusivi di utilizzazione economica dei programmi per computer registrandosi in un apposito albo tenuto dalla SIAE (Società italiana degli autori ed editori). La registrazione è onerosa e prevede l'indicazione del titolo del software, dei dati anagrafici dell'autore, data e luogo di pubblicazione del programma. L'albo può però anche funzionare da *frame* per diverse indagini statistiche.

e) La Guida Monaci fornisce una base di dati relativa a circa 400 mila voci divise tra aziende, enti e persone fisiche ad esse riferite. I prodotti legati alla guida sono un utile strumento per aggiornate la propria clientela potenziale, oltre a favorire operazioni di *marketing* e sondaggi.

Una fase necessaria di ogni trattazione statistica è perciò la definizione di un sistema di riconoscimento ed individuazione delle unità che permetta di distinguerle senza incertezze e consenta altresì, anche solo in via teorica, di raggiungerle singolarmente per poterne acquisire i dati su tutte le variabili di interesse. Tali unità formano la popolazione teorica: $U=\{u_1, u_2, \ldots, u_N\}$

Esempio:
L'Autorità per la tutela della *privacy* ha più volte evidenziato la necessità di garanzie nella predisposizione di misure riguardanti la sfera privata delle persone (ad esempio il riccometro), specie quando tali misure presuppongono l'attribuzione ai soggetti interessati di una carta, di un documento personale con un numero di identificazione. La prudenza quindi suggerisce di scegliere codifiche neutre attraverso delle combinazioni numeriche o alfanumeriche che non facciano riferimento ad informazioni riservate sulle unità, soprattutto se tali informazioni non sono soggette a trattamento statistico.

Per le popolazioni molto numerose è necessario un processo di etichettazione che generi dei codici assegnabili alle unità già esistenti ed in grado di assegnarli anche alle unità che si realizzeranno. Per unità congiunte o non individuabili l'etichettazione è virtuale cioè non un codice, ma una procedura che ne assicura la raggiungibilità a prescindere dalla loro elencazione materiale.

Esercizio_TP130: *il noto caso del Literary Digest (Bradley, 1976, pp. 62-64). Nel 1936 tale rivista attivò un sondaggio postale su dieci milioni di votanti scelti da elenchi telefonici e registri di possessori di auto. Lo scopo era di prevedere il risultato delle elezioni presidenziali: Roosvelt (democratici-progressisti) e Landon (repubblicani-conservatori). Si ottennero 2.4 milioni di riposte: il 57% avrebbe votato Landon ed il 38.5% Roosvelt. Vinse però Roosvelt con il 63%. Gran parte del fiasco è da attribuire ad una scelta inadeguata della lista. Perché?*

Non sempre è facile reperire o stilare la lista delle unità o trovarla depurata da errori ed informazioni non pertinenti. Ad esempio, volendo indagare lo status socio-economico dell'elettorato di un partito politico occorrerebbe conoscere chi lo ha votato alle elezioni, ma il voto è segreto e quindi la popolazione non sarebbe censibile. In prima istanza si potrebbe indagare sui tesserati di quel partito aggiungendo magari le persone notoriamente simpatizzanti per lo stesso, ma anche qui insorgono difficoltà. Innanzitutto la segreteria del partito dovrebbero fornire l'elenco degli iscritti e questo non è affatto garantito; e poi cosa si intende per "simpatizzante"? Esiste una definizione cogente, valida per tutti? D'altra parte, iscritti e simpatizzanti potrebbero essere troppo pochi o troppo peculiari per consentire la copertura degli indecisi che solo all'ultimo momento scelgono quel partito.

Esempi:
a) Nella rilevazione del 1940 il *Bureau of the Census* degli Stati Uniti ha introdotto un campionamento probabilistico chiedendo informazioni aggiuntive al 5% della popolazione censita.

b) La "Indagine campionaria sui bilanci delle famiglie italiane" condotta dalla Banca d'Italia in realtà non si basa sull'universo delle famiglie, ma sulle liste elettorali (disponibili presso ogni comune e presso il Ministero dell'Interno) dato che le anagrafi dei comuni non sono accessibili. Il fatto è che nelle liste elettorali confluiscono tutte le persone che hanno compiuto il 18° anno d'età e una famiglia con più maggiorenni vi ha un rilievo che potrebbe risultare eccessivo.

c) Nel condurre un'analisi di contenuto sugli editoriali comparsi in un quotidiano regionale, diciamo ai tempi della Costituente, si potrebbe scoprire che uno o più numeri sono andati irrimediabilmente perduti.

d) Per valutare l'efficacia di una terapia ci si deve basare sui pazienti attualmente ospedalizzati, ma mancheranno quelli non ricoverati o per i quali la malattia è ancora allo stato latente.

Differenze tra popolazioni effettiva e teorica

Gli esempi fanno intendere che può esserci difformità tra popolazione teorica su cui in astratto si dovrebbe condurre l'indagine e la popolazione effettiva su cui l'indagine può essere concretamente condotta. Si pensi alle rilevazioni congetturali a cui si è costretti nello studio delle popolazioni elusive ed in genere alle indagini su di unità sfocate oppure su popolazioni mobili. Ad esempio, un modello di dichiarazione dei redditi è un indicatore inadatto ad individuare chi non può pagare i servizi sociali perché rispecchia solo il possesso di redditi imponibili, non anche quelli esenti. Poiché condurre l'analisi sulle popolazioni teoriche può rivelarsi problematico, si ricorre ad unità che solo indirettamente -per legami o per analogia- portano a quelle che si vogliono analizzare.

Esempi:

a) Un'indagine sui giovanissimi (13-17 anni) che risultano residenti in un dato comune nell'anno appena trascorso e che abbiano subito condanne penali avrebbe serie difficoltà non solo ad avere dagli uffici giudiziari l'elenco delle unità, ma anche ad ottenere delle risposte. Spesso occorre contentarsi di ciò che raccontano vicini, parenti e/o amici.

b) In un sondaggio tra gli associati alla Dirstat (dirigenti di impresa) sulla possibilità che l'organizzazione entri attivamente in politica, nel caso non si trovasse la Manager si potrebbero ricavare le risposte interrogandone il segretario se è disposto a parlare ovvero con l'addetto alla pulizia della stanza.

c) Nella raccolta di valutazioni sulla qualità dei servizi: di una banca, di un *hard discount*, di un'agenzia di manutenzione si dovrebbero interrogare i clienti "abituali": quelli che al momento dell'indagine si trovano nella sede non sono necessariamente tali; quelli che risultassero dagli elenchi dei pagamenti con assegno o carta di credito potrebbero non esserlo più o esserlo stato solo per un particolare acquisto. Sarà perciò necessario stabilire regole ed opzioni che ripuliscano l'insieme dei due gruppi da coloro che non rientrano negli obiettivi dell'indagine.

d) Per il censimento degli elementi ecologici o bioclimatici di architetture realizzate nelle regioni del Centro-Italia è stata inviata una scheda nonché la richiesta dei progetti agli architetti iscritti all'albo di quelle zone.

e) Per individuare le ditte virtuali che operano con fatturazioni di comodo si cercano discrasie tra i ricavi ed i costi di magazzinaggio e stoccaggio o altri oneri indirettamente legate alle merci. Per la produzione di merci pericolose si possono incrociare i dati sugli acquisti di particolari composti e minerali.

La validità dei risultati dipende dal legame tra popolazione teorica e popolazione effettiva: più è diretto, maggiori saranno le possibilità che ragionamenti e conclusioni condotte per la popolazione effettiva si possano anche riferire alla popolazione teorica. Spesso è necessario far fronte a forti divergenze tra popolazione teorica e popolazione effettiva dovute a carenze della lista (duplicazioni, contraddizioni, contraffazioni), ma soprattutto rispetto al suo aggiornamento. La *frame* contempla solo degli anonimi codici e regole, ma dietro ogni formalismo c'è un organismo che vive, si muove, cambia o subisce trasformazioni che dovrebbero essere sempre monitorate e quindi incorporate dalla lista. A volte questa fornisce una prospettiva talmente angusta da non consentire di abbracciare tutte le proprietà della popolazione teorica rendendosi inutile se non dannosa all'indagine.

Esempi:

a) Un classico caso di popolazioni difficili da analizzare sono quelle che Kish (1965, p. 19) ha chiamato "popolazioni mobili" e cioè formate da unità dotate di estrema dinamicità, difficilmente localizzabili o la cui posizione non può essere desunta da una posizione occupata in precedenza: nomadi, barboni, campeggiatori, animali selvatici, pesci, etc. In queste situazioni si rendono necessari metodi sofisticati quale ad esempio le tecniche cattura-libera-ricattura usate per gli uccelli migratori e la ricerca di scie adoperate nelle indagini campionarie delle balene dove uno *splash* sulle onde è un segnale di presenza.

b) L'individuazione di unità elusive che operano nella finanza è semplificata dall'art. 20, comma 4 della Legge 413/91. Tale disposizione prevede, infatti, la creazione di una lista contenente i dati anagrafici di tutti gli intermediari finanziari compreso il codice fiscale di ogni soggetto che intrattenga con loro rapporti di conto o di deposito. L'accesso a questo tipo di informazioni non è però agevole.

c) Lo schedario dei pazienti di un studio medico è un archivio interessante per indagini sulla salute. Dall'archivio sarebbero però esclusi gli ammalati che non abbiano dato il loro consenso per utilizzare i dati personali.

Esercizio_TP131: *a) Il consumer data base (elenco dei clienti potenziali) è cruciale per aziende che hanno nella promozione (offerte speciali, buoni sconto, finanziamenti personalizzati) un'attività necessaria. Quali elementi possono concorrere a formarlo e quali sono le difficoltà a sfruttarne pienamente il potenziale?*
b) Nelle indagini sulle imprese si dispone di solito di buone liste grazie agli obblighi di legge cui sono sottoposte, ma c'è anche un'altra esigenza. Ridurre il carico statistico sulla singola azienda. In che cosa consiste?
c) La FIAIP ha sottoposto all'Autorità garante per la concorrenza il caso di operatori abusivi nel campo della mediazione immobiliare. Si tratta di soggetti che pubblicano inserzioni su appartamenti da vendere o affittare fingendosi proprietari, ma alla richiesta di vedere l'immobile frappongono il vincolo iscrizione a pagamento in un elenco di clienti. A parte la risibilità dell'applicazione, può essere un metodo per costruire una frame?

Esercizio_TP132: *quali problemi si possono incontrare nel realizzare rilevazioni campionarie in un Paese in via di sviluppo?*

La frame o lista

Tra popolazione teorica e popolazione effettiva si inserisce la *frame* (pron. freim, traducibile con "lista") cioè un sistema di codici identificativi o di norme procedurali con cui le unità diventano visibili o raggiungibili per chi conduce l'indagine.

Esempi:

a) La conoscenza delle realtà locali è fondamentale per lo sviluppo e la programmazione di attività sia pubbliche che private. La banca dati ISETVIEW predisposta dal CERVED (società delle camere di commercio) contiene i dati ufficiali del registro su tutte le attività economiche operanti nel territorio. I dati sono disaggregati fino al livello comunale nonché per ramo e classi di attività economica.

b) L'albo nazionale dei costruttori (ANCE) consente di individuare le aziende che si occupano di edilizia in vari tipi di attività e per classi di fatturato. L'annuario del lavoro autonomo raggruppa tutti i professionisti nel settore dell'ingegneria e dell'architettura cui è possibile affidare la progettazione dei lavori pubblici.

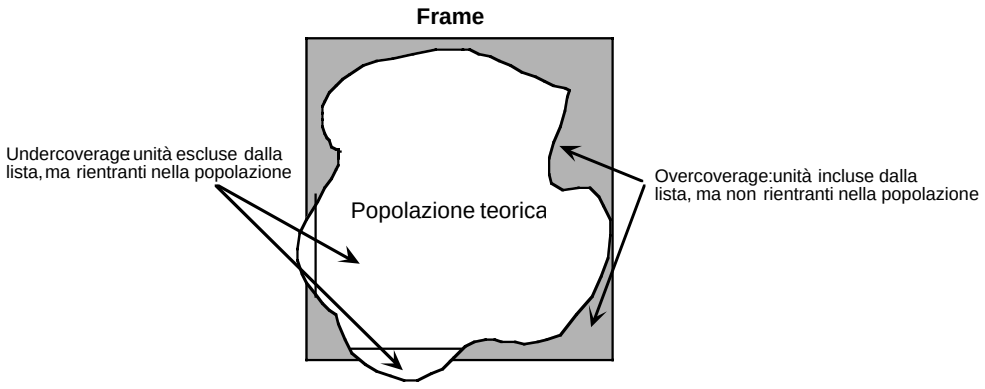
c) Per i vigneti esiste una anagrafe informatizzata gestita dal SIAN (sistema informativo agricolo nazionale) presso il quale debbono essere denunciate -obbligatoriamente- l'estensione e la variazione delle vigne, la quantità di uva destinata alla produzione di vino e le eventuali giacenze di prodotto, le denominazioni di origine e di indicazione geografica tipica. Sono inoltre rilevati i dati identificativi dell'azienda, le informazioni relativi al tipo di produzione e delle tecniche utilizzate.

d) Il 1° comma dell'art. 75 della Costituzione della Repubblica Italiana prevede che l'indizione del referendum popolare abbia luogo quando lo richiedono almeno cinquecentomila elettori. Le firme sono raccolte insieme agli estremi di un documento identificativo dei firmatari. Supponiamo che i moduli di raccolta prevedano 25 righe (quindi un foglio può contenere al massimo 25 firme) e che un certo referendum abbia raccolto un milione di firme su 45,000 moduli (5,000 in più di quelli effettivamente necessari, perché parte dei moduli non è riempita, ci sono firme ripetute per errore o per dimenticanza, oppure firme che prendono più righe). Le firme raccolte formano la popolazione teorica. Per controllarle occorre costruire una *frame*. Ad esempio si potrebbero numerare i moduli e, all'interno di ogni modulo, numerare le righe. In questo modo la *frame* prevederebbe 25x45000=1'125'000 posizioni: {0000001, 0000002,…,1124999, 1125000} di cui 125'000 non corrispondenti ad alcuna unità. Il controllo delle firme è improbo quando riguarda tutti i sottoscrittori e la validità dei documenti di cui sono noti solo gli estremi. Il controllo campionario potrebbe essere altrettanto efficace.

Esercizio_TP133: *il concetto di ordine alfabetico è familiare a tutti. E'usato per dizionari, vocabolari, enciclopedie, annuari, schedari, elenchi di persone o di località. L'idea di fondo è di avere una chiave che permette di determinare la posizione dell'unità in un dato ordinamento a partire dalle informazioni contenute nella chiave.*
a) Perché i bibliotecari sentono il bisogno di spiegare bene le "regole di elencazione".
b) Quali problemi si possono incontrare nell'ordinamento alfabetico?

Errori di lista

La lista è una sovrastruttura imposta alla popolazione teorica allo scopo di tracciare dei confini netti tra la parte che interesserà l'indagine campionaria e la parte che eventualmente rimarrà fuori a causa delle difficoltà di assicurare la partecipazione di tutte le unità *(selection bias)*. La lista ideale elenca o consente di elencare tutte e solo le unità della popolazione, distintamente e solo una volta; di solito, è aggiunta la possibilità di operare per sottoliste assicurando maggiore flessibilità. Poiché i limiti della lista sono spesso artificiali, essa potrebbe incorrere in vari errori:



1) *Overcoverage* (sovracopertura) e cioè includere unità estranee alla popolazione ovvero codici non corrispondenti ad unità di interesse. Ad esempio, volendo condurre una indagine sulle aziende agrarie con allevamenti ed usando un elenco generico di aziende agrarie porterebbe ad esaminare anche aziende prive di capi animali ovvero aziende cessate o in liquidazione.

Le registrazioni multiple della stessa unità che capitano nelle indagini sul campo di uccelli ed insetti rientrano in questo tipo di errore; anche le abitazioni con più numeri di telefono portano ad un eccesso di presenza se le abitazioni sono raggiunte telefonicamente; lo stesso accade ad una famiglia che ha un numero telefonico anche nella casa estiva.

2) *Undercoverage* (sottocopertura) e cioè non includere unità della popolazione. Ad esempio escludere quelle aziende agrarie i cui allevamenti non risultino registrati negli albi delle associazioni oppure non inserire le nuove aziende. In questa tipologia di errore sono da includersi le unità non presenti nella lista perché dimenticate o scomparse oppure che siano state deliberatamente escluse *(cut-off)* perché il costo di ottenimento dei dati non è compatibile con il preventivo delle spese ovvero siano secretate per ragioni di sicurezza interna o militari. L'esclusione avviene anche perché le aziende selezionate hanno già fatto parte di un'indagine ed occorre ridurre il loro onere statistico; Possono, inoltre, mancare le aziende già operanti, ma che non abbiano ancora perfezionato gli adempimenti previsti dalle leggi.

3) *Clustered listings* (unità aggregate). La lista potrebbe non elencare separatamente le unità, ma rifarsi a macrounità composte da un numero variabile di unità. Ad esempio, una ricerca sulle condizione dei soci di cooperative di pulizie potrebbe partire da un apposito albo presso le Camere di commercio e solo successivamente passare al contatto dei soci.

Esempi:

a) Per fare fronte alla crescita dei flussi migratori verso le aree più ricche ed industrializzate si sta sviluppando sempre di più la bio-identificazione cioè un sistema elettronico di verifica e riconoscimento individuale basata su caratteristiche fisiologiche: geometria della mano, contorni dell'iride, impronte digitali che produce un sistema di individuazione molto sicuro.

b) I titolari di conto corrente fiscale sono identificati da un codice composto -nelle prime tre cifre- da una sigla che individua il concessionario della riscossione; le altre cifre riportano il codice fiscale del contribuente. Questi soggetti sono raggiungibili (con il consenso del ministero delle finanze), ma poco si potrà fare per i contribuenti che non sono registrati.

c) Molto graditi sarebbero i dati sulle operazioni che comportano trasmissione o movimentazione di importo superiore ai 20 milioni che gli intermediari finanziari hanno l'obbligo di acquisire e conservare per 10 anni dalla normativa antiriciclaggio. Mancherebbero tuttavia quelli fino ai 20 milioni per i quali non esistono obblighi di registrazione.

d) Pure molto importante sarebbe la costituzione di un archivio informatico in cui inserire i nomi di chi ha emesso assegni non coperti o di coloro cui le banche hanno revocato l'autorizzazione ad utilizzare carte di credito o bancomat (che non sono più reato penale).

e) Gli enti pubblici curano la redazione e la trasmissione alla Presidenza del consiglio di un prospetto in cui indicano numero e valore degli appalti aggiudicati distinti in base alle procedure di aggiudicazione, nazionalità dell'aggiudicatario, categoria di servizi ed altre notizie rilevanti. L'analisi di tale lista potrebbe rivelarsi piuttosto interessante sulla concorrenzialità del mercato degli appalti pubblici.

f) Il Ministero dell'interno -riconoscendo il carattere di piena conoscibilità e di pubblicità delle liste elettorali- ha liberalizzato l'utilizzo dei dati che vi sono contenuti anche a fini commerciali ad esempio per le aziende di *direct marketing* (porta-a-porta, posta, e-mail) che possono richiedere, raccogliere e diffondere le informazioni di carattere personale estratte dalle liste elettorale senza chiedere il consenso degli interessati.

