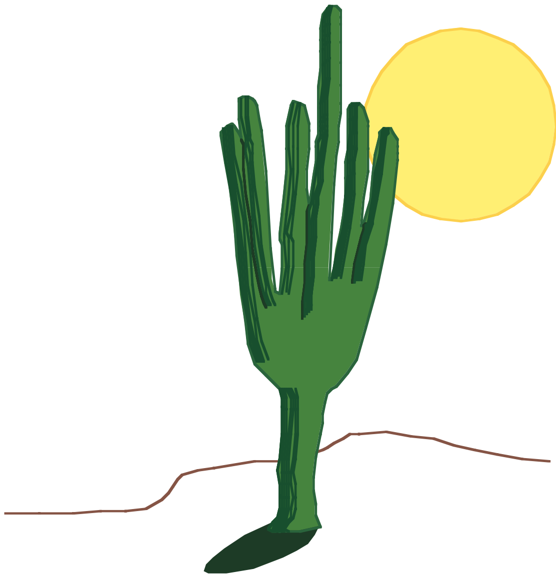
7

Le variabili casuali semplici



Nel capitolo precedente si è privilegiato l'evento e la sua probabilità senza indugiare sulle finalità dell'esperimento e sulle attività connesse alle sue manifestazioni. E' chiaro però che l'esperimento è condotto perché si spera di ricavarne qualche utile indicazione per migliorare lo stato conoscitivo di un problema e si deve perciò stabilire in che modo le conoscenze probabilistiche acquisite possano trasferirsi su uno o più aspetti connessi alla prova. L'esperimento casuale è la sublimazione di una attività che si svolge sotto l'azione della sorte; le variabili casuali di cui ci occupiamo in questo capitolo ne sono aspetti circoscritti che spesso sono comuni a più esperimenti e talvolta vivono al di fuori degli esperimenti. Sono cioè dei modelli da adattare al fenomeno per descriverne e comprenderne il comportamento nel presupposto che le forme disponibili siano abbastanza flessibili ed utili per interpretare i fenomeni reali.

In base alla funzione di insieme dello spazio di probabilità (S,P,W) dell'esperimento si ricava la distribuzione di probabilità della variabile casuale che, simile alla distribuzione di frequenza, verrà sintetizzata con gli indici descrittivi introdotti nel capitolo 3°. Infatti, le variabili casuali riprendono le distribuzioni di frequenza discusse nel capitolo 2° proponendo per loro schemi di studio semplificati ed astratti, in grado di dare risposte non solo in base alla particolare indagine che si effettua, ma mobilitando, informazioni a priori, esperienze precedenti ed il calcolo delle probabilità. Nel primo paragrafo discuteremo le variabili casuali discrete e finite essenzialmente legate ad un numero limitato di prove binarie ripetute (modello binomiale e ipergeometrico). Nel secondo paragrafo verranno rafforzati i postulati di Kolmogorov per gestire aspetti sperimentali enumerabilmente infiniti (modello di Poisson e Pascal). Infine, nel terzo, dopo aver chiarito la tipologia degli eventi di interesse, si tratteranno le variabili casuali continue ed in particolare il modello Normale.

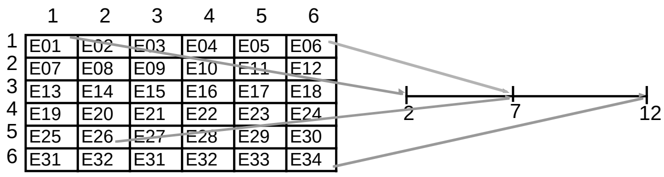
Il livello di astrazione del capitolo è ancora più elevato del precedente, ma si noterà la convergenza verso quanto è stato esposto nella prima parte del testo.

7.1 Dalla probabilità alle variabili casuali

Il verificarsi di un evento non può esser registrato e poi dimenticato, ma deve innescare l'acquisizione e/o l'elaborazione di specifiche informazioni. Non ci si può limitare alla mera esecuzione dell'esperimento o a monitorarne gli eventi, ma si deve tener conto delle loro conseguenze su qualche aspetto osservabile del problema cioè su di una o più variabili.

Esempi:

a) Gioco con il lancio di due dadi in cui vince chi indovina la somma dei due punteggi. Su quale esito giochereste? L'universo degli eventi è formato dalle 6x6=36 coppie di esiti: S={(1,1); (1,2); … (5,6);(6,6)}. Applichiamo agli eventi elementari il modello di equiprobabilità assegnando ad ogni coppia probabilità 1/36. L'esito del lancio non fornisce direttamente il risultato per la scommessa, ma è necessaria una mediazione che da S arrivi a T={2, 3, …,12} cioè l'insieme dei possibili valori della somma dei due dadi.



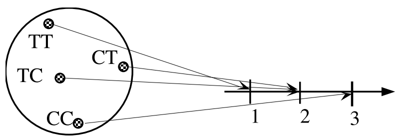
Ci sono 6 eventi elementari che generano il "7" ; gli altri valori si ottengono con un numero inferiore. Addirittura il "2" ed il "12" sono associabili ad un solo evento elementare: (1,1) e (6,6). La scelta quindi dovrebbero ricadere sul "7" in quanto corrispondente ad un evento composto {(1,6); (6,1); (2,5); (5,2); (3,4); (4,3)} che, presupponendo l'equiprobabilità, ha maggiori *chances*.

b) Indichiamo con X il valore della somma dei due dadi e proviamo ad ampliare l'idea di corrispondenza tra eventi composti in W (l'algebra degli eventi generata da S) ed il dominio T delle modalità della variabile.

x=2	(1,1)	1/36
x=3	(1,2) (2,1)	2/36
X=4	(1,3) (2,2) (3,1)	3/36
x=5	(1,4) (2,3) (3,2) (4,1)	4/36
x=6	(1,5) (2,4) (3,3) (4,2) (5,1)	5/36
x=7	(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	6/36
x=8	(2,6) (3,5) (4,4) (5,3) (6,2)	5/36
x=9	(3,6) (4,5) (5,4) (6,3)	4/36
x=10	(4,6) (5,5) (6,4)	3/36
x=11	(5,6) (6,5)	2/36
x=12	(6,6)	1/36

Ad ogni valore della X in T corrisponde un evento in W; tuttavia, la corrispondenza non è biunivoca, dato che ad esempio il "5" si può ottenere in quattro modi diversi.

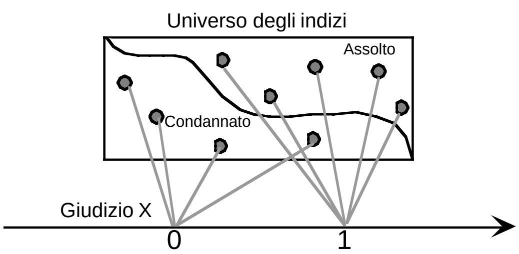
c) Romilda vorrebbe avvicinarlo, ma teme di sbagliare facendo lei il primo passo. D'altra parte quello se ne sta fermo che sembra una statua di sale e Romilda è stanca di aspettare. Si dà comunque una strategia effettuando il seguente "esperimento": lancia due monete e quindi opera con l'universo degli eventi S={(T,T); (T,C); (C,T); (C,C)}; le alternative sono TT=lascio perdere; CC=gli chiedo cosa fa stasera; TC e CT=rilancio le monete. La decisione è quindi una variabile nominale che possiamo così riassumere X=1 (abbandono), X=2 (invito) e X=3 (rilancio). Anche in questo caso si forma una corrispondenza tra S e T={1,2,3}.



X=1	(T,T)	1/4
X=2	(T,C) (C,T)	2/4
X=3	(C,C)	1/4

La strategia di Romilda è una variabile qualitativa, ma che conviene esprimere con dei numeri per uniformare il trattamento con le variabili metriche tenendo ovviamente conto del fatto che, nel caso di variabili nominali, il carattere numerico è solo apparente e, nelle quantitative non metriche, rileva solo per la relazione d'ordine tra le modalità. Da notare che la variabile casuale è la regola di corrispondenza tra modalità e probabilità e non solo la probabilità assegnata ad una data modalità.

d) Un imputato è sotto giudizio. Alcuni indizi portano alla sua assoluzione ed altri alla sua condanna. Se X è l'esito del giudizio poniamo X=1 se "assolto" e X=0 se "condannato". Poiché la valutazione delle prove è in parte soggetta alla sorte la X non assume un valore predeterminato, ma può assumere tutti quelli del suo dominio con probabilità. L'idea è schematizzata nella figura seguente:



La variabile casuale è un concetto introdotto per stabilire ed analizzare i legami tra gli eventi dell'algebra descritti come degli insiemi e le conseguenze delle loro manifestazioni espresse come intervalli di numeri reali.

Esercizio_VC01*: un esperimento teso a saggiare la praticabilità del green di una buca da golf prevede il tiro di 4 palline. L'esito dell'esperimento è incerto in quanto, oltre alla capacità variabile di chi gioca, sono presenti fattori climatici e relativi all'erba del tutto imprevedibili, ma che riteniamo portino alla probabilità del 70% di mandare la pallina in buca. Se indichiamo con Y= "numero di palline in buca" qual'è lo spazio di probabilità dell'esperimento con le palline? Quali probabilità sono associate ai valori della Y?*

7.1.1 Definizione di variabile casuale discreta e finita

In un esperimento casuale è possibile individuare uno o più aspetti -qualitativi o qualitativi- ricondotti, per esigenze di abbreviazione ed elaborazione, ad insiemi di numeri reali. Agli eventi elementari si associano dei numeri che valorizzano le variabili legate alla prova. Ad esempio, nel lancio dei due dadi l'esito induce un valore per la somma delle due facce superiori, ma anche per la differenza in valore assoluto dei due punteggi o per il loro rapporto. In questo capitolo, tuttavia, tratteremo solo il caso univariato.

Variabili ed esperimenti casuali

Sia (S, W, P) lo spazio di probabilità dove S è un universo finito e discreto di eventi elementari, W è l'algebra di tutti i possibili sottoinsiemi di S e P(.) è una funzione di insieme che verifica gli assiomi di Kolmogorov. La variabile casuale è una funzione che assegna un numero reale ad ogni esito di un esperimento casuale. Più formalmente X(e) è un'applicazione reale con dominio S ed a valori in $T=\{x_1, x_2,\ldots,x_n\}\subset R$ tale che

$$X(e): S \to T \subset R, \quad T = \{x | x = X(e) \;\; per \;\; ogni \;\; e \in S\}$$

Una comoda scorciatoia per indicare che l'esperimento ha avuto determinazioni che inducono il valore "a" nella variabile X è: X=a che equivale all'evento composto $\{e\in S|\, X(e)=a\}$. E' ovvio che i valori da collegare agli eventi non sono scelti arbitrariamente, ma discendono in via naturale dagli aspetti che si vogliono considerare e dalla loro definizione operativa. Il fatto che talvolta le modalità siano caratteristiche nominali ovvero siano quantitative, ma non metriche non è in contrasto con la definizione poiché la X può essere interpretata come un indice e, ad esempio, l'insieme $X(e)\leq x$ può essere letto come "considera tutte le modalità il cui indice sia inferiore o uguale ad x".

Esempi:

a) I valori di X in T sono esprimibili in forma di intervallo del tipo

$$E(a,b] = \{x \in T \subset R | a < x \leq b\}$$

La natura discreta e finita del codominio qui considerata comporta che $E(x_i,x_{i+1}]=\{x_i\}$, con $card\{E(x_i,x_{i+1}]\}=1$. Sono inoltre evidenti le seguenti relazioni insiemistiche: $E(x,+\infty)=E(x,+\infty]=S$, $E(x_i,-\infty]=\phi$, $E(-\infty,x_{i+1}]= E(-\infty,x_i]\cup E(x_i,x_{i+1}]$.

b) Il fatto che la funzione X(e) sia reale non è una vera e propria limitazione dato che, in caso di fenomeni rilevati nel campo complesso, si possono utilizzare due variabili casuali: una per la parte reale ed un'altra per la parte immaginaria.

c) Si intende conoscere il numero X di difetti in tre aspetti rilevanti di una scheda logica scelta casualmente. L'universo degli eventi è formato dalle terne S={(000),(001),(010),(011),(100),(101),(110),(111)} dove "0_i" indica che se la l'aspetto i-esimo è accettato e 1_i se difettoso. Il numero di "1" nella terna valorizzerà la variabile X e quindi i valori di questa ricadono nell'intervallo discreto:T=[0,3]; ad esempio, X=2 è la modalità che si produce per gli eventi elementari: (011), (110), (101). Ad ogni elemento dell'universo S corrisponde una ed una sola modalità della variabile X ed ogni modalità della X ha uno o più punti immagine in S.

La variabile X è un congegno automatico che scatta non appena si realizza una manifestazione dell'esperimento. Poiché è casuale la determinazione sperimentale, sarà casuale il valore della X (si aggiunge il termine casuale proprio per sottolineare che il suo valore non è predeterminabile così come non lo è l'esito dell'esperimento a cui è legata). La variabile casuale è quindi una funzione anche se si preferisce il termine "variabile" in quanto, dovendo poi inserire la X in un'altra funzione g(X), si parlerebbe spesso della "funzione di funzione" g[X(e)] con una circolarità che potrebbe ostacolare il flusso espositivo e l'apprendimento.

Esempi:

a) In alcune prove gli eventi elementari sono essi stessi dei numeri reali. Ad esempio la scelta di una cifra nella tabella dei numeri casuali ha come universo degli eventi: S={0, 1, 2, …, 8, 9}. In questi casi X=X(e) è la funzione identità e si porrà X=e per e=0, 1, …,9

b) In un elenco di prenotazioni di un grande albergo se ne scelgono 4 -con reimmissione- per accertare se il cliente sia nuovo oppure è già stato ospite. L'universo degli eventi è: S={(n,v)⊗(n,v)⊗(n,v)⊗(n,v)} che contempla 16 alternative. Se l'attenzione è diretta verso il conteggio dei clienti nuovi la variabile da analizzare è X= "numero di nuovi nei quattro prescelti" ad esempio: X[(n,n,n,v)]=3, X[(v,n,v,v)]=1, X[(v,v,v,v)]=0. L'insieme dei valori della variabili casuale è T={0,1,2, 3, 4). Tutti i possibili eventi composti che coinvolgono questi numeri sono esprimibili come intervalli di interi.

Nel passare dall'esperimento alla variabile casuale l'attenzione si sposta dalla probabilità di eventi composti rientranti nell'algebra W alla probabilità di intervalli di valori in R: nel lancio dei due dadi si passa dall'algebra formata a partire dagli eventi elementari S={(1,1); … ; (6,6)} e si arriva all'intervallo di reali T={2≤x≤12, x intero} e, di fatto, come esito dell'esperimento è considerato il valore $x\in T$ e non l'evento elementare $e\in S$. Da questa angolatura gli elementi di T sono "eventi" del tutto legittimati ad avere assegnata una probabilità, ma non probabilità qualsiasi, bensì quelle compatibili con la funzione di insieme dello spazio originario di probabilità. Le maggiori difficoltà che incontreremo derivano dal fatto che la probabilità è definita per gli eventi composti dell'algebra W mentre la variabile casuale è definita per gli esiti dell'universo degli eventi S. Dovremo mettere in comunicazione questi due mondi.

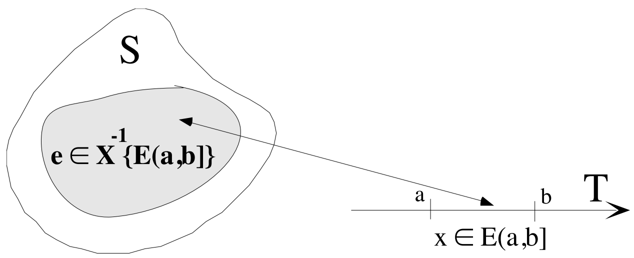
Esercizio_VC02*: un'agenzia ha tre sportelli. Nel momento in cui il cliente-tipo entra in agenzia, ciascuno degli sportelli - indipendentemente dagli altri- ha probabilità 0.50 di essere libero, 0.05 di essere chiuso e 0.45 di essere occupato. L'aspetto che interessa è Y= "numero di sportelli liberi per il cliente tipo che entra in agenzia".*

a) Descrivete il passaggio dallo spazio di probabilità alla variabile casuale;

b) Che caratteristiche hanno le probabilità assegnate ai valori della variabile casuale?

La distribuzione di probabilità

Definita la variabile casuale X dobbiamo dare significato alle asserzioni probabilistiche che la riguardano. Ragioniamo su di un intervallo discreto e finito di valori della variabile casuale X=X(e), diciamo $E(a,b]\subset T$. L'evento costituito con gli eventi elementari di S mandati in E(a,b] dalla X è indicato con $X^{-1}\{E(a,b]\}$



L'insieme E(a,b] in T e la sua controimmagine $X^{-1}\{E(a,b]\}$sono equivalenti dato che ogni elemento del secondo è abbinabile -univocamente- ad un solo elemento del primo: l'uno si verifica quando si verifica l'altro e viceversa. Pertanto, $X^{-1}\{E(a,b])\}$ è un evento dell'algebra W al quale è stata attribuita una probabilità attraverso la funzione di insieme P(.) e risulta quindi logico, tenuto conto della equivalenza, assegnare all'intervallo $E(a,b]\subset T$ la stessa probabilità di cui è dotato $X^{-1}\{E(a,b]\}$ cioè: $P\{E(a,b]\}=P\{X^{-1}\{E(a,b]\}$. Le probabilità così attribuite agli intervalli della X costituiscono la funzione di distribuzione della probabilità:

$$p(x)=\begin{cases} p(X=x_i) & per \quad x\in T=\{x_1,x_2,\dots x_n\} \\ 0 & altrove \end{cases}$$

con cui si esprime la probabilità che la variabile casuale X sia uguale al valore " x" per una data determinazione dell'esperimento casuale. La distribuzione di probabilità è diversa dalla funzione di insieme in quanto p(x) è una funzione puntuale definita sull'intervallo di interi T laddove P(.) è una funzione di insieme definita sull'algebra W dell'esperimento.

Esempio:
Riprendiamo il caso dei clienti dell'albergo ed ipotizziamo una probabilità del 75% per lo *status* di cliente nuovo. Qual'è la probabilità che X>2 cioè qual'è la probabilità da attribuire all'evento E[2,4]={3,4]?

$$(X>2)=E(2,4]\rightarrow\{(n,n,n,v);(n,n,v,n);(n,v,n,n);(v,n,n,n);(n,n,n,n)\}$$
$$p(X>2)=P\{E(2,4]\}=4*\left(0.75^3\right)0.25+0.75^4=0.7383;$$

Calcoliamo anche le probabilità dei singoli valori della X tenendo conto del numero di eventi elementari che li determinano.

$$p(0)=p(X=0)=P\{E(0,0]\}=\binom{4}{0}0.25^4=0.0039;\;\; P(X=1)=P\{E(1,1]\}=\binom{4}{1}\left(0.25^3\right)0.75=0.0469;$$
$$p(2)=p(X=2)=P\{E(2,2]\}=\binom{4}{2}\left(0.25^2\right)0.75^2=0.2109;\;\; P(X=3)=P\{E(3,3]\}=\binom{4}{3}0.25\left(0.75^3\right)=0.4219;$$
$$p(4)=p(X=4)=P\{E(4,4]\}=\binom{4}{4}\left(0.75^4\right)=0.3164;$$

Si può notare che:

$$\sum_{i=0}^{4}p(i)=\sum_{i=0}^{4}P\{E(i,i]\}=P\left\{\bigcup_{i=0}^{4}E(i,i]\right\}=P(S)=0.0039+0.0469+0.2109+0.4219+0.3164=1$$

Se il meccanismo di conversione si estende in modo che tutti gli elementi in S risultino collegati univocamente ad una modalità della X, ovvero ogni sua modalità scaturisce da un evento composto in W (non sempre è possibile dare questa garanzia), allora la funzione di probabilità sarà dotata di particolari caratteristiche. Poniamo:

$$\left(x_i,x_i\right]=\left\{x\in T\middle|x_i<x\leq x_i\right\}\;\; per\;\; i=1,2,\ldots,n\;\;\Rightarrow E_i\cap E_j=\varnothing$$

ed ipotizziamo che la valorizzazione della X avvenga in modo che:

$$X=x_i\;\;\Leftrightarrow\;\; e\in X^{-1}\left(E_i\right),\;\; e\in\bigcup_{i=1}^{n}X^{-1}\left(E_i\right)=S,\;\; X^{-1}\left(E_i\right)\cap X^{-1}\left(E_i\right)=\varnothing$$

Semplici argomenti portano alle relazioni:

$$p\left(x_i\right)=p\left(X=x_i\right)=P\left[e\in X^{-1}\left(E_i\right)\right]\geq 0\quad per\;\; i=1,2,\ldots,n$$
$$\sum_{i=1}^{n}p\left(x_i\right)=\sum_{i=1}^{n}p\left(X=x_i\right)=\sum_{i=1}^{n}P\left[e\in X^{-1}\left(E_i\right)\right]=P\left\{\bigcup_{i=1}^{n}\left[e\in X^{-1}\left(E_i\right)\right]\right\}=1$$

La distribuzione di probabilità eredita la non negatività e la somma unitaria della funzione di insieme P(.)

$$0\leq p(x)\leq 1;\qquad \sum_{x\in T}p(x)=1$$

Il senso della discussione è il seguente: se l'obiettivo dell'esperimento è lo studio della variabile casuale perché non concentrarsi direttamente sulla sua funzione di distribuzione (X, p) che ha lo stesso contenuto informativo invece di coinvolgere l'artefatto intermedio (S,W, P)? Diciamo che i fenomeni soggetti alla sorte non nascono con il loro bravo modello, ma occorre decifrarlo, se possibile, in un opportuno spazio di probabilità. Dopo che ciò è avvenuto e per tutti i casi riconosciuti analoghi si può adottare direttamente il modello di distribuzione.

