Introduzione alla probabilità

Questo è un ciclo di lezioni sulla Sorte, Ma ...

- **Non sarà fornito alcun trucco per vincere al superenalotto, lotto, gratta-e-vinci, bingo e pesche varie**
- **Non verranno dati talismani o filtri per scongiurare eventi funesti**
- **Non saranno indicate formule capaci di attrarre su di noi e sui nostri cari la Buona Ventura**

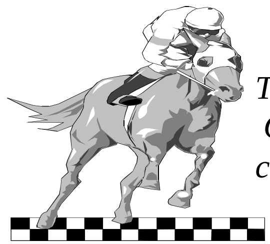
Riflessione

Lo scopo è di illustrare il ruolo della Sorte nella conoscenza scientifica e nel sapere professionale

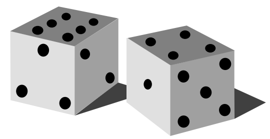
- **L'umanità è progredita grazie alla conoscenza di leggi naturali sempre più numerose ed alla fiducia in esse riposta**

Sappiamo che se si sospende un masso sul piede di qualcuno e poi lo si lascia cadere qualcuno non sarà contento

- **Malgrado il progresso vi sono moltissimi eventi che non siamo in grado di prevedere.**



Tutto è incerto. Questa è l'unica certezza

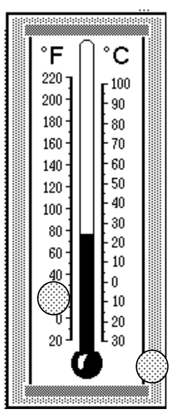


Esperimento deterministico

Ovunque si attivi un processo di osservazione e/o misurazione di un fenomeno soggetto a variazione c'è un esperimento

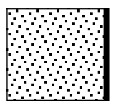
L'esito di un esperimento deterministico è prevedibile con certezza

Scelta dell'asso di bastoni in un mazzo formato da 40 assi di bastone



ESEMPIO: area del quadrato:

$$Y = X^2$$



X=Lato

noto il lato, l'area del quadrato è univocamente determinata: x=5-->y=25

Negli esperimenti di laboratorio relativi a molte leggi fisiche le relazioni sono <u>"quasi-deterministiche"</u> nel senso che gli errori sono imputabili a carenza degli strumenti di misurazione:

Esperimento casuale

Un esperimento casuale è una prova che può essere riproposta -fisicamente o virtualmente- una, due, infinite volte nelle medesime condizioni senza che si possa stabilire quale sarà l'esito della prossima manifestazione.

Casualità ed incertezza

Ciò che è casuale è anche incerto ed in ogni incertezza c'è un elemento di casualità

Significato di casuale

Lessicale: **Accadimento involontario, imprevedibile, accidentale**

Statistico: **Riflessioni e valutazioni su di un fenomeno soggetto all'azione della pura Sorte.**

E' impossibile stabilire a priori quale sarà la sua esatta manifestazione;

Stocastico, aleatorio, erratico,

La Sorte

E' una forza che agisce in tutto l'Universo

ESTRANEA
CAPRICCIOSA
CIECA
CINICA
INACCESSIBILE
INAPPELLABILE
INDIFFERENTE
NEUTRA
IMPREVEDIBILE
DESTRUTTURATA

La Sorte ignora tutto e trascura tutto e garantisce la *par condicio*

E' invocata per la sua trasparenza e per la sua equità.

E' temuta per l'abbandono di ogni discrezione

Superstizioni e furbizie

Si è lanciata una moneta -fisicamente perfetta- per 50 volte e per 50 volte è uscito "testa". Prima della 51ª prova vi viene proposto di scommettere 10'000 lire:

Esce testa e vinci 10'000 altrimenti perdi la puntata

Esce croce e paghi 10'000 altrimenti incassi la puntata

Una moneta perfetta non ha memoria dei lanci precedenti e quindi si deve essere "razionalmente" indifferenti.

Se siete convinti che dopo 50 "teste" sia più favorita la "croce" come vi spiegate che si comprano più biglietti nella ricevitoria in cui si è già vinto?

Percezione della casualità

SCRIVETE, SENZA FARVI VEDERE E SENZA CONSULTARVI, UN NUMERO TRA 1 E 4

Riscontro nel corso

La legge empirica del caso

Sia "S" il dominio di una variabile con un NUMERO FINITO di modalità

$$S = \{E_1, E_2, \ldots, E_k\} \quad \text{con} \quad k < \infty$$

Si supponga che l'esperimento sia ripetuto "n" volte e che le modalità abbiano avuto le frequenze assolute riportate a destra

$$\longrightarrow \begin{array}{cc} E_1 & n_1 \\ E_2 & n_2 \\ \ldots & \ldots \\ E_k & n_k \\ \hline & n \end{array} \quad \text{dove} \quad n = \sum_{i=1}^{k} n_i$$

E' un fatto che se si sceglie a caso una modalità e si ripete la procedura la frequenza relativa si stabilizza su un numero P(E$_i$)

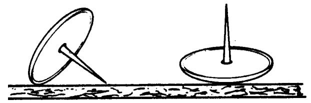
$$P(E_i) = \lim_{n \to \infty} \left(\frac{n_i}{n} \right) \quad \text{per} \quad i = 1, 2, \ldots, k$$

Anche se non si tratta di un vero e proprio limite

La legge empirica del caso/2

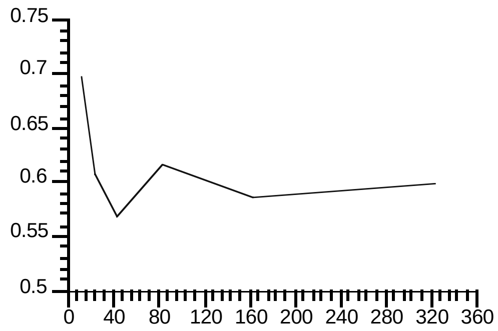
Esperimento della puntina da disegno (Variabile binaria)

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



Numerose verifiche sperimentali corroborano l'idea di convergenza delle frequenze relative.

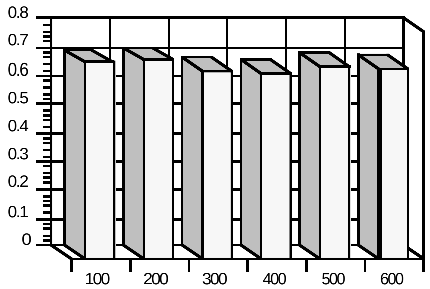
Un esperimento replicato molte volte in condizioni simili, fa emergere una struttura stabile nelle frequenze relative.



La legge empirica del caso/3

Estrazione con reimmissione di una biglia da un'urna contenente 6 biglie di colore diverso in proporzione 2:4

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



L'esito di una singola sperimentazione non può essere previsto con certezza.

Siamo però abbastanza sicuri di ciò che succede nel complesso purché si possa disporre di una serie considerevole di repliche in condizioni stabili

Questo fatto è la legge empirica del caso.

La prova o esperimento

E' un esperimento -indotto o spontaneo- con esito casuale.

- **Il valore di chiusura di un titolo di borsa**
- **Esito del lancio di un dado**
- **Livello di inquinamento da elettrosmog**
- **Esame di laurea**

La PROVA o ESPERIMENTO è una situazione che può essere replicata (almeno in teoria) ed il cui esito è reso imprevedibile dall'azione della Sorte.

Le ripetizioni debbono essere tali da potersi dire repliche di un o stesso esperimento e non relative ad esperimenti diversi.

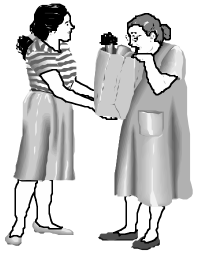
L'evento elementare

E' la descrizione di uno dei possibili risultati della prova. E' "elementare" perché non ulteriormente frazionabile

ESEMPIO: le possibili ragioni di ammanco di inventario sugli scaffali di un supermecato sono:

E_1 = Shoplifting; E_2 = Appropriazione indebita

E_3 = Errore di conteggio; E_4 = Fantasmi



Gli eventi della prova debbono essere ESAUSTIVI ED UNIVOCI:

1) Ogni risultato deve corrispondere ad un "evento"

2) Ogni risultato deve corrispondere ad un solo "evento"

L'universo degli eventi

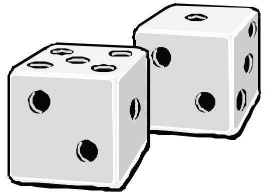
Ad ogni esperimento casuale è associato l'universo degli eventi S: una lista che elenca tutti e solo gli eventi elementari in una data prova.

ESEMPIO: lancio di una moneta

Universo degli eventi S={T,C}

ESEMPIO: lancio di due dadi

Universo degli eventi S={2, 3, 4, 5, 6, 7, 8, 9, 10,11,12}



I giochi d'azzardo aiutano, come in un laboratorio, a sviluppare nozioni e tecniche utili in situazioni più complesse.

Eventi ed Insiemi

Nel 1933 Kolmogorov sviluppò una TEORIA DELLA PROBABILITA' sfruttando le analogie tra insiemi ed eventi.

Per proseguire abbiamo bisogno dei concetti della TEORIA DEGLI INSIEMI applicata agli eventi

L'insieme più importante è l'UNIVERSO DEGLI EVENTI "S" i cui elementi sono gli eventi elementari della PROVA.

$$S = \{E_1,\ E_2,\ \ldots, E_n\}$$

Gran parte dei concetti della insiemistica si estendono agli eventi.

Le illustrazioni usano i DIAGRAMMI DI VENN cioè figure geometriche disegnate sul piano.

Eventi ed Insiemi/2

L'insiemistica impone all'universo degli eventi una precisa struttura:

"S" è un modello della prova di cui riporta <u>tutto e solo</u> ciò che è rilevante.

La struttura però è vincolata. Devono essere eventi univoci ed esaustivi

Non sempre questo è possibile:

- Eventi non separabili
- Eventi sfocati
- Eventi frazionari

Gli aspetti negativi di eventuali forzature sono compensati dal vantaggio di una trattazione snella e rigorosa.

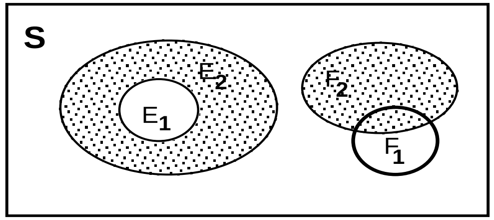
Operazioni sugli eventi

Per gestire le operazioni con gli eventi adoperiamo le regole degli insiemi.

- INSIEMI E SOTTOINSIEMI
- OPERAZIONI BINARIE: coinvolgono due eventi
  - UNIONE
  - INTERSEZIONE
- OPERAZIONI UNARIE: coinvolgono un solo evento
  - NEGAZIONE O COMPLEMENTO

Insiemi e sottoinsiemi

Un insieme "E" è un sottoinsieme di "F", scritto E ⊃ F, se ogni evento in "E" appartiene anche ad "F" e almeno un evento di "F" non è in "E";



Quindi "F" implica "E" perché questo si verifica ogni volta che si verifica "F", ma l'opposto non è necessariamente vero.

Nel diagramma E1 è un sottoinsieme di E2, ma F1 non lo è di F2 dato che non vi è tutto incluso.

Due eventi si dicono UGUALI se, contemporaneamente "E" è un sottoinsieme di "F" e questo è un sottoinsieme di "E".