Esercizio_TP134:
a) Avete ricevuto l'incarico di verificare una graduatoria provinciale per l'insegnamento nella classe 19/A per la quale il Ministero ha ricevuto varie denunce ed esposti. Prima di controllare i singoli incartamenti avete bisogno di una frame. Quali problemi può presentare la graduatoria pubblicata?
b) Nel secondo capitolo è stata discussa la distinzione tra microdato e macrodato e della possibilità di arrivare alle informazioni sul primo avviandone l'individuazione attraverso il secondo. In che modo l'organizzazione per macrodati può complicare la formazione di una lista per i microdati (Sugg. pensate ad caso particolare).
c) Nelle procedure contabili di un'impresa è previsto che le fatture dei fornitori portino la firma del capomagazzino e della responsabile acquisti prima di essere poste in pagamento. In un campionamento per attributi (verifica della conformità alle procedure previste) si analizza un campione di fatture. Quali problemi di frame vi aspettate?

Esercizio_TP135: *individuate il tipo di errore cui potrebbe essere soggetta la frame in un'indagine rivolta ai soggetti indicati.*
a) Un praticante commercialista che non dichiara il rimborso spese percepito;
b) Un inquilino che ha cambiato casa; c) Un collaboratore domestico che risponde a nome della padrona;
d) Un appezzamento di SAU che sia stato occupato da un manufatto per l'energia o per le telecomunicazioni;
e) Possessori di esemplari di specie animali protette;
f) Discariche non controllate.

La discussione precedente potrebbe aver instillato l'idea che la *frame* sia sovrapposta alla popolazione teorica. Essa è invece una struttura sospesa su di essa con possibili distorsioni in modo simile agli effetti della rifrazione dell'aria nelle foto aeree. Ad esempio, un elenco telefonico può essere utilizzato come lista solo in riferimento agli intestatari dell'abbonamento, ma se la popolazione è formata dagli adolescenti sarà difficile che dietro ogni abbonato se ne possa trovare qualcuno. Allo stesso modo, le imprese che effettuano scambi intracomunitari sono tenute a compilare elenchi riepilogativi *(listings)* di tali operazioni che consentono alle amministrazioni fiscali di controllare le operazioni di compravendita. Se però i *listings* sono incompleti, insufficienti o non presentati, le possibilità di indagine si riducono drasticamente. Questi problemi sono all'ordine del giorno quando la lista del campionamento risulta dall'incrocio di più sottoliste o di liste dello stesso livello gerarchico, ma provenienti da fonti diverse o relative a epoche diverse.

Se si dispone delle necessarie informazioni, oltre che di tempo e fondi adeguati, i codici che non corrispondono ad unità della popolazione teorica ed eventualmente i codici incompleti e duplicati (comuni a più unità distinte) possono essere eliminati così come possono essere aggiunti i codici delle unità erroneamente escluse formando una più accurata lista della popolazione effettiva $P=\{e_1, e_2, \ldots, e_N\}$.

Esempi:

a) Uno studio sul personale di un'impresa che dovesse basarsi su documenti ed atti concernenti le loro condizioni psicofisiche troverebbe seri ostacoli se una parte di queste notizie è secretata per le posizioni più preminenti oppure manchi per i nuovi assunti, per i contrattisti a termine (formazione lavoro, apprendistato, reinserimento, etc.). Solo eliminando queste carenze oppure tenendole al minimo, la *frame* è efficace.

b) L'analisi della soddisfazione dei clienti di una catena di negozi potrebbe basarsi sulla lista dei possessori di una carta di fedeltà con cui si concedono sconti, agevolazioni, partecipazioni a concorsi, piccoli doni. Il "buco nero" di questa *frame* sarebbero i possessori di carte duplicate, clienti che non hanno ritirato la carta, clienti che hanno cambiato città, etc. La ripulitura e l'aggiornamento della lista sono condizioni necessaria per poterla usare con profitto.

c) La presenza del proprio nominativo o di un numero distintivo in un elenco è quasi sempre guardata con diffidenza e ostilità soprattutto se la corrispondente *frame* rientra nella sfera di interesse del Fisco. Un esempio lancinante è la partita IVA il cui elenco, se non completato ed aggiornato porta sotto il controllo degli uffici finanziari: defunti, falliti, emigrati, nullatenenti, omonimi, ignari (il cui codice è usato da altri) che sono inutilizzabili ai fini degli accertamenti.

d) Una ricerca statistica sull'albinismo deve basarsi sulla presenza di bambini con tale caratteristica. Così però non sono individuabili le famiglie con entrambi i genitori eterozigoti senza figli o senza figli albini.

e) Per conoscere la verità sull'annosa questione delle quote latte è stato condotto un censimento a tappeto tramite i veterinari (coadiuvati dalla guardia di finanza) delle stalle italiane.

Esercizio_TP136: *l'assenza di frame di qualità adeguatamente certificata ha creato una nicchia di mercato per i List Brokers cioè persone o società che forniscono liste ad hoc per ricerche di vario genere. Effettuate una ricerca sui vari media (particolarmente su Internet) per individuarne almeno uno.*

Esercizio_TP137: *l'attuale diffusione del Web e lo sviluppo che ci si attende sta rendendo disponibili nuove basi di dati che possono servire da frame. Ad esempio il Repertorio fornitori componenti e sottosistemi elettronici di Assodel oppure il Repertorio dell'industria chimica redatto dalla Federchimica. Effettuate una ricerca di disponibilità telematica su un settore di vostro interesse, ad esempio il turismo.*

Esercizio_TP138: *il problema delle mancate risposte ad alcune domande e delle mancate coperture (non coverage) ciò mancate risposte a tutte le domande è uno dei più seri nell'ambito delle indagini statistiche tanto da costringere a rivedere l'impostazione complessiva della logica: popolazione teorica- frame -popolazione effettiva. Tale sottopopolazione infatti non è uniforme sotto questo aspetto, ma deve essere divisa in tre categorie. Oltre alle gradite unità per le quali si può ottenere una risposta al primo contatto si devono considerare:*
a) Le unità che rispondono solo se si insiste un certo numero di volte (call-backs);
b) Le unità dalle quali non è possibile ottenere il dato. Quali problemi comportano?

Tra la lista e la popolazione esiste un rapporto dinamico ovvero, per essere veramente utile la lista deve contenere informazioni esatte, complete, aggiornate, acquisite correttamente ai sensi della Legge 675/1996 e destinate ad usi compatibili con le finalità alla base della loro formazione. Inoltre, le regole di costituzione devono essere note, documentate, trasparenti e individuabili con facilità; sarebbe poi utile la certificazione di qualità e l'assunzione della responsabilità, anche con penale, di chi rilascia l'attestato di validità. Altrettanto capillari ed accurate debbono essere le giustificazioni del perché una certa lista sia stata scelta per analizzare una data popolazione. Si tratta di qualità che è difficile garantire e la costruzione della lista è la fase più onerosa di una indagine statistica.

Esempi:
a) L'indagine sugli sbocchi professionali dei laureati pubblicata dall'ISTAT nel 1990 era riferita alla popolazione di colore che hanno conseguito la laurea nel 1986 in tutte le sedi universitarie italiane. Un problema serio fu quello di disporre di elenchi completi di nome, cognome, indirizzo e corso di laurea per ogni individuo. I dati vennero forniti dalle segreterie studenti su supporto cartaceo preparato *ad hoc* e quindi con dispendio di tempi e risorse.

b) Lo svolgimento di una ricerca su degli immobili che muovesse dalla lista delle proprietà avrebbe il difetto che chi possiede più di una unità immobiliare compare più di una volta. Inoltre potrebbero mancare le costruzioni successive alla data di costituzione della lista e quelle abusive o non censite.

c) L'aggiornamento della *frame* è un suo requisito essenziale: il prelievo forzoso del 6 per mille sui conti correnti postali e bancari che avvenne nell'estate del 1992 si attivò per un periodo determinato e limitato di tempo e interessò lo stato dei conti in quella data indipendentemente dal motivo per cui si trovavano sul conto (furono tassate anche partite di giro e fondi in transito). Qualche conto è sfuggito perché nel corso delle operazioni non si erano perfezionate le operazioni di apertura.

d) Il protesto scatta quando un assegno o un pagherò non vengono saldati oppure quando una tratta non viene accettata dal debitore. I pubblici ufficiali abilitati al protesto (i notai, di solito) sono tenuti a compilare la levata di protesto e a comunicare i dati al presidente della locale Camera di commercio. I dati su protesti e protestati confluiscono su di un apposito bollettino. La necessità di una tempestiva ed esatta compilazione di una tale lista è fondamentale per chi deve prestare denaro e per chi ha bisogno di un prestito.

e) I promotori di un corso di perfezionamento per laureati hanno come popolazione *target* le persone rientranti nelle graduatorie per titoli ed esami dei docenti della scuola secondaria; in particolare, coloro che restano disoccupati. Se, tuttavia la validità delle graduatorie viene estesa oltre il termine naturale i vincitori di concorso difficilmente risponderanno ai promotori.

La copertura è soddisfacente solo se la disamina della popolazione effettiva rende superflua la ricerca di dati sulle unità della popolazione teorica e questo tipo di assicurazione non sempre può essere data. Ad esempio, lo schedario delle imprese in funzione all'ISTAT per diverso tempo, era costituito dalle imprese con almeno 10 addetti che il censimento del 1981 rilevò operanti nell'industria, nel commercio, nei trasporti e in attività di servizi. L'esclusione delle imprese fino a 10 addetti fa perdere la maggioranza delle imprese ed una porzione rilevante degli addetti. Maisel e Hodges-Persell (1996, p.151) suggeriscono di costruire la *frame* definitiva partendo del presupposto che quella di cui si dispone, comunque ottenuta, sia sbagliata e chiedendosi in che modo aumentarne la copertura, ad esempio controllando che tutte le categorie di interesse siano state incluse: un'indagine sul diverso trattamento penale riservato agli extracomunitari rispetto ai cittadini italiani potrebbe mancare i detenuti in attesa di giudizio.

La definizione della lista, non si effettua solo con criteri statistici: ci vuole immaginazione, una conoscenza profonda del problema e la consapevolezza che ogni errore nella costituzione della lista si proietta sulla attendibilità del campione e sulla generalizzazione dei risultati con esso ottenuti. In mancanza di liste complete ed aggiornate ed in mancanza di risorse per ottenerle ci si può muovere con la convinzione che non tutti i soggetti cambiano tutti i giorni allo stesso modo e sperando che questo sia vero per la popolazione che interessa.

Esempi:
a) Una *frame* per gli alunni delle scuole elementari di un dato circolo didattico potrebbe basarsi sui plessi scolastici nel circolo, acquisire per ciascuno la lista degli iscritti composta dall'elenco degli alunni inseriti nel registro di ogni classe. C'è però il rischio che i registri riportino alunni ritirati o trasferiti in altra classe con duplicazione dei nominativi.

b) Una ditta che vende prodotti per bambini potrebbe acquistare dalle banche o direttamente dai gestori delle carte di credito, l'elenco dei clienti che nell'ultimo trimestre abbiano acquistato prodotti trattati dalla ditta o da ditte concorrenti. La risultante *frame* rischia di essere inadeguata perché i soggetti potrebbero non essere più interessati, oppure sono capitati nella lista per una acquisto contingente (magari un regalo), oppure la carta è stata usata da falsari.

c) Il Preside di un istituto superiore che conta più di mille iscritti intende monitorare i progressi degli alunni. Se vuole andare oltre i soliti voti/giudizi (sulla cui validità non si smette mai di dubitare) per cogliere lo sviluppo intellettuale e civile di ragazze e ragazzi del suo istituto ha davanti diverse possibilità. Punto di partenza è l'elenco degli iscritti che andrebbe subito aggiornato e ripulito da errori e ridondanze per poi suddividere gli studenti per sesso, età, condizione sociale, sezione, classe. All'interno di ogni sottopopolazione può individuare una persona rappresentativa e interrogarla personalmente, può inviare un questionario ad un gruppo ragionato per ciascuna ovvero inviare il questionario ad un gruppo scelto casualmente. La strategia più efficace dipende dall'obiettivo dell'indagine, dai rischi di errore e dai costi connessi. Nella teoria dei campioni si apprende come gestire tali questioni.

Esercizio_TP139: *una lista incoerente pone a rischio l'intera indagine. Deming (1960, p. 30) riporta due casi:*
a) Un revisore contabile, dovendo analizzare circa centomila operazioni, apre un registro e seleziona un campione di tali operazioni. L'esame però rivela che nessuna di esse riguarda arrivi di alluminio che invece erano il suo interesse. In effetti ben poche delle transazioni incluse nella lista riguardavano carichi di alluminio cosicché non si può attribuire al campione la loro carenza. Qual' è l'errore del revisore?
b) Il management di un'impresa, per meglio fronteggiare una stagnazione del mercato, decide di approfondire l'andamento degli acquisti e delle vendite. A questo fine sceglie un mese tipico, diciamo gennaio 1994 che viene esaminato nel dettaglio più minuto.

Popolazioni rare

L'indagine statistica può essere ostacolata dalla eventuale rarefazione delle unità (Fabbris, 1995, p. 35). Le unità rare sono quelle presenti in misura minima nel resto della popolazione

Esempi:

a) Praticanti la religione mormone in Calabria;
b) Aziende con più di 20'000 dipendenti;
c) Fabbriche che producono diossina;
d) Supercomputer in funzione in Italia;
e) Parlamentari regionali;
f) Affetti da distrofia muscolare;
g) Partecipanti a giochi televisivi;
h) Titolari di c/c con più di 10 miliardi di disponibilità.

Per localizzare tali unità è necessario una procedura di restrizioni successive passando da elenchi più generali ad elenchi più selettivi che man mano si approssimano alla popolazione di interesse. Più facile a dirsi che a farsi perché se ai livelli primari si trovano liste affidabili sia pur generiche, le altre sono sempre più incerte e fantasiose.

Esempi:

a) Un'indagine era mirata sulla sottopopolazione "utenze telefoniche familiari non in elenco". Circoscritto il distretto ed accertato che i numeri erano composti da sette cifre si è formata una *frame* delle sue zone (i primi quattro numeri del codice in comune). In ogni zona c'erano perciò 1000 codici. Una ricerca sull'elenco informatizzato (in commercio su CD-ROM) ha individuato quelli già assegnati. Per individuare le unità è bastato comporre il numero che a questo punto poteva solo essere: famiglia, affari, libero, fuori uso.

b) I centri informativi dei dipartimenti delle entrate, delle dogane e delle imposte indirette segnalano ai rispettivi uffici ed al comando della guardia di finanza una lista di soggetti la cui attività sia caratterizzata da rilevanti scambi con l'estero. L'uso di tale lista consentirebbe molte investigazioni finanziarie interessanti e produttive.

c) Una banca dati dei sinistri potrebbe evitare truffe e rigonfiamenti dei rimborsi e la corretta determinazione della classe di merito. Un beneficio ancora più grande arriverebbe agli assicurati sui quali non verrebbero scaricati i costi della inefficienza e superficialità delle società nei controlli.

Le popolazioni rare possono essere concentrate in ambiti ristretti (ad esempio i grecanici nella provincia di Vibo Valentia) oppure essere disperse in contesti molto vasti (stabilimenti siderurgici in Italia). Nel primo caso si può sperare una qualche semplificazione, nel secondo saranno necessarie tecniche di alta ingegneria statistica per definire operativamente la popolazione. Se non hanno ragioni particolari per nascondersi, l'uso di elenchi via via più aderenti potrà avvicinarle, ma se sono rare ed elusive sarà necessario l'aiuto dell'investigatore privato.

Esempi:

a) La Camera di Commercio di Palermo, per iniziativa di anonimi funzionari, negli anni scorsi ha inserito in CERVED delle rubriche anagrafiche contenenti dati su circa duemila affiliati alla mafia e su società in odore di riciclaggio operanti nel territorio di competenza. Le informazioni erano di grande valore investigativo, non solo statistico, ed infatti erano solo destinate al rilascio della certificazione antimafia, ma una svista ha fatto in modo che invece di essere archiviate in modo riservato, sono divenute di pubblico dominio. E' così una popolazione rara ed elusiva è divenuta, almeno in parte, trasparente.

b) Una buona possibilità di individuare soggetti in popolazioni rare è l'incrocio delle banche dati. Infatti, lo scambio di informazioni ai fini di controlli incrociati tra enti e amministrazioni diverse nonché le segnalazioni alle organizzazione interessate dei fatti non di propria competenza si sono rilevati efficaci per individuare e ridurre il lavoro nero, l'evasione scolastica, la mano d'opera clandestina).

c) In uno studio sulle vittime di violenza sessuale i soggetti erano donne che avevano subito l'aggressione. A questo fine non potevano essere sufficienti i casi di abusi arrivati al processo oppure denunciati perché non sempre questi reati vedono la luce. L'indagine contattò le vittime attraverso un *network* di conoscenti, conoscenti di conoscenti, annunci sui quotidiani definendo in modo approssimativo, ma utile una base per il campionamento.

Esercizio_TP140: *l'ISTAT organizza sempre più spesso delle indagini omnibus su vasta scala per raccogliere informazioni su vari fenomeni sociali ed economici sia per l'intera collettività, ma anche per un gran numero di sottopopolazioni. Ad esempio "Stili di vita e condizioni di salute. Indagine multiscopo sulle famiglie. Anni 1993-1994". Reperite il testo in biblioteca (o uno più recente) e valutate in che modo ed in che misura possono essere d'ausilio per lo studio delle popolazioni rare.*

Esercizio_TP141: *un'inchiesta sullo stato di attuazione della legge 241/1990 (trasparenza amministrativa) ha dovuto riscontrare che su circa diecimila amministrazioni interessate ha risposto solo il 35%. Alcuni enti (quelli locali, in particolare) non hanno trasmesso alcuna informazione. Il Ministro interessato intende scoprire il perché della mancata risposta piuttosto che spiegare la mancata attuazione. Qual'è la sua popolazione di interesse? Come si può definirne una frame?*

Esercizio_TP142: *un certo materiale è confezionato in sacchi disposti in un magazzino fra questi ce ne sono alcuni che forse sono di un altro prodotto. Come si potrebbe costituire una lista? Come se ne può scegliere un campione?*

6.5.2 L'universo dei campioni

Nel primo capitolo è stato introdotto il campione come utile e pratica semplificazione dell'indagine statistica ovvero come unica soluzione in certe indagini distruttive o su popolazioni scomparse. Perché sia veramente valido deve però essere realizzato secondo una coerente fondazione probabilistica.