Esempi:
a) Tre monete equilibrate sono lanciate simultaneamente e separatamente. L'universo degli eventi è S={E_1=(T,T,T); E_2=(T,T,C); E_3=(T,C,T); E_4=(C,T,T); E_5=(T,C,C); E_6=(C,T,C); E_7=(C,C,T); E_8=(C,C,C)}. La variabile casuale che interessa è X="numero di teste". Per i lanci vigono le condizioni di indipendenza e di equiprobabilità per cui P(E_i)=1/8 per i=1,2,…,8. Le probabilità della X sono riportate in tabella. Il meccanismo della variabile casuale può funzionare allo stesso modo in situazioni diverse. Nel lancio delle monete le stesse probabilità si danno a Y= "numero di croci" per cui X ed Y hanno le stesse modalità e probabilità, ma sono distinte perché afferenti ad aspetti diversi di un esperimento.

$$P(X=0)\Leftrightarrow P\left(E_8\right)=1/8$$
$$P(X=1)\Leftrightarrow P\left(E_5\cup E_6\cup E_7\right)=3/8$$
$$P(X=2)\Leftrightarrow P\left(E_2\cup E_3\cup E_4\right)=3/8$$
$$P(X=3)\Leftrightarrow P\left(E_1\right)=1/8$$

X	p(x)
0	1/8
1	3/8
2	3/8
3	1/8
	1

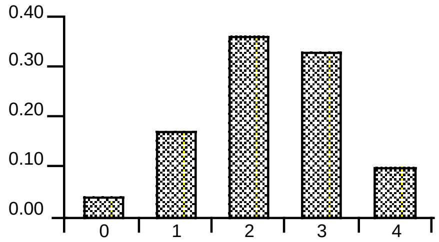
b) Un'indagine campionaria propone due domande ai clienti dei negozi di alimentari: 1) Gli ipermercati determinano riduzioni di prezzo; 2) La piccola distribuzione favorisce l'elusione fiscale. L'esperimento è l'intervista; la variabile casuale di interesse è il numero di sì: YεT={0,1, 2}. Prima di effettuare l'esperimento abbiamo la sensazione che i soggetti intervistati diano risposte secondo lo schema:

y	0	1	2	
p(y)	0.04	0.41	0.55	1.00

Saranno poi i risultati effettivi delle interviste a confermare o sconfermare la nostra ipotesi.

c) Un quiz a risposta multipla comprendente 4 opzioni è strutturato in modo che un numero qualsiasi (anche nessuna) delle opzioni possa essere quella corretta. Ecco un modello e la relativa rappresentazione con ortogramma a colonne.

X	p(x)
0	0.04
1	0.17
2	0.36
3	0.33
4	0.10
	1.00



La distribuzione di probabilità nasce da varie considerazioni ed esiste certamente un esperimento in cui lo schema proposto può essere motivato. Non è però necessario presentarlo per la quasi totalità di analisi riguardanti le variabili casuali.

Esercizio_VC03: *scelta di una coppia di lettere equiprobabili. Sia X =numero di vocali nella coppia.*

ab	*aj*	*bg*	*eg*	*gj*
ae	*ak*	*bj*	*ej*	*gk*
ag	*be*	*bk*	*ek*	*jk*

Costruite la sua distribuzione di probabilità e datene la rappresentazione grafica.

Esercizio_VC04: *la concessionarie di automobili "InStrada" ha come modello di vendite giornaliere la tabella:*

X	0	1	2	3	4	5	6	7	
p(x)	42	102	155	38	29	7	5	2	380

In che modo e in che misura la si può considerare una variabile casuale?

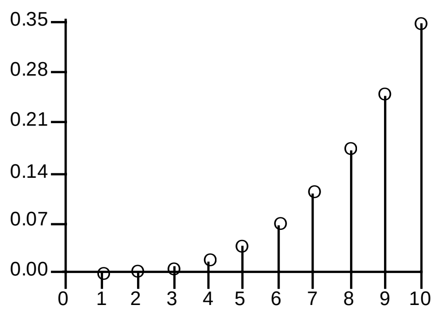
Proprietà della distribuzione delle probabilità

E' la cinghia di trasmissione tra i valori (anche interpretabili come categorie indicizzate) della variabile casuale X e le probabilità del loro verificarsi: p(X=x). L'esistenza dell'esperimento e del relativo spazio di probabilità è però in genere sottinteso e non se ne fa menzione esplicita se non è strettamente necessario; diventa uno scenario appena abbozzato di cui si trascura la genesi per rivolgersi alle quantità ed alle qualità rilevate sulle sue manifestazioni. Naturalmente la distribuzione di probabilità scaturisce dalla funzione di insieme P(.), ma non sempre si ha possibilità o motivo di specificarla (ad esempio, da una stessa P(.) possono nascere diverse p(.) di cui forse ne interessa una sola). E' quindi più comodo indicare le conseguenze della funzione di insieme su di una distribuzione di probabilità e dare i dettagli su quest'ultima piuttosto che assegnare le probabilità agli eventi in W per solo dopo trattare con le sue derivazioni.

Esempi:

a)Il numero di ammanchi accertati in un supermercato con pochi controlli ha come modalità i naturali [1, n]. Invece di gestire la prova con un apparato complesso quale lo spazio di probabilità è più facile escogitare una espressione algebrica che dia le probabilità della variabile casuale recependo ciò che è noto del fenomeno; ad esempio, che le probabilità aumentino con l'aumentare della modalità:

$$p(x)=\frac{(2x-1)^3}{n^2\left(2n^2-1\right)};\quad x=1,2,\ldots,10$$



Il grafico della distribuzione di probabilità mostra la netta asimmetria negativa dovuta alla mancanza di freni che agevola il moltiplicarsi del fenomeno (l'interpretazione è analoga a quanto visto nelle distribuzioni di frequenza).

b) Se il numero delle modalità della variabile casuale non è elevato, la p(.) può essere realizzata elencando i valori con le associate probabilità. Supponiamo che una società armatoriale proponga ad una agenzia di collocare una nuova crociera e che l'agenzia accetti di offrirla sperimentalmente a 4 clienti. In tabella sono riportate le probabilità di accettazione secondo l'armatore:

x	0	1	2	3	4	
p(x)	0.005	0.075	0.22	0.45	0.25	1

cioè la nuova crociera sarà scelta da due clienti (X=2) con probabilità del 22%; sarà rifiutata da tutti (X=0) con 5 *chances* su mille dato che p(0)=0.005..

Esercizio_VC05: *in un'impresa con il 70% di donne si scelgono a caso e con riposizione due persone per una commissione. Determinare la distribuzione di probabilità di X="numero di donne nella commissione".*

La funzione di distribuzione permette di calcolare la probabilità di ogni sottoinsieme di valori del dominio T.

$$
\begin{aligned}
p(a < X \le b) &= P\{E(a,b]\} = P\{(a,b) \cup (b,b]\} = P\{(a,b)\} + P\{(b,b]\} = p(a < X < b) + p(X = b)\\
p(a \le X < b) &= P\{E[a,b)\} = P\{(a,b) \cup (a,a]\}\\
&= P\{(a,b)\} + P\{(a,a]\} = p(a < X < b) + p(X = a)\\
p(a \le X \le b) &= P\{E[a,b]\} = P\{(a,b) \cup (a,a] \cup (b,b]\}\\
&= P\{(a,b)\} + P\{(a,a]\} + P\{(b,b]\} = p(a < X < b) + p(X = a) + p(X = b)
\end{aligned}
$$

Si conviene inoltre di assegnare probabilità zero ai valori estranei al dominio trattandoli come immagine dell'evento impossibile intendendo che la funzione di distribuzione p(.) è comunque definita su tutto l'asse reale anche se la probabilità è positiva solo in alcuni punti specifici.

Esempio:
Nel lancio dei due dadi definiamo X= "differenza in valore assoluto tra i due risultati". La distribuzione di probabilità è la seguente:

X	0	1	2	3	4	5	
p(x)	6/36	10/36	8/36	6/36	4/36	2/36	1

Calcoliamo la probabilità:

1. Che la differenza sia pari, ma non zero; 2. Che sia ≥3; 3. Che sia 1≤X≤4; 4. Che sia inferiore a 1.5

1. $p(X \text{ è pari e } X \neq 0) = p(X = 2 \;\; o \;\; X = 4) = \frac{10}{36} + \frac{4}{36} = \frac{14}{36};$ 3. $p(1 \le X < 4) = p(X = 1 \; o \; X = 2 \; o \; X = 3) = \frac{24}{36};$

2. $p(X \ge 3) = p(X = 3 \;\; o \;\; X = 4 \;\; o \;\; X = 5) = \frac{6}{36} + \frac{4}{36} + \frac{2}{36} = \frac{1}{3};$ 4. $p(X < 1.5) = p(X = 0 \;\; o \;\; X = 1) = \frac{6}{36} + \frac{10}{36} = \frac{16}{36};$

Esercizio_VC06: *la società "Multipol" stabilisce le polizze di responsabilità civile in base alla distribuzione di probabilità di sinistri dell'assicurato secondo i livelli X indicati nella seguente tabella:*

x	0	1	2	3	4	5	
p(x)	45	36	12	4	2	1	100

Calcolare p(X≤4), p(1≤X<4), p(X>0).

Esercizio_VC07: *la variabile casuale X ha distribuzione di probabilità p(x)=x/10 per x=1, 2, 3, 4. Calcolare: a) p(X ≤1); b) p(1≤X ≤3); c) p(X≥3); d) p(1<X<4.5); e) p(X≤0); f)p(X>1.33).*

La funzione di distribuzione, come si è visto negli esempi precedenti, può essere data come tabella oppure costruita algebricamente cioè come una funzione matematica.

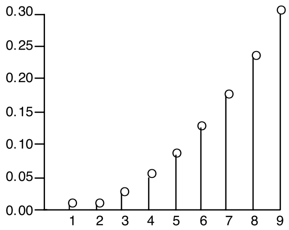
Esempi:
a) Una candidata, in attesa del colloquio per l'ammissione in una *équipe* tecnica ben pagata, ha studiato le commissioni selezionatrici. In base a come si sono susseguite le chiamate ipotizza la seguente tabella di probabilità per la commissione con cui sosterrà l'esame:

Commissione	C1	C2	C3	C4
P(Esame)	0.25	0.40	0.15	0.20

La candidata ha un modello in cui ha maggiori *chances* di essere esaminata dalla commissione C2.

b) Se si ritiene che la probabilità di doversi recare dal medico -nel corso di un mese- per un'adulto aumenti con il numero delle visite, una funzione di probabilità adatta allo scopo è:

$$p(x)=\frac{3(2x-1)^2}{n\left(4n^2-1\right)}\quad per\quad x=1,2,\ldots,n$$



le probabilità sono positive con somma unitaria e riassunte in una espressione analitica compatta.

c) *I Ching* è un antichissimo testo cinese di divinazione. La risposta si basa su esagrammi cioè blocchi di sei linee. Le linee possono essere intere oppure spezzate; possono inoltre essere mobili oppure stabili (senza cerchio al centro). Per ogni domanda si produce uno dei 64 esagrammi possibili e poiché ognuno di questi si può tramutare in qualsiasi altro, gli oracoli possibili sono 64x64=4'096. Per accedere al responso si possono utilizzare due metodi: gli steli di millefoglie oppure le tre monete.

Tipo	Simbolo	Steli	Monete
Linea intera mobile	—o—	1/16	1/8
Linea spezzata stabile	— —	5/16	3/8
Linea intera stabile	——	7/16	3/8
Linea spezzata mobile	— o —	3/16	1/8

La sorte interviene comunque nella scelta, ma in modo diverso a seconda del metodo. In tabella sono presentati i due modelli di probabilità. Chi crede in questo culto ritiene che la sorte sia in grado sintonizzarsi sulle forze che muovono l'universo usando o l'uno o l'altro dei metodi, ed anzi la scelta del metodo è già un entrare in contatto con quelle forze.

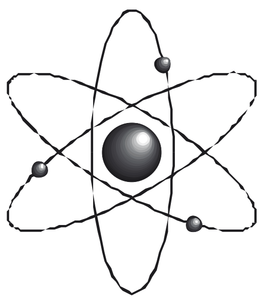
Esercizio_VC08: *verificare che il modello:*

$$p(x)=\begin{cases}\dfrac{(2x-1)^3}{n^2\left(2n^2-1\right)} & per\quad x=1,2,\ldots,n\\ 0 & altrove\end{cases}$$

sia in effetti una distribuzione di probabilità.

Dei modelli

Per proporre una particolare funzione di probabilità si deve dimostrare che certe sue caratteristiche siano consone all'aspetto dell'esperimento analizzato e la sua validità deve aver conferma nella realtà. La distribuzione di probabilità p(x) è una rappresentazione semplificata ed astratta dell'aspetto sperimentale. E' astratta perché non esiste in realtà; non è una rappresentazione in scala come il mappamondo; non esiste in forma analogica come le onde sullo schermo di un modulatore di frequenza; non è proposta in forma fisica come la rappresentazione dell'atomo con il nucleo composto da neutroni e protoni al centro e gli elettroni che gli orbitano attorno (modello peraltro in via di superamento con la teoria delle stringhe);



E' semplificata perché in essa non confluisce tutto ciò che conosciamo o apprendiamo dai dati, ma solo ciò che è rilevante, almeno dal nostro punto di vista. I modelli servono per dare corpo alle congetture suggerite dalle conoscenze acquisite sul fenomeno. Gnedenko (1989, p. 18) evidenzia "... *Nello studio dei fenomeni naturali è d'obbligo trascurare i dettagli non essenziali. La considerazione dei particolari di tutte le relazioni, incluse quelle non pertinenti al fenomeno in esame, non può che portare ad un risultato: l'oscuramento del fenomeno e ne è ritardata la comprensione a cause di complicazioni artificiose".*

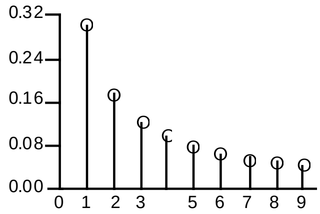
Esempi:

a) La sig.ra Simmons, zia di Harvey nell'omonima commedia e film, dice: *"Quest'inverno ho seguito un corso ed ho imparato la differenza tra una bella pittura ed un lavoro meccanico tipo una fotografia. La fotografia mostra solo la realtà, la pittura invece, oltre la realtà, mostra il sogno che si cela dietro di essa. Sono i sogni che ci sostengono, quelli che ci distinguono dalle bestie".*

b) Tutti i modelli sono sbagliati, ma alcuni di essi sono anche utili (G.E.P. Box).

c) Knuth (1981, pp. 240-241) discute la distribuzione di probabilità della prima cifra significativa nella mantissa "m" di un numero X scritto in notazione scientifica : $X=m*10^a$ dove "m" è un numero razionale tra zero ed uno ed "a" è la potenza da dare al 10 per ottenere X. L'astronomo Newcomb nel 1881 ipotizzò che il modello di probabilità fosse:

$$p(x) = Log_{10}\left(\frac{x+1}{x}\right); \quad x = 1, 2, \ldots, 9$$



Il grafico mostra la diminuzione della probabilità all'aumentare della cifra. Questo modello 8anche noto come legge di Benford) è corroborato da molte esperienze (cfr. Ley, 1996) ed altre potete aggiungerle considerando ad esempio la lunghezza dei fiumi d'Italia o l'altezza delle montagne per cime sopra i 2'500 metri. Come si spiega la prevalenza della cifra "1"?

d) ... *tuttavia, il nostro processo di astrazione non sarà fine a se stesso, e pur partendo dai materiali più semplici non si allontanerà tanto dalla realtà concreta da non consentire di illuminare delle sue conclusioni anche le proprietà dei materiali più complessi. Il processo adatto al nostro scopo è quello che passa attraverso la costruzione di "modelli". E' ad essi che é affidata la mediazione tra l'astrazione teorica e la realtà sperimentale*. G. Segré (1981, pp. 2-3).

La possibilità di rappresentare una situazione reale con un modello nelle indagini statistiche è usuale, ma controversa perché i fenomeni si comportano come credono e nulla sanno dei nostri tentativi di ingabbiarli in schemi precostituiti. Tuttavia, è un fatto che molte rilevazioni sono descritte con accuratezza soddisfacente dalle funzioni discusse in questo capitolo. I modelli sono del tutto generali, svincolati da applicazioni specifiche ed indicano la via naturale per la generalizzazione dei risultati. Il coinvolgimento nel decidere una particolare questione non è essenziale per il modello che vive indipendentemente dalla sua utilizzazione: l'onere della prova della validità non è del modello, ma di chi lo usa. *"Una cosa è costruire un modello, ben altra cosa è indicare le regole per scegliere i fatti a cui quel modello può essere applicato"* (T. Haavelmo). Il modello, perciò, vale fino a che le conclusioni ottenute hanno un impatto sulla realtà. Se questo manca, il modello deve essere cambiato, a meno che -per magia o per gioco- non sia possibile cambiare la realtà.

Modelli e parametri

I modelli sono in genere dotati di uno o più parametri -spesso corrispondenti a costanti caratteristiche del fenomeno- che consentono loro di adattarsi a situazioni diverse pur mantenendo invariate alcuni tratti generali. Il problema dell'indagine statistica non è più allora l'analisi della distribuzione di probabilità, ma l'individuazione dei parametri che definiscono il modello appropriato per i dati già acquisiti oppure quello più plausibile per quelli di cui sono noti solo alcuni comportamenti base.

Esempio:

Modello a probabilità crescenti: $$p(x;d) = \frac{(dx-1)^2}{\left(\frac{n}{6}\right)\left[d^2(n+1)(2n+1) - 6d(n+1) + 6\right]}; \quad x = 1, 2, \ldots, n; \; \text{"}d\text{"} \; parametro$$

Si deve trasmettere il numero di reti segnate per partite di calcio di serie A cioè S={0, 1, 2, 3,…,13,14,≥14) quindi 30 informazioni. Se si scoprisse che le frequenze sono ben modellate dalla p(x;d) basterebbe trasmettere il solo parametro "s" oltre alla formula.

Esercizio_VC09: *una moneta con probabilità "p" di produrre croce è lanciata per 2 volte. Qual'è il modello di probabilità della variabile "facce uguali/facce diverse"? Qual'è il valore minimo della probabilità?*

Esercizio_VC10: *si fa rotolare un dado regolare quattro volte e si considera la faccia rivolta verso l'alto. Se X è la variabile casuale "numero di volte che il pari precede il dispari" qual'è la sua distribuzione di probabilità?*

Gli esempi illustrano un'idea generale: con un'espressione matematica possiamo attribuire probabilità ad ogni elemento del dominio della variabile casuale senza mai materialmente effettuare un solo esperimento perché tutto avviene a livello virtuale. I vantaggi sono formidabili: siamo in grado di rappresentare allo stesso modo fenomeni diversi, ma con uguali presupposti, siamo in grado di riassumere in una formula sintetica il risultato di lunghissime serie di rilevazioni; di descrivere con semplicità il comportamento di un fenomeno ed attuarne la simulazione con il computer; di delinearne le caratteristiche attese o desiderate quali la simmetria o l'allungamento delle code, di colmare lacune o estendere la rappresentazione a valori ipotetici, di ricondurre il confronto tra distribuzioni empiriche alla comparazione -molto più agevole- di funzioni matematiche.

Il processo di astrazione può essere spinto ancora più oltre per ragionare direttamente sui modelli dei fenomeni e trascurare -almeno in parte- i fenomeni stessi. In altre parole, invece di procedere ad una indagine empirica, ad esempio sulla capitalizzazione delle imprese *non profit,* si discute di un modello di distribuzione che ne riassuma i loro aspetti salienti evitando -addirittura- di accertare se mai esistono imprese dell'economia reale simili a quelle postulate dal modello. E' tuttavia un errore grave confondere il modello con ciò che esprime: i due piani debbono sempre essere tenuti distinti e distanti.

Esercizio_VC11: *è noto che un fenomeno ha un andamento monotono con effetti cumulativi cioè la probabilità varia in ragione dell'andamento delle modalità. Ciò che non è chiaro è il tipo di legame: diretto o inverso. Un modello flessibile adatto allo scopo è il seguente:*

$$p(X=x)=\frac{(\theta-1)\theta^{x-1}}{\theta^n-1};\quad x=1,2,\ldots,n;\ \theta>0$$

a) E' una distribuzione di probabilità? b) Come si comporta al variare del parametro θ?