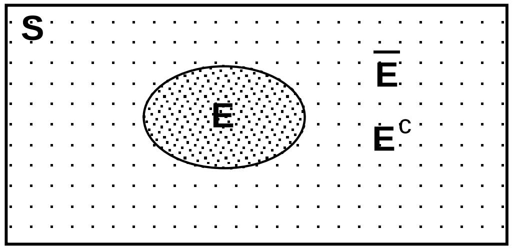
L'evento negazione

Unione ed intersezione si riferiscono ai singoli eventi, la negazione si riferisce all'intera prova.

L'evento negazione o evento complementare di un evento "A", indicato con $\overline{A}$ cioé "non A" si verifica quando non si verifica "A".

ESEMPIO: previsioni economiche

A= inflazione al di sopra del 5%;



Perché si verifichi $\overline{A}$ è sufficiente che l'inflazione sia minore o uguale al 5%,

L'Evento negazione si indica anche con A' o con A^c

Evento certo ed evento impossibile

L'universo degli eventi cambia da prova a prova. Ma in ogni prova sono presenti due eventi

- **L'EVENTO CERTO: si verifica sempre** Tale evento si indica con "S"

  Saremmo disposti a scommettere qualsiasi cifra avendolo a favore e nessuna cifra avendolo contro.

  Nel lancio di un dado esce un numero da uno a sei

- **L'EVENTO IMPOSSIBILE: non si verifica mai** Tale evento si indica con "$\varnothing$"

  Nessuna promessa di vincita, comunque grande, potrebbe indurci a scommettere in suo favore.

  In una gara con dieci concorrenti arrivare undicesimo

$$S^C = \varnothing; \qquad \varnothing^C = S$$

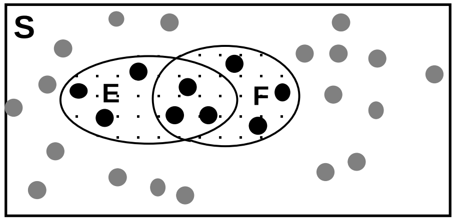
L'evento unione

Dati due eventi "A" e "B" . L'evento unione, "C", si verifica se si verifica o l'uno o l'altro o entrambi gli eventi considerati

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

ESEMPIO: la mano di poker in 4
I giocatori sono: NORD, SUD, OVEST, EST

Poniamo A: Nord ha un tris;
B: Est ha un full (tris+coppia)



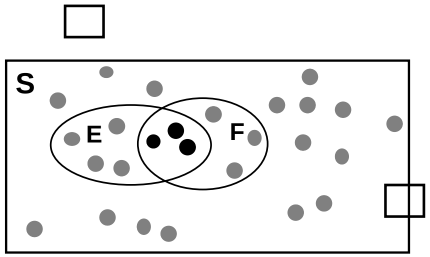
L'evento unione "C" si verifica se Nord ha il tris oppure se Est ha un full, ma si verifica pure se, contemporaneamente, Nord ha il tris ed Est il full (con buona pace di Nord)

Quiz: quando si verifica l'evento: "A oppure B, ma non entrambi" ?

L'evento intersezione

Dati due eventi "A" e "B" . L'evento intersezione, "D", si verifica se accadono entrambi gli eventi considerati

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



ESEMPIO: Età della donna in anni compiuti all'atto del matrimonio

Poniamo A: 28≤ X ≤45; B: 20≤ Y ≤40

L'evento intersezione si verificherà se, contemporaneamente, la donna ha più di 28 anni, ma meno di 40

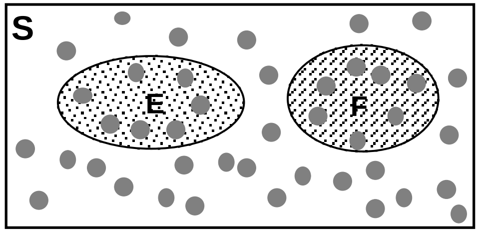
Eventi mutualmente incompatibili

Due eventi che non possono presentarsi insieme sono MUTUALMENTE INCOMPATIBILI (a due a due).

In ogni prova se ne può verificare uno solo e quindi essi non hanno elementi in comune

ESEMPIO: nomina del Presidente in un comitato di cinque membri: M1, M2, M3, M4, M5.

A: Presidente è M1
B: Presidente è M2



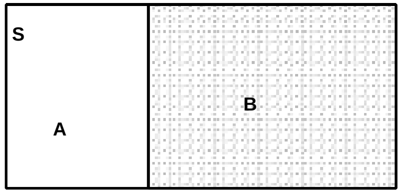
L'intersezione di due eventi incompatibili è un evento impossibile

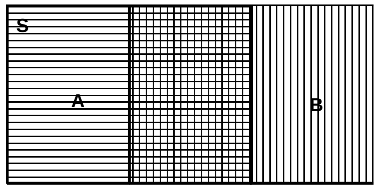
$$E \cap F = \phi$$

Eventi necessari

Se, in una prova, due eventi sono tali che almeno uno dei due si verifica, si dicono NECESSARI.

L'essere necessari vuol dire che, insieme, i due eventi costituiscono l'evento certo





Esempio: *eventi necessari e incompatibili.*

Una classe di 24 studenti consiste di 14 matricole e 10 del 2° anno.

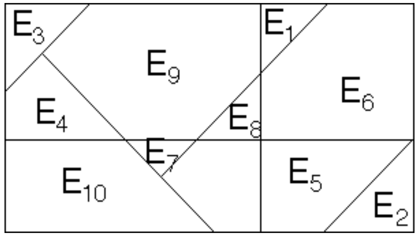
A: Si sceglie una matricola
B: Si sceglie un 2° anno

Esempio: *eventi necessari, ma compatibili*

Il censimento delle persone distingue tra presenti e residenti (uno status non esclude del tutto l'altro)

A: La persona è presente
B: La persona è residente

La partizione



Un insieme di eventi

$$E_1,\ E_2, \ldots, E_k$$

costituisce una PARTIZIONE dell'universo degli eventi se

1) $E_i \cap E_j = \emptyset$ **Per ogni "i≠j"(sono incompatibili a coppie)**

2) $\bigcup_{i=1}^{k} E_i = I$ **Insieme formano l'evento certo**

"unione per i che va da uno a k di e con i"

I pezzi del puzzle non hanno parti in comune, ma insieme formano il disegno

Proprietà delle operazioni

Molte peculiarità dell'insiemistica si applicano all'algebra degli eventi

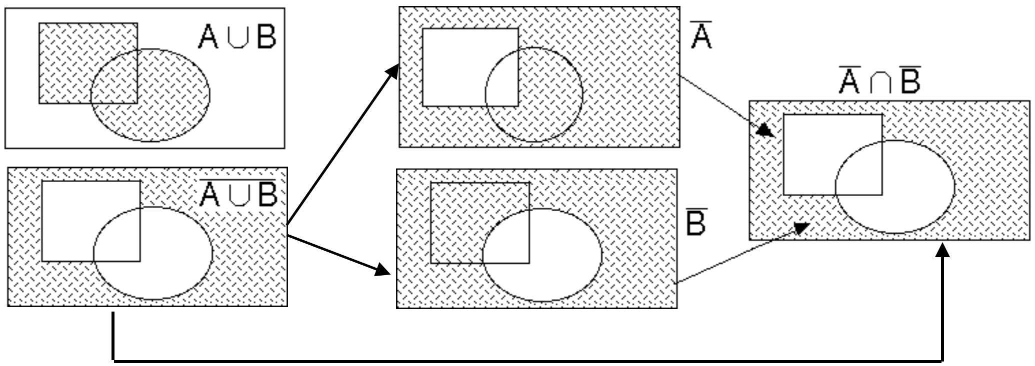
Legge commutativa:	$E \cup F = F \cup E$	$E \cap F = F \cap E$
Legge associativa	$E \cup (F \cup G) = (E \cup F) \cup G$	$E \cap (F \cap G) = (E \cap F) \cap G$
Legge distributiva	$E \cup (F \cap G) = (E \cup F) \cap (E \cup G)$	$E \cap (F \cup G) = (E \cap F) \cup (E \cap G)$
Idempotenza	$E \cup E =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

Leggi di De Morgan

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

1) $\overline{A \cap B} = \overline{A} \cup \overline{B}$ **Complemento ed unione definiscono l'intersezione**

2) $\overline{A \cup B} = \overline{A} \cap \overline{B}$ **Complemento ed intersezione definiscono l'unione**



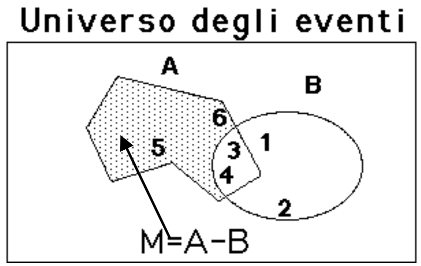
Applicazione: evento sottrazione

Dati due eventi "A" e "B" . L'evento sottrazione, "M = A-B", si verifica se si verificano gli eventi di "A" che non siano anche in "B"

ESEMPIO: lancio di due dadi di colore diverso

L'universo degli eventi è formato dagli interi {1, 2, 3, 4, 5, 6}

A: Dado rosso > 2
B: Dado blù <5



L'evento sottrazione "M" si verificherà se esce un numero più grande di "2", ma che NON sia più piccolo di "5"

E' facile controllare che: $A - B = A \cap \overline{B}$

Le algebre

Con “S” e con l’insiemistica si possono formare tanti altri insiemi;

Tutti questi insiemi formano a loro volta un insieme i cui elementi sono degli insiemi.

Ci interessa un particolare tipo di questi insiemi di insiemi: l’algebra, indicata con W che ha le seguenti caratteristiche:

$$
\begin{array}{lll}
1. & Se \;\; E \in W \;\; e \;\; F \in W & \Rightarrow (E \cup F) \in W; \\
1bis. & Se \;\; E \in W \;\; e \;\; F \in W & \Rightarrow (E \cap F) \in W; \\
2. & Se \;\; E \in W & \Rightarrow E^c \in W;
\end{array}
$$

Un’algebra non contiene eventi elementari, ma solo eventi composti, anche da un solo evento elementare (SINGOLETTO)

Le condizioni 1 e 1bis sono alternative visto che usarle entrambe porterebbe ad una ridondanza di cui non si ha bisogno.

Le algebre/2

Dato il dominio “S” possiamo costruire una algebra attivando un numero finito di operazioni di unione e negazione o di intersezione e negazione.

Se “S” contiene “k” eventi elementari possiamo costruire sottoinsiemi che contengono un solo elemento, che ne contengono due, tre e così via

Il totale degli eventi composti è: 2^k se k è il numero di eventi elementari

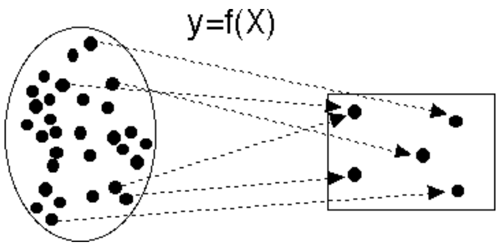
In realtà se ne trattano molti di meno, ma lo schema di Kolmogorov si estende a tutto ciò che è coerente con i postulati e non solo a ciò che interessa.

La funzione di insieme

La funzione $f(X)$ è una legge che associa ad ogni punto di un insieme (DOMINIO) uno ed un sol punto di un altro insieme (CODOMINIO o IMMAGINE)

Nella accezione usuale: "X" è un punto del dominio.