Unità blank ed unità autorappresentative

La lista può includere due tipi di unità che richiedono una attenzione particolare. Si parla di unità *blank* (o estranee) per quegli elementi della lista che sono escluse dal campione. Questo si può verificare quando la *frame* non consente di individuare alcune unità; quando le caratteristiche che interessano sono debitamente rappresentate dalle unità già incluse; perché i dati da essa ricavabili mancano di requisiti essenziali; perché il costo di inclusione è eccessivo rispetto al beneficio che possono dare. In questi casi, la deliberata esclusione *(cut-off)* di alcune unità non provoca alcun danno e semplifica la gestione della lista: ad esempio, seguire le imprese del settore della piccola distribuzione su tutto il territorio nazionale è superfluo; basterà monitorare quelle facenti parte di alcune catene nazionali.

Alcune unità deve necessariamente fare parte del campione. Si tratta in questo caso di unità fondamentali, delle quali non si può fare a meno per dare un'immagine realistica della popolazione. Ad esempio, in una indagine sull'occupazione in Basilicata non si può trascurare la Fiat di Melfi, non si possono tenere fuori le transazioni più cospicue nella revisione di una contabilità né analizzare le università per studenti ignorando "La Sapienza" di Roma (queste unità sono anche dette autorappresentative perché da sole ed uniche rappresentano un preciso segmento della popolazione).

Selezione con reimmissione e senza reimmissione

Una volta che l'unità sia stata scelta (si dice anche estratta) ci sono due alternative: può ancora fare parte del campione oppure è esclusa ogni sua ulteriore comparizione; nel primo caso si parla di estrazione con reimmissione. Ad esempio, dopo che un cliente è stato censito rispetto ad un'acquisto può ancora essere censito ripetendo la rilevazione dell'acquisto di prima ovvero dando peso doppio alle informazioni già ottenute. In generale, se dopo ogni estrazione si ripristina del tutto (a meno di impercettibili e non controllabili variazioni fisiche) la situazione antecedente, si parla di estrazioni con reimmissione (o bernoulliane) e tra due estrazioni di questo tipo non ci può essere alcun legame come si è convenuto nei paragrafi precedenti. Al limite, un campione di ampiezza "n" selezionato con il reinserimento potrebbe essere costituito da una stessa unità ripetuta "n" volte; potrebbe anche essere costituito da unità tutte diverse, ma questo non fa cadere la reimmissione dato che questa caratterizza il meccanismo di selezione e non il suo risultato.

Se invece, una volta estratta , l'unità non può più rientrare nel campione si parla di estrazione senza reimmissione: dopo aver inoculato un vaccino mortale ad una cavia non si può richiamarla in vita per rivaccinarla di nuovo. In questo caso sussiste un legame tra i risultati possibili nelle varie estrazioni dato che alcuni esiti sono impediti; ad esempio, la popolazione deve contenere almeno "n" unità se si deve estrarre un campione di ampiezza "n". Il campionamento senza reimmissione è anche detto campionamento "in blocco" perché è come se le unità fossero prelevate tutte insieme dalla popolazione. Poiché non si può prenderne tante ed in numero esatto si procede estraendole una ad una, ma escludendo quelle già inserite qualora queste si riproponessero per far parte del campione.

Esempi:

a) Si deve estrarre un campione di ampiezza n=12 da una popolazione di ampiezza N=90 costituita dai pesi in kg di tutti i calciatori tesserati da un importante *club* negli ultimi tre anni.

65	64	69	73	70	66	69	72	80	78	74	77	74	79	67	65	77	75
66	68	77	80	72	74	68	79	84	89	87	84	64	79	85	72	65	84
68	75	75	84	73	81	78	87	82	69	74	73	71	81	66	78	85	72
69	76	78	71	74	65	79	77	79	80	69	71	78	65	73	73	68	78
70	78	74	79	77	68	66	83	88	84	75	61	72	89	83	76	72	86

Supponiamo che le unità siano identificate con una coppia di cifre arabe: da "00" a "89" indicanti la posizione d'ordine da esse occupate cominciando a contare dalla prima colonna e procedendo dall'alto verso il basso e da sinistra a destra. Supponiamo che le unità prescelte con reimmissione siano: C={13, 05, 00, 45, 41, 84, 07, 54, 72, 59, 21, 84} corrispondenti ai seguenti valori campionari: {78, 64, 65 ,78 , 84, 72, 75, 75, 66, 61,74,72}. Se la campionatura fosse stata senza reimmissione, la scelta della unità etichettata "84" non poteva essere rifatta ed occorreva sostituire il valore campionario 72 con un altro. Da notare che la reimmissione/non reimmissione riguarda, almeno in questo tipo di applicazione, le unità e non i valori di cui sono portatrici. La presenza di "75" due volte , in caso di mancata reimmissione, vuol solo dire che ci sono unità con una stessa modalità.

b) Una *software house* sta organizzando la raccolta di leggi e regolamenti sul commercio comunitario in un CD da proporre alle imprese. Prima di avviare la duplicazione di massa è importante accertare che il programma di ricerca interna funzioni (OK oppure F) e che il processo non veicoli virus nel sistema (V/NV). Ecco la situazione delle prime 15 scatole da 10 CD:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	NV/OK	NV/OK	NV/OK	NV/OK	NV/F	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK
2	NV/OK	NV/F	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/F	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK
3	NV/OK	NV/OK	NV/OK	NV/OK	V/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	V/OK	NV/OK
4	NV/OK	NV/OK	NV/OK	NV/F	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/F	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK
5	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/F	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK
6	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK
7	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	V/OK	NV/OK	NV/OK	NV/OK	NV/F	NV/OK	NV/OK
8	NV/F	NV/OK	NV/OK	V/F	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	V/F
9	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	V/F	NV/OK	NV/OK	NV/OK
10	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/F	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK	NV/OK

Se i controlli si limitassero ad una sola scatola risulterebbero inefficaci per la 6ª o 11ª scatola; già meglio (perché più esteso) il controllo di un CD in una particolare posizione della scatola anche se, in questo caso, esaminando la 6ª riga non verrebbe riscontrato nulla di anormale. In queste applicazioni la non reimmissione è d'obbligo dato che sarebbe inutile esaminare un CD già controllato.

c) Se si sospetta che unità in posizioni contigue nella lista possano invalidare composizione del campione qualora venissero selezionate, il concetto di non reimmissione potrebbe essere esteso non solo all'unità già estratta, ma ad un intervallo di unità di ampiezza prefissata che abbia la centro quella scelta *(spaced random selection)*: se si sceglie la 15ª e si ritiene che gli effetti arrivino fino alla terza più prossima si possono escludere dalla lista le unità 12,13,14 e 16,17,18 oltre alla 15.

Esercizio_TP143: *una popolazione infinita è composta da due tipi di unità: quelle di tipo A e le altre di tipo B. Del primo tipo però ne esiste solo una. Per un campione di n=2 è importante stabilire che sia con o senza reimmissione?*

Esercizio_TP144: *Ritenete utile distinguere tra campionamento con e senza reimmissione per liste che includano unità blank ed unità autorappresentative?*

La distinzione tra i due tipi di campionamento cade nei casi estremi di n=1 (campione formato da una sola unità), n=N (campione di ampiezza pari alla popolazione) e nel caso di popolazione costante cioè formata da unità che producono la stessa identica modalità. Infine, la distinzione è considerata irrilevante se "n" è molto piccolo rispetto ad N, diciamo nell'ordine di 1 a 5'000, dato che in questi casi la ripetizione delle unità nel campione è sì possibile, ma poco improbabile. A maggior ragione se le popolazione è infinita.

L'esperimento: estrazione campionaria

La formazione di un campione casuale con o senza reimmissione rientra nel modello delle urne e può quindi essere regolato con il calcolo delle probabilità. L'evento elementare è la n-tupla di interi $C_i=(i_1,i_2,\ldots,i_n)$ corrispondenti a posizioni occupate in una lista univoca ed esaustiva delle unità che -per comodità di esposizione- sono in numero finito ed identificate con gli interi naturali (escludiamo la possibilità di campionare unità non distinguibili cioè ogni unità ha un codice unico nell'ambito della *frame*). L'universo degli eventi -indicato con T_n-ha come elemento base il singolo campione di ampiezza "n" ed include tutti i possibili campioni di tale ampiezza $T_n=\{C_1,C_2,\ldots, C_v\}$. La cardinalità di T_n dipende dalla possibilità di reimmettere l'unità e se rileva o no l'ordine di comparizione nel campione ovvero se si tratta di una partizioni. La tabella che segue indica il numero dei possibili campioni secondo le varie ipotesi.

Cardinalità	*c.rim*	*s.rim*	*s.rim – c.rip*	*Part. c.rim – c.rip*	*Part. s.rim – s.rip*	*Part. s.rim – c.rip*
Ordinati	N^n	$N!/(N-n)!$		$\prod_{i=1}^{m}(N_i)^{n_i}$	$\prod_{i=1}^{m}\left[\frac{N_i!}{(N_i-n_i)!}\right]$	
Non ordinati	N^n	$\binom{N}{n}$	$\binom{N+n-1}{n}$	$\prod_{i=1}^{m}(N_i)^{n_i}$	$\prod_{i=1}^{m}\binom{N_i}{n_i}$	$\prod_{i=1}^{m}\binom{N_i+n_i-1}{n_i}$

Il reciproco di queste entrate rappresenta la probabilità di ottenere il singolo campione.

Esempi:

a) Tre tenniste (Ada, Lia, Zoe} debbono decidere chi scenderà in campo per i due singolari (il torneo consente di giocare per due volte). In ragione della strategia adottata il campione (A,L) avrà probabilità 1/9 se la scelta è fatta con reimmissione tra oggetti ordinati; 1/6 se non ordinati, ma è consentita la reimmissione; 1/6 se non è consentita la reimmissione, ma rileva l'ordine in cui si gioca e 1/3 se l'ordine è irrilevante e non è consentita la reimmissione.

	Ordinati	Non ordinati
Con	(A,L) (A,Z) (A,A) (L,A) (L,L) (L,Z) (Z,Z) (Z,L) (Z,A)	(A,L) (A,Z) (L,Z) (A,A) (L,L) (Z,Z)
Senza	(A,L) (L,A) (A,Z) (Z,A) (L,Z) (Z,L)	(A,L) (A,Z) (L,Z)

b) Una società attiva nelle linee *charter* deve noleggiare due aerei per quattro tratte: Londra, Berlino, Parigi, Copenaghen} in modo che il primo vettore serva una linea e l'altro ne serva due. Le partizioni possibili in una scelta casuale sono indicate nella tabella: se il campionamento avviene con ripetizione e trascurando l'ordine, le alternative sono:

$$\binom{4+2-1}{2}\binom{4+1-1}{1}=\binom{5}{2}\binom{4}{1}=10*4=40$$

	Partizioni
Con reimmissione	L-(L,L) B-(L,L) P-(L,L) C-(L,L) L-(L,B) B-(L,B) P-(L,B) C-(L,B) L-(L,P) B-(L,P) P-(L,P) C-(L,P) L-(L,C) B-(L,C) P-(L,C) C-(L,C) L-(B,B) B-(L,B) P-(L,B) C-(L,B) L-(B,P) B-(L,P) P-(L,P) C-(L,P) L-(B,C) B-(L,C) P-(L,C) C-(L,C) L-(P,P) B-(L,P) P-(L,P) C-(L,P) L-(P,C) B-(L,C) P-(L,C) C-(L,C) L-(C,C) B-(L,C) P-(L,C) C-(L,C)
Senza reimmissione	L-(B,P) B-(L,P) P-(L,B) C-(B,L) L-(B,C) B-(P,C) P-(B,C) C-(L,P) L-(C,P) B-(C,L) P-(C,L) C-(P,B)

Se invece si campiona senza ripetizione le scelte sono: C(4,2)C(2,1)=6*2=12

Fabbris (1995, p. 53) osserva che il campionamento senza reimmissione è la norma nelle applicazioni; quello con reimmissione si pratica di rado perché ammette ripetute estrazioni che in molte occasioni non sono possibili o sono illogiche. Tuttavia, per la essenzialità della sua teoria, lo si richiama più spesso di quanto non lo si applichi.

Esempi:

a) Ad un test sull'impatto visivo di un *poster* di 18m^2 sono stati invitati N=50 automobilisti che hanno dato la loro opinione. Di questi, n=7 dovrebbero essere sottoposti -in un ordine qualsiasi- ad un altro test sulla leggibilità delle scritte inserite nel *poster*. Le scelte possibili sono:

$$\binom{50}{7}=\frac{50!}{7!43!}=99'884'400$$

b) Su N=70 sentenze emesse da un collegio giudicante se ne esaminano n=6. La presenza di recidivi legittima la scelta con ripetizione. L'universo dei campioni ha cardinalità pari a:

$$\binom{70+6-1}{6}=\frac{70*71*\ldots*75}{6!}=201'359'550$$

c) Un revisore ha individuato 100 transazioni sospette (che potrebbero dover essere esaminate più volte). Supponiamo che, per ragioni di tempo, ne possa esaminare solo dieci. I campioni possibili sono:

$$\binom{100+10-1}{10}=\frac{109!}{10!(99)!}=42'634'215'112'710;\quad \binom{100}{10}=\frac{100!}{10!90!}=17'310'309'456'440;\quad 100^{10}=10^{20}$$

Il numero è elevatissimo anche per ampiezze piuttosto modeste. Fra queste decine di migliaia di miliardi di campioni possibili ve ne saranno alcuni prossimi alla popolazione altri solo vicini ed altri lontani. Le tecniche di selezione delle unità aiutano a circoscrivere quest'ultima deleteria possibilità.

d) Una società è presente in 19 province meridionali. L'ufficio di controllo vuole effettuare una verifica a campione su quattro filiali -una in ogni provincia- scelte a caso e senza reimmissione. Le filiali sono ordinate per fatturato. Quanti sono i possibili campioni? $D_{SR}(19,4)=19*18*17*16=93'024$

e) Un reparto di N=20 operai è articolato in livelli: 8 "A", 4 "B", 5 "C" e 3 "D". Si scelgono n=11 unità per dai vari livelli secondo la composizione (4,2,3,2). I campioni possibili, non tenendo conto dell'ordine, sono:

$$\binom{8}{4}\binom{4}{2}\binom{5}{3}\binom{3}{2}=\frac{8!}{4!4!}\frac{4!}{2!2!}\frac{5!}{3!2!}\frac{3!}{2!}=12'600$$

f)I candidati alle elezioni politiche di Roccasecca sono raccolti in quattro partiti: stella e corona (24 candidati; falce e spiga (31); torre e gabbiano (27); ulivo e ruota (29). Si deve scegliere un campione di ampiezza n=3 da ciascuna lista per gli *spot* elettorali. Il moderatore ha però fatto confusione con gli elenchi ed ignora la lista di appartenenza. Qual'è la probabilità che ne convochi tre per ciascun partito? Si tratta di un campione senza reimmissione, ma con ripetizione:

$$\frac{\binom{24+3-1}{3}\binom{31+3-1}{3}\binom{27+3-1}{3}\binom{29+3-1}{3}}{\binom{111+12-1}{12}} = 0.018$$

g) Una cura prevede l'impiego di 9 farmaci del gruppo A, 8 del gruppo B e 10 del gruppo C. All'interno di ogni gruppo se ne debbono scegliere 4 così da formare una terapia combinando i vari principi attivi. Quante sono le possibili terapie se conta l'ordine all'interno di ogni gruppo?

$$D_{SR}(8,4) * D_{SR}(9,4) * D_{SR}(10,4) = \left(\frac{8!}{4!}\right)\left(\frac{9!}{4!}\right)\left(\frac{10!}{4!}\right) = 3.84072x10^{12}$$

Esercizio_TP145: *calcolate il numero dei possibili campioni nelle situazioni seguenti*
a)I dipendenti di un call center sono divisi in tre fasce: K con 8 dipendenti, J con 7 e H con 6. Si deve formare un campione senza reimmissione scegliendone 4 da K, 3 da J e 2 da H. L'ordine non è rilevante.
b) I membri di un'assemblea facevano capo a quattro gruppi: progressisti (12), conservatori (8), ambientalisti (7), liberali (13). Tre di ogni gruppo debbono parlare in una seduta; l'ordine degli oratori, all'interno del gruppo, è rilevante.
c) In uno scaffale sono presenti 14 testi di statistica avanzata, 12 di statistica economica, 11 di statistica sociale e 15 di statistica introduttiva. Le richieste di prestito sono state, rispettivamente: 5, 4, 3, 8. Poiché il prestito può riguardare lo stesso testo le scelte si debbono considerare con reimmissione. Anche l'ordine è rilevante.
d) In uno studio medico operano 5 specialisti: Rossi, Neri, Bianchi, Verdi, Bruni con un numero di pazienti fissi di 20, 18, 22, 14, 16. In una data giornata hanno preventivato di ricevere solo cinque ammalati. Le visite di un dottore possono anche riguardare lo stesso paziente. L'ordine delle visite non è considerato;
e) Per accertare la qualità della presentazione le home page di 75 alberghi sono state disposte in una frame dalla quale si scelgono, senza reimmissione, 7 alberghi.
f) Un'ispettore dell'INAIL ha disposto un elenco dei reparti che intende visitare progettando di visitarne 5. Le aziende del suo universo sono 42 ed ognuna ha 6 reparti. L'ispettore non visita mai lo stesso reparto più di una volta anche se può visitare più reparti di una stessa impresa.

Esercizio_TP146: *un'indagine su delle imprese dispone della lista delle N unità della popolazione. Per estrarre un campione si genera una permutazione casuale delle N unità e le "n" unità che si trovano nelle prime "n" posizioni della permutazione costituiranno il campione.*
a) La scelta del campione è casuale? b)Il campionamento è con rimessa? c) Quanti sono i possibili campioni?

Se il campionamento avviene con rimessa da una lista di unità distinte ed equiprobabili allora:
1) Ognuna delle N unità della popolazione ha la stessa probabilità (n/N) di comparire in una qualsiasi delle "n" posizioni del campione;
2) Ogni gruppo di "n" unità ha la stessa probabilità $(1/N)^n$ di costituire il campione. Questo grazie al ripristino integrale delle condizioni di partenza che rende la probabilità di inclusione costante rispetto alla unità da includere ed alla posizione da occupare.