7.1.2 La funzione di ripartizione delle probabilità

Le variabili casuali relative ad aspetti quantitativi della prova espressi con il dominio discreto e finito T={x_1, x_2,…x_n} possono anche essere descritte con le probabilità cumulate acquisendo qualche vantaggio pratico allorché le modalità del dominio siano numerose. Per una modalità generica $x \leq x_i \in T$ si definisce la funzione:

$$F(x)=P(X\leq x)=P\{E(-\infty,x]\}=P\left\{E(-\infty,x_1]\cup\left\{\bigcup_{j=2}^{i}(x_{j-1},x_j]\right\}\right\}=\sum_{x_i\leq x}p(x_i)\quad -\infty<x<\infty$$

che esprime la probabilità che la X sia inferiore o uguale ad un valore prefissato "x" . Tale probabilità è data dalla somma delle probabilità dei singoli valori inferiori o al più uguali ad "x". La F(x) è detta funzione di ripartizione delle probabilità e nasce con impostazione, struttura e finalità analoghe alla funzione di ripartizione empirica introdotta nel capitolo 2.

Esempi:

a) I clienti dell'albergo.

$$F(4)=p(X\leq 4)=P\{E(-\infty,4]\}=P\{T\}=1$$

$$F(3)=p(X\leq 3)=P\{E(-\infty,3]\}=P\{E(-\infty,0]\cup E(0,1]\cup E(1,2]\cup(2.3]\}=P\{E(-\infty,0]\}+P\{E(0,1]\}+P\{E(1,2]\}+P\{E(2,3]\}$$
$$=p(X=0)+p(X=1)+p(X=2)+p(X=3)=0.6836$$

$$F(2)=p(X\leq 2)=P\{E(-\infty,2]\}=P\{E(-\infty,0]\cup E(0,1]\cup E(1,2]\}=P\{E(-\infty,0]\}+P\{E(0,1]\}+P\{E(1,2]\}$$
$$=p(X=0)+p(X=1)+p(X=2)=0.2617$$

$$F(1)=p(X\leq 1)=P\{E(-\infty,1]\}=P\{E(-\infty,0]\cup E(0,1]\}=P\{E(-\infty,0]\}+P\{E(0,1]\}=p(X=0)+p(X=1)=0.0508$$

$$F(0)=p(X\leq 1)=P\{E(-\infty,0]\}=p(X=0)=0.0039$$

b) In uno *staff* operano 4 persone prossime alla pensione, ma ancora abbastanza indecise se rinviare il pensionamento oppure lasciare il lavoro. L'ufficio personale dovendo programmare per tempo le sostituzioni ha messo a punto il seguente schema per la probabilità di ritiro per i prossimi tre anni:

x	0	1	2	3	4	
p(x)	0.15	0.2	0.5	0.1	0.05	1.00

x	0	1	2	3	4
F(x)	0.15	0.35	0.85	0.95	1.00

La funzione di ripartizione può anche essere data in forma tabellare.

Vediamo come la conoscenza della F(.) consenta il calcolo delle probabilità per intervalli di valori in T:

$$p(x_1<X\leq x_2)=P\{E(x_1,x_2]\}=P\{E(-\infty,x_2]-E(-\infty,x_1]\}=P\{E(-\infty,x_2]\}-P\{E(-\infty,x_2]\cap E(-\infty,x_1]\}$$
$$=P\{E(-\infty,x_2]\}-P\{E(-\infty,x_1]\}=F(x_2)-F(x_1)$$

$$p(x_1 \le X \le x_2) = P\{E[x_1, x_2]\} = P\{E(x_1, x_2] \cup E(x_1, x_1]\} = P\{E(-\infty, x_2] - E(-\infty, x_1]\} + P\{E(x_1, x_1]\}$$
$$= F(x_2) - F(x_1) + p(X = x_1)$$
$$p(x_1 \le X < x_2) = F(x_2) - F(x_1) + p(X = x_1) - p(X = x_2);$$
$$p(x_1 < X < x_2) = F(x_2) - F(x_1) - p(X = x_2);$$

in cui la natura discreta del dominio comporta: $P\{E(x_i,x_i]\}=p(X=x_i)$.

Esempi:

a) Larson (1969, p. 79) illustra il gioco “Chuck-a-Luck” in cui si lanciano tre dadi. Si può scommettere una sola posta su uno degli interi (1,2,…,6). Se scommettete sul “5” e questo esce in uno dei tre dadi vincete una posta, se esce su due dadi vincete due poste e ne vincete tre se tutti i dadi mostrano il “5”. Se il “5” non esce il banco ritira la vostra posta. Indichiamo con X la variabile casuale che descrive il vostro incasso netto per una giocata di *Chuck-a-Luck*: T={-1,1,2,3) con probabilità:

$$p(-1) = \frac{125}{216};\; p(1) = \frac{75}{216};\; p(2) = \frac{15}{216};\; p(3) = \frac{1}{216};\; F(-1) = \frac{125}{216};\; F(1) = \frac{200}{216};\; F(2) = \frac{215}{216};\; F(3) = 1;$$

Ecco il calcolo per alcuni intervalli:

$$P(0 < X \le 3) = F(3) - F(0) = 1 - p(-1) = \frac{91}{216};\; P(-1 < X \le 0) = F(0) - F(-1) = \frac{125}{216} - \frac{125}{216} = 0$$

Lo zero dell’ultimo risultato scaturisce dal fatto che la funzione di ripartizione si mantiene costante nell’intervallo (-1,0] in quanto la variabile casuale discreta X non assume qui altri valori ovvero a quei valori è data probabilità nulla.

b) Il prezzo internazionale di un prodotto è fissato in dollari (intero e senza frazioni) in base al modello di probabilità p(x)=1/35 per x=1,2,…,35 ed alla funzione di ripartizione: F(x)=x/35 per x=1,2,…,35. Il cartello che controlla il mercato non interviene finché il prezzo rimane nell’intervallo di [20,30] dollari; se scende sotto l’estremo inferiore riduce le quote in vendita e le aumenta se supera l’estremo superiore. Qual’è la probabilità che il cartello debba intervenire?

$$P\{(-\infty, 20) \cup (30, \infty)\} = 1 - P\{[20, 30]\} = 1 - \{F(30) - F(20) + p(X = 20)\} = 1 - \frac{30}{35} + \frac{20}{35} - \frac{1}{35} = \frac{24}{35}$$

Il calcolo della probabilità dell’intervallo risulta più semplice usando la F(.) in quanto evita di sommare le probabilità per le modalità intermedie, qualunque ne sia il numero.

Esercizio_VC12*: con l’ausilio della funzione di ripartizione data in tabella:*

x	0	1	2	3	4	5
F(x)	0.0459	0.2415	0.5747	0.8585	0.9794	1.0000

Calcolare: 1) p(X≤3); 2) p(x<2); 3)p(1<x≤4).

Esercizio_VC13*: per la funzione di ripartizione:*

$$F(x) = \begin{cases} 0 & se\;\; x = 0 \\ 1/7 & se\;\; 0 < x \le 1 \\ 2/7 & se\;\; 1 < x \le 3/2 \\ 4/7 & se\;\; 3/2 < x \le 5/2 \\ 1 & se\;\; x > 5/2 \end{cases}$$

Calcolare: 1) p(x<3/2); 2)p(x≤5/2); 3) p(1/2<x≤2).

Analizziamo le caratteristiche che rendono una qualsiasi F(.) una funzione di ripartizione di probabilità scegliendo due valori in T: $x_1 \le x_2$ e poniamo:

$$(-\infty, x_2] = (-\infty, x_1] \cup (x_1, x_2] \qquad dove \qquad (x_1, x_2] = \{x \in T | x_1 \le x \le x_2\}$$

L’intervallo $(-\infty,x_2]$ è diviso in due parti senza punti in comune e pertanto:

$$F(x_2) = P\{(-\infty, x_2]\} = P\{(-\infty, x_1] \cup (x_1, x_2]\} = P\{(-\infty, x_1]\} + P\{(x_1, x_2]\} = F(x_1) + P\{(x_1, x_2]\}$$

Poiché la probabilità è non negativa: $F(x_2) \ge F(x_1)$ per $x_2 > x_1$ per cui la F è monotona non decrescente.

Inoltre:

$$F(-\infty)=\lim_{x\to-\infty}F(x)=\lim_{x\to-\infty}P\{E(-\infty,x]\}=P(\emptyset)=0;\quad F(\infty)=\lim_{x\to\infty}F(x)=\lim_{x\to\infty}P\{E(-\infty,x]\}=P(T)=1$$

C'è un'ulteriore caratteristica che deve essere approfondita. Per ogni $x_i \in$ T esistono -finiti- il limite sinistro e quello destro della funzione di ripartizione:

$$F(x_i+)=\lim_{\varepsilon\to0}\left[F(x_i+\varepsilon)-F(x_i)\right];\qquad F(x_i-)=\lim_{\varepsilon\to0}\left[F(x_i)-F(x_i-\varepsilon)\right]$$

Poiché T è discreto i due limiti sono diversi e la F(.) ha un "salto" in x_i di ampiezza $F(x_i+)-F(x_i-)$. Questo, però, è il solo tipo di discontinuità che può presentare la funzione di ripartizione (Chung, 1974, p.2). Il valore della F(.) in una discontinuità è arbitrario, ma soggetto al vincolo: $F(x_i-) \le F(x) \le F(x_i+)$. Fra le scelte possibili abbiamo:

$$\begin{gathered}F(x_i)=\lambda_1F(x_i-)+\lambda_2F(x_i+)\ \ con\ \ \lambda_1,\lambda_2\ge0,\ \ \lambda_1+\lambda_2=1\\ \lambda_1=1,\ \ \lambda_2=0\Rightarrow F(x_i)=F(x_i-)\quad continuità\ \ a\ \ sinistra;\\ \lambda_1=0,\ \ \lambda_2=1\Rightarrow F(x_i)=F(x_i+)\quad continuità\ \ a\ \ destra;\end{gathered}$$

La seconda scelta è la più consueta e sarà così anche nel presente testo. In definitiva, la funzione di ripartizione è monotona non decrescente, limitata nell'intervallo [0, 1], continua a destra

$$F(x_i-)\ne F(x_i+)\quad i=1,2,\ldots,n$$

La continuità a destra implica che il grafico della F(.), come è già avvenuto per le funzioni di ripartizione empirica, raggiunto il livello $F(x_i+)$, vi permanga finché non si arriva ad x_{i+1}; a questo punto scatta a $F(x_{i+1})$ per stazionarvi fino ad un nuovo cambio di modalità. In aggiunta, se il limite superiore dell'intervallo tende all'infinito, si ha:

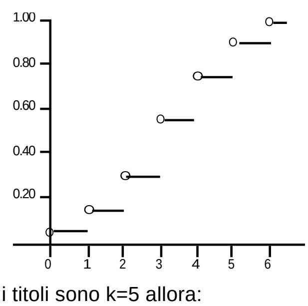
$$p(x_1<X<\infty)=p(X>x_1)=F(\infty)-F(x_1)\ \ =\ \ 1-F(x_1)$$

che quantifica la probabilità da assegnare ai valori maggiori di x e definisce la funzione di sopravvivenza in termini di probabilità.

Esempi:

a) In alcune indagini è importante disporre di uno schema probabilistico che descriva il numero di soggetti con cui una data unità può entrare in contatto. Indichiamo con X il numero di tali contatti ed ipotizziamo che una generica unità abbia i seguenti potenziali contatti:

X	0	1	2	3	4	5	6
p(x)	0.05	0.10	0.15	0.25	0.20	0.15	0.10
F(x)	0.05	0.15	0.30	0.55	0.75	0.90	1.00



Il grafico a destra rappresenta la funzione di ripartizione delle probabilità che, come si vede, ha struttura identica alla funzione di ripartizione delle frequenze relative.

b) La probabilità di un guadagno in borsa è ritenuta proporzionale al numero di titoli in rialzo. Se i titoli sono k=5 allora:

$$p(x)=\frac{x^2}{55};\ \ x=0,1,\ldots,5\Rightarrow F(x)=\frac{x(x+1)(2x+1)}{330}$$

F(x)=0 per x<0, F(x)=1 per x≥5; $F(x+1)=F(x)+(x+1)^2/55$ per x=1,2,…,5.

Esercizio_VC14: *la distribuzione di probabilità della variabile casuale X è:*

$$p(x)=\begin{cases}\dfrac{22-x}{100} & per\ \ x=0,1,\ldots,4\\ 0 & altrove\end{cases}$$

Determinare e rappresentare la sua funzione di ripartizione.

Esercizio_VC15: *Si lanciano tre monete, ma non sono equilibrate: "testa" ha probabilità 2/3 e "croce" 1/3.*
1) Ricostruire la distribuzione della variabile casuale X="numero di teste nei tre lanci";
2) Rappresentare graficamente la funzione di distribuzione e di ripartizione.

Esercizio_VC16: *una cavia si trova in un labirinto dove ha quattro percorsi da seguire. Alla fine di due percorsi trova il cibo, alla fine degli altri due attiva il miagolio di un gatto. Alla cavia sono concessi due tentativi. Se la cavia non impara dall'esito del percorso, qual'è la distribuzione di probabilità e la funzione di ripartizione di X= numero di volte che ha trovato il cibo?*

Esercizio_VC17: *la presentatrice televisiva ha tre numeri di telefono scelti a caso. Sia X il numero di quelli che risponderanno. Se "1" è l'evento che il numero risponda e "0" il contrario, l'universo degli eventi è S={(0,0,0); (0,0,1); (0,1,0); (0,1,1); (1,0,0); (1,0,1); (1,1,0); (1,1,1)}. La X sarà pari alla somma degli "1" presenti nella terna e quindi il suo dominio è T={0,1,2,3}. Ad ogni terna associamo: p[n "1" e (3-n) "0"]=*$0.4^n 0.63^{-n}$*. Utilizzare la F(.) per calcolare* $p(1.2<x<3.6)$, $p(x\leq 2)$, $p(x\geq 2.2)$.

Dalla funzione di ripartizione alla funzione di distribuzione

Se una variabile casuale è nota attraverso la sua funzione di ripartizione, la sottostante distribuzione di probabilità si può ricavare per via analitica. Fissiamo un "x" ricadente nell'intervallo $b-\varepsilon< x \leq b$ con $\varepsilon>0$. Allora:

$$\lim_{\varepsilon\to 0} p(b-\varepsilon < x \leq b) = \lim_{\varepsilon\to 0}\left[F(b)-F(b-\varepsilon)\right] = F(b) - \lim_{\varepsilon\to 0} F(b-\varepsilon) = F(b) - F(b-)$$

Se "b" è un punto di discontinuità allora ha probabilità positiva pari all'altezza del salto p(b). Quindi, in generale:

$$p(x_i) = F(x_i) - F(x_i -) = \begin{cases} F(x_0) & per \quad x \leq x_0 \\ F(x_i) - F(x_{i-1}) & per \quad x_{i-1} < x \leq x_i \end{cases} \qquad i = 0,1,\ldots,n$$

Esempio:

Un'indagine sull'intensità del traffico in una strada particolarmente frequentata ha rilevato il numero di veicoli che ogni 5 minuti svoltano in una traversa posta a metà percorso.

X	2	4	6	8	10	12	
F(x)	0.05	0.15	0.30	0.40	0.50	1.00	
p(x)	0.05	0.10	0.15	0.10	0.10	0.50	1.00

La probabilità della prima modalità coincide con il valore della funzione di ripartizione dato che F(0)=0; le altre si ricavano per differenza tra valori consecutivi della F(.).

Esercizio_VC18: *data la funzione di ripartizione:* $F(x)=x^2[0.5n(n+1)]^2$, $x=1,2,\ldots,n$ *determinare la corrispondente distribuzione di probabilità.*

La funzione di ripartizione può essere riformulata in termini della distribuzione di probabilità. Infatti, poiché $F(x_i-)=p(X\leq x_{i-1})=F(X\leq x_{i-1})$ ne consegue che: $p(x_i-)=p(X=x_i)=F(x_i)-F(x_{i-1})$ da cui segue la relazione ricorsiva:

$$F(x_i) = F(x_{i-1}) + p(x_i); \qquad F(x_0) = p(x_0) = 0$$

Esercizio_VC19: *la distribuzione di probabilità della somma dei punti nel lancio di due dadi regolari è:*

$$p(i) = \begin{cases} \dfrac{i-1}{36} & per \quad i = 2,3,\ldots,6 \\ \dfrac{13-i}{36} & per \quad i = 7,8,\ldots,12 \end{cases}$$

Determinare la funzione di ripartizione.

Esercizio_VC20: *esprimere la funzione di ripartizione F(x) come somma delle probabilità che la funzione di ripartizione assegna alle modalità inferiori o uguali ad x.*

Trasformazioni di variabili casuali discrete finite

Le variabili casuali sono spesso trasformate per studiare delle altre variabili legate funzionalmente alle prime: si rileva il numero di posti occupati in un cinema X, ma poi si analizza l'incasso Y=4.5X. Se i valori della X sono assunti con probabilità lo stesso accade a Y=g(X) per effetto di concatenamento. Le funzioni di variabili casuali sono delle nuove e diverse variabili casuali rispetto a quelle da cui derivano e per determinarne la funzione di distribuzione di probabilità si deve ritornare all'universo degli eventi dell'esperimento ed individuare i valori della variabile trasformata per poi riunire i valori eventualmente ripetuti.

Esempi:

a) La distribuzione di probabilità del numero di blocchi di confezioni ritirate per superamento data è descritta nella tabella a sinistra; i blocchi devono essere allineati per il trasporto automatico e serve conoscere la distribuzione di y=√x.

X	1	4	9	16	25	36	49	
p(x)	0.03	0.09	0.16	0.44	0.16	0.09	0.03	1.00

Y	1	2	3	4	5	6	7	
p(y)	0.03	0.09	0.16	0.44	0.16	0.09	0.03	1.00

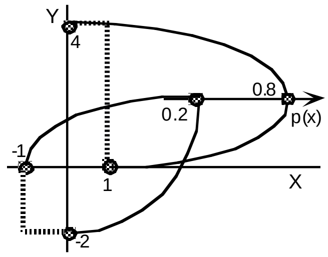
Dato che per ogni "x" c'è uno ed un solo "y" si modificano le modalità, ma non le probabilità del modello.

b) Studiamo la variabile casuale X che ha distribuzione di probabilità:

$$p(x)=\frac{4-|x-4|}{16};\quad x=1,2,\dots,9$$

ed analizziamo Y=1/X. Si vede subito che è una trasformazione biunivoca; inoltre, se la modalità X=2 ha probabilità del 6.25% lo stesso dovrà succedere per la sua trasformata Y=1/2 visto che niente è cambiato nel quadro probabilistico dell'esperimento. Per ottenere le probabilità della Y basta collocare la relazione inversa X=1/Y nella funzione di distribuzione:

$$P(y)=\frac{4-\left|\frac{1}{y}-4\right|}{16};\; y=1,\frac{1}{2},\dots,\frac{1}{9}$$



c) Consideriamo la variabile casuale X a valori in T=(-1,1) con probabilità (0.2,0.8) e trasformiamola in Y=3X+1. La nuova variabile casuale può assumere due soli solo valori (-2, 4) ed esiste una corrispondenza uno-a-uno X ed Y e variabili per cui le probabilità della Y sono le stesse della X. Poiché P(Y≤y)=P[X≤(y-1)/3] le funzioni di ripartizione sono:

$$F(x)=\begin{cases}0 & se\;\; x<-1\\ 0.2 & se\;\; -1\le x<1\\ 1 & se\;\; x\ge 1\end{cases}\;\Rightarrow F(y)=\begin{cases}0 & se\;\; y<-2\\ 0.2 & se\;\; -2\le y<4\\ 1 & se\;\; y\ge 4\end{cases}$$

Le trasformazioni pongono il seguente problema: determinare la funzione di distribuzione di una variabile casuale Y allorché sia nota quella di X legata alla precedente dalla relazione: Y=g(X). Se X è una variabile casuale discreta e finita con dominio $T=\{x_1, x_2,\dots, x_k\}$ e g(.) una applicazione continua e con un numero di inverse alpiù enumerabile, allora Y=g(X) è una variabile casuale discreta con distribuzione di probabilità e ripartizione:

$$p(y)=p(y=y_i)=p\left[x\in\left\{g^{-1}(y_i)\right\}\right];\qquad F(y)=p\left[x\le g^{-1}(y_i)\right]\qquad i=1,2,\dots,n$$

dove $\{g^{-1}(y_i)\}$ per i=1,2,…,k è l'insieme dei punti x_i in T tali che $g(x_i)=y_i$. Trattandosi di una variabile discreta la probabilità di $Y=y_i$ può essere calcolata sommando le probabilità dei valori della X che generano y_i.

$$p(y=y_i)=\sum_{x_i\in\left\{g^{-1}(y_i)\right\}} p(x=x_i)$$

Esempi:

a) Sia X=0, 2, 3 la variabile casuale che descrive il punteggio ottenibile con un tiro a canestro ed ipotizziamo le probabilità: {0.40, 0.35, 0.15}. Qual'e la distribuzione da associare a Y=2X-1 che rileva i punteggi partendo dal negativo e raddoppiando i dislivelli tra valori? Y=-1, 3, 5; le probabilità sono invariate rispetto alla X. Qual'è la funzione di ripartizione di Y? F(y) = p(2X-1≤y) = p[X≤(y+1)/2].

b) Sia X=-3 o X=3 con P(-3)=0.6 e P(+3)=0.4 e consideriamo la trasformazione $y=x^2$. La variabile casuale Y in questo caso è degenere dato che Y=9 è un evento certo: $P(Y=9)= P[X=(-3)^2]+P[X=3^2]=0.6+0.4=1$.

c) La squadra di calcio che gioca in casa ottiene punti 0 se perde, 1 se pareggia e 3 se vince. Ipotizziamo che P(3)=0.60, P(1)=0.30, P(0)=0.10. Qual'è la distribuzione di probabilità della media inglese? La media inglese assegna punti zero alla squadra che vince in casa, -1 se pareggia e -2 se perde. La trasformazione che collega i due punteggi è:

$$y = x - 2 + d \quad con \;\; d = \begin{cases} -1 & se \;\; x = 3 \\ 0 & altrimenti \end{cases}$$

Nessuna trasformazione è necessaria per le probabilità.