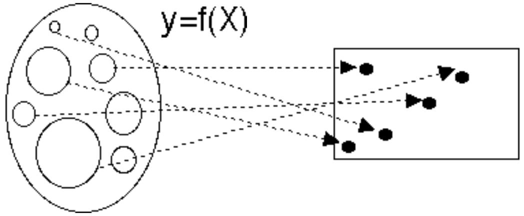
Esempio: $y = 2X$



La nozione si estende al caso in cui il dominio è costituito da INSIEMI.

FUNZIONE DI INSIEME: l'argomento "X" è un insieme.

Esempio: dominio dei cerchi nel piano con associata la circonferenza



Teoria elementare della probabilità

Esistono diverse presentazioni della probabilità. Noi seguiamo la teoria assiomatica di KOLMOGOROV

- **Introduzione dei CONCETTI PRIMITIVI: *(PROVA, EVENTO, PROBABILITA')***
- **In base a questi si stabiliscono i POSTULATI cioè le regole per ragionare sulle probabilità**
- **Da questi e solo da questi segue il CALCOLO DELLE PROBABILITA'**

La teoria delle probabilità forma un modello matematico astratto usato per amministrare razionalmente sensazioni di fiducia, speranza, timore.

L’approccio di Kolmogorov

Perché la teoria non sia una mera raccolta di concatenazioni logiche, ma sia "scienza" i postulati sono attinti dalla realtà osservabile.

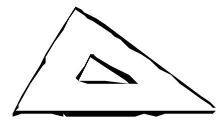
In particolare debbono essere:

- **COERENTI: non devono generare contraddizioni interne**
- **UTILI: devono essere subito operativi**
- **NON RIDONDANTI: non devono poter essere dedotti da altri postulati**

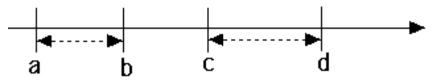
L’approccio di Kolmogorov/2

Un vantaggio dei postulati di Kolmogorov è che essi derivano in modo diretto dall'esperienza.

In fondo, quello che ha fatto Kolmogorov altro non è che allargare agli eventi la misura di una grandezza fisica.



ESEMPIO:



Le lunghezze (b-a) e (d-c) sono sempre non negative.

La lunghezza congiunta dei due intervalli è (b-a)+(c-d) perché non ci sono punti in comune,

LA NON NEGATIVITA' E L’ ADDITIVITA' ricorrono nei postulati sulla probabilità.

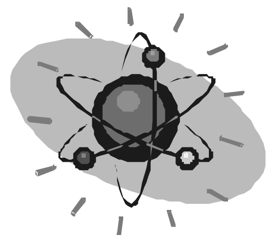
Postulato_0: gli eventi formano un’ algebra

Dall'universo degli eventi "S" si possono ricavare tanti sottoinsiemi, incluso l'evento certo ed impossibile.

Indichiamo con “W” l’insieme di tutti gli insiemi comunque derivati da “S”

Tale insieme è detto algebra se:

1. $S \subseteq W$;
2. Se $A \in W \Rightarrow \overline{A} \in W$
3. Se $A_1 \in W$ e $A_2 \in W \Rightarrow A_1 \cap A_2 \in W$
4. Se $A_1 \in W$ e $A_2 \in W \Rightarrow A_1 \cup A_2 \in W$



ciò che l'unione FINITA e la negazione di eventi in W (e quindi anche le altre operazioni) generano altri eventi che ricadono pure nell’algebra W

Postulati_1,2: limiti della probabilità

Secondo Kolmogorov ad ogni evento $A \in W$ si può associare un numero reale P(A) variabile in un intervallo limitato

- **La probabilità di un evento "A" è maggiore o uguale a zero**

$$P(A) \geq 0$$

- **La probabilità dell'evento certo ”S" è pari ad uno**

$$P(S) = 1$$

Postulato_3: additività

La probabilità dell'evento unione di due o più eventi incompatibili è data dalla somma delle probabilità degli eventi.

$$\text{Se } A_i \cap A_j = \varnothing \;\; \forall i \neq j \Rightarrow P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

per ogni serie, FINITA di eventi.

La probabilità P è una funzione di insieme NON NEGATIVA, ADDITIVA E LIMITATA

La probabilità può essere espressa come decimale: 0.25, percentuale 25%, frazione: 1/4: casi contro e a favore 1 a 3

Probabilità di eventi elementari e composti

Consideriamo solo universi formati da un numero finito di eventi elementari.

In questo caso basta assegnare la probabilità ai singoli eventi elementari e poi sfruttare il 3° postulato per assegnare la probabilità agli eventi composti

Se l'evento composto "E" è formato da: $E = \{A_1, A_2, \ldots, A_n\}$

La probabilità di "E" è pari alla somma delle probabilità degli eventi elementari che in esso ricadono.

$$P(E) = P(A_1 \cup A_2 \cup \ldots \cup A_n)$$
$$= P(A_1) + P(A_2) + \ldots + P(A_n)$$

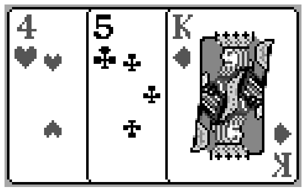
Esempi

Impostare il calcolo della probabilità di avere una carta rossa (cuori o quadri) in una mano di poker

$$Probabilità\{carta\ rossa\} = Probabilità\{cuori\ o\ quadri\}$$
$$= Probabilità\{cuori\} + Probabilità\{quadri\}$$

Impostare il calcolo della probabilità di avere una figura (“J”, “Q”, “K”) in una mano di poker

$$Probabilità\{figura\} = Probabilità\{"J",\ "Q",\ "K"\}$$
$$= Probabilità\{"J"\} + Probabilità\{"Q"\} + Probabilità\{"K"\}$$



La probabilità dell'evento impossibile è zero

Dal fatto che $S \cup \varnothing = S \quad e \quad S \cap \varnothing = \varnothing$

Consegue che $P(S \cup \varnothing) = P(S) + P(\varnothing) = 1 + P(\varnothing)$

Tuttavia $P(S \cup \varnothing) = P(S) = 1$

E quindi $P(\varnothing) = 0$

Una parte su un milione di miliardi del bicchiere di un ottimo vino è mortale. Berreste?

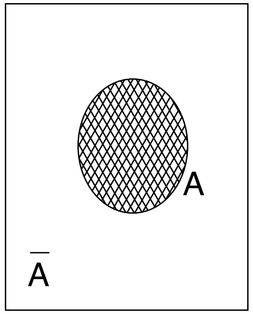
Probabilità zero implica l’evento impossibile?

Non sempre: lo zero può essere l’approssimazione di un numero talmente non distinguibile dallo zero, ma positivo.

La probabilità dell'evento negazione

è il complemento ad uno della probabilità dell'evento che si nega

S



Dal fatto che: $A \cup \overline{A} = S$ e $A \cap \overline{A} = \varnothing$

consegue che: $P(A \cup \overline{A}) = \boxed{P(S) = 1}$

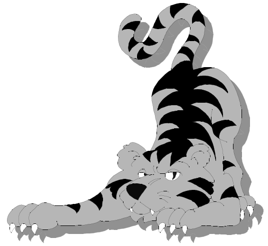
ma anche che: $P(A \cup \overline{A}) = P(A) + P(\overline{A})$

Quindi $P(\overline{A}) = 1 - P(A)$

Le due scommesse
10,000 che esce il 27 sulla ruota di Napoli al 10%
10,000 Non esce il 27 su Napoli al 90%

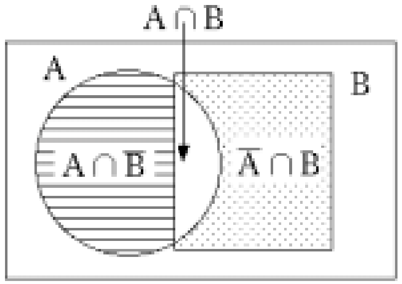
Devono risultare indifferenti

Legge additiva



Teorema_4: probabilità totale. Dati due eventi qualsiasi si ha

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



$$(A \cap \overline{B}) \cup (A \cap B) = A \cap (B \cup \overline{B}) = A \cap I = A$$

$$(A \cap \overline{B}) \cap (A \cap B) = A \cap (B \cap \overline{B}) = A \cap \varnothing = \varnothing$$

D'altra parte

$$(A \cap B) \cup (\overline{A} \cap B) = B \text{ e } (A \cap B) \cap (\overline{A} \cap B) = \varnothing$$

Quindi, l'unione è stata scomposta in tre eventi incompatibili. Applicando ora il 3° postulato

aggiungo e tolgo

$$\begin{aligned} P(A \cup B) &= \underline{P(A \cap \overline{B}) + P(A \cap B)} + \underline{P(\overline{A} \cap B)} \\ &= P(A \cap \overline{B}) + P(A \cap B) + P(\overline{A} \cap B) + P(A \cap B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

Esempio

In un processo produttivo si sceglie a caso un item. Indichiamo con

- A: L'item è difettoso sul peso — P(A)=0.38
- B: L'item è difettoso nella forma — P(B)=0.33
- A « B L'item è difettoso sia nella forma che nel peso — $P(A \cap B)=0.26$

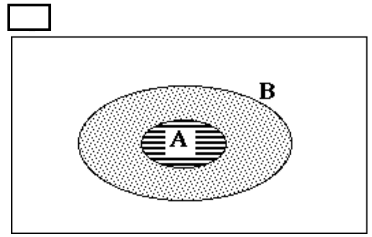
La probabilità che l'item sia difettoso per uno o entrambi i motivi è

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.38 + 0.33 - 0.26 = 0.45 \end{aligned}$$

Probabilità di un sottoinsieme

Se A è un sottoinsieme di B allora

$$P(B) \geq P(A)$$



Scriviamo l'evento "B" come

$B = A \cup (B \cap \overline{A})$ **con** $A \cap (B \cap \overline{A}) = \varnothing$

Richiamando il postulato degli eventi incompatibili avremo

$$P(B) = P(A) + P(B \cap \overline{A}) \geq P(A)$$

poiché la probabilità è non negativa.