Esempio:
Ipotizziamo che il sesso alla nascita sia equiprobabile. Per verificare tale congetture esaminiamo un campione di n=3 famiglie con cinque figli tra le N=100 che risultano nella popolazione di interesse. In un'urna sono inserite 100 biglie indistinguibili se non per il cognome del/la capofamiglia scoperto solo dopo l'estrazione. L'urna è agitata per tanto tempo e in un modo che sia impossibile localizzare una qualsiasi delle biglie. La biglia è estratta; individuata la famiglia attraverso l'etichetta contenuta nella biglia si osserva il numero di femmine presenti nella prole. La biglia è poi reimmessa nell'urna.

$$P\left(Fam_i \;\; in \;\; posizione \;\; j\right) = \frac{100*100}{100*100*100} = \frac{1}{100}; \quad P\left(Fam_{i_1}, Fam_{i_2}, Fam_{i_3} \in C\right) = \frac{1}{100*100*100}$$

La probabilità che la famiglia i-esima entri nel campione è:

$$P\left(\bigcup_{j=1}^{n} Fam_i \;\; in \;\; posizione \;\; j\right) = \frac{1}{100} + \frac{1}{100} + \frac{1}{100} = \frac{3}{100}$$

Esercizio_TP147: *Earl Dumarest, eroe di una space saga molto nota ai lettori di fantascienza, ha sottratto una formula segreta che permetterebbe ai suoi nemici - i cyclani- di dominare l'universo. La formula si compone di 15 elementi da provare nei vari ordinamenti ed ogni prova richiede almeno una settimana di vita terrestre perché sia testata in modo adeguato. I cyclani, oltre a tentare di catturare Dumarest per farsi rivelare la giusta sequenza, svolgono dei tentativi estraendo dei campioni casuali di permutazioni. Sono attivi 1'000 laboratori.*
a) Qual'è la probabilità che una singola permutazione faccia parte del campione? b) Qual'è la probabilità che un gruppo di 1'000 permutazioni formi il campione? c) *Quanto tempo sarebbe necessario per provarle tutte?*

Nel campionamento semplice senza reimmissione il numero della lista corrispondente all'unità già estratta non è considerato valido in caso di riuscita. Ebbene, anche in questo caso la probabilità di occupare una data posizione è la stessa per tutte le unità e gruppi qualsiasi di "n" unità hanno tutti la stessa probabilità di formare il campione purché la selezione sia casuale. Questo, a prima vista non sembra convincente, perché qualcuna delle posizioni potrebbe già essere occupata ovvero che se l'unità è collocata in una posizione non è più ricollocabile in un'altra.

Esempio:
La probabilità condizionata può servire per verificare che la probabilità di inclusione di una qualsiasi unità della popolazione nel campione casuale semplice -senza reimmissione- è pari alla frazione di campionamento f=n/N qualunque sia la posizione del campione da occupare. Sia E_j= "L'unità i-esima compare nel campione in posizione j-esima". Per n=1 è evidente che $P(E_1)=1/N$ per l'ipotesi di equiprobabilità (sottinteso al termine "casuale"). Per n=2, la comparsa dell'unità i-esima è segnalata dal verificarsi dell'evento:

$$E_1 \cup \left(E_1^c \cap E_2\right) \;\; con \;\; E_1 \cap E_2 = \varnothing \Rightarrow E_1^c \cap E_2 = \varnothing \Rightarrow E_1 \cap \left(E_1^c \cap E_2\right) = \varnothing$$

La cui probabilità è data da: $P\left[E_1 \cup \left(E_1^c \cap E_2\right)\right] = \;\; P[E_1] + P\left[\left(E_1^c \cap E_2\right)\right] = P[E_1] + P\left[E_1^c\right] * P\left[\left(E_2 | E_1^c\right)\right] = \frac{1}{N} + \frac{N-1}{N}\frac{1}{N-1} = \frac{2}{N}$

Per n=3 si ha:

$$P\left[E_1 \cup \left(E_1^c \cap E_2\right) \cup \left(E_1^c \cap E_2^c \cap E_3\right)\right] = P[E_1] + P\left[E_1^c\right] * P\left[\left(E_2 | E_1^c\right)\right] + P\left[\left(E_2^c | E_1^c\right)\right] * P\left[\left(E_3 | E_1^c \cap E_2^c\right)\right] = \frac{2}{N} + \frac{N-1}{N}\frac{N-2}{N-1}\frac{1}{N-2} = \frac{3}{N}$$

Nel costruire il campione ignoriamo quale sarà la posizione occupata da una data unità: ci si trova come se, con gli occhi bendati, dovessimo inserire delle biglie in varie urne disposte -a nostra insaputa- alla rinfusa su di un tavolo; una data biglia può capitare ovunque ed una data buca potrà essere occupata da una qualsiasi delle biglie: non bisogna fermarsi alla singola collocazione, ma occorre considerare l'intero processo. In questo senso è logico che la probabilità di entrare nel campione si esplichi indipendentemente dalla posizione da occupare.

Esempi:
a) La probabilità che la famiglia "i" compaia al 1° posto del campione è 1/(99*98) dato che il 1° posto è ora bloccato dalla i-esima lasciando le altre due posizioni per le rimanenti unità. Bloccata la 2ª, la 3ª può essere occupata da 99 famiglie e la 1ª da 98. Lo stesso succede per le altre posizioni perché se la i-esima deve comparire in 3ª posizione, la 1ª può essere occupata in 99 modi diversi e la 2ª in 98:

$$P\left(Fam_i \;\; in \;\; posizione \;\; j\right) = \frac{99*98}{100*99*98} = \frac{1}{100}$$

Scelta la prima famiglia su N=100 a far parte del campione, la 2ª è scelta su 99 e la 3ª su 98. Qualunque famiglia può essere la prima, la seconda o la terza. Ne consegue che:

$$P\left(Fam_{i_1}, Fam_{i_2}, Fam_{i_3} \in C\right) = \frac{1}{100*99*98}$$

b) La famosa scienziata ha intuito che una combinazione di cinque elementi scelti -senza reimmissione- tra 20 e disposti nella giusta sequenza, può risolvere un serio problema genetico. Quanti sono i campioni possibili? 20!/15!=1'860'480.

Esercizio_TP148: *una sperimentazione clinica interessa N pazienti ai quali possono essere praticati "r" trattamenti. I pazienti sono assegnati casualmente ai trattamenti. In quale caso applichereste le probabilità:*

$$1) \;\; \frac{\binom{N}{n_1, n_2, \ldots, n_r}}{N^r}; \qquad 2) \;\; \frac{1}{\binom{N+r-1}{N-1}}; \qquad 3) \;\; \frac{1}{\binom{N}{r}}$$

Esercizio_TP149: *la direttrice di un'agenzia per il lavoro interinale vuole conoscere la destinazione di alcuni curricula che non hanno avuto contatti nell'ultimo anno. Fatta una lista dei 2272 nominativi decide di estrarne un campione casuale di ampiezza n=5 usando per ogni unità da campionare un numero formato accostando il primo estratto di due ruote del gioco del lotto e dividendo il risultato per 4. Se, ad esempio, il 1° estratto di Torino è 24 ed il 1° estratto di Venezia è 28 si forma il numero 2428 che diviso per quattro fornisce la posizione 607 che individuerà la persona da intervistare. E' un campionamento con o senza reimmissione? Consente la scelta equiprobabile?*

6.5.3 Rappresentatività del campione ed errore campionario

Quando si considera un campione di unità, l'interesse non è limitato a queste perché hanno caratteristiche speciali, ma perché sono rappresentative della popolazione: un risultato ottenuto su di esse dovrebbe essere valido, almeno entro certi limiti statisticamente stabiliti, per tutta la popolazione. L'efficacia di un'indagine parziale è commisurata alla capacità di mimare e miniare la rilevazione completa a cui si sostituisce. Tale capacità è la rappresentatività del campione. Tuttavia, sul singolo campione non è possibile pronunciarsi ed infatti la rappresentatività non riguarda il campione prescelto, ma il modo in cui è stato formato, cioè le potenzialità di errore e non l'errore vero e proprio. Le rilevazioni sulle unità campionate danno una certa immagine dello stato informativo di un problema; quale corrispondenza ci si può aspettare con l'immagine ottenibile se il campione fosse ripetuto ovvero se si considerasse un campione più grande o l'intera popolazione?

Esempi:

a) Friedman (1972, p. 14) ricorda che la rappresentatività del campione può essere determinata solo in relazione alle caratteristiche in esame. I giocatori di una squadra di basket non sono una rappresentanza tipica della popolazione rispetto all'altezza, ma potrebbero esserlo rispetto alle capacità di apprendimento o al metabolismo basale. Se le altezze sono un fatto cruciale da analizzare nelle unità allora la squadra di basket è un campione sbagliato ed occorre una selezione più trasversale.

b) Schofield (1972, p.10) avverte: "*Questi sono i risultati per il campione esaminato.* Tale dichiarazione è usata deliberatamente per avvertire chi legge il rapporto che i risultati possono non corrispondere alla realtà in generale, che il campione è troppo piccolo per generalizzarli e che le conclusioni non possono essere assunte come dei fatti acquisiti riguardo alla popolazione totale.

c) Ragharavao (1988, p. 47) riporta il seguente esempio: un'impresa nel ramo del legno deve disporre -per ogni area di lavoro- di una stima delle piante lasciate in piedi dopo una campagna di tagli (ci sono degli obblighi di legge in questo senso). In passato si era sempre affidata ad una unità di personale molto esperta che forniva la stima richiesta esaminando delle foto aeree. Quando andò in pensione l'impresa rifiutò di sostituirla e si affidò ad un sofisticato (e molto costoso) processo meccanico di campionamento.

Esercizio_TP150: *Vianelli S. e Ingrassia G. (1986) osservano che, in inglese, il termine "campione" è tradotto con la parola "sample" che deriva, attraverso il francese antico "essample", dal vocabolo latino "exemplum" da cui hanno origine anche l'italiano "esemplare" ed "esemplificativo".*
a) Ritenete che questo rispecchi il significato proprio della rilevazione parziale?
b) Cercate dei sinonimi per il termine "campione" e "popolazione".

Esercizio_TP151: *nelle seguenti situazioni di indagine vi sembra che il progetto di selezione possa portare a campioni rappresentativi?*
a) Una pubblicazione elencava le prime 500 società di assicurazioni in Italia. Per valutare la sincerità dei dati di bilancio su cui si basa la graduatoria si scelgono le prime dieci banche con sede legale in uno dei comuni del Nord Est.
b) Per stabilire "chi comanda in borsa" si decide di esaminare la proprietà di un campione delle società quotate. Il campione è formate da quelle società la cui denominazione si compone di almeno due parole;
c) Per localizzare delle discariche abusive si dispone dell'elenco di tutte le cave, fosse, miniere dismesse e prive di un progetto di riqualificazione. Dall'elenco si isolano quelle vicine a grossi centri abitati e fra queste ne viene selezionato un campione casuale.

La simulazione con il campione può essere più o meno fedele, più o meno accurata, anche se non è possibile quantificarne l'errore ("errore" non va inteso alla lettera: non significa che c'è qualcosa di sbagliato nel campionamento, ma solo che ci si attendono delle divergenze tra ciò che da esso risulta e ciò che potrebbe risultare dall'intera popolazione). Per specificare l'errore campionario bisognerebbe conoscere in dettaglio la popolazione, ma questo renderebbe inutile il campione. Sembra un circolo vizioso, ma verrà subito spezzato.

L'obiezione principale alle indagini campionarie è che esse abbiano valore solo nell'ambito in cui si effettuano e non dovrebbero mai esondare da tali limiti: non è possibile sostituire calcoli e congetture alla concreta osservazione o sperimentazione dei fenomeni. Tra gli svantaggi del campionamento c'è infatti l'impossibilità di dare informazioni su tutte le unità e non può quindi essere utilizzato per i conteggi esaustivi necessari ad alcune attività amministrative: anagrafe, liste elettorali, leva militare, albi professionali, elenco delle imprese di un settore. Inoltre, se ripetuto, può dar luogo a esiti diversi togliendo alla Statistica la rassicurante replicabilità di cui beneficiano (però solo apparentemente) scienze più esatte. Soprattutto contiene degli errori ed i suoi risultati non possono essere trasposti meccanicamente al complesso delle unità. In particolare, gli errori sono dovuti al fatto che non si esaminano tutte le unità ed alle fluttuazioni campionarie cioè la naturale variazione dei fenomeni rilevati tra le unità produce campioni diversi ed il confronto tra due di essi darà delle differenze in parte attribuibili all'errore campionario e solo in parte ad un cambiamento del fenomeno.

Esempio:
Una “cacciatrice di teste” ha di fronte N=7 persone ed intende saggiare quale sia il numero di posizioni lavorative in media occupate dagli aspiranti onde tarare i colloqui. Per non allarmarli decide di porre la domanda solo a n=2 di loro. Supponiamo che la situazione della popolazione sia quella indicata dalla tabella. La scelta è fatta ovviamente senza reimmissione; anche l’ordine non ha importanza dato che la somma (e quindi la media) non cambia se si cambia l’ordine degli addendi. Il numero dei possibili campioni è: C(7,2)=21.

Aspirante	Alba	Bice	Ciro	Dino	Emma	Febo	Gina
Lavori	0	1	2	3	4	5	6

	Bice	Ciro	Dino	Emma	Febo	Gina
Alba	0.5	1.0	1.5	2.0	2.5	3.0
Bice		1.5	2.0	2.5	3.0	3.5
Ciro			2.5	3.0	3.5	4.0
Dino				3.5	4.0	4.5
Emma					4.5	5.0
Febo						5.5

Il calcolo della media, per ciascuno dei possibili campioni di n=2 unità, è riportato nella tabella a destra. Il valore vero della media, cioè quello relativo a tutte le unità è 3, che si ottiene solo per i campioni: {Ciro, Emma}; {Bice, Febo}, {Alba, Gina}. In questi casi il campione darebbe la misura esatta della media della popolazione; in tutti gli altri casi c’è un errore. Vediamo l’intero spettro dei valori campionari:

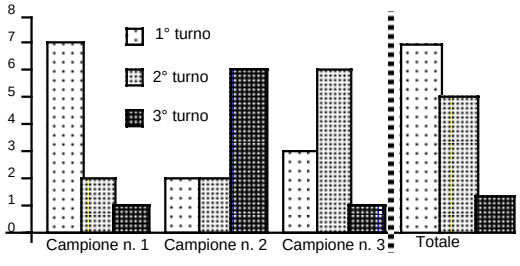
Media	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5
N.campioni	1	1	2	2	3	3	3	2	2	1	1

La responsabile non ha ancora deciso come sceglierà le due unità: due maschi, due femmine, uno ed una, i primi due sull’elenco, gli ultimi due, il primo e l’ultimo, etc. Sa solo che una volta ottenuto il campione deciderà esclusivamente rispetto alle sue risultanze: se le capitano i campioni {Gina, Febo} o {Alba, Bice} ne sarà falsata l’impostazione dei colloqui in quanto il valore campionario è molto lontano da quello della popolazione. In breve, fissata a n=2 l’ampiezza del campione e scelta l’estrazione senza reimmissione, l’unico fattore che controlla è il meccanismo della scelta delle unità. Ne esiste uno ottimale?

L’errore campionario rientra nei fattori rilevanti, ma non controllabili di un problema. Qualunque sia la conclusione raggiunta a mezzo del campione essa include un errore; non solo, il successo del campione nel riprodurre i risultati della popolazione può solo corroborare psicologicamente la validità della procedura per il passato, che magari ci sembrerà più convincente, ma poco di significativo può aggiungere sulla conoscenza del suo comportamento futuro.

Esempio:
Un’imprenditrice controlla uno stabilimento con 100 dipendenti. In vista di una prossima contrattazione vuole conoscere le disponibilità rispetto ai turni di lavorazione (I, II, III). Nella tabella è fotografata la situazione delle attese di ciascun dipendente. Ipotizziamo che, per ragioni di riservatezza, l’imprenditrice possa dialogare solo con n=10 persone e consideriamo tre possibili scelte: i dieci nella prima colonna, i dieci dell’ultima colonna ed i dieci della terza riga.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*8*	*9*	*10*
1	1	1	1	3	1	1	2	3	1	1	2
2	1	1	1	3	2	3	1	1	2	2	1
3	2	2	2	2	2	2	2	1	1	1	3
4	3	3	2	1	3	2	1	2	1	2	3
5	3	1	3	1	1	2	1	1	1	3	3
6	1	1	1	2	2	1	2	2	1	1	2
7	1	3	2	3	1	1	2	2	2	1	1
8	1	1	1	1	2	2	1	1	2	1	3
9	1	2	3	1	2	2	1	3	3	2	3
10	1	2	2	1	2	1	1	1	3	2	3



L’esito è diverso in ciascun campione ed ognuno di essi sbaglia nel riprodurre le scelte del totale dei dipendenti: se si ragiona con il campione “1” si sposterà il peso della produzione sul 1° turno perché questo risulterebbe maggiormente gradito; se invece capita il campione “2” si spingerà sul 3° turno; nel caso del campione “3” si dovranno attivare degli incentivi per spostare qualcuno al 3° turno che sembrerebbe poco gradito. Per ogni campione c’è una strategia che può risultare perdente perché basata su dati distorti. Per evitare il rischio di considerare valido un campione eccezionale (ovvero non rappresentativo) sarebbe opportuna la ripetizione dell’esperimento, anche più di una volta se i tempi e i costi lo consentono.

Esercizio_TP152: *si supponga di voler conoscere la somma complessiva dei numeri riportati nella tabella.*

7	13	5	5	10
2	8	5	4	1
6	10	11	1	12
1	7	8	4	8
2	3	3	1	3

Si sceglie una riga o una colonna (un campione), si calcola la loro somma e la si moltiplica per cinque (proporzionamento alla popolazione). In quale caso il campione dà la stima esatta o ed in quale il massimo errore.

Come si misura la rappresentatività del campione? A questo fine è fondamentale la variabilità che ci si aspetta di riscontrare nella popolazione: se le unità fossero tutte uguali basterebbe un campione di ampiezza n=1 (ad esempio per controllare la qualità di una pezza di stoffa è sufficiente un campioncino di pochi centimetri quadrati), ma se le modalità sono due o più non si potrà essere certi che entrambe siano rappresentate nel campione a meno che non si abbia n=N. Solo che la diversificazione della variabile non è nota, anzi spesso è uno degli scopi della ricerca.

La Statistica ha elaborato diverse tecniche (ad esempio la stratificazione ed il raggruppamento delle unità) che consentono di ottenere una ottima rappresentazione della realtà purché la loro selezione avvenga secondo determinati schemi: i piani di campionamento (*sample design*). Il piano di campionamento è un insieme di tecniche mirate alla migliore selezione delle unità che rendono i risultati più efficaci cioè più prossimi a quelli che si sarebbero ottenuti considerando l'intera popolazione e più efficienti ovvero che non si possano ottenere risultati superiori -a parità di tempi e costi- scegliendo altre unità.