Esercizio_VC21: *la società che gestisce una linea di autobus accetta prenotazioni fino a 4 posti oltre la capienza. In tabella è data la distribuzione della richiesta X dei posti.*

X	0	1	2	3	4	
p(x)	0.50	0.20	0.15	0.10	0.05	1.00

Esercizio_VC22: *se l'incasso è pari al doppio del quadrato del numero dei posti: $Y=2X^2$, determinare p(y) e F(y).*
b) Per la variabile casuale con distribuzione di probabilità di (0.1, 0.2, 0.3, 0.3, 0.1) rispettivamente associate a $x\in$ {-2,-1,0,1,1}, determinare la funzione di distribuzione e di ripartizione di $Y=(X+3)/(X-3)^2$.

Teoricamente, è possibile che Y=g(X) trasformi la variabile discreta e finita in discreta ed enumerabile (ad esempio con una funzione periodica). In questo caso trattazione della Y rientra nei casi studiati nel paragrafo 7.2.

7.1.3 Descrizione delle variabili casuali discrete finite

Sia $T=\{x_1, x_2,\ldots,x_n\}$ il dominio della variabile casuale X e sia p_i, i=1,…,n la sua distribuzione di probabilità. Tutte le misure di centralità viste nel paragrafo 3.1 possono essere applicate alle variabili casuali sottolineando però la modifica di significato: la distribuzione empirica è quella costruita "dopo" la rilevazione, la distribuzione di probabilità è costruita "prima", ma forse dopo altre che hanno portato ad ipotizzare quel particolare modello.

Centralità delle discrete finite

La moda nella rilevazione è la modalità che si è verificata più spesso; la moda di una variabile casuale è il valore più probabile. Lo stesso ragionamento vale per i quantili, per le medie e tutti gli altri momenti.

Esempi:
a) Determinare la mediana per una variabile casuale che ha distribuzione di probabilità:

$$p_i = \frac{12i(i+1)^2}{n(n+1)(n+2)(3n+5)}; \quad i = 1,2,\ldots,n$$

Se "n" è dispari sarà $M_e=(n+1)/2$ altrimenti ogni valore interno a [n/2,n/2 +1] è una mediana.

b) Determinare la moda nel modello:

$$p_i = \frac{4i\left(m^2 - i^2\right)}{n(n+1)\left[2m^2 - n^2 - n\right]}; \quad i = 1,2,\ldots,n; \;\; m \geq n; \qquad \frac{p_{i+1}}{p_i} = \left(\frac{i+1}{i}\right)\left[1 - \frac{(2i+1)}{m^2 - i^2}\right]$$

Il quoziente tra probabilità successive ricorre spesso nello studio delle variabili discrete. La moda è individuata da un rapporto pari ad uno: in questo caso per i=n dato che i rapporti di probabilità sono crescenti.

c) Per una distribuzione di probabilità $p(x_i)$ i=1,2,…,n determiniamo 1° e 3° quartile in base alla formula del paragrafo 3.1.3:

$$X_{0.25} = (1-\gamma)X_{(i)} + \gamma X_{(i+1)}, \;;\;\; i \leq 0.25n < i+1; \;\; \gamma = \begin{cases} 0.5 & se \;\; [0.25n] = 0.25n \\ 1 & se \;\; [0.25n] < 0.25n \end{cases};$$

$$X_{0.75} = (1-\gamma)X_{(i)} + \gamma X_{(i+1)}, \;;\;\; i \leq 0.75n < i+1; \;\; \gamma = \begin{cases} 0.5 & se \;\; [0.75n] = 0.75n \\ 1 & se \;\; [0.75n] < 0.75n \end{cases}$$

Esercizio_VC23: *una scatola contiene 5 biglie di cui 3 bianche e 2 rosse. Ipotizziamo che la scelta delle biglie avvenga a caso e senza reimmissione.*
a) Definire la distribuzione di probabilità di X="numero di estrazioni per ottenere le due biglie rosse";
b) Calcolare moda e mediana di X.

La media aritmetica o valore atteso della variabile casuale X, indicata con E(X) è:

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

dove "E" sta per "Expectation" cioè aspettativa, valore atteso.

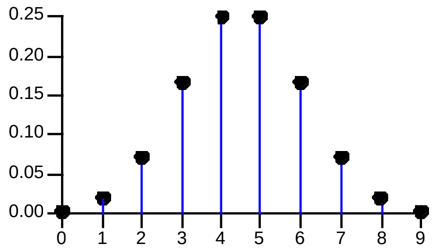
Esempi:
a) un dado è truccato in modo che la probabilità di uscita sia legata al punteggio (questo può succedere se i punti del dado sono realizzati con dei rilievi). Qual'è la media aritmetica?

$$Truccato: \ \mu = \sum_{i=1}^{6} \frac{i^2}{21} = 4.33; \qquad Non \ \ truccato: \ \mu = \sum_{i=1}^{6} \frac{i}{6} = 3.5 \qquad p_i = \begin{cases} \frac{i}{21} & se \ \ i = 1,2,\ldots,6 \\ 0 & altrimenti \end{cases}$$

Il trucco può essere scoperto realizzando che la media è ora più elevata rispetto a ciò che ci si aspetterebbe da un dado non truccato.

b) Un modello assegna le probabilità ai valori secondo la formula di seguito riportata e di cui è dato un esempio in figura per n=9.

$$p(i) = \frac{\binom{n}{i}}{2^n} \quad per \quad i = 0,1,2,\ldots,n; \Rightarrow \mu = \sum_{i=0}^{n} i \frac{\binom{n}{i}}{2^n} = \frac{n2^{n-1}}{2^n} = \frac{n}{2}$$



In questo caso μ=M_e. Se "n" è pari la distribuzione è unimodale e la moda è pari alla media ed alla mediana.

c) Port (1994, cap. 24) discute un interessane esempio di come la Statistica consenta di risparmiare risorse. Una popolazione di N persone deve essere sottoposta a test per accertare la presenza di un *marker* la cui probabilità "p" di presenza è costante ed è indipendente tra una persona e l'altra. Invece di provare su tutti, i soggetti sono si suddivisi in "k" gruppi di uguale numerosità "m" (uno dei gruppi potrebbe essere meno numeroso degli altri, ma per semplificare ipotizziamo lo stesso numero di soggetti in ogni gruppo); i prelievi degli "m" individui sono mischiati ed il test è effettuato sul solo prelievo aggregato. Se il *marker* è assente non si procede oltre; se invece è presente si cerca il *marker* in ciascuno dei prelievi individuali. Se si pone X_i=1 allorché il test aggregato dell'i-esimo gruppo è positivo e 0 altrimenti. Il numero di accertamenti svolti nel gruppo i-esimo è N_i=1+mXi. Poiché X_i=0 se nessun soggetto del gruppo è portatore allora $P(X_i=0)=(1-p)^m$ e quindi $P(X_i=1)=1-(1-p)^m$. Il valore atteso del numero di test da effettuare per l'i-esimo gruppo è:

$$E(N_i) = E(1+mX_i) = 1 + mE(X_i) = 1 + m\left\{0(1-p)^m + 1\left[1-(1-p)^m\right]\right\} = 1 + m\left[1-(1-p)^m\right]$$

Generalizzando il valore atteso a tutti "k" gruppi ed usando il fatto (approfondito nella parte dedicata alle variabili casuali multidimensionali) che $E(\sum x_i)=\sum E(X_i)$, il valore atteso del totale dei test è:

$$E(T) = E\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} E(N_i) = k\left\{1 + m\left[1-(1-p)^m\right]\right\} = \frac{N}{m}\left\{1 + m\left[1-(1-p)^m\right]\right\} = N\left[\frac{m+1}{m} - (1-p)^m\right]$$

La suddivisione in gruppi risulta una strategia migliore dell'analisi totale (che richiede N test) se l'argomento in parentesi è minore di uno e maggiore di zero. Data la probabilità di presenza "p" esiste un valore ottimo di "m" che rende minimo E(T). La tabella che segue ne riporta il valore per alcuni livelli di "p" (altri sono ottenibili facilmente con il foglio elettronico).

p	0.2	0.15	0.1	0.05	0.01	0.005	0.001	0.0005	0.0001
m	2	2	3	5	11	15	32	45	101

Esercizio_VC24: *per il modello di Newcomb:*

$$p(x_i) = Log_{10}\left(\frac{x_i + 1}{x_i}\right) \quad x_i = i, \ \ i = 1,2,\ldots,9$$

a) Determinare moda, mediana e valore atteso; b) Qual'è la differenza interquartilica?

Esercizio_VC25: *una cliente dell'albergo ricorda il piano della sua stanza, ma non il numero che peraltro non è leggibile dalla sua scheda. La cliente decide di tentare con tutte le "n" porte del piano scegliendo a caso le porte su cui tentare senza però ripetere il tentativo più di una volta per ogni stanza. Sia X indica il numero di tentativi necessari per aprire la porta. Poiché la procedura è la stessa della estrazione senza reimmissione di una particolare biglia da un'urna che ne contiene "n". Allora:*

$$P(X > x) = \frac{\binom{n-1}{x}}{\binom{n}{x}} = \frac{n-x}{n}$$

a) Calcolare E(X); b) Come cambia E(x) se la cliente non riesce a ricordare i tentativi già fatti?

Scommesse e guadagni previsti

La media aritmetica delle variabili casuali si ammanta di un nuovo significato: quello di speranza matematica che costituisce il punto di riferimento di molte applicazioni della probabilità. Nelle scommesse i *bookmakers* propongono l'espressione "l'evento è dato x:y (x a y)" cioè se l'evento si verifica si riceveranno "x" puntate rischiandone "y" per un incasso totale di x+y unità di conto. E' prassi indicare per prima il guadagno in caso di vincita. Un gioco d'azzardo si dice equo se le poste che si pagano per parteciparvi sono proporzionali alle probabilità di vincita. In altre parole, un gioco equo non favorisce nessuno nel ripetersi prolungato delle giocate.

Esempi:

a) Nel lancio del dado l'uscita singola è data 5:1. Per una puntata di 5'000 lire, in caso di vincita dovrei incassare 30'000 (le mie 5 mila più 25'000 di vincita) ovvero se perdo, il banco trattiene solo 1'000 lire e restituisce 4'000 per compensare le sue maggiori probabilità:

$$5'000\left(\frac{1}{6}\right) - 1'000\left(\frac{5}{6}\right) = \frac{5'000}{6} - \frac{5'000}{6} = 0$$

La speranza matematica è quell'importo certo che si è disposti a pagare per ricevere in cambio un importo aleatorio descritto dalla variabile casuale.

b) Nella *roulette* americana decidete di giocare lire X sul nero con P(N)=18/38 e P(N^c)=20/38 (in questo tipo di *roulette* ci sono lo "0" ed il "00" di colore verde). Se esce il nero ricevete 2X. E' un gioco equo?

$$Y * \frac{18}{38} - X\frac{20}{38} = 0 \Rightarrow Y = \frac{10}{9}X$$

Se il gioco fosse equo puntando 27'000 se ne dovrebbero incassare 57'000: 30'000=27000*10/9 di vincita più la giocata. Il fatto che il casinò dia 54'000 (27'000 di vincita più la giocata) si spiega per le spese organizzative, di manutenzione e gestione, ma lo scarto dell'11% è alto.

In generale, se "p" è la probabilità di vincere una scommessa G in cui vi sia la promessa di vincere un importo "x" con probabilità "p" e di perdere un importo "y" con probabilità (1-p) l'esito atteso è:

$$E(G) = xp - y(1-p) = xp - y + yp = (x+y)p - y$$
$$E(G) = 0 \Rightarrow (x+y)p - y = 0 \Rightarrow p = \frac{y}{x+y},\ 1-p = \frac{x}{x+y}$$

Questa espressione, detta speranza matematica deve essere nulla per un gioco equo e ciò avviene quando la probabilità di vincere coincide con la quota dell'importo perso sul totale della posta in gioco

Esempi:

a) La tassa sulla stupidità. Ciccillo si gioca un ambo secco (5, 25) sulla ruota di Cagliari (perché comincia come il suo nome). La sua speranza matematica, tenuto conto delle combinazioni a favore: C(2,2)*C(88,3) e delle combinazioni possibili: C(90,5), è:

$$10'000\left(\frac{20}{8010}\right) - 10'000\left(\frac{7090}{8010}\right) = 25 - 8851 = -8'826$$

Per equilibrare il maggior rischio di Ciccillo il banco dovrebbe pagargli, in caso di vincita 3'545'000=7090*10'000/20. Ciccillo invece riceve 250 volte la posta cioè 2'500'000. La differenza è in parte da attribuire a spese organizzative e di gestione, ma una parte -cospicua- è la tassa sulla dabbenaggine dei giocatori che trova parziale giustificazione nell'interesse pubblico con cui si impiegano i fondi così ottenuti.

b) A tre giornate dalla fine del campionato la retrocessione della Reggina è data 5:2. Qual'è la probabilità soggettiva che i *bookmakers* attribuiscono all'evento?

$$E(G) = 5 * p - 2(1-p) = 0 \Rightarrow p = \frac{2}{7} \cong 28.6\%$$

Questo però per assicurare il gioco equo. Tuttavia, l'allibratore deve rientrare delle spese ed avere un ragionevole profitto per cui la probabilità di retrocessione della Reggina è in realtà ritenuta più bassa del 28.6%.

c) La Società Anonima di Assicurazioni offre polizze vita alle donne in età 23-26 anni che, a fronte di un premio annuale di 120 euri, paga 120 mila euri in caso di morte. Il premio è equo? Il dati censuari segnalano 1'200'000 donne in quella fascia d'età e la frequenza di morte nell'anno del censimento è di 700. L'incasso atteso della polizza è:

$$E(I) = 120'000 * \frac{700}{1'200'000} = 70$$

Su ogni polizza la SADA incassa annualmente 120 euri e si aspetta di sborsarne 70.

Esercizio_VC26: *state risparmiando per comprare casa. Riuscite a mettere da parte dieci milioni l'anno e vi si prospettano due strategie:*
1) Entro il 4 gennaio di ogni anno depositate in banca i soldi con un tasso fisso del 4% e di inflazione dell'8%;
2) Nei primi di gennaio di ogni anno vi recate in un casinò e giocate i 10 milioni sul "12" (o altro numero a vostra scelta). Quale strategia è più conveniente?

Esercizio_VC27: *Roberts (1993, pp139-140) illustra il gioco del Keeno, simile alla tombola. I giocatori scelgono da 1 a 15 numeri compresi tra 1 e 80. Il gestore estrae, senza reimmissione, 20 numeri dall'urna che contiene gli 80 possibili. Le vincite sono proporzionali al numero di abbinamenti tra le scelte del giocatore e le estrazioni del gestore. Definiamo X="numero di abbinamenti" ed indichiamo con "n" il numero di scelte". Se n=1 allora p(x=1)=20/80=1/4 con pagamento equo in ragione di 3 a 1, ma il gestore paga 2.2:1. Per n=2 si ha p(x=2)=(20/80)(19/79)=3/50 cioè 16:1 contro un payoff di 12:1. La tipica cartella del Keeno ha 10 numeri.*
a) Calcolare le probabilità di vincita semplici e cumulate per n=5, 6, e 7;
b) Tenuto conto che il gestore paga, rispettivamente, 1:1, 17:1, 179:1 valutate l'equità del gioco.

Valore atteso di funzioni di variabili casuali discrete finite

Segue dalla discussione fatta in precedenza che se X è una variabile casuale discreta anche Y=g(X) è una variabile casuale discreta purché g(.) sia continua ed abbia un numero di inverse finito o enumerabile; inoltre, la sua distribuzione di probabilità può essere ricavata da quella della X. Il valore atteso della trasformata è:

$$E(Y) = \sum_{i=1}^{k} y_i p(Y = y_i) = \sum_{i=1}^{k} g(x_i) p\left[X = g^{-1}(y_i)\right] = \sum_{i=1}^{k} g(x_i) p\left[X = x_i\right]$$

cioè la somma dei prodotti $y_i\ p(x_i)$ per tutte le x_i del dominio che inducono y_i.

Esempi:

a) Sia X una variabile casuale discreta con valori in {-2,-1,1, 2} e p(X=-2)=p(X=2)=0.20 e p(X=-1)=p(X=1)=0.30. Calcoliamo il valore atteso di $Y=X^2$. Tale trasformazione genera una variabile pure discreta con valori: {1, 4}. Possiamo determinare la distribuzione di probabilità della Y tenendo conto che i diversi valori della X sono incompatibili:

p(Y=1)=p[(X=-1)U(X=+1]=p(X=-1)+p(X=+1)=0.60;
p(Y=4)=p[(X=-2)U(X=+2)]=p(X=-2)+p(X=2)=0.40.

A questo punto è semplice calcolare il valore atteso della Y: E(Y)=1*0.6 + 4*0.4=2.2.

b) Una variabile casuale discreta X ha modello di distribuzione:

$$p(x) = \begin{cases} \dfrac{6 - |x-7|}{36} & per \quad x = 2, 3, \ldots, 12 \\ 0 & altrove \end{cases}$$

Qual'è il valore atteso di y=Ln(x)? Tenuto conto che $x=e^y$ si ha:

$$p(y) = \begin{cases} \dfrac{6 - \left|e^y - 7\right|}{36} & per \quad y = Ln(2),\ Ln(3), \ldots, Ln(12) \\ 0 & altrove \end{cases} \Rightarrow E(y) = 1.875$$

c)Se la distribuzione di probabilità di x è p(x)=x/55, per x=1,2,…,10 con valore atteso μ_x=7 allora il valore atteso di y=3x-2 sarà μ_y=3*7-2=19 dato che:

$$E(Y)=\sum_{i=1}^{k} y_i p(y_i)=\sum_{i=1}^{k}(ax_i+b)p(x_i)=\sum_{i=1}^{k} ax_i p(x_i)+\sum_{i=1}^{k} bp(x_i)=a\sum_{i=1}^{k} x_i p(x_i)+b\sum_{i=1}^{k} p(x_i)=a\mu_x+b$$

Esercizio_VC28: *Dougherty (1990, pp.100-101) propone un esempio sulla compressione delle immagini. Se il pixel ha otto livelli di grigio, la trasmissione del suo colore richiede 3 bit (2^3=8). Se i livelli fossero ridotti a quattro basterebbero 2 bit con un notevole risparmio. Si abbia la distribuzione dei livelli originari di grigio:*

x	0	1	2	3	4	5	6	7	
p(x)	0.05	0.10	0.25	0.20	0.19	0.12	0.06	0.03	1.00

L'algoritmo di compressione prevede l'accoppiamento di livelli successivi Y=0 se X=0 o X=1, Y=1 se X=2 o X=3, Y=2 se X=4 o X=5, Y=3 se X=6 o X=7.
a) Determinare la distribuzione di probabilità della Y;
b) Calcolare il valore atteso della X e della Y. Cosa si può dire sul risultato della compressione?

Esercizio_VC29: *una variabile casuale ha modello:*

$$p(x)=\begin{cases} k-\dfrac{x}{10}, & x=1,2,3,4 \\ 0 & altrove \end{cases}$$

a) Per quale valore di k è in effetti una distribuzione di probabilità;
b) Determinare il valore atteso di X; c) Determinare il valore atteso di y=|x-2|.

La procedura con cui si ottiene il valore atteso di una trasformata è chiara nel suo sviluppo, ma lascia perplessi sul fatto che una trasformazione più complessa possa essere ragionevolmente affrontata in questo modo. Davenport (1970, p.216) propone una possibile semplificazione: calcoliamo E(y) applicando la trasformazione y=g(x) ad ogni valore e moltiplicando poi il risultato ottenuto per la probabilità associata alla modalità di X senza coinvolgere esplicitamente la distribuzione della nuova variabile Y.

$$E(Y)=\sum_{r=1}^{k} y_j p\left[X\in\left\{g^{-1}(y_r)\right\}\right] \;=\; \sum_{r=1}^{k} g(x_j) \sum_{x_j\in\left\{g^{-1}(y_r)\right\}} p(X=x_j) \;=\sum_{r=1}^{k} g(x_j)p(X=x_j)$$

Pertanto, la procedura prima delineata dà tutte le garanzie di generalità.