Misura della probabilità

Esiste una classe di eventi per la quale possiamo assegnare le probabilità in base ai soli postulati

Assumiamo che gli eventi $E_1, E_2, \ldots E_k$ **siano una partizione**

$$1)\ E_i \cap E_j = \varnothing \quad \forall\ i \neq j$$

$$2)\ \bigcup_{i=1}^{k} E_i = I$$

Teorema:
se gli eventi sono equiprobabili, la loro probabilità è data dal reciproco del numero di eventi.

$$1 = P\left(\bigcup_{i=1}^{k} E_i\right) = \sum_{i=1}^{k} P(E_i)$$

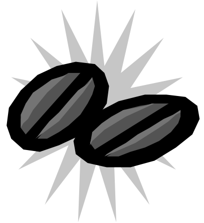
Eventi equiprobabili significa che $P(E_i) = \alpha \quad \forall i$ **e quindi** $1 = \sum_{i=1}^{k} \alpha = k\alpha \Rightarrow \alpha = \frac{1}{k}$

In questo caso non ci sono circostanze oggettive né soggettive che ci indicano quale debba essere la probabilità, ma questa è la conseguenza automatica della SIMMETRIA degli eventi e dei postulati

Perché il calcolo combinatorio

Basato sulle idee primitive di distinzione e di classificazione, stabilisce in quanti modi diversi si possono combinare degli oggetti e torna molto utile nell'enumerazione dei casi possibili di una prova

ESEMPIO:
Indagine sul consumo di caffè



Miscela $\{A, C, M\}$
Confezione $\{B, S\}$
Formato $\{S, D, F, B\}$

ABS	ABD	ABF	ABB
ASS	ASD	ASF	ASB
CBS	CBD	CBF	CBB
CSS	CSD	CSF	CSB
MBS	MBD	MBF	MBB
MSS	MSD	MSF	MSB

Ogni terna è un possibile soggetto di studio

Non sempre è facile catalogare le possibilità. C'è bisogno di una regola generale

Prove e sottoprove

L'evento può risultare da più sottoprove anche con dominio diverso

Comunque si produce una n-tupla che può essere:

$$\textit{Non ordinata}: \ E = \{x_1 \in S_1,\ x_2 \in S_1,\ \ldots,\ x_n \in S_n\}$$

$$\textit{Ordinata}: \ E = (x_1 \in S_1,\ x_2 \in S_1,\ \ldots,\ x_n \in S_n)$$

Nella prima ogni alterazione dell'ordine genera un evento distinto

Nella seconda dove, per avere un evento diverso, è necessario modificare almeno un elemento.

Scelta con e senza ripetizione

Con ripetizione. Uno stesso elemento può essere ripetuto più volte nella stessa n-tupla

Esempio: nella roulette esce il 12 e poi riesce il 12

Senza ripetizione (o in blocco). Gli elementi della n-tupla sono tutti distinti

Sulla ruota di Venezia esce il 12 al 1° estratto, ma lo stesso 12 non può uscire nelle altre estrazioni

La moltiplicazione combinatorica

Nell'esempio, l'entità è formata da tre "caselle": la 1^ può variare in 3 modi, la 2^ in 2 e la 3^ in 4. Il numero totale di entità è quindi

$$N = 3 * 2 * 4 = 24$$

Supponiamo che

1) l'entità sia formata dall'accostamento di "k" categorie: $\boxed{1 \mid 2 \mid \ldots \mid k-1 \mid k}$

2) Ogni categoria abbia un numero definito di livelli: $\boxed{n_1 \mid n_2 \mid \ldots \mid n_{k-1} \mid n_k}$

Le possibilità saranno in tutto: $n = n_1 * n_2 * \ldots * n_k = \prod_{i=1}^{k} n_i$

Questa è una regola fondamentale del calcolo combinatorio e permette di stabilire il numero di possibilità risultanti dalla interazione di scelte diverse.

L'albero degli abbinamenti

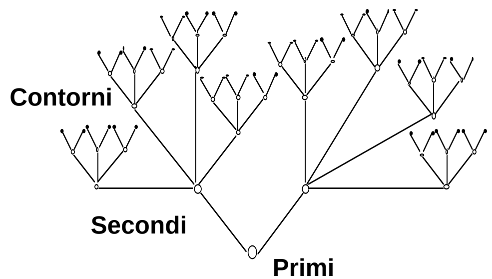
E' una utile rappresentazione: da una radice comune partono tanti rami per quante sono le categorie della prima variabile classificatoria

Il ramo termina con un nodo e da ognuno di questi partirà un nuovo ramo per ogni categoria della seconda variabile e così via

Esempio:

Un menù 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?

La risposta si può ottenere contando i rami terminali: 48.



Esempio

ESEMPIO: indagine sul consumo di caffé

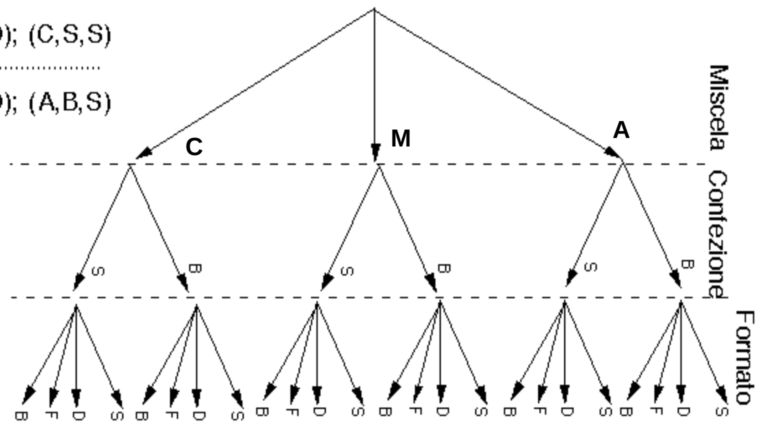
$n = 3 * 2 * 4 = 24$

(C,S,B); (C,S,F); (C,S,D); (C,S,S)

....

(A,B,B); (A,B,F); (A,B,D); (A,B,S)

$E_1, E_2, \ldots, E_{24}$



Disposizioni senza ripetizione

Un'urna prevede N bussolotti. L'esperimento consista nell'estrarne -senza reimmissione- "n" .

L'evento elementare è una n-tupla di elementi.

Si ammette che l'ordine sia importante e cioè che {A, B} è diverso da {B,A} anche se ha gli stessi elementi.

ESEMPIO: $S = \{A, E, C, N\}$ **Se n=1 le scelte sono 4:** (1), (2), (3), (4)

Scelte possibili per n=2

AE	AC	AN	EC	EN	CN
EA	CA	NA	CE	NE	NC

12

Scelte possibili per n=3

AEC	EAC	CAE	NAE	NCE	ECN
ACE	ECA	CEA	NEA	NEC	ENC
AEN	EAN	CEN	NAC	CAN	ACN
ANE	ENA	CNE	NCA	CNA	AEC

24

Disposizioni senza ripetizione/2

Scelte possibili con n=4:

AECN	AENC	ANEC	ANCE	ACEN	ACNE
EACN	EANC	ENAC	ECAN	ENCA	ECNA
CEAN	CENA	CNEA	CNAE	CANE	CAEN
NACE	NAEC	NECA	NEAC	NCAE	NCEA

$\boxed{24}$

DISPOSIZIONI SENZA RIPETIZIONE DI "N" OGGETTI PRESI "n" ALLA VOLTA

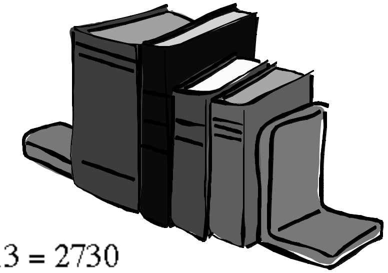
$$D_{sr}(N,n)$$

l'evento è formato da n caselle: | 1 | 2 | ... | n |

Nella prima casella possiamo porre uno qualsiasi degli "n" oggetti; nella seconda uno dei restanti cioè "(N-1)" perché il primo non si può ripetere; nella terza"(N-2)" perché le prime due sono già impegnate e così via.

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

Esempi



I volumi numerati di un'opera in N=15 tomi sono riposti a caso su di uno scaffale. In quanti modi i primi tre posti possono accogliere i volumi?

$$D_{15,3} = 15*(15-1)*(15-3+1) = 15*14*13 = 2730$$

$$3 = 15-3+1$$

In quanti modi possono venir ordinatamente estratti i 5 numeri della ruota di Roma fra gli N=90 disponibili del gioco del lotto

$$D_{90,5} = 90*89*88*87*86 = 5,273,912,160$$

$$\text{NB:}\quad 86 = 90-5+1$$

Disposizioni con ripetizione

Trattiamo ora il caso in cui sia permessa la ripetizione degli oggetti. Quello che cambia è che, qualunque sia la casella, ci sono sempre "N" possibilità

$$D_{cr}(N,n) = \overbrace{N*N*\ldots*N}^{\text{n volte}} = N^n$$

Avendo "N" biglie da collocare in "n" celle (in modo che ogni cella possa contenere anche più di una biglia) le disposizioni con ripetizione esprimono il numero possibile di scelte

ESEMPIO: $S = \{A, E, C, N\}$

Disposizioni con ripetizione di 4 oggetti presi a 2 alla volta

$$D_{cr}(4,2) = 4^2 = 16$$

AA	AE	AC	AN
EA	EE	EC	EN
CA	CE	CC	CN
NA	NE	NC	NN

Esempi

Nel codice ASCII un carattere è rappresentato da una disposizione di 8 bit ognuno dei quali può assumere valore "0" oppure "1". Quanti sono i possibili codici:

$$\overbrace{2*2*2*2*2*2*2*2}^{8\text{ volte}} = 2^8 = 256$$

Nei concorsi TOTIP si ottiene la vincita massima indovinando n=14 risultati legati ad altrettanti concorsi ippici.

Ogni risultato può essere: "1", "X", "2" quante sono le possibili colonne?

$$3^{14} = 4\ \ 782\ \ 969$$

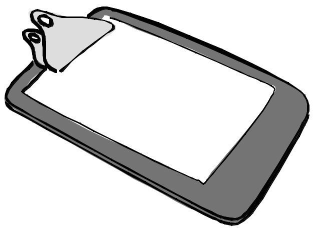
Le permutazioni

Quante sono le scelte possibili se un diverso ordinamento delle unità genera una scelta diversa?

La risposta è data dal numero di disposizioni senza ripetizione per tutte le "n" posizioni

$$P_{sr}(n,n) = n*(n-1)*(n-2)*\ldots*2*1 = n!\ (fattoriale\ \ di\ \ "n")$$

detta anche PERMUTAZIONE senza ripetizione degli "n" oggetti



ESEMPIO:

Un ordine del giorno è stato circoscritto a n=7 argomenti. Quante sequenze di discussione sono possibili?

$$P_{sr}(7) = 7*6*5*4*3*2*1 = 5040$$

Digressione sui fattoriali

Per ogni intero positivo "n" il simbolo "n!" indica il fattoriale di "n" cioè il prodotto dei primi "n" interi naturali.

$$n! = \prod_{i=1}^{n} i = 1*2*3*\ldots*(n-1)*n$$

Da notare che $n! = n*(n-1)!$

Perché questa valga anche per n=1 si pone convenzionalmente $0! \equiv 1$

Per avere un'idea della rapidità con cui cresce il fattoriale di un numero si consideri la seguente tabella:

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

Permutazioni e Disposizioni S.R.