Allo scopo di avere un'idea di come agiscono le tecniche di campionamento presentiamo uno dei piani di campionamento più semplici e che costituisce il nucleo di altri più complessi: il campione casuale semplice (scelta randomizzata delle unità) in cui la scelta delle unità avviene esclusivamente in base a sorteggio garantendo che tutti i possibili campioni di ampiezza "n" ricavabili -con reimmissione o senza reimmissione- dalle N unità della popolazione abbiano la stesse opportunità di essere prescelti. Forse non è inutile sottolineare che il termine casuale è attribuito al meccanismo di scelta delle unità e non all'esito della scelta: come si è appreso dalla teoria della probabilità, nessuna sequenza finita di numeri può dirsi rigorosamente casuale. Smith (1991, p.315) sottolinea l'apparente contraddizione di tale schema che non manca di suscitare perplessità tra gli studenti: da un lato c'è l'esigenza di ottenere un campione che possa sostituirsi alla popolazione e poi per realizzarlo si propone di scegliere le unità affidandosi alle bizzarie della sorte. Se disporre di un campione rappresentativo è così importante, perché non si cerca di ottenerlo con metodi più sicuri? Nello studio delle tecniche di campionamento si approfondisce la questione del perché una selezione casuale delle unità sia preferibile alla scelta discrezionale. Qui faremo solo considerazioni generiche agganciate al postulato empirico del caso discusso all'inizio del capitolo.

Esempio:
Una popolazione è formata da tre tipi di unità: A, B, C di cui è nota la proporzione nella popolazione: p(A)=50%, p(B)=30%, p(C)=20%. Dalla popolazione sono prelevati con reimmissione dei campioni di varia ampiezza per valutare le percentuali campionarie.

Ampiezza	A	B	C
n=10	0.6	0.2	0.2
n=100	0.51	0.29	0.2
n=1000	0.512	0.293	0.195
n=10000	0.5017	0.2983	0.2
n=100000	0.50047	0.302	0.19573
n=1000000	0.500482	0.301929	0.197589
Popolazione	0.5	0.3	0.2

La tabella mostra che l'approssimazione migliora all'aumentare dell'ampiezza campionaria, ma il miglioramento non è uniforme: per certe ampiezze peggiora cioè anche per campioni più numerosi non si ottiene un avvicinamento, anzi si ha un allontanamento, sia pure di scarsa entità (nell'ipotesi che gli errori di arrotondamento abbiano avuto la cortesia di rimanere fuori dalla porta da questo calcolo).

L'esperimento, come tanti altri dello stesso genere, mostra che in un campione casuale abbastanza grande, le unità sono guidate dalla sorte a comparire nelle medesime proporzioni con cui sono presenti nella lista della popolazione. Il singolo campione può avere una conformazione più o meno simile a quello della popolazione, ma non è dato sapere in che misura (a meno che non si tratti di una simulazione in cui la popolazione sia conosciuta). Ciò che tranquillizza gli utilizzatori della Statistica è che un sorteggio corretto delle unità e per ampiezze campionarie elevate tenderà a riprodurre le caratteristiche della popolazione.

Esempio:
Riprendiamo il problema posto nell'esercizio CB32 in cui la responsabile della "Italian Camping" doveva stimare il totale delle persone con esigenze di diete particolari su di un totale di 300 soggetti ed avendo il tempo di consultarne solo un campione di ampiezza n=10. L'universo dei campioni -senza reimmissione- contiene 1400 miliardi di miliardi elementi. Il totale relativo alla popolazione è 3015. Ecco i risultati di alcune simulazioni con il computer.

n(10)	20	40	80	160	320	640
T	3051.00	3095.63	2894.06	3046.97	3001.59	3018.61

Il simbolo n(10) indica il numero di campioni di ampiezza 10 scelti casualmente dall'universo dei campioni. La stima del totale è stata ottenuta calcolando il totale medio sugli n(10) campioni e moltiplicando poi per 30 il risultato. Anche in questo caso si può notare la confortante convergenza verso il valore vero quando la procedura è ripetuta molte volte.

Se c'è differenza tra ciò che risulta dalle unità selezionate e quello che risulterebbe dalle unità selezionabili, la causa è la sorte e non fattori sistematici ancora da scoprire. Inoltre, all'aumentare di "n" si avrà la riduzione tendenziale dell'errore campionario (postulato empirico del caso).

Esercizio_TP153: *un modo per verificare la rappresentatività del campione è di confrontare popolazione e campione rispetto a caratteristiche note (ad esempio persone per età, professione, residenza), ma diverse da quelle oggetto d'indagine. Se il campione è simile alla popolazione rispetto a tali profili dovrebbe risultarlo anche rispetto a quelli di nostro interesse.*
a) Vi sembra una procedura plausibile? b) Come procedereste per assicurare la rispondenza del campione alla struttura nota (ad esempio quella per sesso ed età) della popolazione? c) Fissati gli strati si sceglie un campione dello strato di ampiezza proporzionale alle unità nello strato. E' ragionevole?

Sono anche possibili errori non campionari dovuti sia a sviste nell'acquisizione dei dati che all'uso di informazioni imprecise, incomplete o incomprensibili. Ad esempio, nei sondaggi telefonici, in dipendenza dell'ora in cui si telefona, si raggiungono unità diverse persino sulla stessa utenza. Se una lista per intervistare persone sposate è formata elencando solo le mogli, la probabilità che un marito vi sia incluso è zero (unità *blank*) se la moglie non è selezionata ed uno (unità autorappresentativa) se la moglie è stata scelta. Del resto, non tutte le unità che si è previsto di esaminare sono di fatto esaminabili o disposte a fornire i dati richiesti ed a fornirli veritieri; inoltre, le informazioni possono essere inavvertitamente o volutamente alterate da chi li rileva o dagli strumenti impiegati. Questi sono errori che, a differenza di quelli campionari possono essere evitati e se non si provvede non scompaiono neanche analizzando tutte le unità della popolazione.

Esempio:
Le proiezioni elettorali sono un caso privilegiato di campionamento in cui si riescono poi a conoscere i valori esatti a livello di popolazione (almeno per i voti validamente espressi). In televisione sono comunicati gli esiti parziali forniti dagli istituti di ricerca nonché quelli "ufficiali" del Ministero dell'interno. Questi ultimi si sono quasi sempre mostrati lontani dall'esito definitivo pur riguardando porzioni cospicue dell'elettorato. La ragione, suggerisce S. Draghi (1995) è la distorsione sistematica dei criteri di afflusso dei dati che arrivano dopo il perfezionamento dell'iter di spoglio; le stime ottenute con un campione rappresentativo di poche sezioni sono già prossime ai valori finali. Non sempre però: sono ben noti alcuni casi in cui tecniche affidabili e ben consolidate hanno avuto clamorosi insuccessi.

E' evidente quindi che non sempre il campione fornisce un'idea esatta o anche solo moderatamente buona della popolazione e che l'errore varia in modo imprevedibile al variare del campione. Si sa però che la rappresentatività è influenzata positivamente dal numero di unità prelevate e dal modo in cui avviene il prelevamento.

Esercizio_TP154: *supponiamo che la popolazione sia infinita. Può esistere un campione rappresentativo?*

L'ampiezza del campione

E' la determinante essenziale anche se non esclusiva della rappresentatività del campione. Se il pino di campionamento è efficiente l'aumento dell'ampiezza può solo migliorare l'attendibilità dei risultati, ma se è sbagliato, l'aumento delle unità potrebbe essere inutile se non dannoso. L'effetto dell'ampiezza si esplica attraverso i due rapporti: frazione di campionamento f=n/N e intervallo di campionamento h=N/n che indicano, rispettivamente, la quota di unità inclusa nel campione ed il numero di unità escluse comprese tra due unità incluse. Se la popolazione è infinita questi due rapporti perdono di significato; se è invece indeterminata diventano dei parametri incogniti da stimare. I due rapporti variano secondo la popolazione indagata e le finalità dell'indagine. Una ricerca riguardante i residenti in Lombardia può avere una frazione di 4 a 10 mila; un sondaggio tra gli abitanti di una circoscrizione di Pavia può dover considerare una frazione di 2 a 10 per cogliere gli aspetti più interessanti.

Esempi:
a) Se la popolazione include 3'000 soggetti e tra questi si sceglie un campione di 300 la frazione di campionamento è pari a 300/3000=0.1 o 10% e l'intervallo di campionamento è 3000/300=10 cioè due unità incluse sono mediamente divise da dieci unità escluse.

b) Per il prelievo di campioni di benzina da analizzare rispetto al contenuto di benzene e di idrocarburi aromatici, l'ampiezza del campione è fissata per legge: 5 litri immessi immediatamente in cinque contenitori di contenuto non inferiore al mezzo litro.

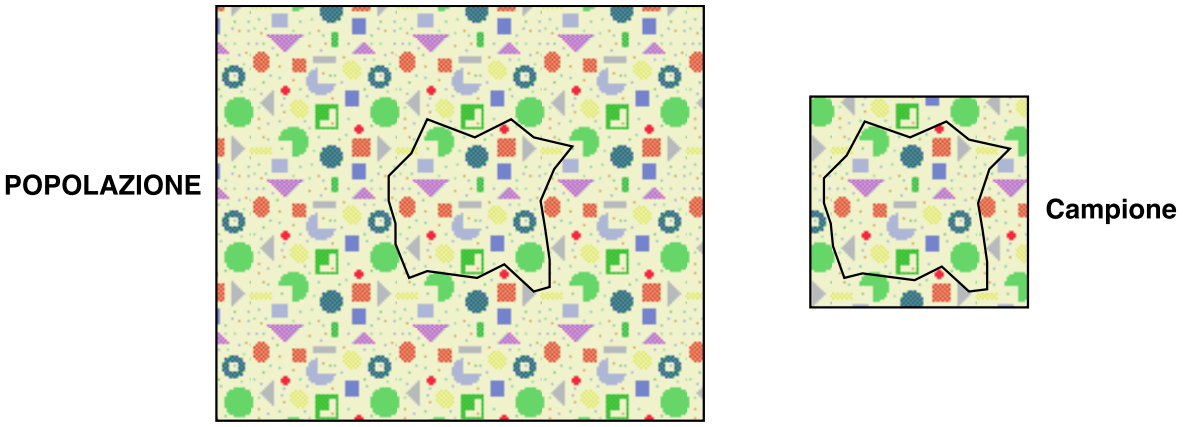
c) Il piano di rimborso di un prestito obbligazionario è spesso attuato progressivamente stabilendo fin dall'inizio il numero di obbligazioni rimborsate ogni anno. I titoli sono poi sorteggiati.

d) Esistono situazioni in cui l'ampiezza del campione è indeterminata in quanto si selezionano unità fino a raggiungere un certo ammontare di una variabile: volume, peso, etc. Questo perché le singole unità non sempre sono osservabili o distinguibili al punto da poter essere contate separatamente: particelle o pezzetti di minerale, fertilizzanti, cementi. Se tali particelle non sono regolari rispetto a ciò che ne controlla l'acquisizione l'ampiezza del campione è sconosciuta.

e) Non sempre è opportuno fissare a priori l'ampiezza del campione. Ad esempio nei test distruttivi o nelle sperimentazioni cliniche si preferiscono procedure che lascino aperta la possibilità di sospendere la sperimentazione se questa sembra andare in una direzione pericolosa, senza danneggiare o invalidare i risultati già stabiliti.

La determinazione dell'ampiezza del campione è un elemento base del piano di campionamento e deve essere gestito con attenzione: un campione troppo piccolo potrebbe non rappresentare adeguatamente la popolazione, un campione troppo grande rischierebbe di sprecare risorse. Una certa ampiezza è adatta per alcuni scopi, ma poi il campione non può essere riutilizzato per finalità di dettaglio: le 5'000 famiglie seguito dall'Auditel per rilevare l'ascolto televisivo possono andar bene per emittenti nazionali, ma non per quelle locali. In breve, stabilire quanto debba essere "n" dovrebbe essere il primo passo del piano di campionamento ed è invece l'ultimo.

Esempio:
Ripreso da Wilburn (1984, p. 36). La figlia chiede vuole un vestito con il medesimo disegno di quello della madre. Che campione si dovrà portare al negozio di stoffe?



Deve essere abbastanza piccolo per evitare di impacchettare l'intero vestito, ma deve anche essere abbastanza grande da includere il motivo ricorrente della stoffa; non solo, dovrebbe essere limitato a questa ampiezza senza ripetere, neanche in parte il motivo.

Esercizio_TP155: *nei campioni fortuiti o accidentali le unità sono incluse indipendentemente dalla volontà di chi forma il campione: resti preistorici, i casi di una malattia mai ancora diagnosticata, le galassie scoperte dagli astronomi, gli edifici risparmiati da una catastrofe naturale.*
a) Stabilite frazione ed intervallo di campionamento;
b) In che cosa differiscono dal campione casuale semplice?

Ammesso che nella popolazione ci siano unità sufficienti per formare un campione di ampiezza inappuntabile, ciò che ne governa la numerosità è il rapporto costo/qualità. Il *budget* per l'acquisizione dei dati è limitato e deve essere ben amministrato. Il singolo dato ha un costo di accesso che è in parte fisso e riguarda allo stesso modo tutte le unità, tanto che può essere imputato ad esse in parti uguali (spesa per la modulistica, tempo *standard* per la compilazione, tariffa di entrata in banche dati). Un'altra parte è variabile e cambia da unità ad unità (ad esempio il loro costo di reperimento) di modo che diventa possibile confrontare l'inserimento di una nuova unità nel campione con l'aumento o la diminuzione di rappresentatività (riduzione -potenziale- dell'errore campionario) che ne deriva. E' qui che si esce dalla ragioneria e si entra nella Statistica.

6.5.4 Sorteggio delle unità

La selezione delle unità di una popolazione da inserire nel campione si può basare come abbiamo visto su di un idea essenziale: la scelta casuale o sorteggio. E' possibile stabilire che la sorte stia agendo in modo corretto? L'equità è il punto essenziale; la forza parificatrice della sorte applicata ad un insieme di oggetti identici garantisce che ognuno di essi sia scelto con la stessa frequenza purché la sua azione possa esplicarsi abbastanza a lungo nelle medesime condizioni.

Esempi:
a) Se ci capita di fare da scrutatori, prima di procedere allo spoglio sarà bene mescolare diffusamente le schede in modo da non cominciare dalle ultime deposte (oppure le prime, se l'urna viene capovolta) perché votate da gruppi non rappresentativi dei votanti: solerti o tiratardi. Se l'urna e ben mischiata, le prime schede danno una buona idea di come sono andate le votazioni nel vostro seggio.

b) La forza del cemento o la potenza di alcuni esplosivi dipendono strettamente da una adeguata mescolatura degli elementi.

c) L'assegnazione casuale è la procedura con la quale i 2n soggetti di una sperimentazione sono scelti, in numero di "n", per la somministrazione di un trattamento: scelte casualmente le prime "n" anche le rimanenti "n" non scelte sono determinate casualmente.

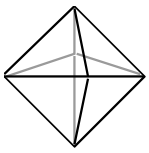
d) L'estrazione dei biglietti vincenti in una lotteria può avvenire con i bambini bendati o con mezzi elettronici, sotto il controllo effettivo e costante dei rappresentanti del ministero delle finanze. Il sistema usato in una trasmissione televisiva venne integrato con colpi sul retro dell'urna che smossero una biglia rimasta incastrata. Il fatto diede luogo a polemiche roventi e danni per l'erario.

Esercizio_TP156: *C. Gini propose il caso dello studio di un carattere antropometrico di una popolazione di maschi ventenni formando un campione scegliendo le unità aventi un cognome che iniziava con una fissata lettera iniziale. Ritenete casuale questa tecnica di selezione?*

Il comportamento della sorte può essere simulato in molti modi. Il lancio di oggetti di foggia regolare: monete, dadi, astragali, ha una tradizione antica.

Esempi:
a) L'ottaedro ha 8 facce uguali in forma di triangolo che possono essere numerate da 0 a 7. Se fatto rotolare su di una superficie piatta e liscia finirà col poggiarsi su una delle facce. Il numero sulla faccia nascosta sarà il prescelto. Se il lancio è ripetuto per 5 volte possiamo inglobare le 5 uscite in un solo numero disponendo in ordine i valori. Tale numero non è subito utilizzabile dato che è espresso in base ottale; è però facile trasformarlo in base decimale:

$$D = \sum_{i=1}^{m} C_i 8^{m-i}$$



dove C_i è la cifra in posizione "i" ed "m" il numero di cifre. Se ad esempio la sequenza dei lanci dell'ottaedro è stata: 6,0,5,3,2 il numero decimale ottenuto è: $6*8^4+0*8^3+5*8^2+3*8+2=24'922$. L'unità etichettata con questo numero farà parte del campione.

b) Inoue ed al. (1983) utilizzano -con un apposito apparato- i raggi gamma emessi da un nucleo radioattivo sfruttando il fatto che i nuclei decadono separatamente, l'energia dei raggi gamma è sufficiente a farli distinguere dal rumore di fondo e la turbolenza indotta da altre radiazioni non è rilevante. Inoltre, se p_i è la probabilità del verificarsi della cifra i-esima si è accertato che:

$$p_i = \frac{1+\varepsilon_i}{10};\ i = 0,1,2,\ldots,9; \quad |e_i| \le 5x10^{-6}$$

c) Kendall e Babbington-Smith nel 1939, per costruire la loro famosa tavola di un milione di cifre casuali, hanno adoperato un disco diviso in dieci settori, fatto ruotare e fermato a caso.

d) Le cifre casuali decimali possono essere ottenute rotolando un cilindro con 10 sfaccettature di uguale superficie oppure ripetendo le cifre due volte sulle 20 facce dell'icosaedro. Un'altro metodo lo indica Bradley (1976, p. 59). Si lancia un dado regolare: se esce il 6 si ripete il lancio e si continua a lanciare finché non esce X≠6. A questo punto si lancia una moneta con due facce equiprobabili: se esce croce il numero casuale è Y=X se esce testa il numero casuale è Y=(X+5)mod 10 cioè il resto della divisione di X per dieci.

Ogni processo fisico che ricalchi l'equo sorteggio può essere adoperato per selezionare le unità del campione. Nella evoluzione dei computer sono noti diversi dispositivi -basati sul comportamento di alcuni tipi di diodi- che generano numeri casuali ed è da diverso tempo che essi sono presenti nelle calcolatrici tascabili e nei *videopoker.* Il problema, con questi dispositivi, è che il loro comportamento non è stabile nel corso del tempo: fintanto che includono componenti soggette ad usura la loro affidabilità è destinata a diminuire con l'uso ed il loro mantenimento in condizione di corretta operatività richiede manutenzione continua ed un costante monitoraggio.