Esempio:
Riconsideriamo gli elementi dell'esempio precedente:

$$E(Y)=(-2)^2 p(X=-2)+(-1)^2 p(X=-1)+(1)^2 p(X=1)+(1)^2 p(X=2)=4*0.2+1*0.3+1*0.3+4*0.2=2.2$$

I momenti delle discrete finite

Analoga estensione avviene per i momenti che hanno qui le stesse finalità di sintesi che in statistica descrittiva.

$$Rispetto\ \ all' origine:\ E\left(X^r\right)\ \mu_r=\sum_{i=1}^{k} x_i^r p_i; \qquad Rispetto\ \ alla\ \ media:\ E\left(\left|X-\mu\right|^r\right)=\mu_r^{'}=\sum_{i=1}^{k}(x_i-\mu)^r p_i$$

Esempi:
a) Per le tre distribuzioni di probabilità date in tabella:

	1	2	3	4	5
p(X=x)	0.05	0.10	0.70	0.10	0.05
p(Y=y)	0.20	0.20	0.20	0.20	0.20
p(Z=z)	0.40	0.10	0.00	0.10	0.40

Calcoliamo, per ciascuna, l'indice di entropia di Backman-Paternoster: E_4(X)=0.2435, E_4(Y)=0.40, E_4(Z)=0.33. L'interpretazione passa per il grado di incertezza con cui si può prevedere l'esito di una determinazione della variabile casuale. Nel caso della Y è massima e se fosse una scommessa dovremmo razionalmente scegliere di giocare con la X in cui la previsione è meno incerta.

b) Un gioco consiste nel lanciare un dado con le seguenti puntate: (1,2) si vincono 20 euri, (3) se ne vincono 25 e (4,5,6) si vincono 10 euri. La posta è di 20 euri. Calcoliamo la varianza per ogni giocata:

$$\sigma^2 = \sum_{i=1}^{3} x_i^2 p_i - \mu^2 = 20^2\left(\frac{2}{6}\right) + 25^2\left(\frac{1}{6}\right) + 10^2\left(\frac{3}{6}\right) - \left[\frac{40+25+30}{6}\right]^2 = 36.806$$

Si potrà notare che il gioco non è equo. In questo senso basterebbe una puntata di poco meno di 16 euri.

c) La deviazione standard è una misura del rischio connesso con la previsione di eventi incerti. Consideriamo gli investimenti in tabella.

				$E(x)$	$\sigma(X)$
1	$+5*1$	$0*0$		5	0
2	$+44*\frac{1}{4}$	$-8*\frac{3}{4}$		5	22.52
3	$+60*\frac{1}{3}$	$+9\frac{1}{3}$	$-54*\frac{1}{3}$	5	46.62
4	$+35*\frac{1}{5}$	$-2\frac{2}{5}$	$-3\frac{2}{5}$	5	14.81

Dal punto di vista del guadagno atteso non sono distinguibili. Chi rifugge dal rischio sceglierà l'investimento "1", chi ha una moderata avversità al rischio opterà per la "4", chi è disposto a farsi carico di maggiori rischi sceglierà la "3"; chi è indifferente rispetto al rischio sarà indifferente rispetto alla scelta. In realtà, tutte le misure di variabilità possono essere interpretate come funzioni di rischio di quell'importo aleatorio di cui μ (o altra misura di centralità) costituisce il valore di riferimento.

d) Calcoliamo la deviazione media per il modello:

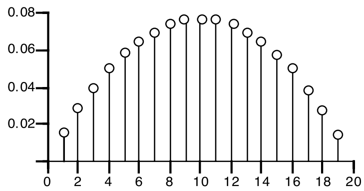
$$p(x) = \frac{10}{26\left(x^2+1\right)}; \quad x = -3, -2, -1, 0, 1, 2, 3$$

La formula della deviazione media già data per misurare la dispersione nelle rilevazioni empiriche si modifica per le f_i sostituite con le p_i:

$$S_\mu = \sum_{x=1}^{n} \left|x_i - \mu\right| p(x) \Rightarrow \sum_{i=-3}^{3} \frac{10\left|i-\mu\right|}{26\left(i^2+1\right)} = 2\sum_{i=1}^{3} \frac{10*i}{26\left(i^2+1\right)} = 0.9231$$

Esercizio_VC30: *alcuni fenomeni si manifestano con probabilità speculari rispetto al centro. Un modello che risponde a tale requisito è:*

$$p_i = \frac{i*(n-i)}{\left(n/6\right)\left(n^2-1\right)}; \quad i = 0, 1, 2, \ldots, n$$



a) Calcolare il σ e S_{Me} per n=20; b) Come si interpretano i valori ottenuti?

Esercizio_VC31: *i tassi di rendimento annuali attesi di tre blu chips sono i seguenti:*

Titolo	13%	12%	11%	10%	9%	8%	7%	6%	5%
Giat	0.01	0.05	0.09	0.15	0.18	0.29	0.09	0.08	0.06
Mirelli	0.08	0.15	0.26	0.1	0.1	0.1	0.1	0.07	0.04
Nocetti	0.04	0.06	0.21	0.19	0.27	0.1	0.08	0.03	0.02

Adoperate il coefficiente di dispersione (S_{Me}/M_e) per comparare il rischio dei tre investimenti.

Trasformazioni lineari

Una delle trasformazioni più ricorrenti è quella lineare: Y=aX+b dove "a" e "b" sono delle costanti finite. Il fatto che X sia una variabile casuale significa che comunque si formi $A \subset T$, la sua controimmagine $X^{-1}(A)$ ricade nell'algebra W di un esperimento probabilistico. In particolare, dato che T è discreto e finito, certamente $X^{-1}\{(a, x]\} = \{e \in S | X(e) \leq x\} \in W$ è cioè un evento e quindi probabilizzabile e lo è qualsiasi sottoinsieme di valori di T.

$$\left\{e \in S \middle| aX(e) + b \leq x\right\} = \left\{aX \leq x - b\right\}$$

$$a>0 \Rightarrow \left\{X \le \frac{x-b}{a}\right\} \in W;\ \ a<0 \Rightarrow \left\{X < \frac{x-b}{a}\right\}^c \in W;\ \ a=0 \Rightarrow \begin{cases} S \in W & se\ \ x \ge b \\ \varnothing \in W & se\ \ w < b \end{cases}$$

Quindi, la trasformazione lineare crea una nuova variabile casuale di cui possiamo calcolare, ad esempio, valore atteso e scarto quadratico medio:

$$\mu_y = \sum_{i=1}^{k} y_i p_i = \sum_{i=1}^{k} \left(ax_i + b\right) p_i = a\mu_x + b;\quad \sigma_y = \sqrt{\sum_{i=1}^{k} \left(ax_i + b - a\mu_x - b\right)^2 p_i} \quad = |a|\sigma_x$$

Il valore atteso della trasformata lineare è la trasformata del valore atteso; lo scarto quadratico medio riproduce solo la parte moltiplicativa.

Esempio:

$$X_i = i,\ \ i = 2,3,\ldots,12;\quad p_i = \frac{1}{\alpha\left(i^2-1\right)};\ \ \alpha = 0.66987179;\ \ U_i = \frac{X_i - 2}{12-2}$$

$$\mu_x = \sum_{i=2}^{12} \frac{i}{0.66987179\left(i^2-1\right)} = 3.508;\ \ \mu_y = 0.1508\ \ \sigma_x = \sqrt{\sum_{i=2}^{12} \frac{i^2}{0.66987179\left(i^2-1\right)} - 3.508^2} = 2.261;\ \ \sigma_y = 0.2261$$

Esercizio_VC32:
a) Se X una variabile casuale discreta con valori T={-3,-2,-1,0,1} e con probabilità: $p(x)=x^2/15$. Calcolare il valore atteso di Y=X(X-1).
b) Riprendere la trasformazione in variabile standardizzata discussa nel capitolo 3° e dimostrare che la variabile casuale $Z=[X-E(X)]/\sigma(x)$ ha valore atteso nullo E(Z)=0 e scarto quadratico medio uguale ad uno: $\sigma(Z)=1$.

Esercizio_VC33: *le misure di variabilità definite come scarti medi da un valore di riferimento:*

$$\sqrt[r]{E\left(|X-A|^r\right)} = \left[\sum_{i=1}^{k} \left(x_i - A\right)^r p_i\right]^{\frac{1}{r}}$$

sono state interpretate come misura dell'aleatorietà della vincita attesa di un gioco o del rischio connesso al rendimento medio di un portafoglio titoli. Quali sono i valori di A che rendono minimo il rischio se r=1 e se r=2?

Disuguaglianza di Tchebycheff

L'interpretazione dello scarto quadratico medio nell'ambito dei modelli probabilistici può ancora avvantaggiarsi della disuguaglianza di Tchebycheff:

$$p\left[|X - E(X)| \ge b\sigma(X)\right] \le \frac{1}{b^2} \Rightarrow p\left[|X - E(X)| \ge c\right] \le \left[\frac{\sigma(X)}{c}\right]^2 con\ \ c = b\sigma(X)$$

la cui dimostrazione e condizioni sono identiche a quelle svolta nel 3° capitolo per le frequenze relative (queste, ovviamente, sostituite con le probabilità); in particolare, l'esistenza finita di σ(X) è scontata nel caso di variabili discrete e finite.

La disuguaglianza di Tchebycheff, come si è già allora osservato, ha scarsa o nulla rilevanza pratica, ma è fondamentale come strumento di analisi teorica e nel corso di questo capitolo sarà richiamato per controllare un importante risultato sul postulato empirico del caso. Se lo scarto quadratico medio è piccolo, la probabilità che la variabile casuale si allontani dal valore atteso per più di una costante "c" tende a zero con l'aumentare di "c"; d'altra parte, per una fissata distanza "c" tra X ed E(X), la probabilità che la variabile assuma valori all'interno dell'intervallo [E(X)-c,E(X)+c] tende ad uno al ridursi dello scarto quadratico medio:

$$p\left[|X - E(X)| < c\right] \ge 1 - \left[\frac{\sigma(X)}{c}\right]^2 \qquad ovvero \qquad p\left[|X - E(X)| \ge c\right] \le \left[\frac{\sigma(X)}{c}\right]^2$$

Si può quindi affermare ch, $\sigma(X)$ misura l'addensamento dei valori della X intorno al valore atteso $E(X)$ fornendo anche la misura del rischio (probabilità avversa) per valori remoti rispetto al centro.

Esempi:
a) La variabile casuale discreta X ha la distribuzione riportata in tabella:

X	0	1	2	3	4	5	6	7
p(x)	0.05	0.08	0.14	0.27	0.20	0.13	0.11	0.02

$\mu = 3.42$; $\sigma=1.66$

Proviamo a dare un valore di soglia per l'intervallo [0.63, 6.21] cioè per c=1.68. La disuguaglianza di Tchebycheff propone $P(0.63<x<6.21)\geq 0.02$ che è molto minore della probabilità effettiva attribuita all'intervallo: 0.08+0.14+0.27+0.20+0.13+0.11=0.93. Non dobbiamo però trascurare che l'esito della Tchebycheff è ottenibile con pochi calcoli e poche informazioni.

b) Una società di assicurazione ha accertato che il tempo di risposta dei vigili del fuoco ha μ=15 minuti e σ= 3 minuti. Con quale probabilità massima risponderanno tra 10 e 20 minuti?

$$p\left[|X-E(X)|<b\sigma(X)\right]\geq 1-\frac{1}{b^2} \Rightarrow p\left[|X-15|<\frac{5}{3}3\right]=p(10<x<20)\geq 1-\left(\frac{3}{5}\right)^2=0.64$$

Esercizio_VC34: *una variabile casuale ha valore atteso E(X)=15 e momento secondo all'origine: $E(X^2)=230$.*
a) Determinare un valore di soglia inferiore per la probabilità dell'evento: $9<x<21$;
b) Determinare un valore di soglia superiore per la probabilità dell'evento $|X-15|\geq 7$;
c) Che considerazioni suggerisce il valore $E(X^2)=225$.

Esercizio_VC35: *la variabile casuale X assume valori in {0, 1, 2, 3} con rispettive probabilità 7/16, 3/16, 5/16, 1/16. Determinate il valore di soglia superiore della disuguaglianza di Tchebycheff per $p(|X-\mu|\geq 2)$.*

7.1.4 Modelli di variabili casuali discrete e finite

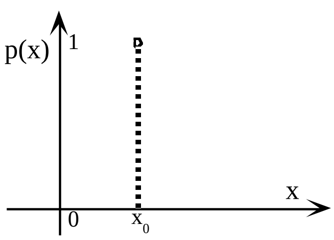
I modelli scaturiscono in genere da osservazioni di fenomeni reali, ma sono relativamente pochi quelli che hanno ricevuto consenso sufficiente per potersi estendere oltre le applicazioni che hanno portato alla loro costruzione. In questo paragrafo discutiamo i modelli più diffusi per probabilizzare gli aspetti di un esperimento casuale riconducibili ad un dominio discreto e finito: $T=\{x_1, x_2,\ldots,x_n\}$.

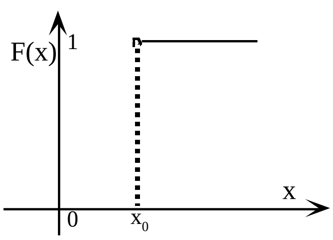
La variabile casuale degenere (modello di Dirac)

Una formulazione inusuale della distribuzione di probabilità è la variabile casuale di Dirac:

$$p(x)=\begin{cases}1 & se \;\; x=X_0\\ 0 & altrimenti\end{cases}$$

La densità si annulla dappertutto ad eccezione del punto $X=X_0$. Con questo non si vuole affermare che X sia costante, ma che l'evento $X \neq X_0$ in questo modello ha probabilità nulla. La differenza è sottile, ma c'è. Nonostante l'apparente inutilità, il modello di Dirac rappresenta bene il caso concreto di una rilevazione fatta senza errori ovvero di errori troppo piccoli perché valga la pena prenderli in considerazione. La funzione di ripartizione ha un unico salto a $X=X_0$. Infine, il valore atteso e lo scarto quadratico medio sono di calcolo immediato:





$$E(x)=X_0*1=X_0$$
$$\sigma(x)=\sqrt{(X_0-X_0)^2*1}=0$$

Nella parte del corso dedicata alla statistica inferenziale scopriremo che la distribuzione degenere è il modello ideale per la distribuzione delle statistiche chiamate a stimare i parametri incogniti di altre distribuzioni.

La variabile casuale binaria o dicotoma (modello di Bernoulli)

E' la variabile casuale non degenere più semplice in quanto il dominio include due sole modalità e non è suscettibile di ulteriore suddivisione.

$$X = \begin{cases} x_1 \quad con \quad probabilità \quad p \\ x_2 \quad con \quad probabilità \quad 1-p \end{cases}$$

Questo modello è di base per molte altre variabili casuali. Esso infatti interviene ad uno stadio preliminare in cui si colgono le caratteristiche più evidenti di un esperimento accertandone intanto la presenza o l'assenza per poi procedere ad una analisi più sofisticata cercando e analizzando maggiori dettagli anche con una successione di esperimenti binari. Di solito le due modalità sono riferite alla esistenza/non esistenza di uno *status*: un'auto ha gli *airbag* laterali con probabilità "p" oppure non li ha con probabilità (1-p), la biblioteca offre un servizio oppure no, un azionariato ha un investitore estero oppure no. Le condizioni dell'esperimento potranno determinare il valore del parametro "p": se c'è simmetria sarà p=1-p=1/2, ma altre indicazioni potrebbero venire da sperimentazioni passata oppure informazioni qualitative sull'esperimento. Altre volte la dicotomia si riferisce alle fasi *on/off* di un flusso: un prodotto è venduto oppure rimane sugli scaffali, un'impresa assume o non assume, la banca concede o non concede il fido. Spesso, sono aspetti complementari o contrapposti: positivo/negativo, promosso/riprovato, entrate/uscite, attivo/passivo, conservatore/innovatore, sinistra/destra, bianco/nero, perde/vince, sì/no, successo/insuccesso, vero/falso.

Esempi:

a) Le variabili *dummy* sono variabili fittizie che consentono di trattare con dei numeri veri e propri anche le variabili qualitative. La *dummy* si definisce con una espressione logica vero/falso

$$X(L) = \begin{cases} 1 \quad se \quad L \text{ è presente} \\ 0 \quad se \quad L \text{ è assente} \end{cases}$$

dove L è la modalità che si è verificata nella prova. Se X=Paese UE X(Svizzera)=0 e X(Grecia)=1.

b) La variabile dicotoma è usata come domanda filtro nei questionari (ISTAT, 1989, p. 48) per isolare sottogruppi di rispondenti cui proporre successivamente domande più specifiche. E' inutile chiedere che tipo di letture gradisce una data persona se prima non si è accertato che acquista un libro almeno una volta l'anno come è pure imbarazzante chiedere la destinazione estera preferita per le vacanze a chi non se le può permettere.

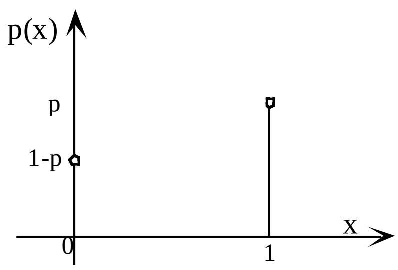
La funzione di distribuzione della variabile binaria è molto semplice:

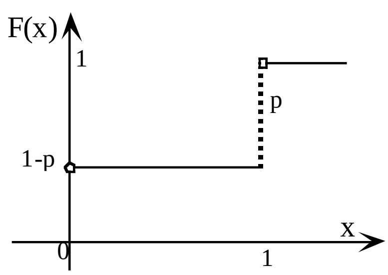
$$p(x) = \begin{cases} p^x (1-p)^{1-x} \quad per \quad x = 0,1 \\ 0 \qquad altrove \end{cases}$$

In questa formulazione si è posto $x_1=1$ e $x_2=0$, ma altre scelte sono possibili come si vedrà negli esempi e negli esercizi.

Anche la funzione di ripartizione di questo modello ha struttura elementare:

$$F(x) = \begin{cases} 0 \quad se \quad x < 0 \\ 1-p \quad se \quad 0 \le x < 1 \\ 1 \quad x \ge 1 \end{cases}$$





Valore atteso e varianza rispecchiano l'essenzialità del modello:

$$E(x)=\sum_{i=0}^{1} ip_i = 1*p+0*(1-p)=p;\quad \sigma^2(x)=\left(\sum_{i=0}^{1} i^2 p_i\right)-p^2 = 1*p+0*(1-p)-p^2 = p(1-p)$$

L'aspettativa è pari alla probabilità di successo e la varianza è proporzionale sia alla probabilità di successo che a quella di insuccesso. Il valore atteso non coinciderà mai con uno dei due valori a meno che il modello non degeneri e rimanga possibile una sola modalità e il nome di "valore atteso" è almeno inopportuno visto che non potrà mai realizzarsi e quindi è inutile aspettarselo. Analoga difficoltà interpretativa subentra se la dicotomia (0,1) è riferibile ad un aspetto in scala nominale. Rizzi e Fraire (1988, pp. 47-48) segnalano che le variabili binarie sono un tipo ibrido che possiede le proprietà di varie scale senza però possedere tutte quelle di una in particolare se non quella nominale. Bisogna immaginare un continuo tra i due aspetti della dicotomia per stabilire che -in base a "p" - il modello recepisce una tendenza a favorire la prima denominazione se $p>0.5$ e la propensione è tanto maggiore quanto più "p" si avvicina ad uno. La mediana del modello infatti, finisce sull'uno se $p\geq 0.5$ e sullo zero se $p<0.5$.