Tra permutazioni e disposizioni S.R. esiste un legame che torna molto utile

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

il senso è che ogni disposizione S.R. di "N" oggetti presi "n" alla volta" può essere abbinata alla permutazione dei restanti (N-n) oggetti

il numero di disposizioni S.R. si può ricavare dal numero di permutazioni:

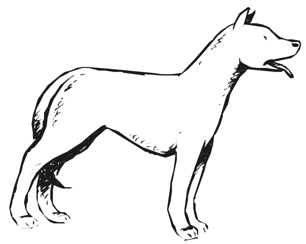
$$D_{SR}(N,n) = \frac{N!}{(N-n)!} = \frac{D_{SR}(N,N)}{D_{SR}(N-n,N-n)}$$

ESEMPI:

$$D_{SR}(6,3) = \frac{6!}{(6-3)!} = \frac{6*5*4*3*2*1}{3*2*1} = \frac{6*5*4}{1} = 120$$

$$D_{SR}(7,5) = \frac{7!}{(7-5)!} = \frac{7*6*5*4*3*2*1}{2*1} = \frac{7*6*5*4*3}{1} = 2520$$

Esempi



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 esistono?

$$\frac{12!}{10!} = 12*11 = 132$$

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

$$\frac{18!}{15!} = 18*17*16 = 4896$$

Combinazioni

Un modo frequente di costruire "S" è la scelta in blocco di "n" oggetti tra "N" disponibili sena badare all'ordine (Combinazioni)

ESEMPIO:

P={E, N, O, S} che implica: {E,N}, {E,O}, {E,S}, {N,O}, {N,S}, {O,S}

Le scelte {O,S} e {S,O} coincidono e contano per una

Le disposizioni sarebbero : $D_{sr}(N,n) = N!/(N-n)!$

ma per ogni disposizione ve ne sono n! identiche che contano come se fossero una sola perciò:

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

Esempio

Nel solito caso: $S = \{A, E, C, N\}$ consideriamo

Disposizioni senza ripetizione

{A,E}	{A,C}	{A,N}	{E,C}	{E,N}	{C,N}
{E,A}	{C,A}	{N,A}	{C,E}	{N,E}	{N,C}

Combinazioni: $\boxed{\{A,E\}\ \{A,C\}\ \{A,N\}\ \{E,C\}\ \{E,N\}\ \{C,N\}}$

$$C(4,2) = \frac{4!}{2!(4-2)!} = \frac{4*3*2*1}{2*1*2*1} = 6$$

Esercizi

Da una lista di N persone si deve formare un comitato di "n" membri tra cui sarà nominato un presidente. Quante sono le opportunità di scelta se N=15 e n=6?

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

Prima si formare il comitato e poi, ad ogni comitato, si abbina la nomina del presidente:

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

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

i coefficienti binomiali

il numero di combinazioni ha un altro simbolo:

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

detto coefficiente binomiale.

Ecco alcune sue proprietà importanti

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

a) e b) affermano che c'è un solo modo di scegliere tutti gli elementi e c'è pure un solo modo di non scegliere alcun elemento: lasciare intatto l'insieme di base.

c) Discende dal fatto che ogni scelta di "n" implica una scelta di (N-n)

Teorema del binomio

In molte occasioni ricorre la seguente uguaglianza:

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

Poichè $(a+b)^n = \overbrace{(a+b)*(a+b)*\ldots*(a+b)}^{n \text{ volte}}$

il termine a^n **non può che essere ottenuto prendendo il fattore "a" da tutti i binomi e questo può essere fatto in un solo modo** $\binom{n}{0} = 1$

il successivo $a^{n-1}b$ **si ottiene prendendo il fattore "a" in (n-1) binomi e "b" nel restante e ciò può farsi in** $\binom{n}{1}$ **modi diversi**

il termine generico $a^{n-i}b^{i}$ **può aversi scegliendo "a" in "(n-i)" binomi e "b" negli altri "i" e questo si farà in** $\binom{n}{i}$ **modi.**

Applicazioni

Calcolo combinatorio ed equiprobabilità sono al centro di gran parte delle applicazioni del nostro corso

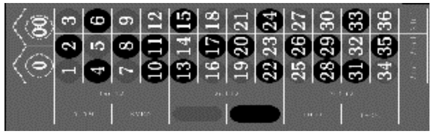
Poiché la formula per stabilire la probabilità di un evento composto è

$$P(E) = \frac{\textit{Numero eventi elementari in E}}{\textit{Numero di eventi nell'Universo degli eventi}}$$

Basterà conoscere quanti e quali sono gli eventi elementari pertinenti ad E per determinarne la probabilità

Esempio:
Probabilità di un carré

$$P(E) = \frac{4}{36} \quad \textit{ovvero} \;\; 1:8$$



Lo schema ipergeometrico

Separazione di "N" oggetti in due gruppi

uno con N_1 elementi "speciali" perché hanno una certa proprietà

un altro di (N-N_1) elementi "comuni" per i quali la proprietà non è vera.

L'esperimento consiste nella scelta casuale <u>senza reimmisione</u> di "n" elementi di cui n_1 speciali ed i restanti (n- n_1) comuni.

Qual'è la probabilità che la scelta contenga proprio n_1 speciali?

La scelta di questi può avvenire in $C(N_1, n_1)$ modi diversi. Ognuna delle combinazioni di speciali può abbinarsi con le combinazioni di (N- N_1) comuni presi a blocchi di (n-n1) e quindi i casi favorevoli sono:

$C(N_1, n_1)*C(N- N_1, n- n_1)$ con $C(N,n)$ casi possibili.

Quindi

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

Esempio

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 in coppie, anche di membri dello stesso genere

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

$$P(D=2, U=2) = \frac{\binom{6}{2}\binom{6}{2}}{\binom{12}{4}} = 0.4546$$

2. Qual'è la probabilità che siano scelte più ragazze che ragazzi?

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

Probabilità condizionata

Dati due eventi "A" e "B" come si valuta la probabilità di "B" supponendo che "A" si sia già verificato?

la probabilità di "B" supponendo che si verifichi dopo "A" con P(A) > 0 è

Definita come

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Probabilità congiunta di "A" e "B"

Probabilità marginale di "A"

la linea significa "dato" non diviso

il verificarsi di "A" determina un restringimento dell'universo degli eventi che debbono essere ora riconsiderati nel nuovo contesto "A"

Universo originario

Esempio

Supponiamo che le facce di un dado siano equiprobabili. Abbiamo perciò le probabilità:

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	$1/2$

Se sappiamo che "è uscito un dispari" questo modifica la prova : alcuni eventi sono ancora possibili, altri no.

Le probabilità ridefinite alla luce di ciò che ciascuno aveva in comune con "A" e scalate in modo da sommare ad uno (probabilità dell'evento certo

E	1	2	3	4	5	6	
$P(E)$	$\cancel{1/6}\ \cancel{1/2}$	0	$\cancel{1/6}\ \cancel{1/2}$	0	$\cancel{1/6}\ \cancel{1/2}$	0	$\cancel{1/2}\ \cancel{1/2}$

$E\|A$	1	3	5	
$P(E\|A)$	$1/3$	$1/3$	$1/3$	1

Probabilità nelle partizioni

Consideriamo una partizione (eventi necessari e incompatibili) dell'universo degli eventi

$$H_1, H_2, \ldots, H_k$$

Partizione significa che di questi eventi, in ogni prova, se ne verifica uno e solo uno.

Poiché $E = E \cap I = E \cap \{H_1 \cup H_2 \cup \ldots \cup H_k\}$

$$= (E \cap H_1) \cup (E \cap H_2) \cup \ldots \cup (E \cap H_k)$$

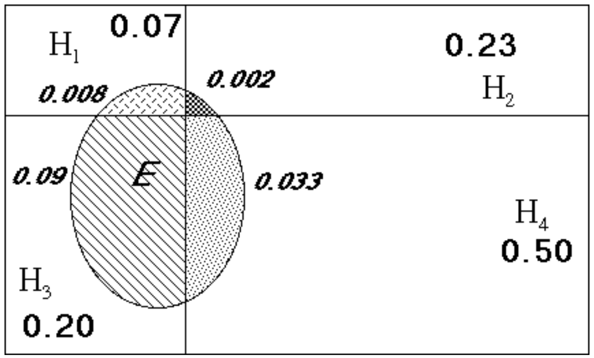
Ne consegue che

$$P(E) = P(E \cap H_1) + P(E \cap H_2) + \ldots + P(E \cap H_k)$$
$$= P(H_1) * P(E|H_1) + P(H_2) * P(E|H_2) + \ldots + P(H_k) * P(E|H_k)$$

La probabilità di "E" è pari alla somma PONDERATA delle probabilità condizionate di "E" . I pesi sono le probabilità incondizionate degli eventi elementari

Esempio con i diagrammi di Venn

Universo degli eventi



Scegliamo le seguenti probabilità

$P(H_1) = 0.07 \quad P(E \cap H_1) = 0.008$
$P(H_2) = 0.23 \quad P(E \cap H_2) = 0.002$
$P(H_3) = 0.20 \quad P(E \cap H_3) = 0.090$
$P(H_4) = 0.50 \quad P(E \cap H_4) = 0.033$

$$P(E) = 0.133$$

$$E = (H_1 \cap E) \cup (H_2 \cap E) \cup (H_3 \cap E) \cup (H_4 \cap E)$$

Ne consegue:

$$P(E) = P(E \cap H_1) + P(E \cap H_2) + P(E \cap H_3) + P(E \cap H_4)$$
$$= P(H_1) * P(E|H_1) + P(H_2) * P(E|H_2) + P(H_3) * P(E|H_3) + P(H_4) * P(E|H_4)$$
$$= 0.07 * \frac{0.008}{0.07} + 0.23 * \frac{0.002}{0.23} + 0.20 * \frac{0.0.09}{0.20} + 0.50 * \frac{0.033}{0.50} = 0.133$$

Teorema di Bayes

Le condizioni sperimentali non sempre consentono soluzione intuitive. Possiamo però dimostrare il seguente teorema

Data una partizione $H_1, H_2, \ldots, H_k$ **dell'universo degli eventi la probabilità a posteriori rispetto all'evento "E" di** H_i **è data dalla formula**

<u>Formula</u>

$$P(H_i|E) = \frac{P(H_i)*P(E|H_i)}{\sum_{j=1}^{k} P(H_j)*P(E|H_j)}$$

<u>Logica</u>

Si sa già il risultato della prova e si cerca quale ne sia la causa fra quelle possibili (PRINCIPIO DELLA PROBABILITA' INVERSA)

Da quanto detto sulle partizioni

$$P(E) = \sum_{j=1}^{k} P(H_j)*P(E|H_j); \qquad P(H_i \cap E) = P(H_i)*P(E|H_i)$$

Per cui

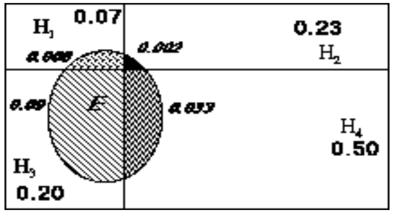
$$P(H_i|E) = \frac{P(H_i)*P(E|H_i)}{\sum_{j=1}^{k} P(H_j)*P(E|H_j)} = \frac{P(H_i \cap E)}{P(E)}$$

Esempio

Ritorniamo al caso illustrato con i diagrammi di Venn e determiniamo la causa più probabile di "E"

$$P(H_1|E) = \frac{0.008}{0.133} = 0.0602 \qquad P(H_3|E) = \frac{0.090}{0.133} = 0.6767$$

$$P(H_2|E) = \frac{0.002}{0.133} = 0.0150 \qquad P(H_4|E) = \frac{0.033}{0.133} = 0.2481$$



La causa più probabile è allora "H3" come il diagramma mostra con chiarezza: se, in una scommessa, tutti gli eventi dessero luogo alla stessa vincita, la logica ci impone di scegliere H3.

Nella formula di Bayes il denominatore è costante per cui spesso si scrive

$$P(H_i|E) = \frac{P(H_i)*P(E|H_i)}{P(E)} \propto P(H_i)*P(E|H_i)$$

che esprime la probabilità a posteriori come proporzionale alla probabilità a priori per la cosiddetta Verosimiglianza (LIKELIHOOD) ovvero la probabilità, sotto Hi, che si verifichi "E".

Esercizio

Può un ciuccio superare un esame?

Dati:
1) il 75% di chi si presenta all'esame, supera l'esame.
2) il 70% di chi supera l'esame è bravo.
3) il 90% dei bocciati è ciuccio.