Esercizio_TP157: *il cubo è il più classico dei poliedri usati per i giochi di sorte. Supponete di effettuare 7 lanci e di affiancarne i valori per costituire le cifre del numero casuale.*
a) In che base è espresso il numero? b) Qual'è il minimo ed il massimo che si può ottenere in base decimale? c) I poliedri regolari sono cinque. Quali sono gli altri? Perché si è diffuso solo l'uso del cubo?

Esercizio_TP158: *una concessionaria ha in giacenza 97 veicoli classificati in base alle difficoltà di vendita (gli stessi dati usati per introdurre il diagramma a punti nel capitolo 2). Per ottenere un campione con reimmissione si può usare il foglio elettronico Excel richiamando l'aggiunta denominata analisi di dati e nel conseguente sottomenu scegliere campionamento e proseguire con le istruzioni ottenendo un campione di ampiezza 12.*

La tavola dei numeri casuali

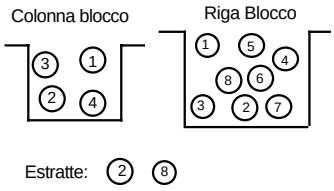
Molti esperimenti hanno dimostrato che è praticamente impossibile garantire le condizioni della pura sorte: mischiare le carte, gettare dei dadi, agitare le biglie, mescolare dei biglietti, uniformare la rappresentazione fisica delle unità in modo da non dare vantaggi inappropriati è un'attività che richiede controlli permanenti ed accurati nonché regolari interventi di manutenzione perché ne sia garantito il funzionamento stabile nel tempo come meccanismo equo ed imprevedibile. Le condizioni non possono essere tenute costanti come si vorrebbe, soprattutto quando il numero degli oggetti tra cui scegliere è elevato. D'altra parte, se la popolazione è molto numerosa ci si può trovare nella impossibilità di mettere in opera un meccanismo fisico di selezione delle unità. Per aggirare tali problemi si usano le tavole di numeri casuali (un esempio è dato in appendice) che hanno più ampia applicabilità ed hanno una discreta tradizione nel campionamento (una delle prime pubblicazioni che li riguarda sono i *Random Sampling Numbers* di L. Tippett nel 1927).

Le tavole dei numeri casuali sono formate da sequenze di cifre da 0 a 9 variamente raggruppate (l'organizzazione in blocchi ha il solo scopo di facilitarne l'uso) e caratterizzate dall'assenza di una qualsiasi legge di successione o di ordinamento. Il numero che si trova in una data posizione non ha alcuna relazione con quelli adiacenti o di altra zona e nella tavola il movimento è libero purché regolare: per riga o per colonna, in diagonale, da sinistra a destra e viceversa, dall'alto verso il basso e al contrario. La costruzione della tavola garantisce che le cifre da "0" a "9" tendono ad avere tutte la stessa frequenza, così come uguali frequenze relative hanno le coppie da "00" a "99", le terne: "000"…"999" e così via. Lo stesso accade per tali sequenze quando si considerino le foro frequenze condizionate ad una particolare lunghetta: "00|11" ha la stessa frequenza di "00|22", "00|33", etc. Una sequenza di cento "0", di cento "1", di dieci sottosequenze "0123456789" o di dieci blocchi "0000000000", "1111111111" sono selezionabili su basi di assoluta parità, almeno per tavole abbastanza estese, anche se ciò potrebbe non apparire a prima vista. Le proprietà anzidette rimangono valide anche se i numeri si scambiano di posizione con (ad esempio il primo con l'ultimo, il secondo con il penultimo e così via). I numeri casuali sono forse uno dei primi esempi di globalizzazione della cultura: possono essere letti da sinistra a destra come nel mondo occidentale, da destra a sinistra come nel mondo arabo, dall'alto verso il basso come in cinese o in coreano.

Si voglia estrarre, ad esempio, un campione di n=1 unità da una popolazione di N=3754 unità. Poiché N>1000 la scelta richiede gruppi di quattro cifre. Decidiamo di procedere per riga e che il blocco sia formato dalla prima cifra e dalle tre immediatamente a destra; se si arriva al termine di riga si passa all'inizio della riga successiva. Precisiamo inoltre da quale riga e colonna (o da quale cifra) si comincia. Anche queste dovrebbero essere scelte casualmente: se così non fosse ci si potrebbe ricordare che la 6ª riga e la 12ª colonna della tabella si incrociano sulla cifra "1" che è seguita da "5", "0" e "8" per cui la unità prescelta sarebbe la n. 1508 e questa scelta non può certo dirsi casuale. In alternativa, si potrebbe formare il numero da estrarre scegliendo ogni sua cifra in una diversa riga o in una diversa colonna, ma il problema del ricordo rimarrebbe.

Esempio:

Scriviamo i numeri dei blocchi su dei bigliettini inseriti all'interno di biglie indistinguibili dall'esterno e collochiamole in due scatole. Agitiamo le scatole a lungo, con movimenti energici e sapienti come un barista che prepari un *cocktail* per un tavolo di commensali importanti. Da ciascuna delle scatole si sceglie una biglia il cui numero fornirà il blocco prescelto: supponiamo che siano "2" per la colonna e "8" per la riga (analoga tecnica deve essere usata per scegliere la pagina nel caso la tavola sia articolata in più pagine). Il numero in alto a sinistra è la posizione nella lista dell'unità campionata: "1" e "3", "0", "5" sono le cifre a destra per cui l'unità prescelta è quella in posizione 1305. Se il campione è con reimmissione allora il ripetersi del numero rigenera la stessa unità già inserita nel campione; se la campionatura è senza reimmissione il numero già prescelto deve essere scartato.



Esecizio_TP159:

a) Per utilizzare una tavola di numeri casuali sono sconsigliabili procedure naif quali quelle di colpire alla cieca il foglio con una matita oppure lanciare uno spillo per cominciare da dove è caduta la punta. Perché?
b) Riuscite ad immaginare una tavola, anche estesa, formata da un egual numero di "0", "1", "2",…, "9" e, nello stesso tempo, assolutamente non casuale?

Se il numero ottenuto dalla tavola è superiore ad N ci sono due possibilità: a) Escluderlo e procedere ad una nuova selezione in un modo prestabilito. b) Sottrarre ripetutamente il numero massimo N dal numero trovato "r" fino ad ottenere un risultato inferiore o uguale a N. Ciò equivale a prendere il resto della divisione di "r" per N. L'estratto sarebbe quindi:

$$i = Resto(r, N) \;=\; r - \left[\frac{r}{N}\right] * N$$

dove [.] indica, come sempre, la parte intera del suo argomento. Se, nell'esempio precedente, si fosse scelta la colonna "3" avremmo trovato 9307 ed invece di abbandonarlo l'avremmo trasformato nel numero:

$$i = 9307 - \left[\frac{9307}{3754}\right] * 3754 = 9307 - 2 * 3754 = 1799$$

Questo metodo è utile per risparmiare estrazioni quando N è appena superiore ad una delle potenze di dieci. Dobbiamo però avere l'accortezza di escludere i numeri superiori a $[10^k/N]*N$ dove "k" è il numero di cifre di N, in quanto la loro presenza dà una possibilità in più ai numeri piccoli. Infatti, il resto uno può essere ottenuto da 3755 e da 7509; invece, il resto 3754 si può avere solo da 7507. Nell'esempio, si devono eliminare i numeri maggiori di

$$\left[\frac{10^4}{3754}\right] * 3754 = 2 * 3754 = 7508$$

Esempi:

a) Nella tabella che segue è descritta la composizione degli addetti dell'industria aeronautica nel periodo 1981-1995. Siamo interessati al valore campionario del rapporto operai/impiegati. Scegliamo un campione, senza reimmissione, di n=3 pari al 20% della popolazione totale. Sorteggiamo il blocco di partenza e troviamo che è il (3,2) che dà 57 (trasformato in 13), 99 (scartato perché superiore a [100/15]*15=90), 16 (trasformato in 2), 96 (scartato), 56 (trasformato in 12). Il campione finale è C={1982, 1993, 1994}.

$$\frac{\frac{24200}{18400}+\frac{16400}{20600}+\frac{15000}{2000}}{3} = 0.5355$$

Anno	Operai	Dir. e Imp.
1981	24500	17500
1982	24200	18400
1983	23900	18500
1984	23400	19300
1985	22700	20600
1986	22500	22200
1987	22800	24200
1988	23600	25900
1989	24200	26300
1990	24100	26600
1991	21800	25200
1992	18800	23200
1993	16400	20600
1994	15000	20000
1995	14000	19500

Nella popolazione il rapporto è 0.9969 per cui l'approssimazione campionaria risulta insoddisfacente. In effetti, l'ampiezza del campione è troppo piccola per poter dare una valore presunto attendibile.

b) Riprendiamo i dati del compito SD126 in cui 75 pazienti erano monitorati rispetto alle dosi ricevute di uno psicofarmaco. Si ritiene necessario un controllo su n=20 soggetti per verificare la presenza di effetti collaterali imprevisti. Ad ogni paziente è assegnato un numero da 1 a 75 procedendo per riga da sinistra verso destra nella tabella del compito. Scelto casualmente il blocco di inizio della tavola di numeri casuali si ottiene la sequenza di coppie di cifre (blocco 7.1, 1ª riga): 53 81 29 13 39 35 01 20 71 34 62 33 74 82 14 53 73 19 09 03 da questa dobbiamo escludere 81 e 82 in quanto numeri fuori lista e non possiamo neanche recuperarli operando in modulo in quanto [100/75]*75=75. Anche il 2° 53 deve essere escluso dato che il campionamento è senza rimessa. Nella riga successiva troviamo utili 5, 32 e 68. Poiché la lista comincia dal numero 1 le unità prescelte sono: 54 52 30 14 40 36 02 21 72 35 63 34 75 33 15 69 74 20 10 04 che portano ai valori campionari:

1 40 5 84 129 322 8 13 737 18 37 127 119 163 573 22 5 122 32 56

Tale scelta non garantisce che il campione sia rappresentativo, ma esclude distorsioni: non è formato dai pazienti meglio in salute o di salute più cagionevole, più anziani o più giovani, più uomini che donne, che meglio rispondono alla terapia o che l'avvertono poco. Hollander e Proschan (1984, pp. 90-92) discutono un esempio analogo ed osservano che, guardando ex post, l'esito della scelta non è difficile trovare dei controsensi nei numeri casuali (rispetto all'equiprobabilità ed alla casualità). Ad esempio il "3" è presente dieci volte su di un totale di 80 cifre (due in più) ed il "2" solo quattro volte (quattro in meno). Inoltre i dispari sono 14 ed i pari solo 6. Questo non è sorprendente ed infatti sulla singola serie (peraltro di numerosità limitata) non possiamo pronunciarci: è il meccanismo che conta e tavola dei numeri casuali si è di consolidata affidabilità.

Esercizio_TP160: *la dott.ssa Molinaro Emilia conduce una ricerca sull'alimentazione infantile ed utilizza una lista di quattro cifre. Per sbrigarsi usa una sua variante della tavola dei numeri casuali: fissato un punto di avvio, invece di leggere in sequenza i numeri della tavola considera le quattro cifre che precedono e seguono tale punto e, se possibile, quelle immediatamente sopra e quelle sotto. Vi sembra corretto?*

Esercizio_TP161:

a) Sia N=1103 e si estragga un campione di ampiezza n=7 partendo dal blocco (7,2) leggendo da sinistra a destra e procedendo di riga in riga verso il basso;

b) Quali accorgimenti occorrerà adottare se le unità sono numerate a partire da un numero diverso da zero o da uno?

c) Supponendo di dover estrarre un campione molto numeroso da una popolazione di un milione di unità in che modo allarghereste la tavola della pagina precedente per rendere possibile le estrazioni?

d) L'esito delle estrazioni del lotto potrebbe essere usato per formare una tavola di numeri casuali?

e) Per scegliere un campione casuale 20 da una popolazione di 100 si esaminano le coppie di numeri che occupano una posizioni multiple del 13 nella tavola fissando la prima posizione in alto a sinistra. E' casuale?

Le tavole dei numeri casuali non sono limitate alla estrazione di interi, ma possono simulare anche l'estrazione di una frazione casuale. Basta fissare il numero di cifre e poi procedere come per gli interi scegliendo il blocco ed il modo di proseguire dopo la prima scelta; al numero così ottenuto si premette "0." ed ecco la frazione casuale. Ad esempio, per la prima riga del blocco (1,1) e per una frazione di 5 cifre abbiamo: 0.53742 e 0.39967 con l'unico difetto dell'impossibilità di estrarre in questo modo una frazione pari ad uno che sarà di scarsa importanza se le cifre del numero sono parecchie.

Esempi:

a) La frazione casuale risolve efficacemente il problema di scegliere da una lista che abbia codici compresi in un intervallo [a, b] estremi inclusi. Infatti, data la frazione casuale "q" si può utilizzare la relazione lineare: p=a+(b-a)q se poi "p" deve essere un intero in [a, b] si applica la formula: p=[a+(b-a)q]

b) Knuth (1981,vol. 2, p. 121) suggerisce la procedura seguente per estrarre -senza reimmissione- un campione di ampiezza "n" da un popolazione di "N" unità (numerate da 1 a N). Sia "t" l'unità prossima da considerare e sia "m" il numero di posizioni del campione già occupate.
1) Si seleziona una frazione casuale "U" dalla tabella dei numeri casuali (o con un altro metodo).
2) Se (N-t)*U≥(n-m) l'unità t-esima non è inclusa. Si aumenta "t" di 1 e si ritorna al punto 1.
3) Se (N-t)*U<(n-m) l'unità t-esima è inclusa nel campione. Si aumenta "m" di 1 e se m≤n si ritorna all'inizio del punto 2.
Applichiamo l'algoritmo di Knuth alla selezione di cinque nominativi tra i quelli riportati in tabella ed ottenendo i numeri casuali dalla tavola data in precedenza a partire dalla prima riga del blocco (5,4) per gruppi di due cifre continuando per righe dello stesso blocco.

1	Bellini Vincenzo	16	Moffo Anna	31	Rossini Geoacchino
2	Boccanegra Simone	17	Montesano Enrico	32	Tebaldi Renata
3	Buzzati Dino	18	Moricone Ennio	33	Toscanini Arturo
4	Callas Maria	19	Nazzari Amedeo	34	Ughi Uto
5	Carducci Giosuè	20	Nenni Pietro	35	Verdi Giuseppe
6	Caruso Enrico	21	Occhini Ilaria		
7	De Curtis Antonio	22	Orlando Silvio		
8	De Rosa Sergio	23	Paganini Nicolò		
9	Del Monaco Mario	24	Pascoli Giovanni		
10	Donizetti Gaetano	25	Pavarotti Luciano		
11	Jotti Nilde	26	Petacci Claretta		
12	Leopardi Giacomo	27	Ponchielli Amilcare		
13	Lucherini Armando	28	Proietti Gigi		
14	Mascagni Pietro	29	Puccini Giacomo		
15	Mastrocinque Camillo	30	Rendano Alfonso		

1.N=35, n=5, t=1, m=0. U_1=0.42; Poiché 34*0.42=14.28>5 la 1ª unità non è inclusa; t=2, U_2=0.35, 33*0.35=11.55>5; …, t=9,U_9=0.18, 26*0.18=4.68<5; "Del Monaco Mario" fa parte del campione.
2. t=10, m=1, U_{10}=0.51, 25*0.51=12.75>4; t=14,U_{14}=0.06, 21*0.0.6=1.26<4; "Mascagni Piero" fa parte del campione.
3. t=15, m=2, U_{15}=0.07, 20*0.07=1.4<3; "Mastrocinque Camillo" è incluso
Di seguito saranno inclusi "Orlando Silvio" e "Pascoli Giovanni".

Esercizio_TP162: *Bissell (1996) propone un algoritmo per estrarre un campione casuale senza reimmissione di ampiezza "n" da una popolazione di "N" unità che risulta tre volte più rapido di quello suggerito da Knuth.*

0. Definiamo r=N-n, m=N, k=0.
1. Si ottiene un numero casuale uε [0,1] e si pone p=1.
2. Si calcola p=pr/m.
3. Se p≤u allora l'unità in posizione N-m+1 della lista entra nel campione.
4. Porre k=k+1; m=m-1; se k<n tornare al punto 1 altrimenti stop.
5. Se p>u porre m=m-1, r=r-1. Tornare al punto 2.

Applicate la tecnica di Bissell alla selezione di un campione 4 sedimi (scali aerei) usando come numeri casuali quelli del blocco (6,2) per gruppi di due cifre procedendo da sinistra a destra, riga per riga.

Num.	Sedime	Movimenti	Num.	Sedime	Movimenti
1	Alghero	6'991	18	Napoli	43'429
2	Ancona	14'652	19	Olbia	19'799
3	Bari	11'636	20	Palermo	29'080
4	Bergamo	26'933	21	Perugia	2'028
5	Bologna	45'901	22	Pescara	3'587
6	Brindisi	5'591	23	Pisa	17'124
7	Cagliari	19'851	24	Reggio C.	4'988
8	Catania	32'037	25	Rimini	3'827
9	Crotone	2'638	26	Roma C.	19'085
10	Cuneo	15'063	27	Roma F.	194'007
11	Firenze	25'712	28	Ronghi	74'747
12	Foggia	4'864	29	Torino	39'799
13	Forlì	6'167	30	Trapani	1'838
14	Genova	23'503	31	Treviso	4'324
15	Lametia T.	7'146	32	Venezia	43'424
16	Milano L.	70'496	33	Verona	23'155
17	Milano M.	164'492			

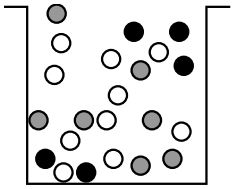
Selezione delle unità con probabilità ineguali

Il campionamento casuale potrebbe essere realizzato dando a ciascuna unità "i" una propria probabilità p_{ij} di essere inclusa nel campione nella posizione j-esima. In effetti, l'equiprobabilità di inclusione non è essenziale per assicurare la casualità delle scelte né se si forma con reimmissione né se si forma senza reimmissione; anzi, il campionamento con probabilità ineguali può garantire -a parità di ampiezza campionaria- livelli di rappresentatività maggiori del campionamento casuale semplice (cfr. Fabbris, 1993, cap. 2).