Esempi:

a) L'intepretazione del valore atteso come speranza matematica aveva come base un modello binario. La partecipazione ad un gioco dà luogo ad una variabile di questo tipo: si verifica l'evento su cui si è scommesso -che ha probabilità "p"- e si incassa la somma "a"; non si verifica -con probabilità (1-p) e si perde la somma giocata "b". L'aspettativa di guadagno è E(X)=ap -b(1-p). Il gioco è equo se E(X)=0 cioè se (b/a)=[p/(1-p)] cioè l'incasso in caso di vincita deve essere un multiplo della giocata esattamente pari al multiplo delle *chances* contro rispetto a quelle a favore. Il giocatore accorto giocherà sempre se la speranza matematica è positiva (per alcune persone -veramente poche- lo sono i giochi misti di abilità ed azzardo) e, in caso di scelta, preferirà il gioco con speranza matematica maggiore.

b) Una nuova concessionaria di auto nel primo anno di attività può avere due soli risultati: "fallimento" oppure "pareggio". La situazione del settore è tale che ci sono quattro possibilità su dieci di fallire. Santina De Rose intende avviare una nuova concessionaria. Qual'è la probabilità che sia ancora attiva all'inizio del 2° anno? Quali sono media e varianza dell'attività dopo un anno? Poniamo 0= fallimento, 1=pareggio; p(1)=0.6, p(0)=0.4, E(X)=0.60, σ^2(X)=0.60*0.40=0.24. Per interpretare il risultato può tornare utile l'idea di probabilità come convincimento sul succedere di un evento utilizzando il modello di Bernoulli come ragionevole astrazione del mercato. Il valore atteso μ=0.60 è il punto di equilibrio fra le ragioni a favore e le ragioni contro con una leggera preferenza per le prime.

c) Una versione del modello di Bernoulli è così strutturata: P(X=5)=1/3 e P(X=-5)=2/3. Il valore atteso riflette la tendenza più spiccata verso i valori negativi,

$$\mu = 5\left(\frac{1}{3}\right)-5\left(\frac{2}{3}\right)=-\frac{5}{3};\quad \sigma=\sqrt{25\left(\frac{1}{3}\right)+25\left(\frac{2}{3}\right)-\frac{25}{9}}=\sqrt{22.22}=4.71$$

Esercizio_VC36: *per il modello di Bernoulli con probabilità di successo "p".*
a) Calcolare i momenti intorno allo zero fino al 4° ordine;
b) Calcolare gli stessi intorno alla media aritmetica.

Esercizio_VC37: *lo scarto X da uno standard può essere -2 o +2 con probabilità, rispettivamente, di 0.45 e 0.55.*
a) Calcolare valore atteso e scarto quadratico medio della X;
b) Calcolare valore atteso e scarto quadratico medio di $Y=(X-1)^2$;
c) Qual'è la distribuzione di probabilità di Y=(X-1)(X+1)?

La variabile casuale uniforme discreta

Deriva direttamente dal modello di probabilità uniforme: tutte le modalità hanno la stessa probabilità:

$$X\in\{x_1,x_2,\ldots,x_n\};\qquad p(x_i)=\frac{1}{n}\quad per\ \ i=1,2,\ldots,n$$

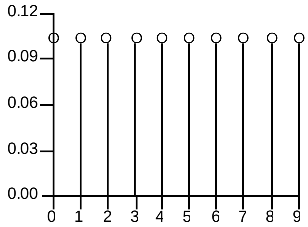
La singola probabilità è determinata dal numero di valori distinti della variabile casuale.

Esempi:

a) Il tetraedro è un poliedro con quattro facce formate da triangoli equilateri di uguale superficie (è uno dei precursori del dado, già noto nell'antica Roma come astragalo). Le simmetrie dell'esperimento: "lancio del tetraedro" portano ad assegnare probabilità 1/4 a tutte le facce.

b) Un selettore di numeri telefonici sceglie casualmente l'ultima cifra del numero. Senza ulteriori specificazioni si deve intendere che le 10 cifre {0, 1, … ,9} sono equiprobabili. Detta X la cifra selezionata, la sua distribuzione è:

$$P(x) = \frac{1}{10} \quad per \quad x = 0,1,\ldots,9$$



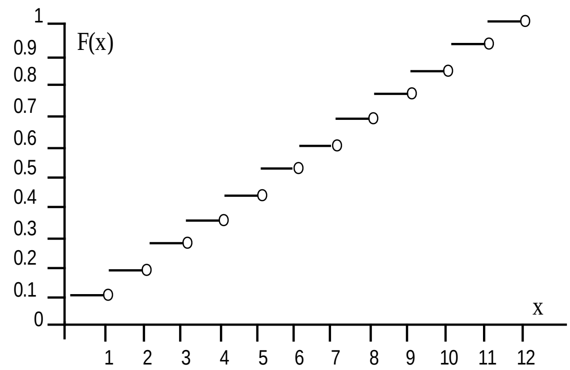
E' appena il caso di ricordare che il carattere metrico del dominio è richiamato solo se necessario. La natura di variabile casuale e in particolare di variabile casuale uniforme discreta non si modificherebbe se le cifre fossero delle etichette distintive in scala nominale oppure scalini di una graduatoria.

La funzione di ripartizione della uniforme discreta presenta "n" salti di altezza $p_i=p(x_i)=1/n$ corrispondenti alle modalità $\{x_i\}$.

Esempio:
L'astrologo Sibillinus ha costruito degli oroscopi base con 12 livelli sulla scala: "meglio che non uscite di casa/Il mondo è vostro". Ogni giorno permuta casualmente oroscopi e segni dello zodiaco integrando i primi con delle banalità perché non siano mere fotocopie dei prececedenti; pertanto, ogni segno ha probabilità 1/12 di ricevere un oroscopo base. Costruiamo la funzione di ripartizione. Le ascisse sono i livelli di ottimismo dell'oroscopo e le ordinate sono date -per ogni "i"- dalla somma delle probabilità dei livelli precedenti o uguali ad "i":

$$F(i) = \frac{i}{n}; \quad i = 1, 2, \ \ldots, \ n$$

La funzione è a gradini come è d'obbligo per le variabili casuali discrete che non hanno altre modalità nell'intervallo (x_i, x_{i+1})



Esercizio_VC38: *il nuovo tipo di scarpe è disponibile solo in 5 taglie: $X_i=35+2i$. Ipotizzando l'equiprobabilità della richiesta rappresentare la distribuzione di probabilità e la funzione di ripartizione della X.*

Esercizio_VC39: *un dodecaedro ha 12 facce uguali numerate da 1 a 12. Il materiale e la fabbricazione garantiscono la equiprobabilità dei risultati. Determinate:*
a) $p(X\leq6)$, b) $p(X>8)$; c) $p(2\leq X<9)$
Dove X è il valore della faccia coperta dopo aver lanciato casualmente il poliedro.

Nel modello uniforme discreto il valore atteso e lo scarto quadratico medio sono:

$$\mu = E(x) = \sum_{i=1}^{n} x_i p(x_i) = \sum_{i=1}^{n} x_i \left(\frac{1}{n}\right) = \frac{\sum_{i=1}^{n} x_i}{n}; \qquad \sigma = \sqrt{E(x-\mu)^2} = \sqrt{\sum_{i=1}^{n} x_i^2 p(x_i) - \mu^2} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \mu^2}$$

che ricalcano le analoghe espressioni per le rilevazioni empiriche che, lo scopriamo ora, non sono affatto neutrali, ma presuppongono uno specifico modello di probabilità: il modello uniforme che ha sì delle indicazioni, ma non è sempre il più adatto.

Esempi:
a) Uno studio di consulenza tributaria ha tariffato 5 tipi di servizi secondo le indicazioni date in tabella dove si riportano anche le aspettative di richieste secondo il modello uniforme. Calcoliamo μ e σ.

Servizio	X=Incasso	P(X=x)	F(x)
1	120	0.2	0.2
2	60	0.2	0.4
3	20	0.2	0.6
4	80	0.2	0.8
5	70	0.2	1.0

$$\mu = \frac{120+60+20+80+70}{5} = 70$$

$$\sigma = \sqrt{\frac{120^2+60^2+20^2+80^2+70^2}{5} - 70^2} = \sqrt{1040} = 32.249$$

b) Modello uniforme ed entropia. Consideriamo un modello di variabile casuale discreta finita $\{x_i, p_i; i=1,2,\ldots,n\}$. La moda del modello cioè la modalità che ci si aspetta più spesso nelle eventuali repliche dell'esperimento, è quella associata alla probabilità più grande: $\max\{p_i\}; i=1,2,\ldots,n$. La sua "tipicità" è massima in caso di variabile degenere dato che c'è un solo valore possibile e l'esito dell'esperimento è predeterminato con probabilità uno. Il grado di indeterminatezza cambia con le probabilità $\{p_i\}$ e per misurarlo potremmo richiamare uno degli indici di eterogeneità discussi nella prima parte del testo (par. 3.2). Tra questi, ha avuto maggiore risalto l'entropia che, per una variabile casuale, è espressa da:

$$h(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i Ln(p_i)$$

La $h(p_1,p_2,\ldots,p_n)$ è nulla per la distribuzione degenere: $h(p_1)=0$ e raggiunge il suo massimo $Ln(n)$ per la uniforme discreta $h(1/n,1/n,\ldots,1/n)=Ln(n)$. In altre parole, prevedere il valore della uniforme discreta in base alla rappresentatività del suo valore modale, è più difficile (in termini di entropia) che prevedere il valore di qualsiasi altra variabile discreta e finita.

Esercizio_VC40: *ottenere il momenti 3° e 4° intorno all'origine della uniforme discreta e verificare che si tratti di una distribuzione simmetrica. Sugg. Adoperate le formule delle progressioni degli interi.*

Esercizio_VC41: *i biglietti di una lotteria hanno una numerazione a base esadecimale: 0,1,2,…,9,A,B,C,D,E,F. Per la scelta di ciascuna cifra della serie sono stati inseriti 16 bussolotti -ognuno contenente uno dei numeri esadecimali- fatti poi ruotare vorticosamente grazie ad un congegno elettronico. Se con X indichiamo il primo numero della serie qual'è la media e qual'è lo scarto quadratico medio?*

La variabile casuale uniforme discreta è anche esprimibile con i numeri naturali: $x_i=i$, $i=1,2,\ldots,n$; per questa formulazione, valore atteso e varianza richiamano alcune semplificazioni:

$$E(X) = \frac{\sum_{i=1}^{n} i}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2};\; E(X^2) = \frac{\sum_{i=1}^{n} i^2}{n} = \frac{n(n+1)(2n+1)}{6} \Rightarrow \sigma^2(X) = \frac{n^2-1}{12}$$

Esempio:
Un'urna contiene sette biglie numerate da 1 a 7. L'esperimento consiste nello scegliere a caso una biglia e di annotare l'esito della scelta come variabile casuale X. In mancanza di altre informazioni riteniamo equiprobabili i valori della X. Valore atteso e varianza sono: (7+1)/2=4 e (49-1)/12=4.

Esercizio_VC42: *la finale di un torneo si gioca al meglio dei cinque e cioè: non appena una delle squadre ha raggiunto la terza vittoria è dichiarata vincitrice. La finalista A ha probabilità "p" di vincere una qualsiasi delle gare e le gare sono indipendenti.*
a) Se N è la variabile casuale che indica il numero di partite giocate, qual'è il suo valore atteso?
b) Per quale valore di "p" E(N) è massimo?

Esercizio_VC43: *un dato può avere un numero di bit X ricadente nell'insieme X={2, 4, 8, 16, 32, 64}. Se le lunghezze in bit fossero equiprobabili:*
a) Quali sarebbero valore atteso e scarto quadratico medio della X;
b) Quale media sarebbe più appropriata per sintetizzarne la centralità?

La variabile casuale binomiale

Consideriamo un esperimento con risultato binario S={0,1} e p(1)=p e p(0)=1-p. Una "sequenza di prove bernoulliane" è un esperimento composto da "n" sub-esperimenti indipendenti la cui funzione di probabilità non cambia nelle "n" prove (ed ognuna potrebbe essere descritta da ua variabile casuale di Bernoulli). La variabile casuale binomiale X è generata in tale contesto e descrive il numero di successi potenziali nelle "n" prove. Per costruirne la distribuzione di probabilità partiamo dal fatto che l'universo degli eventi è formato dalle 2^n sequenze ordinate di valori binari.

Ad una sequenza formata da x "1" nelle prime x prove e da (n-x) "0" nelle restanti è assegnata, secondo il modello moltiplicativo, la probabilità:

$$P\Big(\overbrace{1\cap 1\cap\ldots\cap 1}^{x\ volte}\cap \overbrace{0\cap 0\ldots\cap 0}^{n-x\, volte}\Big) = p^x(1-p)^{n-x}$$

La stessa probabilità deve essere attribuita ad ogni sequenza con lo stesso numero di “1” e di “0” dato che ciò che rileva il numero complessivo di successi e non la loro sequenza.

$$p(x) = \begin{cases} p(x) = \binom{n}{x} p^x (1-p)^{n-x} \; per \quad x = 0,1,2,\ldots,n \\ 0 \qquad\qquad\qquad\qquad altrove \end{cases}$$

Nella formula sono presenti due parametri “n” e “p” cioè numero di prove e probabilità di successo nella singola prova. Una notazione sintetica è B(x;n,p).

Esempi:

a) Per fronteggiare i virus rinforzati comparsi negli ultimi tempi si sperimenta un antibiotico di nuova generazione su “n” pazienti affetti da influenza. L’esito può essere “successo: il paziente migliora” o “insuccesso: il paziente rimane stazionario”. Se i pazienti non hanno rapporti di parentela e sono ospitati in modo da entrare in contatto solo con personale medico immune, possiamo ipotizzare l’indipendenza delle prove e quindi arriviamo all’esperimento in cui è definita la binomiale. C’è però ancora l’ostacolo logico della probabilità di guarigione costante. E’ un evidente forzatura: ogni paziente è in condizioni fisiche diverse, ha un metabolismo particolare ed una sua anamnesi che, a rigor di logica, dovrebbe portare a tante “p” quanti sono i pazienti. L’applicazione del modello binomiale è giustificata se le differenze tra le varie probabilità di guarigione sono poco rilevanti.

b) Lo stato di un prodotto è indicato con 1=difettoso, 0=idoneo; inoltre: P(1)= p= 0.05. Qual’è la probabilità di individuare 2 difettosi in una successione di n=10 prodotti?

$$P(X=2) = \binom{10}{2} 0.05^2 (0.95)^8 \cong 7.5\%$$

c) Una coppia di dadi regolari è lanciata per sei volte. Qual’è la probabilità che il “7” compaia esattamente tre volte?

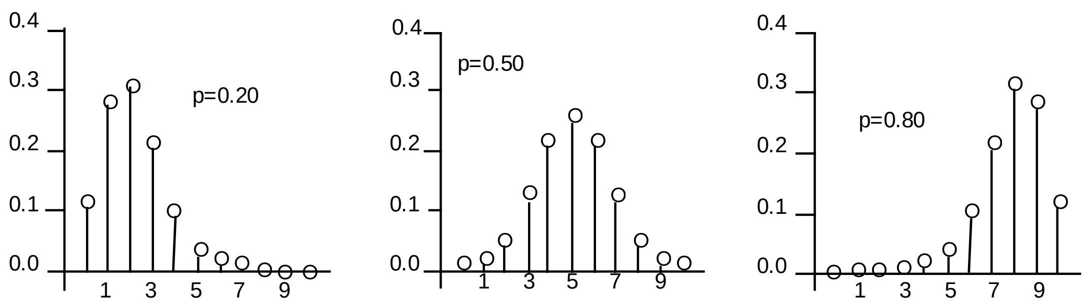
$$P(X=3) = \binom{6}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^3 \cong 5.36\%$$

Il modo con cui è stata ottenuta la funzione di distribuzione potrebbe far sorgere il dubbio che non si tratti di una vera e propria distribuzione di probabilità o che questo possa dipendere dal valore di “p”. Per accertarsene basta constatare che p(X=x)≥0 visto che sono coinvolti solo prodotti di numeri non negativi ed inoltre:

$$\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{0}(1-p)^n + \binom{n}{1} p\;(1-p)^{n-1} + \ldots + \binom{n}{n-1} p^{n-1}(1-p) + \binom{n}{n} p^n$$
$$= \left[p + (1-p)\right]^n = 1^n = 1$$

Esempio:

Tre grafici della distribuzione di probabilità della binomiale per n=10.



Il grafico presenta asimmetria a sinistra se p<0.5 ed asimmetria a destra per p>0.5 (pur nella particolare valenza che l’asimmetria assume nelle distribuzioni discrete). Per p=0.5 la distribuzione binomiale è simmetrica. Infatti:

$$p(n-x) = \binom{n}{n-x} 0.5^{n-x} 0.5^x = \binom{n}{x} 0.5^x 0.5^{n-x} = p(x)$$

Esercizio_VC44: *per calcolare le probabilità della binomiale è possibile escogitare una formula ricorsiva tale che p(x+1)=k(x)*p(x). Qual'è il fattore k(x)?*

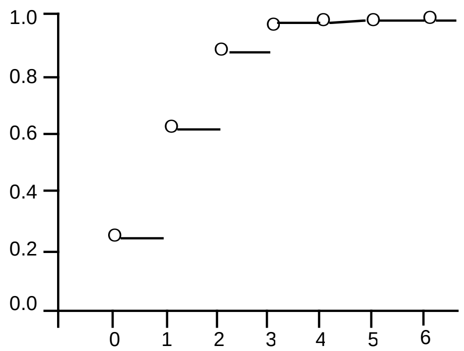
La funzione di ripartizione della binomiale esprime la probabilità di ottenere un numero di successi inferiori o uguali ad una soglia prefissata:

$$F(x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}, \quad x = 0,1,\ldots,n$$

Esempi:

a) Una scatola contiene 5 dischetti (non visibili dall'esterno) di cui uno di colore nero e 4 di colore bianco. Si estraggono -con reimmissione- 6 dischetti. Sia X= "numero di volte che si estrae il dischetto nero". Quali solo le probabilità che X≤0, 1,2,…,6? Basta applicare la formula della probabilità binomiale per ogni valore di "i" e poi procedere all'accumulo ricorsivo delle probabilità:

$$F(x) = \sum_{i=0}^{x} \binom{6}{i} \left(\frac{1}{5}\right)^i \left(\frac{4}{5}\right)^{6-i}$$



b) Un test basato su quiz a scelta multipla è articolato su 25 domande ognuna con quattro risposte di cui solo una è corretta. Il test si supera rispondendo bene ad almeno 13 quiz (la metà più uno). Isidoro è totalmente impreparato, ma -tenuto conto che non ci sono penalità nelle risposte sbagliate- vorrebbe tentare lo stesso rispondendo a casaccio a tutte le domande, ma in modo che la risposta data ad un quiz non sia influenzata da quelle già date e non influenzi quelle ancora da dare. Qual'è la probabilità che Isidoro passi il test?

$$P(X \geq 13) = 1 - P(X < 13) = 1 - F(12) = 1 - \sum_{x=0}^{12} \binom{25}{x} 0.25^x 0.75^{25-x} = 1 - 0.99663 = 0.33\%$$

Su mille che tentano -nel ripetersi infinito dei mille tentativi- il numero di chi supera il test rispondendo a caso è tre. Isidoro spera che ciò avvenga nel suo migliaio e di essere uno dei tre.

Esercizio_VC45: *la probabilità di sopravvivenza di un'albero in un nuovo impianto è del 90%. Quanti alberi occorre piantare perché ne sopravvivano almeno 6 con una probabilità del 99%?*

Esercizio_VC46: *una commissione è formata da 15 membri, ma opera validamente con la maggioranza qualificata di 2/3 presenti all'atto di una votazione. Se la probabilità che ogni membro sia presente è del 75% e le assenze/presenze sono indipendenti, qual'è la probabilità che un voto abbia il numero legale?*

Esercizio_VC47: *all'incrocio tra le due file di un supermercato arrivano 18 clienti. Con probabilità del 35% il cliente svolta verso i prodotti tecnologici e del 65% verso gli altri reparti (si può ritenere che i clienti agiscano indipendentemente). Nel settore tecnologico è necessario un operatore se i clienti sono almeno 7 ed è sufficiente fino a 13 (da 14 in poi ne sono necessari due). Qual'è la probabilità che un solo operatore bisogni e basti?*

Momenti fattoriali

La determinazione di media e varianza è facilitata dalla formula dei momenti fattoriali, che, in generale, sono:

$$\mu_{r!} = \sum_{i=1}^{k} \prod_{j=0}^{r-1} (x_i - j) p_i; \quad \mu_{1!} = \mu; \quad \mu_{2!} = \sum_{i=1}^{k} x_i (x_i - 1) p_i = \sum_{i=1}^{k} x_i^2 p_i - \sum_{i=1}^{k} x_i p_i = \mu_2 - \mu;$$
$$\mu_{3!} = \mu_3 - 3\mu_2 + 2\mu; \quad \mu_{4!} = \mu_4 - 6\mu_3 + 11\mu_2 - 6\mu;$$

Vediamo come possono servire per il calcolo dei momenti della binomiale.

$$\mu_{1!} = \sum_{i=0}^{n} i\left[\frac{n!}{i!(n-i)!}\right]p^i(1-p)^{n-i} = \sum_{i=0}^{n} i\left[\frac{n(n-1)!}{i(i-1)!(n-i)!}\right]pp^{i-1}(1-p)^{n-i} = np$$

$$\mu_{2!} = \sum_{i=0}^{n} i(i-1)\left[\frac{n!}{i!(n-i)!}\right]p^i(1-p)^{n-i} = \sum_{i=0}^{n} i(i-1)\left[\frac{n(n-1)(n-2)!}{i(i-1)(i-2)!(n-i)!}\right]p^2p^{i-2}(1-p)^{n-i}$$
$$= n(n-1)p^2$$

$$\mu_2 = \mu_{2!} + \mu = n(n-1)p^2 + np \Rightarrow \sigma^2 = \mu_2 - \mu^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p)$$

Esempi:

a)Il monitoraggio di una linea di comunicazione ha asseverato che la probabilità di turbolenza in un messaggio è del 10%; scegliendo a caso n=20 trasmissioni, qual'è la media delle comunicazioni disturbate che ci si deve aspettare con quella percentuale di "successo"? p=0.1, n=20 e μ=np=2. con scarto $[np(1-p)]^{0.5}$=1.34.

b) La corsia di un ospedale prevede "n" posti letto e la probabilità che uno dei ricoverati abbia bisogno della stanza singola è "p". Supponendo vigenti le condizioni del modello binomiale, la probabilità che "r" pazienti debbano essere isolati è:

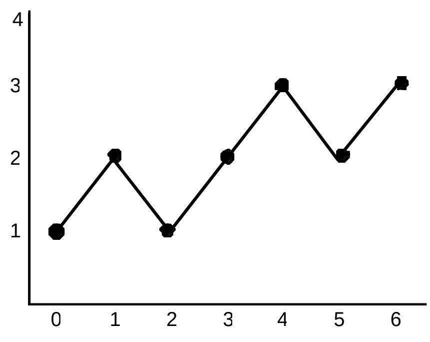
$$p_r = \binom{n}{r}p^r 0.5^{n-r}$$

$$S = \sum_{r=0}^{m} rp_r + \sum_{r=m+1}^{n} mp_r$$

m	E_d	E_u
1	0.28	0.98
2	0.39	0.94
3	0.54	0.87
4	0.64	0.77
5	0.70	0.67
6	0.72	0.57
7	0.73	0.50

Se si dispone di "m" stanze singole, il numero medio di pazienti che può esservi accomodato è S. Il primo termine è la media delle richieste soddisfatte e l'altro la media di quelle inevase. Bailey (1977, pp.162-163) chiama efficienza di disponibilità ed efficienza d'uso i due rapporti: E_d=S/(np) e E_u=S/m. Se p=0.25 e n=20 abbiamo la tabella riportata a destra. La disponibilità di m=5 è quella che genera minore divario tra l'accomodamento dei pazienti da isolare (70%) e l'uso standard dei posti letto (67%). Questo può essere migliorato isolando un posto letto in meno, ma con aumento del 6% per il rischio di non soddisfare le richieste.

c) Le particelle sospese in un fluido hanno un movimento erratico dettato dagli impatti con le molecole del fluido (il movimento delle particelle è noto come moto browniano). Un modello rudimentale di tale processo è il *random walk* (passeggiare a caso) in cui ogni passo è una prova bernoulliana in un sistema di prove indipendenti e probabilità di successo costante. Applichiamolo alle decisioni economiche. Un'impresa nel periodo "0" è in attesa di notizie dai concorrenti. Se le notizie sono buone alzerà il prezzo di una unità (+1), se sono cattive lo diminuirà di una unità (-1); comunque, il prezzo non rimarrà invariato. Ogni decisione è una prova bernoulliana ritenendosi che gli episodi avvengano in condizioni troppo diverse per influenzarsi e la probabilità "p" di un aumento rimane invariata.