S="Superato", R="Respinto", B="Bravo", C="Ciuccio"

$$1) P(S) = 0.75, \qquad 2) P(B|S) = 0.70, \qquad 3) P(C|R) = 0.9$$

E' richiesto il calcolo di P(SIC).

$$P(S|C) = \frac{P(S \cap C)}{P(C)} = \frac{P(S)P(C|S)}{P(S)P(C|S)+P(R)P(C|R)} = \frac{P(S)[1-P(B|S)]}{P(S)P(C|S)+P(R)P(C|R)}$$

$$= \frac{0.75*0.30}{0.75*0.30+0.25*0.90} = 0.5$$

Uso del teorema di Bayes

il 5% degli abitanti di un paese è affetto da una malattia.

Poniamo:

$$A_1 = \text{ha la malattia}; \quad A_2 = \text{non ha la malattia};$$

Supponiamo di disporre di un test clinico la cui SENSITIVITA', cioè la probabilità che esso sia positivo (B) dato che la persona è ammalata, sia pari a:

$$P(B|A_1) = 0.90$$

e con probabilità di FALSO POSITIVO (la persona è sana, ma il test indica il contrario)

$$P(B|A_2) = 0.15$$

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?

$$P(A_1|B) = \frac{P(A_1)*P(B|A_1)}{P(A_1)*P(B|A_1)+P(A_2)*P(B|A_2)} = \frac{0.05*0.90}{0.05*0.90+0.95*0.15} = 0.24$$

Probabilità a priori e a posteriori

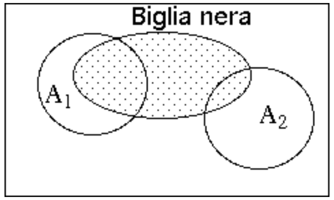
1) si sceglie a caso l'urna;
2) Si scegli a caso la biglia.

La prova è stata effettuata e risulta "biglia nera". Da dove proviene?

$P(A_1) = 0.5$ $\qquad$ $P(A_2) = 0.5$

La probabilità assegnata ad A1 e A2 prima dell'esperimento è detta A PRIORI.

Come si modifica alla luce del fatto è stata scelta una biglia nera?



il verificarsi dell'evento "N" limita l'attenzione alla sola intersezione di A1 con N

$$P(A_1|N) = \frac{P(A_1 \cap N)}{P(N)} = \frac{0.35}{0.45} = 0.78$$

la probabilità dell'evento dopo il verificarsi di un altro è detta A POSTERIORI

Esercizi

1.a) il lancio di due dadi non truccati ha prodotto almeno un "3". Qual'é la probabilità che la somma sia "7"?

$S = \text{La somma é "7"}; \qquad E = \text{é uscito almeno un "3"}$

$$S = \{(1,6);(2,5);(3,4);(4,3);(5,2);(6,1)\}$$

$$E = \left\{\begin{matrix}(3,1);(3,2);(3,3);(3,4);(3,5);(3,6)\\(1,3);(2,3);(4,3);(5,3);(6,3)\end{matrix}\right\}$$

$$P(S|E) = \frac{P(S \cap E)}{P(E)} = \frac{\frac{2}{36}}{\frac{11}{36}} = \frac{2}{11}$$

1.b) qual'é la probabilità che si ottenga una somma pari a "7" nel lancio di due dadi non truccati uno dei quali o entrambi ha prodotto un "3"?

$$P(S \cap E) = P(E) * P(S|E) = \frac{11}{36} * \frac{2}{11} = \frac{2}{36} = \frac{1}{18}$$

2) Un mazzo di carte francesi ha 52 carte di cui 4 sono assi. Se si estraggono due carte senza che la prima estratta venga reimmessa prima della seconda estrazione. Qual'é la probabilità che siano entrambi degli assi?

$A_1 = \text{asso alla prima}$
$A_2 = \text{asso alla seconda}$

$$P(A_1 \cap A_2) = P(A_1) * P(A_2|A_1) = \frac{4}{52} * \frac{3}{51} = 0.0045$$

Indipendenza

DUE EVENTI A E B SONO INDIPENDENTi SE IL VERIFICARSI DELL'UNO NON ALTERA LA PROBABILITA' DELL'ALTRO

$$P(A|B) = P(A)$$

Tale interpretazione è coerente con l definizione di probabibilità condizionata

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow \frac{P(A) * P(B)}{P(B)} = P(A)$$

L'indipendenza è una relazione BILATERALE: se "E" è indipendente da "F" allora è vero anche il viceversa purché "E" non sia impossibilie"

$$P(E|F) = P(E) \;\Rightarrow\; P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E) * P(F)}{P(E)} = P(F)$$

Esempi

Si supponga che gli eventi "A" e "B" siano indipendenti e che si abbia

$$P(A) = 0.45 \quad e \quad P(B) = 0.80$$

a) Calcolare $P(A \cup B)$

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A) * P(B) = 0.45 + 0.80 - 0.36 = 0.89}$$

b) Calcolare $P(\overline{A}|\overline{B})$

$$\boxed{P(\overline{A}|\overline{B}) = \frac{P(\overline{A} \cap \overline{B})}{P(\overline{B})} = \frac{P(\overline{A \cup B})}{1 - P(B)} = \frac{1 - P(A \cup B)}{0.20} = \frac{0.11}{0.20} = 0.55}$$

Soluzione di problemi con le probabilità

1) Individuare i dati del problema e tradurli in simboli.

2) Delimitare le richieste del problema ed esprimerle in simboli

3) applicare le regole del calcolo delle probabilità

In molti casi può essere utile la seguente formula

$$P\left(E|F\right)=\frac{P(E)}{P(F)}P\left(F|E\right)$$

Che consente di scambiare il ruolo degli eventi tra condizionato e condizionante

Esempi

Un controllo di qualità rivela:

1) il 20% delle componenti è difettoso.
2) Il 90% delle componenti passa il controllo.
3) Il prodotti privi di difetti passano il test nel 95% dei casi.

Qual è la probabilità che una componente non risulti difettosa una volta superato il controllo?

Poniamo E="La componente è difettosa"; F="La componente passa il test"

Il problema ci suggerisce

$$1)P(E)=0.20, \qquad 2)P(F)=0.90, \qquad 3)P\left(F|E^c\right)=0.95$$

E' richiesto il calcolo di $P(E^c|F)$.

$$P\left(E^c|F\right)=\frac{P\left(E^c\right)}{P(F)}P\left(F|E^c\right)=\frac{\left[1-P(E)\right]}{P(F)}P\left(F|E^c\right)=\frac{0.8*0.9}{0.95}=0.7579$$

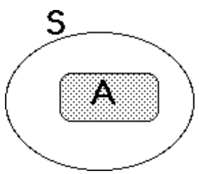
Interpretazione della probabilità

La probabilità è un modello numerico delle relazioni che intercorrono tra le possibili occorrenze degli eventi e le proprietà fisiche dell'esperimento

Sia n(A) il numero di eventi elementari in A e sia n(S) il numero totale di eventi semplici.

La probabilità dell'evento composto è data da

$$\mathrm{P(A)}=\frac{\mathrm{n(A)}}{\mathrm{n(S)}}=\frac{\text{casi favorevoli}}{\text{casi possibili}}$$



La presenza di "simmetrie" negli esperimenti consente una assegnazione oggettiva delle probabilità

Principio della ragione insufficiente: gli eventi sono equiprobabili a meno che non si dimostri il contrario

Interpretazione /2

Una lotteria ha venduto 4750 biglietti. Ciccillo ne ha comprati 4. Se tutti i numeri hanno la stessa probabilità di vincere allora:

$$P(Ciccillo\ \ vince)=\frac{4}{4750}=0,000842=8.42\ \ per\ \ mille$$

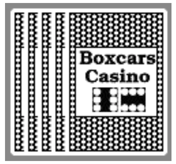
PREGI: E' "naturale". Nel valutare il verificarsi di un evento eseguiamo a mente il rapporto tra le circostanze a favore e quelle contro

DIFETTI: 1) Include una tautologia: "ugualmente possibili" è già una definizione di probabilità.

2) Non può essere richiamata se si ignora la struttura fisica della prova e come questa influenza gli eventi.

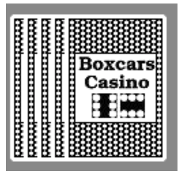
Interpretazione/3

Presenta delle contraddizioni:



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 di colore nero?

- **POISSON: i casi possibili sono: (N_1,N_2), (N_1,R_2), (R1,N2) e (R_1,R_2). Se è la prima è nera allora restano solo 3 casi di cui due a favore. Perciò:**



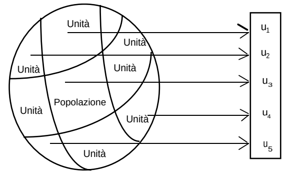
$$P(N_2|N_1) = 2/3$$

- **von Kries: le due scelte sono indipendenti per cui la scelta della 2^ carta può ignorare la scelta della 1^. Quindi:**

$$P(N_2|N_1) = P(N_2) = \frac{26}{52} = \frac{1}{2}$$

Popolazione teorica ed effettiva

In ogni rilevazione di dati è necessario un sistema di riconoscimento delle unità



Non sempre è facile fare la lista delle unità: popolazioni elusive od incerte in cui non sappiamo chi sono e quante sono le unità

Dobbiamo quindi distinguere tra

- **POPOLAZIONE TEORICA: quella che vorremmo esaminare**
- **POPOLAZIONE EFFETTIVA: unità effettivamente raggiungibili**

La *frame* o lista

Tra TEORICA ed EFFETTIVA si inserisce la *frame* o lista cioè un sistema di codici che identificano le unità

La lista è una sovrastruttura imposta alla popolazione.

Per essere utile deve risultare:

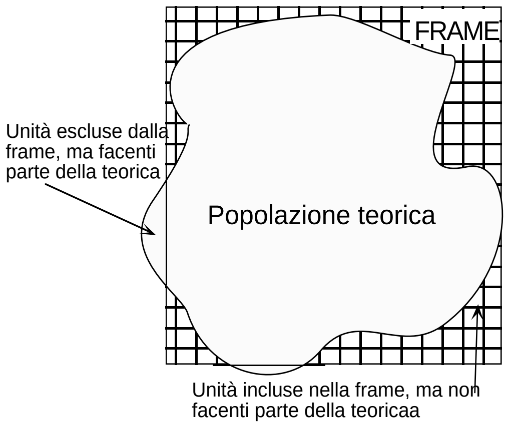
ESATTA

AGGIORANTA

COMPLETA

DOCUMENTATA

con regole note e disponibili



Le ragioni della scelta di una certa popolazione effettiva per una data popolazione teorica devono essere capillari ed accurate.