La diversificazione della probabilità di inclusione (attitudine, tendenza, propensione, *chance*) a comparire nel campione per l'unità i-esima consente di precisare due casi estremi utili nella realizzazione del campionamento. Si ha $p_{ij}=0$ per le unità *blank* e $p_{ij}=1$ per le unità autorappresentative. Per le altre unità la probabilità di inclusione può coincidere con una proprietà fisica delle unità, ma può anche essere una caratteristica loro attribuita da chi deve formare il campione (ad esempio dalla disponibilità a rilasciare dei dati). Se il campione deve servire per valutare una numerosità, si può pensare ad un contenitore: un'urna, un cappello, una scatola, etc. nella quale sono posti tanti bussolotti quanti sono i conteggi di pertinenza delle varie unità.

Esempi:

a) La probabilità di inclusione in una popolazione di imprese potrebbe essere misurata dal numero di addetti. In questo caso si metterebbero nell'urna tanti bigliettini con il nome dell'impresa (oppure tante biglie di un colore univocamente associato all'impresa) per quanti sono i suoi addetti.



b) La verifica campionaria della disponibilità di insegnanti nelle scuole elementari e materne dovrebbe scegliere le scuole con probabilità di inclusione legata ai posti in organico. Se $\{d_i, i=1,2,\ldots,k\}$ è la serie delle disponibilità nei "k" istituti, $N=\sum d_i$ il totale dei posti e $\alpha=n/N$ la frazione di campionamento, la probabilità di inclusione dell'unità i-esima dovrebbe essere $d_i n/N$. Ipotizziamo che per ogni scuola esista una lista in cui gli insegnanti sono elencati in un ordine definito ed associati ai numeri progressivi da 1 a d_i.

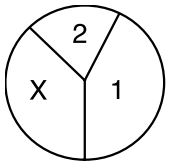
d_i	d_i/N	d_i*n/N	arrot.
38	0.14615	3.8	4
61	0.23462	6.1	6
44	0.16923	4.4	4
25	0.09615	2.5	3
92	0.35385	9.2	9
260	1.00000	26.0	26

Ad esempio N=5 e d={38, 61, 44, 25, 92} con T=260; poniamo inoltre n=26. A questo punto il campione si forma scegliendo 4 numeri casuali interi nell'intervallo [1, 38], 6 in [1, 61], 4 in [1, 44] e così via per poi far confluire nel campione i nominativi corrispondenti delle varie liste di istituto.

L'immagine che meglio può rappresentare la ripartizione dell'attitudine legata ad una variabile continua (ad esempio l'estensione delle unità areali nel caso di serie territoriale) è quella di un disco suddiviso in settori di arco proporzionale alla probabilità di inclusione delle unità.

Esempi:

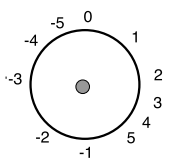
a) Nel decidere quale segno annerire nella schedina del totocalcio, ci si potrebbe aiutare con una trottola la cui circonferenza sia divisa in tre parti con i vari segni "X", "1", "2". L'arco attribuito ad ogni segno può essere commisurato alla maggiore o minore convincimento della sua presenza: il segno "2" (sconfitta della squadra che gioca in casa) dovrebbe avere una misura inferiore. Ad esempio i settori potrebbero rispettare le proporzioni 6:5:2 che è la combinazione di "1, "x", "2" che si è verificata più spesso da quando è nato il totocalcio.



b) Sono molto diffusi i campionamenti di entità territoriali *(area sampling)* in cui le unità derivano da una suddivisione, naturale o artificiale, del territorio in unità areali di ampiezza differente. Spesso, per questo tipo di campione si fa coincidere l'attitudine ad entrare nel campione con la superficie territoriale ovvero in ragione inversa. Invece, il *quadrat sampling* (Rao, 1985) usato per indagini sulla fauna selvatica è realizzato dividendo una regione in maglie quadrate (che quindi ignorano la curvatura terrestre) per poi accertare il numero di capi presenti all'interno di un campione di *quadrat* che hanno tutti uguale attitudine.

c) Nella roulette, il *crupier* piazza la pallina in un punto e le imprime una forza di rotazione non controllabile e non sincronizzabile con l'eventuale moto della base rotante. Se non ci sono trucchi (magnetizzazioni, canalizzazioni, spigoli arrotondati, scorrimenti, vernici speciali) e se i settori rappresentanti i numeri hanno uguale lunghezza d'arco allora la selezione delle unità è equiprobabile.

Esercizio_TP163: *la calibrazione del disco riflette le probabilità di inclusione delle unità misurate da una variabile continua nell'intervallo (-4,5). Si supponga inoltre che al disco sia imposta una forza incontrollabile e imprevedibile che lo faccia ruotare in senso orario.*



Qual'è l'intervallo che tenderà ad apparire più spesso e qual'è quello meno favorito?

Campionamento per aggregati

In diverse applicazioni non è possibile contare o enumerare le unità della popolazione: le particelle di principio attivo in un formulato; le parti di alcool in un liquore, la presenza di alcuni minerali in un terreno, l'ammontare di una esposizione o di una linea di credito articolata per saldi e transazioni. In questi casi il campione casuale viene formato ipotizzando che gli elementi siano scelti uno alla volta ed il loro ammontare progressivamente cumulato.

Esempi:

a) Nella tecnica di campionamento a valanga (*snowball sampling* o *network sampling*) adoperato nelle popolazioni elusive si procede ad individuare alcuni soggetti che fanno parte del campione iniziale. Da ciascuno di questi si tenta di ottenere notizie di altri soggetti aventi le caratteristiche di interesse e si includono nel campione questi nuovi soggetti ai quali si richiede di indicare altri soggetti e così via fino a raggiungere la dimensione campionaria prefissata. Se il soggetto è isolato farà parte del campione solo se cade nella scelta iniziale; se ha molti legami le sue possibilità di far parte del campione aumentano. Poiché la struttura dei legami è solitamente sconosciuta il campionamento non dà la stessa probabilità di comparire a tute le unità e la casualità dipende dalla scelta del primo nucleo e dalla eventuale esclusione di qualche legame quando qualche unità ne presenta troppi.

b) Una società dubita di aver pagato tasse non dovute in quanto parte della sua attività si è svolta all'interno di patti territoriali esonerati per legge da alcune contribuzioni. La popolazione consiste di N=950'000 operazioni realizzate in un periodo di 826 giorni. L'ufficio di consulenza fiscale ha scelto un campione casuale di transazioni cumulandone gli importi fino a raggiungere un valore nominale pari al 5% dell'importo totale delle operazioni (cfr. Sully, 1973).

Esercizio-_TP164: *l'ufficio di recupero crediti deve sollecitare i "k" debitori dell'azienda a versare il dovuto. Per contenere le spese opera su di un campione casuale scelto come segue. Si Indica con x_i i giorni di ritardo alla scadenza del debito i-esimo; si forma una lista in cui ogni debitore è presente tante volte quanti sono i suoi giorni /debito. I debitori sono disposti in ordine: i numeri della lista in [1, x_1] sono attribuiti al debitore 1° nell'ordinamento prescelto, quelli in [x_1+1, x_1+x_2] al 2°, quelli in [x_1+x_2+1,x_1+x_2+x_3] al 3° e così via fino all'intervallo [x_1+x_2+…x_{k-1}+1,T=$\sum x_i$]. A questo punto si ottiene un numero casuale nell'intervallo [1, T] includendo nel campione il debitore che occupa tale posizione nella lista; la procedura è riportata tante volte quante sono le unità da campionare.*
a) Il campionamento avviene con o senza rimessa? b)Come si stabilisce l'ampiezza del campione?

Selezione cumulativa

Non è semplice realizzare il meccanismo dell'urna scossa o quello del disco rotante fermato a caso se la popolazione è numerosa. Si potrebbe infatti verificare una situazione in cui il settore di spettanza ad una unità sia più sottile della linea di demarcazione dei settori. Per evitare il problema esiste un metodo, detto cumulativo, molto semplice e di vasta applicabilità basato sull'uso di frazioni casuali. Sia A la variabile che governa la probabilità delle unità ad entrare nel campione e sia a_i il valore pertinente l'unità i-esima; supponiamo inoltre che le unità della popolazione siano numerate da 1 ad N. Il metodo cumulativo prevede la ripartizione dell'intervallo (0, A_N), dove A_N è il totale noto delle attitudini, in sottointervalli di lunghezza proporzionale all'attitudine cumulata delle unità a partire da quella più piccola. Sucessivamente, si determina una frazione casuale "q" sufficientemente precisa (diciamo con quattro cifre decimali) e si sceglierà l'elemento "i" tale che:

$$A_{i-1} \le qA_N \le A_i$$

unità	*attitudine*	A_i	*Scelta*
1	a_1	$A_1 = a_1$	$]0, A_1]$
2	a_2	$A_2 = a_1 + a_2$	$]A_1, A_2]$
3	a_3	$A_3 = a_1 + a_2 + a_3$	$]A_2, A_3]$
:	*:*	*:*	*:*
N	a_n	$A_N = \sum_{i=1}^{N} a_i$	$]A_{N-1}, A_N]$

Esempio:

Dagli elenchi S.C.A.U. del comprensorio di Imola sono stati tratti i dati sulle famiglie presenti nel comune. Per scegliere un comune selezioniamo una frazione casuale ad esempio lanciando 4 volte un dado, formando un numero accostando le 4 uscite posposte a "0." ed ottenendo una frazione tra 0 ed 1 con (numero-0.1111)/0.5555. Ad esempio le uscite 6, 3,1,5 portano alla frazione casuale (0.615-0.1111)/0.5555=0.9368 ed alla scelta di 0.9368*1825=1709.66 e cioè Mordano.

Comuni	Famiglie	A_i	scelta
Borgo Tossignano	60	60	[0, 60]
Caslafiumanese	106	166	[61, 166]
Castel del Rio	61	227	[167, 227]
Castel Guelfo	176	403	[228, 403]
Dozza	97	500	[404, 500]
Fontanelice	80	580	[501, 580]
Imola	1079	1659	[581, 1659]
Mordano	166	1825	[1660, 1825]

Esercizio_TP165: *Paesi per prodotto interno lordo. Selezionatene uno a caso in base al PIL.*

Paese	PIL	Paese	PIL
CAN	27909	USA	305690
MEX	19635	BRA	27474
CIN	30250	JAP	119000
IND	18413	URS	156300
GER	87436	ESP	20424
BEL	10388	ITA	38223
FRA	62731	POL	14561
ARG	12743	UNG	7846

Per la frazione casuale scegliete come prime due cifre il 1° estratto della ruota di Napoli (se è tra 1 e 9 premettete uno zero) e come secondo blocco di due cifre il 1° estratto della ruota di Roma.

La scelta del modello fisico delle attitudini è argomento troppo intricato per poter essere affrontato compiutamente in un manuale di base; anche il meccanismo che collega le probabilità di inclusione delle unità in estrazioni diverse è molto complesso (tranne che nel caso di equiprobabilità e di reimmissione). E' forse per questo che, sebbene tale impostazione sia l'approccio più generale alla selezione delle unità è invece emarginata come caso particolare (campionamento proporzionale all'ampiezza: *sampling proportional to size* o, in breve, *pps)* in molti testi di Statistica e nei manuali di tecniche campionarie.

La generazione di numeri pseudo-casuali

Le tavole dei numeri casuali sono un utile strumento didattico, ma poco fruibili professionalmente. Inadatte di certo per il Fisco che deve effettuare selezioni dell'ordine di 500'000 cartelle su 22 milioni di contribuenti. Nello studio del comportamento di alcune statistiche non è raro che si debbano simulare ad esempio mille campioni di ampiezza cinquemila adoperando perciò cinque milioni di numeri casuali il ché è impraticabile con i mezzi finora citati. A questo fine si usa una tecnica che, avviata nei primi anni del secondo dopoguerra, ha ormai raggiunto una ampia diffusione: la simulazione di numeri casuali con il computer. Precisiamo subito che si parla di numeri "pseudo-casuali" perché basati su sequenze che scaturiscono da ben definite relazioni funzionali che, pur conservando un precipuo carattere deterministico, mostrano un comportamento assimilabile a quello di una sequenza casuale. La presenza di legami è considerata ininfluente poiché non affiora nella selezione campionaria (si veda la discussione all'inizio del capitolo).

L'evoluzione dei generatori di numeri pseudo-casuali ha una sua pietra miliare nei generatori congruenziali lineari introdotti da Lehmer nel 1949. La loro formula è

$$X_i \equiv \left(aX_{i-1} + c\right) \; mod \; m = Resto\left(aX_{i-1} + c, m\right); \quad i = 1, 2, \ldots,$$

Si tratta di una formula ricorsiva semplice, di agevole traduzione nel linguaggio macchina di ogni computer e risulta di rapidissima esecuzione. In essa compaiono quattro costanti (tutti numeri interi)

X_0	Valore di partenza	$X_0 \geq 0$ se $c>0$ e $X_0 > 0$ se $c = 0$;
a	Moltiplicatore	$1 < a < m$;
c	Incremento	$0 \leq c < m$;
m	Modulo	$m > 2$;

Se c=0 i generatori sono detti puri in contrasto alla denominazione di misti con c >0. Un limite dei generatori congruenziali lineari è che possono produrre al massimo "m" numeri pseudo casuali diversi; giunto all'm-esimo la formula entra -necessariamente- in ciclo riprendendo dal numero iniziale X_0. La conseguenza è la riproducibilità di ogni successione: per ripetere integralmente una stessa successione o sottosuccessione di numeri pseudo casuali basta conservare il loro valore iniziale o finale. La simulazione è anche reversibile: per ogni numero si possono conoscere altrettanto bene i suoi susseguenti ed i suoi antecedenti.

Esempi:

a) Un computer binario con una parola di (n+1) bit è in grado di effettuare operazioni aritmetiche in numeri interi purché il loro risultato, in valore assoluto, non sia superiore a 2^n-1. Un PC a 32 bit è in grado di gestire operazioni con interi fino a 2^{31}-1=2'147'483'647. Un ottimo schema proposto da Fishman (1996, p.604) è il seguente:

$$X_i = Resto\left(950'706'376X_{i-1}, 2'147'483'647\right); \quad i = 1, 2, \ldots,$$

b) Lo schema congruenziale:

$$X_i = Resto\left(314'159'221X_{i-1} + 211'324'863, 1'000'000'000\right); \quad i = 1, 2, \ldots,$$

è adatto per calcolatrici tascabili basata sull'aritmetica decimale (a differenza dei computer basati sull'aritmetica binaria).

Esercizio_TP166: *determinate la formula che permette di calcolare qualsiasi altro che interverrà in successione nell'algoritmo di Lehmer.*

Periodo dei generatori

Il numero "m" di valori diversi prodottisi prima del riavvio costituisce il "periodo" del generatore. Un periodo elevato è un requisito essenziale dei generatori; in particolare, il periodo dovrebbe essere maggiore del quadrato dei numeri da usare, dovrebbe cioè rispettare la regola di Ripley $m \geq 200n^2$. Se si deve estrarre un campione casuale di 5'000 numeri il periodo del generatore dovrebbe essere superiore a 5 miliardi e qui insorge un inconveniente: $5x10^9$ è più grande di 2^{31} che è il limite massimo per la rappresentazione in molti PC. In base alla regola di Ripley, un PC che effettua in precisione semplice le operazioni intere (32 bit) può generare sequenze attendibili non più lunghe di 3'200 valori. Per uno studio serio del comportamento campionario di calcoli relativi all'evasione fiscale nei quali si selezionano 500'000 unità sarebbe necessario un periodo dell'ordine di 2^{46} che non è ottenibile con la virgola fissa della stragrande maggioranza dei computer oggi in uso. Quando però il periodo dei generatori è molto elevato si corre un altro rischio: quello di imbattersi in sottosequenze con struttura non casuale della cui presenza non è facile rendersi conto.

Esempi:

a) Il foglio elettronico EXCEL della Microsoft include una funzione per la generazione di numeri interi pseudo-casuali: CASUALE.TRA(min;max) che può essere utilmente richiamata per estrarre un campione con reimmissione.

b) Se disponete su foglio elettronico tipo Excel di una *mailing list* di possibili contatti e desiderate estrarne un campione casuale semplice senza reimmissione potete procedere come segue: posizionatevi sulla prima cella della prima colonna libera dopo i campi di riferimento degli indirizzi e scrivete =Casuale(). Ricopiate in basso la funzione fino ad arrivare all'ultimo elemento della lista. Selezionate tutto il data set e scegliete il comando ordina per disporre le righe in ordine secondo la colonna in cui ci sono i comandi "CASUALE()". I primi "n" elementi di questi nuova lista sono il vostro campione casuale semplice. Per un campione casuale semplice con reimmissione si può eseguire la stessa procedura, ma applicata ad un listone ottenuto replicando "n" volte la lista originale.

c) Dodge (1996) suggerisce di immagazzinare in un DVD diverse decine di miliardi di cifre dello sviluppo decimale di π e di utilizzarle come numeri pseudo-casuali. L'idea è valida, soprattutto in vista dei test di casualità che le cifre hanno superato, ma contrasta con la relativa lentezza delle operazioni di I/O che sono ancora lente per un dispositivo laser attualmente avanzato quale il DVD.

L'uso della formula congruenziale lineare di rado coinvolge direttamente gli interi che da essa scaturiscono. Piuttosto, poiché $0 \leq X_i < m$ si opera con le frazioni:

$$q_i = \frac{X_i}{m}; \quad i = 1, 2, \ldots, m$$

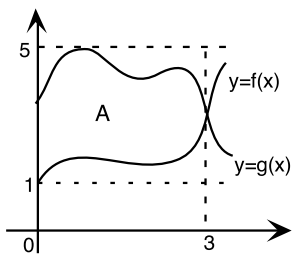
Se poi insorge l'esigenza di numeri reali (in verità, di loro approssimazioni) ricadenti in un intervallo limitato, si procede alla trasformazione:

$$y_i = a + bq_i \quad a \leq y_i \leq b, \; i = 1, 2, \ldots, m$$

Esempio:

I metodi Monte Carlo forniscono soluzioni approssimate, ma valide a molti problemi a mezzo del campionamento computerizzato. Per darne una breve illustrazione supponiamo di dover calcolare l'area della regione A delimitata dalle curve f(x) e g(x) nonché dalla retta X=0. Per ottenere il valore di A si generano "n" coppie di numeri casuali $0 \leq X_i \leq 3$ e $1 \leq X_i \leq 4$ e si contano quelli che ricadono in A sul totale. L'area sarà perciò:

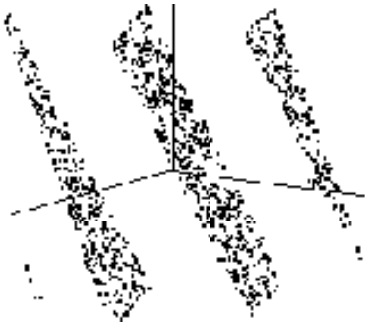
$$Area(A) = \left[\frac{numero \;\; di \;\; coppie \;\; in \;\; A}{numero \;\; di \;\; coppie}\right] * 12$$



dove 12 è l'area del rettangolo che racchiude A ed all'interno del quale ricadono tutti i punti generati con i numeri casuali.