Qual'è la probabilità che dopo "n" decisioni il percorso del prezzo si trovi "r" punti sopra lo zero? Ciò si verifica se il numero n_1 di segni "+" soddisfa n_1-(n-n_1)=r cioè n_1=(n+r)/2. La probabilità è data da B([X=(n+r)/2];n,p]. Per la traiettoria in figura con n=7 e p=0.4 la probabilità è: B(5;7,0.4)=7.74%. In media il prezzo si muove di μ=7*0.4=2.8 cioè circa 3 punti con uno scarto σ=1.23.

d) L'esame di Ragioneria è basato su "n" quiz. La risposta corretta ad uno dei quiz aggiunge "r" punti e quella sbagliata sottrae "w" punti. Indichiamo con X il numero di risposte corrette. Il valore atteso del punteggio nell'ipotesi di risposte casuali e indipendenti è:

$$E[rX-(n-X)w] = rE(X) - nw + wE(X) = (w+r)E(X) - nw$$

Se le opzioni di ogni risposta sono "k" si ha E(X)=n/k. La composizione equilibrata del test impone che il punteggio atteso sia zero (assenza di vantaggi o svantaggi sistematici per chi lo sostiene alla cieca). Ciò implica:

$$(w+r)\frac{n}{k} - nw = 0 \Rightarrow w = \frac{r}{k-1}$$

se r=2 e k=4 allora la risposta sbagliata deve sottrarre w=2/3. Si osservi che il risultato non dipende dal numero dei quiz "n". E' chiaro che se si riesce ad eliminare una opzione per ogni domanda l'aspettativa diventa positiva: (2/3+2)(n/3)-n(2/3)=2n/9 e converrà rispondere invece di lasciare in bianco le risposte.

Esercizio_VC48: *usare i momenti fattoriali per determinare il momento terzo centrale e verificare che è nullo per p=0.5. Con la stessa tecnica determinare il coefficiente* $\gamma_2 = E[(X-\mu)/\sigma]^4$ *ed indicare il valore di convergenza per "n" tendente all'infinito;*

Esercizio_VC49: *per quale valore di "p" è massima la varianza della binomiale?*

Esercizio_VC50: *la conoscenza dei parametri "n" e "p" del modello binomiale consente non solo di stabilire il valore atteso, ma anche il rischio (variabilità) connesso con l'esito medio. Calcolare la probabilità che nel modello binomiale con n=16 e p=0.5 le determinazioni della X rimangano nell'intervallo [6, 10];*

Variabile casuale frequenza di successi

La variabile casuale H=X/n è la frequenza relativa del numero di successi su "n" prove:

$$P(H = x) = P\left(\frac{X}{n} = x\right) = P(X = nx)$$

dove X è una binomiale B(x;n,p). Media e scarto quadratico non presentano difficoltà:

$$E(H) = \sum_{i=0}^{n} \frac{i}{n}\binom{n}{i} p^i (1-p)^{n-i} = \frac{1}{n}\sum_{i=0}^{n} i\binom{n}{i} p^i (1-p)^{n-i} = \frac{np}{n} = p; \;\; \sigma(H) = \sqrt{\frac{p*(1-p)}{n}}$$

Il valore atteso della frequenza è la probabilità di successo e la deviazione standard si riduce in ragione inversa del numero delle prove. La portata di questo risultato è notevole.

Teorema di Bernoulli

Riprendiamo la disuguaglianza di Tchebycheff:

$$P\left[\left|H - p\right| < \varepsilon\right] \geq 1 - \frac{p*(1-p)}{n\varepsilon^2}$$

Per quanto piccolo sia ε sarà sempre possibile trovare un numero di prove "n" tale che la probabilità di uno scarto tra frequenza relativa e probabilità, inferiore ad ε, tenda ad uno (evento certo).

$$\underset{n\to\infty}{Lim} P\left(\left|H_n - p\right| \geq \varepsilon\right) < \frac{p*(1-p)}{n\varepsilon^2} = 0$$

Questo è noto come teorema di Bernoulli o legge dei grandi numeri. Sembra quasi banale, ma Bernoulli - ignorando il risultato di Tchebycheff che arriverà dopo più di un secolo e mezzo- vi lavorò per vent'anni. Esso non deve essere però inteso come un avallo dell'approccio frequentista alla probabilità, ma un segno di coerenza dell'impalcatura di Kolmogorov. Da notare infine che il limite riguarda la probabilità dell'evento e non l'evento: il teorema di Bernoulli stabilisce che la probabilità che H_n differisca da "p" più di un ammontare prefissato tende a zero e non che la differenza $|H_n-p|$ tende a zero perché questo può essere vero o falso per ogni "n", anche per successioni infinitamente grandi di prove bernoulliane.

Esercizio_VC51: *determinare la distribuzione di K=1-H cioè della frequenza di insuccessi in una successione di "n" prove bernoulliane.*

Esercizio_VC52: *dimostrare che:*

$$P\left[\left|H - p\right| \geq \varepsilon\right] \leq \left(\frac{1}{2n\varepsilon}\right)^2$$

La variabile casuale ipergeometrica

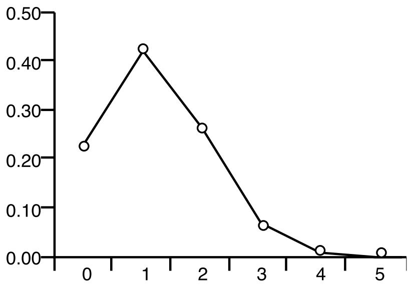
Il modello binomiale descrive la scelta -con reimmissione- di un numero finito di "n" biglie da un'urna che ne contiene un numero altrettanto finito N ed in cui le biglie possono essere di due tipi: un certo numero N_1 è speciale perché verifica una qualche proprietà ed il resto sono comuni. Qual'è il modello appropriato se la selezione è senza reimmissione? Poiché ricorre lo schema ipergeometrico la probabilità cercata è:

$$p(x) = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}; \quad x = 0,1,2, \ldots, n;$$

Tale modello è definito variabile casuale ipergeometrica a causa dell'analoga funzione matematica che si affaccia allorché si debbano definire i momenti della distribuzione.

Esempi:

a) Il monitoraggio di una produzione prevede la selezione casuale -senza reimmissione- di "n" prodotti da un lotto di N. Se si trova uno o più prodotti difettosi il lotto è scartato. Sia X il numero di prodotti difettosi. Il fatto che X=0 non implica che siano assenti i prodotti difettosi perché la sorte potrebbe aver scelto solo quelli privi di difetti. La probabilità che nella selezione di "n" item ci siano esattamente 0, 1, 2, …, n prodotti difettosi dipende da N_1 cioè dai difettosi presenti nella partita.



Nel grafico è riportato l'andamento della probabilità per N=37, N_1=7 ed n=5. In particolare la probabilità che il lotto sia accettato nonostante la severità della regola ed i 7 prodotti difettosi è del 23%.

b) In un controllo casuale su n=1000 lavoratori extracomunitari non in regola, alcuni (r=200) sono identificati ed espulsi. In un successivo controllo, su m=400 irregolari, k=80 avevano già subito il decreto di espulsione. Quanti sono i cosiddetti "sans papier"? L'incognita è l'ampiezza N della popolazione (si veda il paragrafo 1.2.1 sulle popolazioni indeterminate ed elusive). Si ipotizza che il campione casuale semplice sia ottenuto senza reimmissione. Le unità speciali sono i "già fermati" e quelle comuni i "mai fermati" . La distribuzione di probabilità delle persone rifermate è ipergeometrica con:

$$P(X=80) = \frac{\binom{1000}{80}\binom{N-1000}{320}}{\binom{N}{400}} = \frac{1000!\,200!\,(N-1000)!\,400!\,(N-400)!}{80!\,920!\,320!\,(N-1320)!\,N!}$$

Esiste un valore di N che rende massima tale probabilità? Questo è un problema di statistica inferenziale la cui soluzione è di una semplicità disarmante: N=n(r/k) che, nell'esempio, significa N=2'500.

Esercizio_VC53: *una lotteria gratta-e-vinci presenta sestine formate da cerchi azzurri e cerchi rosa: la vincita è proporzionale al numero di cerchi. Le sestine sono formate con scelte senza reimmissione di 6 biglie da un'urna che ne contiene N=45; 6 di esse sono rosa e azzurre le altre. Costruite e rappresentate la distribuzione di probabilità e di ripartizione delle possibili vincite.*

Esercizio_VC54: *per un difetto nella colla le etichette di 40 scatole di legumi si sono staccate per cui non si conosce più il loro contenuto. Dalla fattura di apprende che il 25% era di piselli ed il 75% di fagioli. Per controllo ne vengono aperte 10 e si scopre che ve ne sono 8 di piselli. Che la fattura sia sbagliata?*

Esercizio_VC55: *fra le 100 città che il movimento politico "GrandeJuve" sta considerando per tenere i prossimi 5 meeting elettorali 56 sono nel Nord e 44 nel Centrosud. Per evitare questioni campanilistiche si lascia decidere alla sorte. Qual'è la probabilità che siano scelte in maggioranza città del Nord?*

Le probabilità fornite dal modello ipergeometrico sono non negative, ma sommano ad uno? Per accertarsene occorre richiamare un risultato sui coefficienti binomiali già presentato nell'esercizio TP50:

$$\sum_{i=0}^{n}\binom{N_1}{i}\binom{N-N_1}{n-i}=\binom{N}{n}$$

poiché la somma dei numeratori è pari al denominatore, il modello ipergeometrico è una legittima distribuzione di probabilità.

La sua funzione di ripartizione è:

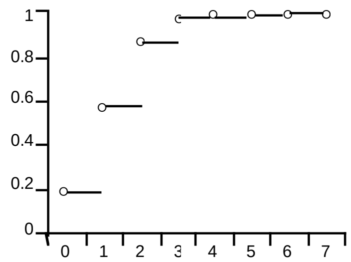
$$F(x)=\frac{\sum_{i=0}^{n}\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}};\quad 0\le x\le n$$

con qualche difficoltà di calcolo in più, rispetto alla binomiale, per la presenza di tre coefficienti binomiali e dei conseguenti fattoriali.

Esempi:

a) In una *mailing list* ci sono 50 indirizzi di cui 10 sono di società. In un sondaggio di opinione si scelgono a caso e senza reimmissione n=7 indirizzi. Rappresentiamo la funzione di ripartizione del numero di società nella scelta.

$$F(x)=\frac{\sum_{i=0}^{n}\binom{10}{x}\binom{40}{7-x}}{\binom{40}{10}};\quad x=0,1,\dots 10$$



Anche per la F(x) riferita alle probabilità, la ripidità di un gradino è segno dell'importanza della modalità. Nel caso in esempio si vede chiaramente che per X=2 si ha l'incremento maggiore di probabilità e quindi X=2 è la modalità più probabile della variabile casuale.

b) Durante le operazioni di scrutinio si è giunti alla 350ª scheda; la lista "Calabria Unita" ha ottenuto 214 voti e la lista "Calabria Libera" 136. Un guasto all'impianto elettrico del seggio interrompe l'illuminazione per un'ora. Alla ripresa dello spoglio si riscontrano altre 200 schede per un totale di 550 votanti con esito finale: Calabria Unita 270 voti, Calabria Libera 280. Gli unitaristi protestano, gridano all'imbroglio e fanno ricorso. In attesa del responso della commissione elettorale calcoliamo la probabilità che con schede scrutinate a caso, gli unitaristi ottengano voti inferiori o uguali a 56. Usiamo il modello ipergeometrico con N=550, N_1=270, n=200

$$P(X\le 56)=\sum_{x=0}^{56}\frac{\binom{270}{x}\binom{280}{200-x}}{\binom{550}{200}}=0.00093$$

C'è solo una probabilità su mille che il ribaltamento dell'esito sia dovuto al caso. Forse il controllo non è inutile.

Esercizio_VC56: *in una confezione di 21 CD ne sono presenti -per errore- 5 non riscrivibili. Se ne vengono scelti a caso e senza reimmissione 7 per realizzare un back-up qual'è la probabilità che, fra questi, i non riscrivibili siano tre o più?*

Esercizio_VC57: *la MacBartel S.a.d. gestisce un corso di preparazione per un importante concorso. Il corso è per 28 persone che hanno pagato cifre esorbitanti. Una coincidenza porta l'amministratore della società a presiedere una commissione del concorso ed un'altra coincidenza porta i 28 corsisti ad essere esaminati da questa. Tra gli 80 candidati esaminati dalla commissione risultano 20 idonei e tra questi 18 hanno fruito del piano formativo MacBartel.*
a) Se l'idoneazione fosse equiprobabile rispetto al corso, quale sarebbe la probabilità di un tale evento?
b)Qual'è la probabilità che i frequentanti MacBartel siano la maggioranza?

Il valore atteso e la varianza della ipergeometrica possono essere ottenute con la stessa procedura adoperata per la binomiale:

$$E(X)=\sum_{x=0}^{n} x\frac{N_1!(N-N_1)!n!(N-n)!}{x!(N_1-x)!(n-x)!(N-n_1-n+x)!N!}=n\left(\frac{N_1}{N}\right)\sum_{x=0}^{n-1}\frac{\binom{N_1-1}{x-1}\binom{N-N_1}{n}}{\binom{N-1}{n-1}}=n\left(\frac{N_1}{N}\right)$$

$$\sigma^2(X)=n\left(\frac{N_1}{N}\right)\left(\frac{N-N_1}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Esempi:

a) La variabile casuale X assume solo valori interi con probabilità data da

$$p(x)=\frac{\binom{3}{x}\binom{4}{3-x}}{35};\quad x=0,1,2,3;$$

Calcolare il valore atteso e la deviazione standard di X. Poiché si tratta di una variabile casuale ipergeometrica con N_1=3, N=7 ed n=5 la risposta è semplice: μ=5*3/7=2.14; σ=$(2.14*8/42)^{0.5}$=0.64.

b) Fra gli n=20 voti più alti di un corso universitario erano incluse 15 donne e 5 maschi (l'attestato di frequenza era stato rilasciato a 85 donne e 50 maschi). Qualcuno mormorò sulla discriminazione a danno dei maschi. Se i buoni voti fossero dati a caso rispetto al sesso il valore atteso del numero di donne nei primi 20 posti è μ=20*85/135=12.6 con σ=2. In pratica, le 15 donne sono leggermente superiori alla soglia di allerta di un *boxplot* formato con media e scarto. La lamentela non è seriamente fondata.

Esercizio_VC58: *la criminologa Julia ha individuato 11 possibili colpevoli del delitto al ristorante cinese, ma ha il tempo di occuparsi solo di 6 (è accertato che il delitto è stato commesso da 3 persone). Si ipotizza che Julia scelga a caso i suoi sospetti.*
a) Rappresentare la distribuzione di probabilità del numero di colpevoli che individua correttamente;
b) Calcolare valore atteso e deviazione standard dei colpevoli individuati.

Approssimazione binomiale della ipergeometrica

C'è una forte somiglianza tra ipergeometrica e binomiale che diventa più evidente se si pone p=(N_1/N) cioè probabilità di successo pari alla probabilità di scelta di una biglia speciale:

$$E(X)=np;\qquad \sigma^2(X)=np(1-p)\left(\frac{N-n}{N-1}\right)$$

Il valore atteso coincide; la varianza è invece corretta per un fattore dovuto alla mancata rimessa.

Esempi:

a) In un mazzo di carte francesi con N=52 si scelgono -senza reimmissione- 10 carte. Quali sono media e varianza del numero di carte di picche nella scelta? N_1=13, n=10 e quindi μ=10*(13/52)=2.5 cioè da due a tre carte di picche in media per ogni mano di dieci carte. La varianza è: 10*(13/52)(39/52)*42/51=1.54.

b) La preselezione per alcune posizioni di vice-dirigente prevede 24 argomenti tra i quali ne verranno scelti -ovviamente senza reimmissione- 8 per il colloquio. Si è ammessi alla fase successiva se si risponde bene ad almeno 4 argomenti. Lucia non si impegna (lei aspira almeno alla dirigenza) e prepara 4 argomenti tra i più facili. Qual'è il numero medio e lo scarto quadratico degli argomenti scelti da Lucia? μ=8*4/24=1.33, σ^2=8*(4/24)*(20/24)*(16/24)=0.773.

Esercizio_VC59: *le fatture emesse dalla "Cartiera s.r.l." verso la società "Lower & C." sono state 50. Si sospetta che il 22% sia falso. Il revisore dei conti ne sceglie 9 fra le 50 senza reimmissione per l'analisi completa.*
a) Calcolare il valore atteso e la varianza del numero X di fatture false presenti nel campione;
b) Qual'è la soglia minima che la disuguaglianza di Tchebycheff assegna all'evento $0\le X\le 4$*?*

Il modello ipergeometrico è poco gestibile per ampiezze elevate a causa della crescita dei fattoriali. Se però N diventa grande nei confronti di "n" il fattore di correzione della varianza tende ad uno: per N=1000 e n=10 è pari a 0.991 per cui media e varianza della ipergeometrica differiscono poco da quelle della binomiale.