Le rilevazioni totali

Le RILEVAZIONI TOTALI (O CENSIMENTI) sono quelle in cui sono enumerate o misurate tutte ed indistintamente le unità della popolazione

All'interno delle totali si hanno:

- **RILEVAZIONI GENERALI: riguardano la rilevazione di tutte le unità rispetto alle variabili di interesse (POPOLAZIONE)**

<u>Esempio:</u> un'indagine sul voto che si rivolga a tutti gli elettori di qualsiasi sesso e regione di residenza

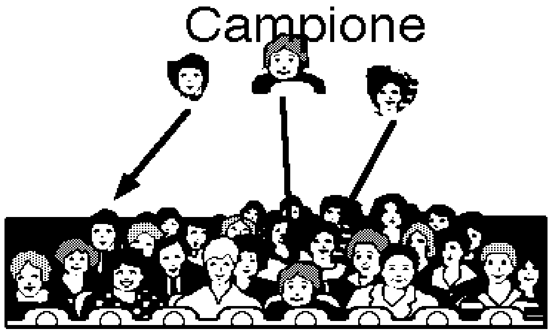
- **RILEVAZIONI SPECIALI: riguardano la rilevazione delle sole unità rispondenti a certe specifiche (SOTTOPOPOLAZIONE)**

<u>Esempio:</u> un'indagine sul voto che si rivolga a tutti, ma i soli iscritti alle camere di commercio come "artigiani"

Le rilevazioni parziali

Sono limitate solo ad una parte della popolazione scelta in base ad opportuni criteri. La parte esaminata si chiama CAMPIONE.

TOTALE/PARZIALE NON E' UNA COTRAPPOSIZIONE, MA UNA COMPLEMENTARITA'



Esperienze consolidate in molti paesi e in molte discipline dimostrano che si può dare pieno affidamento ai campioni purché scelti bene.

Le ragioni del campione

Nel corso di un'indagine ci si può accorgere che la RILEVAZIONE TOTALE non è praticabile perché:

- **HA UN COSTO ECCESSIVO O RICHIEDE GRANDI ORGANIZZAZIONI**

  <u>Esempio:</u> il censimento generale si realizza ogni 10 anni

- **RICHIEDE TROPO TEMPO**

  <u>Esempio:</u> l'intervista di tutti i lavoratori dipendenti richiederebbe tanti anni che una volta finita la popolazione attuale sia molto diversa dalla censita

- **E' TEORICA: PARTE DELLE SUE UNITA' NON ESISTE ANCORA o NON ESISTE PIU'**

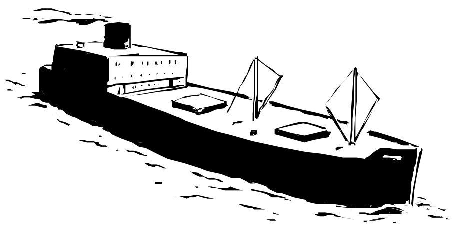
  <u>Esempi:</u> il controllo della qualità dovrebbe riguardare anche le unità non ancora prodotte.
  Le vestigia di antiche civiltà

Il campione

- **L'analisi del campione è meno costosa, più precisa, più asettica, più controllabile e più rapida dell'esame della rilevazione totale.**
- **I censimenti generali si limitano alle variabili fondamentali lasciando ai campioni il compito di scendere nei dettagli.**

La popolazione è la nave che quando naviga lascia vedere solo la parte che galleggia: il campione.

Osservando e analizzando la parte visibile si conoscerà anche la parte che è sotto l'acqua.



Applicazioni del campione

- **Sondaggi elettorali; gradimento delle amministrazioni locali; consenso alle scelte politiche governative.**
- **Ricerche di mercato: accettazione di un nuovo prodotto; apprezzamento della modifica di un prodotto conosciuto; desiderio di un nuovo servizio.**
- **Controllo della qualità: aderenza agli standard di un item; verifica della integrità di una fornitura; certificazione della composizione di un prodotto.**
- **Indagini di laboratorio: efficacia di un fertilizzante; pericolosità di un farmaco; validità di terapie comportamentali; tolleranza ad un prodotto.**
- **Imprenditoria: pagamento di *royalties*; diffusione di quotidiani e settimanali; *audience* televisiva; revisione dei conti.**

Errori non campionari

DISTORSIONI NELLA SCELTA DELLE UNITA'

Il meccanismo di estrazione delle unità agisce solo su alcune parti e ne esclude altre (undercoverage e over coverage)

ESEMPIO:

i sondaggi telefonici: in dipendenza dell'orario in cui si telefona si raggiungono unità diverse. Persone che non hanno telefono o il cui telefono non è in elenco non potranno mai essere raggiunte.

DISTORSIONI NELLA RILEVAZIONE DELLE UNITA'

Non sempre è possibile garantire l'accuratezza delle misurazioni ovvero non sempre le unità sono disposte a farsi rilevare o a rispondere con sincerità (noncoverage e nonresponse)

ESEMPIO:

Rilevazione degli affiliati a "Cosa Nostra”

Il noto caso del Literary Digest

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?

L'ampiezza del campione

il numero di unità “n” considerate nel capione è detto AMPIEZZA del campione. L'ampiezza della popolazione è indicata con "N"

Frazione di campionamento: $\frac{n}{N}$ — **Frazione di unità della popolazione "rappresentata"nel campione**

Intervallo di campionamento: $\frac{N}{n}$ — **Ogni quante unità della popolazione si sceglie una unità del campione.**

Anche, fattore di proporzionalità.

Se la popolazione è infinita questi due rapporti perdono di significato.

Errori campionari e variabilità

Se le unità fossero uguali basterebbe un campione di ampiezza n=1

" " " " " " " "

" " " " " " " "

" " " " "

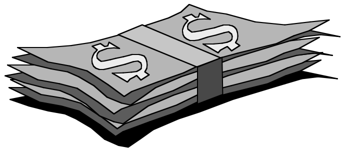
Se le modalità sono due o più saremo certi che entrambe siano nel campione solo se n=N.

Ma la variabilità non è nota. Spesso è uno degli scopi di una ricerca.

Determinare "n"

E’ un elemento fondamentale del piano di campionamento. Sulla scelta incidono...

- **Obiettivo dell'indagine**
- **Variabilità attesa nella popolazione (controllo errori campionari)**
- **Costo dell'acquisizione**
- **Controllo errori non campionari**

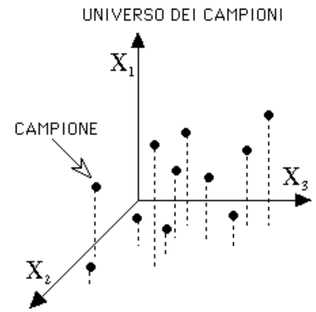


La determinazione di "n" è molto complessa e verrà ripresa in un corso successivo

L'universo dei campioni

Fissata l'ampiezza campionaria "n" definiamo UNIVERSO DEI CAMPIONI (di ampiezza "n") l'insieme di tutti campioni di tale ampiezza che possono essere ottenuti da una data popolazione "P"

$$T_n(P) = \{C_1, C_2, \ldots, C_i, \ldots\}$$



L'universo dei campioni può anche essere considerato a sua volta una POPOLAZIONE le cui unità sono i campioni di ampiezza "n"

Cardinalità dell’universo dei campioni

Dipende dalla:

- **Dalla possibilità di ripetere o no la stessa unità**
- **Se rileva o no l’ordine di comparizione nel campione.**

Se non c’è reimmissione ed i campioni sono considerati uguali purché formati dalle stesse unità allora la cardinalità è il coefficiente binomiale:

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

Esempio:
Ad un test sull’impatto visivo di un poster 6x3 metri sono stati invitati N=50 automobilisti che hanno dato la loro opinione. Di questi, n=7 dovrebbero essere sottoposti ad un altro test sulla leggibilità delle scritte.

Le scelte possibili sono:

$$\binom{50}{7} = \frac{50!}{7!43!} = 99'884'400$$

Cardinalità2

Se le unità possono ripetersi fermo restando che l’estrazione è senza reimmissione e che l’ordine non conta, la cardinalità è:

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

Esempio:
Si analizzano le N=70 sentenze emesse da un collegio giudicante con un’intervista di n=6 condannati. La presenza di recidivi può provocare la ripetizione delle unità.

I campioni possibili sono

$$\binom{70+6-1}{6} = \frac{70*71*\ldots*75}{6!} = 201'359'550$$

Cardinalità/3

Se è NON consentita la reimmissione e l'ordine diverso rende diversi due campioni con uguali unità allora i campioni possibili sono

$$\frac{N!}{(N-n)!}$$

Esempio:
La famosa scienziata ha intuito che un trattamento di 5 farmaci scelti tra 20 principi attivi e somministrati nel giusto ordine può curare una fastidiosa patologia.

Scegliendo a caso i principi attivi quanti campioni sono possibili?

$$\frac{20!}{15!} = 20*19*18*17*16 = 1'\,860'\,480$$

Cardinalità/4

Se è consentita la reimmissione e l'ordine diverso rende diversi due campioni con uguali unità allora l'universo dei campioni ha un numero di elementi pari a:

$$N^n$$

Nei due esempi precedenti avremmo:

$$50^7 = 781'\,250'\,000'\,000; \quad 70^6 = 117'\,649'\,000'\,000$$

il numero di elementi dell'universo dei campioni è quasi sempre troppo elevato - anche con i supercalcolatori- perché valga la pena di studiare il comportamento delle statistiche su tutti.

Effetti del Campionamento

Se si considerano tutte le unità di una popolazione, il problema della selezione delle unità non si pone.

Se non è possibile effettuare un'indagine completa ci saranno unità effettivamente esaminate ed altre no.

Il risultato è che ci si trova di fronte a dei dati che sono quelli, ma avrebbero potuto essere altri.

Cosa si può dire allora sui risultati ottenuti?

Errori campionari

L'uso del campione introduce un errore dovuto alle differenze tra risultati nel campione e risultati POTENZIALI ottenibili dall'esame di tutta la popolazione.

Gli errori sono dovuti a fluttuazioni in parte attribuibili alla naturale variabilità campionaria: i dati sono quelli, ma potevano essere altri

ESEMPIO
Vogliamo conoscere il totale dei valori della tabella (popolazione).

Si sceglie una riga o una colonna di cinque numeri (campione)

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

Riga	Stima	Errore camp.	Colonna	Stima	Errore camp.
40	200	60	18	90	-50
20	100	-40	41	205	65
40	200	60	32	160	20
28	140	0	15	75	-65
12	60	-80	34	170	30

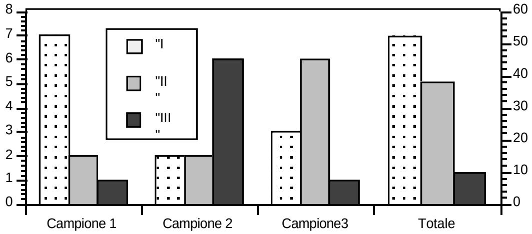
Solo se si sceglie la 4ª riga non c'è errore campionario

Errori campionari/2

Qualunque sia la conclusione raggiunta a mezzo del campione essa è dominata dall'incertezza.

il suo successo può corroborare la validità della procedura per il passato, ma ben poco può aggiungere sulla conoscenza del suo comportamento futuro.

ESEMPIO:
una variabile può assumere tre soli valori: 1,2,3. Per stimare la sua distribuzione di frequenza: per campione si sceglie: prima colonna, ultima colonna e terza riga.