Marsaglia (1968) ha dimostrato che n-tuple di valori consecutivi di U_i ricadono in un numero limitato di iperpiani paralleli. Il numero e la densità degli iperpiani determina la qualità del generatore.

Esempio:
Il generatore RANDU in uso sui computer IBM serie 360/370 operativi fino alla prima metà degli anni ottanta aveva formula X_i=resto($65539X_{i-1}+41, 2^{31}$) e $U_i=X_i/2^{31}$.



Tenuto conto che $65539=2^{16}+3$ si arriva (cfr. Fishman, 1996, pp. 619-620) alla relazione $U_{i+2}-6U_{i+1}+9U_i \equiv 0 \bmod 1$ per cui le terne sono vincolate a giacere su piani paralleli del tipo illustrato in figura. I larghi vuoti che si vedono evidenziano la poca affidabilità di questo schema.

L'essenzialità della tecnica ed i buoni risultati ottenuti hanno indotto una ampia ricerca sulla definizione di criteri che portino alla scelta dei parametri (X_0, a, c, m) tali da assicurare il periodo pieno del generatore per valori elevati del modulo e la "casualità" delle serie da essi generate. Una variante che si è ormai consolidata nelle applicazioni prevede l'uso congiunto di più generatori, preferibilmente del tipo puro (la cui esecuzione è più rapida comportando meno operazioni). Fra i molti programmi pubblicati su varie riviste scientifiche merita attenzione quello suggerito da Wichmann e Hill (1982). L'algoritmo, noto con la sigla AS183, si basa sulla combinazione di tre generatori congruenziali

$$X_i \equiv (171X_{i-1}) \ mod \ 30269; \quad Y_i \equiv (170Y_{i-1}) \ mod \ 30307; \quad Z_i \equiv (172Z_{i-1}) \ mod \ 30323;$$

La sua esecuzione richiede aritmetica intera fino a 2^{24} e può quindi "girare" anche sui personal computer meno dotati. Fissati i tre valori di partenza: X_0, Y_0 e Z_0 , l'algoritmo procede generando ogni volta tre numeri pseudo-casuali da ognuna delle formule per poi sommarli: $W_i=(X_i/30269)+(Y_i/30307)+(Z_i/30323)$. Le tre frazioni componenti di W_i sono frazioni casuali sull'intervallo unitario (estremi esclusi). Tale sarà anche la frazione $q_i \equiv W_i \bmod 1$ che è poi il numero pseudo casuale adoperato. Il periodo del generatore Wichmann-Hill è 2^{44} (circa 28 mila miliardi) che lo rende idoneo in molte applicazioni.

Esempio:
Ecco la codifica in *Future Basic* sottoposta a molti test ed applicazioni rivelando un comportamento soddisfacente.

```
'Valori iniziali
IX=23311:IY=13367:IZ=26317
LOCAL FN AS183 (X, N)
FOR    I=1 TO N
       IX=171*IX MOD 30269: IY=172*IY MOD 30307: IZ=170*IZ MOD 30323
       X(I)=FRAC(IX/30269+IY/30307+IZ/30323)
NEXT I
END FN
```

La routine calcola "n" numeri pseudo casuali unitari che è più rapido di "n" generazioni di un numero casuale.

Esercizio_TP165: *Zeisel (1986) ha dimostrato che l'algoritmo di Wichmann ed Hill equivale allo schema congruenziale puro:*

$$X_i \equiv 16'555'425'264'690 * X_{i-1} \ Mod \ \ 27'817'185'604'309$$

a) Provate ad applicarlo e verificate perché è conveniente ancora la formula originale.
b) Quanti numeri si possono generare tenendo conto del limite di Ripley?

6.5.5 La selezione sistematica

Supponiamo che la lista assegni un numero d'ordine progressivo alle unità della popolazione, diciamo da 1 ad N. Si deve estrarre un campione di ampiezza "n" con un intervallo di campionamento h=N/n. Se non si vuole perdere troppo tempo ovvero la composizione della lista è difficile oppure se la tavola dei numeri casuali è insufficiente per coprirla tutta, si può procedere in modo sistematico e cioè selezionando a caso solo la prima unità e poi, a partire da questa, inserirne un'altra ogni "h" unità non inserite.

Vediamo prima il caso di "h" intero cioè quando la numerosità della popolazione N è un multiplo esatto della numerosità del campione "n". Si sceglie, adoperando la tabella dei numeri casuali o altro ogni mezzo ritenuto idoneo, la prima cifra o il primo gruppo di cifre "c": se questo è inferiore ad "h" si pone r=c+1 (si ricordi che nelle cifre casuali c'è anche lo "0") e l'unità che nella lista occupa tale posizione sarà la prima ad entrare nel campione. Se "c" è maggiore o uguale ad "h" sarà trascurato e si sceglierà un altro numero. La posizione valida "r" rappresenterà l'inizio del nostro campione e l'unità che vi compare sarà la prima ad essere inserita; la seconda scelta sarà l'unità occupante la posizione r+h, la terza quella in posizione r+2*h e, in generale, l'i-esima unità che finisce nel campione è:

$$u_j \text{ entra nel campione in posizione i-esima se } j=r+(i-1)*h$$

fino a completare il campione di "n" unità.

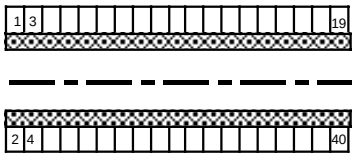
Esempio:
Sia N=20 ed n=5 con h=20/5=4. Supponiamo che il blocco della tabella dei numeri casuali prescelto sia in riga 7 e colonna 3 e che si cominci dalla cifra del blocco in alto a sinistra riservandosi di proseguire in basso lungo la colonna in caso di cifra superiore ad (h-1): subito troviamo un "6" che non è utilizzabile, dopo c'è "4" che è pure da escludere; la prima cifra utile è il "2" per cui abbiamo: c=2, r=3 e le unità che formano il campione sono: 3, 7, 11, 15, 19.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**y**	0	0	0	**y**	0	0	0	**y**	0	0	0	**y**	0	0	0	**y**	0	0	0
2	0	**w**	0	0	0	**w**	0	0	0	**w**	0	0	0	**w**	0	0	0	**w**	0	0
3	0	0	**x**	0	0	0	**x**	0	0	0	**x**	0	0	0	**x**	0	0	0	**x**	0
4	0	0	0	**z**	0	0	0	**z**	0	0	0	**z**	0	0	0	**z**	0	0	0	**z**

Lo schema di questa selezione è meglio illustrato con la tabella in cui sono riportate tutte le posizioni campionabili in base alla scelta della prima posizione: poiché l'intervallo di campionamento è 4, i campioni possibili sono appunto soltanto 4. E' facile verificare che i diversi campioni possibili con la selezione sistematica non hanno alcuna unità in comune per cui, ad esempio la probabilità di entrare nel campione è 0.25 per le unità (3,7), ma è nulla la probabilità che vi compaiano insieme l'unità in 3ª ed in 4ª posizione. I dubbi sulla rappresentatività ci sono, ma i vantaggi di rapidità e semplicità non sono da trascurare.

I possibili impieghi di questa tecnica sono tantissimi: nella scelta dei fotogrammi da campionare per il controllo di un filmato o per la sua memorizzazione, l'interruzione di una catena mobile per prelevare una confezione, i tempi di accertamento dello stato di usura di una macchina, etc.

Esercizio_TP167: *Marietta Monarca lavora a part-time come rilevatrice ed ha avuto l'incarico di visitare 10 famiglie residenti in una certa strada. Sulla strada si affacciano 40 isolati (20 da un lato e 20 dall'altro numerati nella sequenza indicata in figura).*



Marietta decide di intervistare una famiglia per ogni isolato e che si fronteggiano ai due lati della strada. Se effettua una selezione sistematica con "3" come numero casuale iniziale, quali di essi costituiranno il campione?

La selezione sistematica non deve limitarsi a trovare una unità per intervallo, ma può considerare un blocco di unità collocate in "k" posizioni contigue senza che siano compromessi i requisiti di semplicità e di casualità della scelta (peraltro il blocco potrebbe non essere formato dallo stesso numero di unità ovvero il numero di unità del blocco selezionato casualmente). Supponiamo si dividere l'ampiezza del campione in "m" blocchi con "k" unità ciascuno. Le unità del campione sono quelle collocate nelle posizioni:

$$c+(i-1)*k*h+j \text{ per } j=1,2,\ldots,k;\ i=1,2,\ldots,m \text{ con } c\leq n-k$$

Esempi:

a) Si vuole valutare la variazione media negli N=36 fondi lussemburghesi in una data giornata borsistica. Ipotizziamo che il campionamento richieda n=12 (quindi con h=N/n=36/12=3) fondi scelti per blocchi sistematici di m=4 unità. Supponiamo che il= blocco della tavola dei numeri casuali sia il (6,3) che ci propone un c=8 ed un r=c+1=9 come punto di partenza casuale. Le unità che entrano nel campione sono quelle in posizione 9ª, 10ª, 11ª, 12ª. Il secondo blocco sarà quello dopo la posizione 8+1*4*3 cioè 21ª, 22ª, 23ª ,24ª ed infine il 3° blocco formato con le unità posizionate a partire dalla 8+2*4*3 cioè la 33ª, 34ª, 35ª e 36ª che fornisce un valore medio di 0.00 piuttosto distante dalla media di tutta la popolazione: 0.077.

	Fondo	Var%		Fondo	Var%
1	Rominv. Dm B. Portfolio	-0.23	19	Fonditalia Eq. Italy	2.27
2	Rominv. Dm Short term	-0.31	20	Fonditalia Eq. Japan	0.12
3	Euroras F Bond	0	21	Fonditalia Eq. USA	-0.47
4	Rominv. Frech Bonds	-0.24	22	Fonditalia Lira	0.04
5	Rominv. Frech Index	0.56	23	Fonditalia Float rate lit	-1.04
6	Rominv. Frech Short t.	-0.23	24	Fonditalia Yen	0.06
7	Rominv. German index	0.02	25	Interfund	0.05
8	Euroras F Equity	0	26	International sec.	0.03
9	Capital Italia	0	27	Italfortune cat. A	-0.99
10	Euroras F Dollar	0	28	Italfortune cat. B	0.08
11	Euroras F Mark	0	29	Italfortune cat. C	0.81
12	Fonditalia	0.14	30	Italfortune cat. D	0
13	Fonditalia B. Lira	0.44	31	Italfortune cat. E	0.1
14	Fonditalia Dir	-0.61	32	Italfortune cat. F	-0.73
15	Fonditalia Dmk	-0.36	33	Rominv.Ecu. Short T.	0
16	Fonditalia Em.mk.Asia	-0.12	34	Rominv. It. Bond	0.18
17	Fonditalia Eq. Brit.	0.32	35	Rominv. Univ.	-0.25
18	Fonditalia Eq. Europa	0.58	36	Rominv. Univ. Med. T.	0.45

b) Per adoperare il campionamento sistematico non è necessario che le unità della frame siano numerate. Se ad esempio sono riportate regolarmente su di un supporto: la scelta di un campione di record in una popolazione costituita da righe disposte in 100 pagine di 60 righe può avvenire selezionando una riga per pagina o una ogni due pagine. In genere, il campionamento sistematico è più rapido del campionamento casuale semplice se la lista non è automatizzata.

Esercizio_TP168: *le amministrazioni pubbliche, prima di procedere all'apertura delle buste delle offerte ammesse richiedono ad un numero di offerenti non inferiori al 10% arrotondato all'unità superiore, scelti con sorteggio pubblico, di comprovare il possesso di alcuni requisiti di idoneità già autocertificati per la gara. Supponete che le offerte regolari siano 80 numerate progressivamente a partire da uno.*
a) Costruite un campione sistematico di ampiezza n=16 procedendo per unità singole ed ipotizzando che il primo numero casuale sia c=3;
b) Costruite il campione sistematico procedendo per blocchi contigui di 2 unità.

Il campionamento sistematico è senza reimmissione in quanto nessuna unità può comparirvi più di una volta. Il numero di campioni possibili, come si è visto negli esempi, è ridotto: invece delle usuali combinazioni si dispone solo di "h" scelte. Non si tratta però di una limitazione seria purché tale piano non si adoperi con un intervallo di campionamento legato alla formazione della lista. Se le unità da campionare sono inserite nella *frame* in ordine alfabetico e la variabile da esaminare non ha alcuna relazione con la denominazione delle unità il campionamento sistematico non ha controindicazioni. Nel caso opposto possono insorgere dei problemi: se in una fila di alberi è stato piantato un platano ogni dieci pioppi, una selezione sistematica che partisse da un platano e procedesse di passo dieci darebbe un campione formato di soli platani denotando una sorprendente assenza di pioppi.

Esempi:

a) Durante una verifica fiscale si pone l'esigenza di esaminare il rullino del registratore di cassa di due giornate lavorative che contiene 400 battute. Nel primo giorno alla cassa era addetto un cassiere integerrimo, nel secondo giorno gli incassi erano gestiti dalla disinvolta proprietaria. Nell'impossibilità di tagliare e mescolare le singole registrazioni si opta per un campione. Un campione casuale semplice di 40 battute potrebbe facilmente dar luogo ad una selezione composta interamente da battute di uno dei due giorni lasciando intravedere situazioni non realistiche in entrambi i casi. Un campione sistematico di una battuta ogni dieci sarebbe più fedele alle operazioni delle due giornate.

b) Stephan (1969) propone di aggirare il problema della periodicità con un metodo ibrido. La struttura è data dalla selezione sistematica, ma dopo che l'unità è stata prescelta, la nuova posizione non è quella determinata aggiungendo l'intervallo di campionamento "h" , ma si sposta di tante posizioni in avanti quante ne indica un numero casuale "k", con $1\leq k\leq m$ dove "m" è scelto è fissato in relazione al tipo di indagine ed al correttivo di sequenza che si vuole apportare. In pratica, se la 1ª posizione è $L_1=r$, la 2ª sarà $L_2=L_1+h+k_1$, la 3ª è $L_3=L_2+h+k_2$ e così via. Per semplicità si può porre $k_1=k_2=\ldots$

Esercizio_TP169*: le rilevazioni giornaliere della produzione di alcuni reparti sono state elencate in una lista univoca ed esaustiva per disporre alcuni controlli campionari. Per ovviare alla evidente periodicità (ed altre presenti, ma non visibili) si decide di operare con una selezione sistematica non di singole unità, ma di blocchi di unità. Ad esempio per un campione di ampiezza 16 con un intervallo di campionamento di 22 a partire da r=3 si tratterebbe con le unità dalla 3ª alla 18ª, dalla 41ª alla 56ª, dalla 89ª alla 104ª, etc. La singola unità da campionare nel blocco verrebbe poi scelta con un numero casuale tra 1 e 16. E' una procedura convincente?*

La procedura di selezione sistematica si complica se l'intervallo di campionamento è frazionario. Supponiamo di avere N=120 e n=13 per cui h=120/13=9.23. Se "h" è approssimato all'intero inferiore, i campioni possibili in caso di cifra casuale "0" o "8" e conseguente prima posizione tra "1" e "9" sono:{1, 10, 19, 28, 37, 46, 55, 64, 73, 82, 91, 10, 109}; {9, 18, 27, 36, 45, 54, 63, 72, 81, 90, 99, 108, 117} e le unità che occupano le ultime tre posizioni: 118ª,119ª e 120ª vedono ridotta a zero le loro possibilità di entrare nel campione.

	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0
1	x								y	x								y	x								y	x		
2						y	x								y	x														
3			y	x								y	x									y	x							y
4	x								y	x								y	x								y			

Se l'approssimazione è all'intero superiore, l'ultima posizione campionata potrebbe trovarsi fuori lista; infatti, se h=10 sono possibili: {1, 11, 21, 31, 41, 51, 61, 71, 81, 91, 101, 111, ?}; {9, 19, 29, 39, 49, 59, 69, 79, 80, 99, 109, 119, ?}. Il problema si può aggirare scegliendo a caso un nuovo numero casuale tra 1 e 120 e se la posizione così individuata non è già stata usata sarà chiamata a colmare la lacuna.

Esempio:

In *"Understanding robust and exploratory data analysis"* di D.C.Hoaglin, F. Mosteller e J.W. Tukey (1983) l'ultimo autore è presente nell'indice delle citazioni 57 volte alle pagine indicate in tabella.

6	8	30	40	44	55	76	93
	130	156	162	163	164	176	190
	192	200	205	206	211	212	223
	242	243	268	273	274	278	279
	280	281	296	313	321	333	334
	335	336	349	350	363	376	387
	388	390	391	392	395	398	399
	406	412	413	423	426	427	428

Per valutare l'argomento delle citazioni se ne sceglie un campione di n=5. L'intervallo di campionamento 11.4 è approssimato a h=12. Stabiliamo di scegliere casualmente il blocco su cui cercare il primo numero casuale (di due cifre in questo caso) e di procedere, all'interno del blocco, per righe successive. Supponiamo che la selezione casuale (con le biglie) di righe e di colonna indichi il blocco (1,2) che porterà alla scelta della prima unità in lista in quanto corrispondente allo "00" della *frame*. Pertanto, il campione sarà formato dalle unità:{1, 13, 25, 37,49} con le conseguenti citazioni: {pag.6, 164, 268, 280, 398}. In questo caso non si è resa necessaria la selezione dell'unità aggiuntiva per il fuori lista.

***Esercizio_TP170**: i codici di addebito degli acquisti sono un multiplo di 4. Si analizzi la lista in tabella con un campione sistematico di n=17 e h=7. Se un codice è sbagliato l'esame si ferma per procedere ad altri controlli.*

312	256	352	108	136	320	196	216	26	32	120	304	112	208	148	8	324	256	348	216	304	284	80	228	164	276	312
340	168	172	188	124	96	20	80	216	4	384	152	348	92	136	200	20	72	32	264	196	100	132	80	224	380	48
300	388	120	160	72	236	28	28	312	32	232	188	180	220	184	52	140	368	336	320	4	80	220	168	108	216	102
224	124	96	20	128	128	200	120	352	176	40	152	312	260	120	228	320	264	180	192	256	340	312	292	88	68	92

Usiamo il blocco (6,1) per i numeri casuali (per quello iniziale e per colmare le lacune).

a) Quali valori saranno riscontrati?

b) Qual'è la probabilità che una sia errata?

c) Qual'è la probabilità che un campione sistematico trovi l'errore?

d) E' più alta o più bassa che con il campione casuale semplice?