Popolazione dei valori

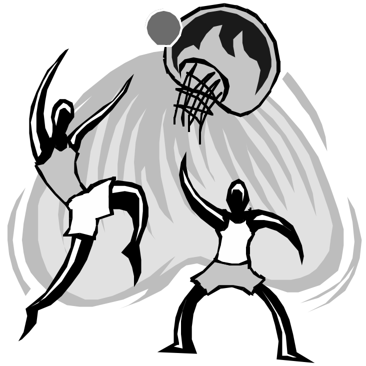
	1	2	3	4	5	6	7	8	8	9	10
1	1	1	1	3	1	1	2	3	1	1	2
2	1	1	1	3	2	3	1	1	2	2	1
3	2	2	2	2	2	2	2	1	1	1	3
4	3	3	2	1	3	2	1	2	1	2	3
5	3	1	3	1	1	2	1	1	1	3	3
6	1	1	1	2	2	1	2	2	1	1	2
7	1	3	2	3	1	1	2	2	2	1	1
8	1	1	1	1	2	2	1	1	2	1	3
9	1	2	3	1	2	2	1	3	3	2	3
10	1	2	2	1	2	1	1	1	3	2	3

Per ciascuno dei campioni si prenderà una decisione sbagliata

La rappresentatività

il campione deve essere RAPPRESENTATIVO della popolazione da cui è estratto, cioè assicurare che i risultati qui ottenuti si estendano a tutta la popolazione. Almeno in relazione alla caratteristica in esame

I giocatori di una squadra di basket non sono rappresentativi della popolazione per l'altezza, ma potrebbero esserlo per capacità di apprendimento.

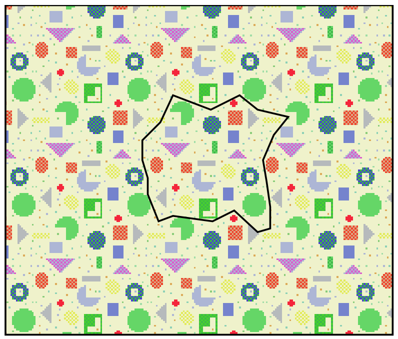


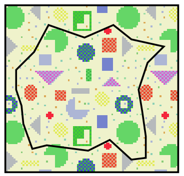
Solo la rilevazione totale è certamente rappresentativa ovvero la selezione di un numero qualsiasi di unità da una popolazione invariabile, ma così è inutile il campione.

La rappresentatività/2

La figlia vuole un vestito con il medesimo disegno di quello della madre. Che campione si dovrà portare al negozio di stoffe?

POPOLAZIONE





Campione

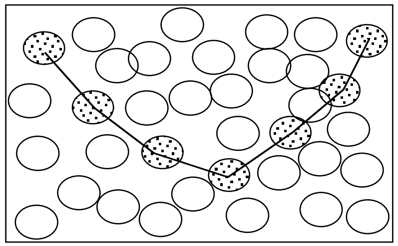
Il campione deve essere abbastanza piccolo per evitare di impacchettare l'intero vestito, ma deve anche essere abbastanza grande da includere il motivo ricorrente della stoffa.

La rappresentatività/3

Ci sono tanti campioni possibili

Alcuni sono rappresentativi cioè coincidono con l'intera popolazione per le variabili che ci interessano

Altri non sono rappresentativi e danno una idea distorta delle variabili di interesse.



Due fattori possono incidere sulla rappresentatività

L'ampiezza del campione "n"

La scelta casuale tra le unità della popolazione

Casualità e campionamento

E' casuale il meccanismo di scelta e non il campione ottenuto.

Perché scegliendo a caso dalla popolazione posso ottenere un campione rappresentativo?

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%.

All'aumentare dell' ampiezza del campione, il meccanismo casuale di scelta porta a campioni che riproducono la popolazione

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

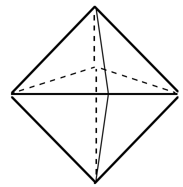
Questo è il postulato empirico del caso

Sorteggio delle unit

il modello della pura sorte può essere simulato in molti modi: monete, dadi.

Ogni processo fisico che nel suo funzionamento segua lo schema del sorteggio tra unità identiche può servire a formare il campione

ESEMPIO: l'ottaedro ha 8 facce uguali a forma di triangolo. Se fatto rotolare su di una superficie liscia finirà col poggiarsi su di una faccia. Se è ben costruito risulta imprevedibile la faccia su cui si poggerà



Accostando le uscite si ottiene il numero casuale in base ottale: 7201 convertibile in base decimale: $7*8^3+2*8^2+0*8+1= 3713$ o tra zero ed uno dividendo per 4095

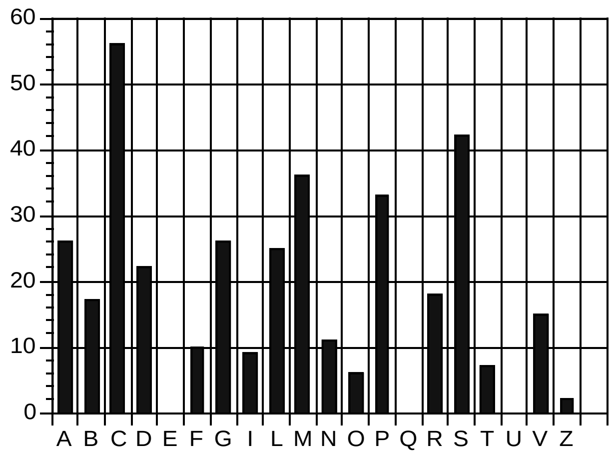
I giochi di sorte sono emblematici: le uscite sono casuali se nessun giocatore -per quanto furbo- riesce a determinare una regola che gli consenta di scommettere meglio che alla pari.

La casualità dei cognomi

Ritenete casuale una selezione di unità che avvenga scegliendo a caso lettera dell'alfabeto italiano per includere tutte le unità che hanno un cognome che inizia con quella lettera?

Lettera	f.a.	f.r
A	26	7.18%
B	17	4.70%
C	56	15.47%
D	22	6.08%
E	0	0.00%
F	10	2.76%
G	26	7.18%
I	9	2.49%
L	25	6.91%
M	36	9.94%
N	11	3.04%
O	6	1.66%
P	33	9.12%
Q	0	0.00%
R	18	4.97%
S	42	11.60%
T	7	1.93%
U	0	0.00%
V	15	4.14%
Z	2	0.55%
	361	

Qual'è la lettera più diffusa per i vostri cognomi?



Estrazione con i numeri pseudo-casuali

Ci sono formule matematiche come i generatori congruenziali lineari, che, per opportune scelte dei parametri producono numeri tra 0 ed (m-1) aventi comportamento simile ai numeri casuali

$$X_{i+1} \equiv (aX_i + c)\text{Mod } m \qquad \text{con} \begin{cases} a = 293 \\ c = 1 \\ m = 1024 \\ X_1 = 68 \end{cases}$$

68	469	202	819	352	737	902	95	188	813	642	715	600	697	446
631	564	389	314	867	80	913	246	399	172	221	242	251	840	361
302	423	36	309	426	915	832	65	614	703	156	653	866	811	56
25	158	215	532	229	538	963	560	241	982	1007	140	61	466	347
296	713	14	7	4	149	650	1011	288	417	326	287	124	493	66
907	536	377	894	823	500	69	762	35	16	593	694	591	108	925

In molti computer si usa $m = 2^{64} = 1.8x10^{19}$

Se se ne usano un miliardo al secondo per finirli ci vorrebbero più di 200 mila anni

Numeri pseudo-casuali

I numeri pseudo-casuali sembrano prodotti dalla sorte, ma sono noti a priori

$$X_k = Resto\left[aX_0 + \frac{(a^k - 1)c}{(a-1)}, m\right] \qquad [\ ]\text{=parte intera}$$

Basta conoscere il primo e tutti gli altri sono noti.

La sequenza è ciclica: dopo "m" valori i numero si ripetono nello stesso ordine

Il periodo "m" deve essere grande rispetto al campione da estrarre: Regola di Ripley

$$m \geq 200n^2 \Rightarrow se \quad n = 10'\,000 \quad m \geq 20'\,000'\,000'\,000 > 2^{31}$$

Campionamento casuale semplice

Un tipico esperimento che rientra nel calcolo combinatorio è la scelta casuale di "n" unità da una popolazione finita di "N".

L'evento elementare è la n-tupla di interi $C_i=(i_1,i_2,\ldots,i_n)$ corrispondenti alle posizioni di una lista univoca ed esaustiva delle unità della popolazione.

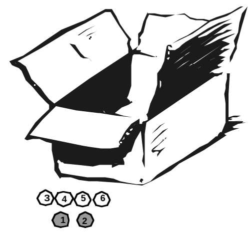
1	2	K	$i-1$	i	$i+1$	K	$n-2$	$n-1$	n
u	u		u	u	u		u	u	u

Se La popolazione è grande si usano tecniche di selezione computistiche.

Per ampiezze più piccole basta una scatola con delle biglie di uguale volume, superficie, temperatura, porosità, colore, lucidatura, etc.

Prima di ogni estrazione la scatola è ben scossa con moto sussultorio e ondulatorio.

Si mette in opera ogni accorgimento per assicurare la equiprobabilità delle biglie nella scatola.



Campione casuale semplice con reimmissione

Scelta di n=3 famiglie in una lista di N=100. Supponiamo che, dopo ogni estrazione, la biglia sia rimessa nell'urna che poi è adeguatamente scossa

La procedura descritta assicura che:

Ognuna delle N famiglie della popolazione ha la stessa probabilità di comparire in una qualsiasi delle n posizioni del campione:

$$P(Fam_i \;\; sia \;\; in \;\; posizione \;\; j) = \frac{100*100}{100*100*100} = \frac{1}{100} \qquad \frac{casi \;\; favorevoli}{casi \;\; possibili}$$

Ne consegue che ogni gruppo di n famiglie ha la stessa probabilità di costituire il campione:

$$P(Fam_{i_1},\ Fam_{i_2},\ Fam_{i_3}) = \frac{1}{100*100*100}$$

Campione casuale in blocco

Dopo ogni estrazione, la biglia NON è rimessa nell'urna.

La probabilità che la famiglia "i" compaia al 1° posto del campione è 1/(99*98)

Fam_i	2°	3°

il 1° posto e bloccato, il 2° può essere occupato dalle 99 restanti famiglie ed il 3° in 98 modi diversi dato che due famiglie impegnano già i primi due posti.

La stessa cosa succede per tutte le altre posizioni

1°	Fam_i	3°
99	1	98

1°	2°	Fam_i
99	98	1

Ognuna delle N famiglie della popolazione ha la stessa probabilità di comparire in una qualsiasi delle n posizioni del campione:

$$P(Fam_i \;\; sia \;\; in \;\; posizione \;\; j) = \frac{99*98}{100*99*98} = \frac{1}{100}$$

Campione casuale in blocco/2

Scelta la prima famiglia su N=100, la seconda è scelta su 99, la terza su 98.

Qualunque famiglia può essere la prima, la seconda o la terza. Ne consegue che:

Ogni gruppo di n famiglie ha la stessa probabilità di costituire il campione:

$$P\left(Fam_{i_1},\ Fam_{i_2},\ Fam_{i_3}\right) = \frac{1}{100 * 99 * 98}$$

E' chiaro che se una famiglia compare in una posizione non può comparire in un'altra dato che la scelta è senza reimmissione.

Ogni scelta, tranne la prima, dipende da quella e da quelle precedenti.

Proprietà del campione casuale

Il campione casuale ha le proprietà seguenti:

1) Ogni unità può comparire in una qualsiasi delle posizioni del campione.

2) Se le unità sono equiprobabili, ogni gruppo di unità ha la stessa probabilità di formare il campione

3) Se le unità sono scelte con reimmisione le singole scelte sono indipendenti

4) Se il campione è piccolo rispetto alla popolazione la differenza dovuta alla reimmissione/non reimmissione diventa trascurabile