E' possibile che le probabilità della ipergeometrica siano ben approssimate dalla B(x;p, n) se il rapporto n/N è piccolo? La risposta è positiva visto che la discriminante tra i due modelli passa per la reimmissione/non reimmissione e che questa diventa ininfluente quando "n" è piccolo rispetto ad N. Ipotizziamo che, per un fissato valore di "n", N_1 ed N tendano all'infinito, ma in modo che N_1/N converga ad un valore definito: $0 \le p \le 1$.

$$\frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} = \frac{N_1!(N-N_1)!}{x!(N_1-x)!(n-x)!(N-N_1-n+x)!} * \frac{n!(N-n)!}{N!}$$

$$= \binom{n}{x} \frac{\left[\overbrace{N_1(N_1-1)\ldots(N_1-x+1)}^{x \; fattori}\right] * \left[\overbrace{(N-N_1)(N-N_1-1)\ldots(N-N_1-n-x+1)}^{n-x \; fattori}\right]}{\underbrace{N(N-1)\ldots(N-n+1)}_{n \; fattori}}$$

$$= \binom{n}{x}\left[\left(\frac{N_1}{N}\right)\left(\frac{N_1}{N}-\frac{1}{N}\right)\ldots\left(\frac{N_1}{N}-\frac{x-1}{N}\right)\right] * \left[\left(1-\frac{N-N_1}{N}\right)\left(1-\frac{N-N_1}{N}-\frac{1}{N}\right)\ldots\right.$$
$$\left.\left(1-\frac{N-N_1}{N}-\frac{n-x-1}{N}\right)\right] * \left[\frac{N^n}{N(N-1)\ldots(N-n+1)}\right]$$

Man mano che N ed N_1 diventano grandi (purché si abbia la convergenza di N_1/N a "p") l'ultima espressione si avvicina all'unità perché è formata dal rapporto di "n" fattori dello stesso ordine di grandezza. Inoltre, per ogni "x" gli addendi (x/N) si riducono a zero dato che $x \le n$ ed "n" è fisso; infine, N_1/N e $(N-N_1)/N$ tendono a "p" e (1-p). Quindi, per valori grandi, l'ipergeometrica è approssimabile con la binomiale:

$$\frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} \to \binom{n}{x}\left[\overbrace{\left(\frac{N_1}{N}\right)\left(\frac{N_1}{N}\right)\ldots\left(\frac{N_1}{N}\right)}^{x \; vole}\right] * \left[\left(1-\frac{N_1}{N}\right)\left(1-\frac{N_1}{N}\right)\ldots\left(1-\frac{N_1}{N}\right)\right]^{n-x}$$

$$\to \binom{n}{x} p^x (1-p)^{n-x}$$

Esempi:

a) L'ufficio imposte ha riscontrato 500 dichiarazioni IRPEF di una categoria a rischio e 350 hanno dovuto pagare tasse in più. L'ispettrice Mariacarmela Turco decide di approfondire la posizione di 10 dichiarazione per un controllo di 2° livello. Qual'è la probabilità che ne trovi 6 già riscontrate evasive?

$$Ipergeometrica: \; p(X=6) = \frac{\binom{350}{6}\binom{150}{4}}{\binom{500}{10}} = 0.2016; \;\; Binomiale: \; p(X=6) = \binom{10}{6} 0.7^6 0.3^4 = 0.2001$$

I risultati ottenuti con i due modelli sono praticamente uguali.

b) Il movimento politico GrandeJuve! ha ottenuto il 23% (115 mila voti su mezzo milione di votanti) al primo turno elettorale raggiungendo però solo la 3ª posizione dovendosi così coalizzare con uno dei primi due partiti . Si estrae -senza reimmisione- un campione casuale di n=300 simpatizzanti dichiarati per GrandeJuve! Qual'è la probabilità che il 51% sia a favore del primo partito? Il numero X dei grandjuventini è una variabile casuale ipergeometrica con N=500'000, N_1=115'000, n=300, N_1/N=0.23

$$p(X \ge 153) = 1 - \sum_{i=0}^{147} \frac{\binom{115000}{i}\binom{385000}{300-i}}{\binom{500000}{300}} \cong p(X \ge 153) = 1 - \sum_{i=0}^{147} \binom{300}{i} (0.23)^i (0.77)^{300-i}$$

L'approssimazione con B(300,0.23), legittimata da 300/500'000=0.0006, è più malleabile, ma non è ancora pratica (vedremo nel prossimo paragrafo una ulteriore approssimazione).

c) In una circoscrizione di Roccasecca risiedono 2000 cittadini. A 20 di essi, scelti con il metodo del campione casuale semplice senza reimmissione, è stata inviata una cartella esattoriale palesemente sbagliata per constatarne il grado di reattività. Di solito, il 75% dei residenti protesta esplicitamente contro una cartella “pazza”. Qual’è la probabilità che tale percentuale sia la stessa o superiore?

$$P(x \geq 11) = 1 - \sum_{x=0}^{10} \frac{\binom{1500}{x}\binom{500}{20-x}}{\binom{2000}{20}} = 0.9866; \quad Appr.Bin. \; P(x \geq 11) = 1 - \sum_{x=0}^{10} \binom{20}{x} 0.75^x 0.25^{20-x} = 0.9861$$

d) Riconsideriamo l’esempio della lotteria e dei biglietti analizzato con lo schema ipergeometrico nel paragrafo 6.3.1. 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 il numero medio di vincite? I biglietti vincitori sono le unità speciali scelte senza reimmissione dal totale k^2 delle unità. Quindi: $\mu = k*k/k^2 = 1$

Esercizio_VC60: *una associazione venatoria deve ripopolare una zona con N=10’000 esemplari. Per controllare l’effettiva immissione gli ispettori regionali catturano un campione di N_1=200 capi, li marcano -con metodo gentile- per una successiva identificazione e li lasciano liberi nella zona in cui sono stati catturati. Dopo un certo tempo, necessario per consentire la mescolanza tra marcati e non marcati, si ricattura un altro campione di n=150 capi. Se la selezione delle unità è casuale, se la popolazione di animali è stabile e se le targhette non vanno disperse, il numero di esemplari già targati nel secondo campione segue una distribuzione ipergeometrica.*
a) Calcolare il suo valore atteso e la sua varianza.
b) Calcolare la probabilità che nel secondo campione si trovi almeno il 20% di esemplari già targati (n_1>30);
c) Se il numero di capi già targati fosse n_1=4 riterreste confermata l’immissione di diecimila esemplari nella zona? N.B. Una stima di N nota nella letteratura del settore è:

$$\hat{N} = \frac{(N_1+1)(n+1)}{n_1+1} - 1$$

Esercizio_VC61: *un file di 500 dichiarazioni dei redditi ne contiene 10 con elusioni che potrebbero mettere in discussione la capacità del consulente che le ha preparate. La commissione tributaria di controllo estrae un campione senza reimmissione di n=5 dichiarazioni per sottoporle ad una revisione accurata. Qual’è la probabilità che il buon nome del/la consulente non ne sia compromesso? N.B. Calcolate la probabilità con il modello ipergeometrico e valutatene l’approssimazione con il modello binomiale.*

Variabile casuale ipergeometrica negativa

Un esperimento consiste nell’inserire con equiprobabilità e indipendentemente blocchi di “n” biglie scelte tra N in due urne inizialmente vuote. L’esperimento si ferma non appena un’urna, diciamo la prima, abbia raggiunto la soglia di “r” biglie. La variabile casuale che ci interessa è il numero di biglie X complessivamente inserite nelle due urne allorché se ne trovino “r “ nella prima. La distribuzione di probabilità si determina considerando che ci sono in tutto C(N,n) scelte possibili senza reimmissione. Si raggiungono le “r” biglie nella 1ª urna all’x-esimo inserimento se nella 1ª urna ve ne sono già (r-1) e queste possono essere scelte in C(x-1,r-1); se la x-esima biglia va nella 1ª, nella 2ª urna vi si troveranno (n-r) biglie scelte fra le (N-x) cioè C(N-x,n-r). La distribuzione di probabilità è perciò:

$$p(X=x) = \frac{\binom{x-1}{r-1}\binom{N-x}{n-r}}{\binom{N}{n}} \quad r \leq x \leq N-(n-r); \qquad F(x) = \sum_{i=r}^{x} \frac{\binom{i-1}{r-1}\binom{N-i}{n-r}}{\binom{N}{n}}; \; r \leq x \leq N-n+r$$

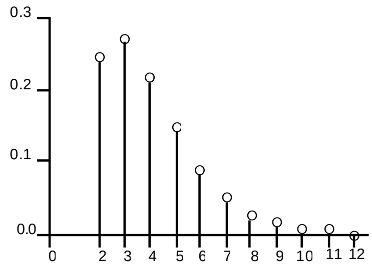
La somma delle probabilità è uno dato che:

$$\sum_{x=r}^{N-n+r} \binom{x-1}{r-1}\binom{N-x}{n-r} = \sum_{y=0}^{N-n} \binom{N-r-y}{n-r}\binom{r-1+y}{r-1} = \binom{N}{n}$$

come si è visto nel paragrafo dedicato al calcolo combinatorio (compito TP52d).

Esempi:

a) Ripreso da Berry e Lindgren (1990, p. 150). Si sta sperimentando su N=20 cani un nuovo vaccino a rischio di gravi effetti collaterali tali che la probabilità di morte è p=0.80. Una possibilità è di somministrare il vaccino a n=10 animali e registrare il numero di sopravviventi: l'informazione è ottenuta, ma con costi elevati e con seri rischi di denunce da parte della Lega antivivisezione (8=10*0.80 morti attese). Un'altra strategia è di scegliere blocchi di 10 animali e somministrare il vaccino ad un animale alla volta interrompendo l'esperimento alla morte, diciamo del secondo animale.



Come si vede dal grafico, le probabilità delle morti diminuiscono all'aumentare del numero di somministrazioni se si opera con uno schema ipergeometrico negativo.

b) Formulazione alternativa del modello. Poniamo E={nelle prime (x-1) prove vanno (r-1) biglie nella 1ª urna}, G={alla x-esima prova 1 biglia va nella 1ª urna}; quindi P(X=x)=E∩G ovvero P(X=x)=P(G)*P(E|G). La prima è la probabilità di scegliere, tra le C(N,x-1) combinazioni di (x-1) elementi, una contenente (r-1) unità speciali (biglie nella 1ª urna) fra le "n" possibili. La seconda è un semplice rapporto casi favorevoli su casi possibili:

$$P(G)=\frac{\binom{n}{r-1}\binom{N-m}{x-r}}{\binom{N}{x-1}};\quad P(E|G)=\frac{n-(r-1)}{N-(x-1)}\Rightarrow P(X=x)=\frac{\binom{n}{r-1}\binom{N-m}{x-r}}{\binom{N}{x-1}}*\frac{n-(r-1)}{N-(x-1)}=\frac{\binom{x-1}{r-1}\binom{N-x}{x-r}}{\binom{N}{n}}$$

Per quanto riguarda valore atteso si usano passaggi analoghi alla ipergeometrica positiva

$$\begin{aligned}E(X)&=\sum_{x=0}^{n}x\frac{(x-1)!(N-x)!N!}{n!(r-1)!(N-n)!(x-r)!(n-r)!(N-x-n=r)!}\\&=r\left(\frac{n+1}{N+1}\right)\sum_{x=0}^{n}\frac{x!(N-x)!(N+1)!}{(n+1)!r!(N-n)!(x-r)!(n-r)!(N-x-n=r)}=r\left(\frac{n+1}{N+1}\right)\end{aligned}$$

Per la varianza in cui si cerca: E[(X+1)X]=E(X^2)+E(X). In questo modo si ottiene:

$$\sigma^2=r\left(\frac{N+1}{n+1}\right)\left[(r+1)\left(\frac{N+2}{n+2}\right)-\frac{(n+1)+r(N+1)}{(n+1)}\right]$$

Esempi:

a) Nel caso della sperimentazione sui cani si raggiunge un notevole risparmio di vite: μ=2(21/11)=3.82; σ=√2.603=1.613. Analoghi ragionamenti si possono estendere ad ogni tecnica di sperimentazione distruttiva o costosa.

b) In un'indagine telefonica su di una popolazione di N=50 persone da contattare si decide di intervistarne un campione di n=20, ma di fermare l'indagine non appena 5 dei rispondenti si saranno dichiarati contrari al quesito proposto. Quanti contatti sono necessari in media? Qual'è lo scarto quadratico medio? Qual'è la probabilità che l'indagine si interrompa prima del 10° contatto? μ=12.75, σ=3.55, P(X<10)=0.2474.

Esercizio_VC62: *le persone in fila per la mensa sono N=100. Rimangono r=15 porzioni di dolce. Nella fila la proporzione di coloro che gradirebbero il dolce è del 40%.*
a) Calcolare media e deviazione standard per il numero di clienti serviti prima di esaurire il dolce;
b) Calcolare la probabilità che il dolce finisca prima di aver sbrigato metà della fila.

Esercizio_VC63: *una scatola contiene N biglie di cui "n" rosse. dopo una energica scossa, si estrae-senza reimmissione- una biglia dall'urna; se è rossa l'esperimento termina, se non è rossa l'esperimento continua con ulteriori estrazioni senza reimmissione finché non si ottiene una biglia rossa. A questo punto l'esperimento si conclude. Se si indica con X il numero di estrazioni complessivamente effettuate allorché compare la biglia rossa si può ritenere che la X abbia distribuzione ipergeometrica negativa?*

7.2 Variabili casuali enumerabili

Il dominio di molte variabili è l'insieme dei numeri naturali $\{1, 2, \ldots\}$ richiamato ad esempio nelle prove ripetute in cui ci si chiede quanti tentativi effettuare prima di raggiungere un certo obiettivo: dal numero di lanci necessari per ottenere una "testa", al numero di cicli di produzione completati prima che si presenti un lotto con il 10% di pezzi difettosi. E' anche possibile considerare insiemi di naturali che partono da un certo intero in poi: ad esempio quante interviste fare per raccogliere almeno "r" opinioni favorevoli: {r,r+1,r+2,…}. Altri aspetti sperimentali includono lo zero {0,1,2,…,}; ad esempio il numero di eventi che si verificano in una data unità di tempo o su di un determinato soggetto. In questi casi le manifestazioni possibili sono un numero arbitrariamente grande che è ovviamente finito nel mondo reale, ma che conviene interpretare senza limite superiore o inferiore per tenere conto della tendenza alla continua crescita o contrazione dei fenomeni.

7.2.1 Estensione della teoria elementare della probabilità

Con i postulati della teoria elementare non siamo in grado di probabilizzare eventi inseriti in un universo con la cardinalità dell'enumerabile (quale ad esempio i naturali dispari o multipli del tre); non siamo cioè in grado di gestire eventi scomponibili in un numero infinito di eventi elementari. Per poter ampliare le applicazioni del calcolo delle probabilità dobbiamo rafforzare il 1° e il 4° postulato. E' possibile fornire una descrizione unica degli esperimenti in cui le determinazioni possibili formano un insieme discreto di modalità allargando la sfera di applicabilità dei postulati di Kolmogorov senza abbandonare i risultati già ottenuti con il discreto finito e senza dover considerare o luna o l'altra situazione un caso a parte.

Dall'algebra alla sigma-algebra

Ecco la formulazione del primo postulato che si adatta sia al nuovo che al vecchio universo degli eventi.

1bis. S si compone di un numero finito o enumerabile di eventi elementari $S=\{e_1,e_2,\ldots,\}$ e come insieme di eventi probabilizzabili consideriamo la σ-algebra W formata con tutti i possibili sottoinsiemi di S:

$$1)\ \ S \in W; \qquad 2)\ \ Se\ \ E \in W \Rightarrow E^c \in W; \qquad 3)\ \ Se\ \ E_i \in W,\ i = 1,2,\ldots, \Rightarrow \left(\bigcup_{i=1}^{\infty} E_i\right) \in W$$

Il prefisso "sigma" o la sua abbreviazione "σ" con cui si denota ora l'algebra dell'esperimento segnala che la chiusura riguarda unione e negazione di una classe enumerabile e non solo finita di eventi.

Esempio:
Negli insiemi enumerabili gli elementi sono in corrispondenza biunivoca con l'insieme dei numerati naturali positivi. Lo stesso accade ad alcune operazioni realizzate con tali insiemi (Khuri, 1993, pp. 7-10; Scozzafava, 1989, pp. 141-146):
1. Ogni sottoinsieme infinito di un insieme enumerabile è anche enumerabile;
2. L'unione di due o più eventi enumerabili è enumerabile;
3. Se T_1 e T_2 sono due insiemi di cui uno enumerabile lo è pure il loro prodotto cartesiano $T_1 \otimes T_2$.
4. L'insieme dei numeri razionali è enumerabile;

Una σ-algebra è anche un'algebra, ma il contrario non è sempre vero per cui la σ-algebra è una restrizione dell'algebra. A prima vista la differenza non è impressionante, ma non è da sottovalutare perché ora sono possibili sviluppi teorici affascinanti. La coppia (S,W) costituisce lo spazio misurabile dell'esperimento cioè la classe di eventi composti generata da S ai quali è può essere attribuito un numero reale in base ad una funzione di insieme. A partire da S si possono costruire tante classi di suoi eventi composti: alcune saranno delle σ-algebre ed altre no (noi però tratteremo solo con queste). In particolare, sono ottenibili due speciali σ-algebre. La prima è $W^*=\{E|E \subseteq S\}$ formata inserendo ogni possibile sottoinsiemi di S e la seconda è $W_*=\{S,\varnothing\}$ formata solo dall'evento certo e dall'evento impossibile; quest'ultima è troppo ristretta per essere di utilità, la prima è invece del tutto operativa nel caso di un S enumerabile ed infatti l'abbiamo scelta come base di lavoro per gli esperimenti che portano a questo tipo di universo.

Esempi:

a) Sia S={1,2,…,} l'insieme degli interi positivi legato all'esperimento di lanciare e continuare a lanciare una moneta finché non esca croce e poniamo E={1, 3, 5, 7, 9, …} cioè l'evento che sia necessario un numero dispari di tentativi. Allora W={∅, S, E, E^c} è una σ-algebra che consente di dare risposta alla domanda "si è verificato il tale evento?" per qualsiasi evento in W.

b) Poniamo ora E_i={1+(i-1)*5, i*5}. E' evidente che l'unione finita o enumerabile degli {E_i} è una copertura di S per cui l'algebra formata da ∅, S e dall'unione finita o enumerabile di eventi E_i costituisce una σ-algebra.

***Esercizio_VC64:** verificare che, se l'universo degli eventi S ha un numero enumerabile di elementi, allora la classe formata da tutti i possibili sottoinsiemi finiti di S è un'algebra, ma non una σ-algebra.*

Breiman (1968, p.35) osserva che la σ-algebra include comunque tanti eventi ai quali deve poter essere assegnata una probabilità anche se la stragrande maggioranza non avrà mai alcuna utilizzazione. Questo disturba l'abbinamento intuitivo e naturale di evento e probabilità tanto che sono in corso raffinamenti e semplificazioni del concetto di σ-algebra che però non hanno ancora sufficiente consenso per essere studiati in un corso di base.

Additività completa (σ-additività)

Se S è enumerabile, le possibilità sono due: o assegniamo probabilità solo ad un numero ridotto di essi e agli altri diamo probabilità zero (ma se sono tutti impossibili perché considerarli separatamente?) oppure le probabilità debbono scaturire da una successione convergente. Ecco due alternative:

$$1.\ \ p_i = \frac{\pi^2}{6i^2},\ \ i = 1,2,\ldots;\qquad 2.\ \ p_i = \frac{1}{(3*i+2)};\ \ i = 0,1,2,\ \ldots$$

Nella prima, la funzione di insieme sarebbe costituita da numeri non negativi per ogni "i", ma questi raggiungono la somma unitaria -condizione indispensabile per parlare di probabilità- solo se i $\to \infty$. Nella seconda già per i=5 la somma supera l'unità ed è divergente; in questo caso non si tratta di una idonea funzione di insieme.

Esempi:

a) Il numero di spostamenti che una particella subisce dal passare dallo stato liquido a quello gassoso è descritto dal modello:

$$p_i = \frac{0.23365}{(2i+1)^2};\quad i = 1,2,\ldots$$

E' una funzione di distribuzione? La non negatività è evidente. Vediamo la convergenza. A questo fine applichiamo il criterio stabilito da J.L. Raabe nel 1832:

$$Sia\quad \frac{p_i}{p_{i+1}} = 1 + \frac{a}{i} + o\left(\frac{1}{i}\right);\ \ se\ \ i \to \infty\ \ allora \begin{cases} se\ \ a<1\ \ \sum p_i\ \ diverge \\ se\ \ a>1\ \ \sum p_i\ \ converge \\ se\ \ a=1\ \ il\ \ test\ \ non\ \ è\ \ conclusivo \end{cases}$$

Il simbolo o[g(x)] vuol dire che si sta considerando -per ogni x- un termine il cui ordine di grandezza è inferiore alla funzione g(x)

$$f(x) = o[g(x)] \Rightarrow \lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$$

Nel caso proposto la convergenza è assicurata dato che a=2.

b) Esiste un modello in cui le probabilità da assegnare ai naturali diminuiscono in progressione armonica?

$$p_i = \frac{k}{i},\ \ i = 1,2,\ldots;\ \ 0 < k < 1$$

Poiché p_i/p_{i+1}=1+1/i con a=1 la serie rientra nel caso inconclusivo del criterio di Raabe. Possiamo però applicare il criterio di Gauss:

$$Sia\quad \frac{p_i}{p_{i+1}} = 1 + \frac{a}{i} + O\left(\frac{1}{i^{\delta+1}}\right);\ \ \delta > 0\ \ se\ \ i \to \infty\ \ allora \begin{cases} se\ \ a \le 1\ \ \sum p_i\ \ diverge \\ se\ \ a>1\ \ \sum p_i\ \ converge \end{cases}$$

dove O[g(x)] esprime -per ogni x- una quantità non eccedente g(x):

$$f(x) = O[g(x)] \Rightarrow |f(x)| \le Mg(x)\ \ con\ \ M>0\ \ per\ \ ogni\ \ x$$

Nel rapporto che riguarda la serie armonica il 3° addendo è sempre zero perciò la serie è divergente (anche se ogni termine è compreso tra zero ed uno) e non può essere usata per assegnare probabilità ad un dominio infinito anche se discreto.

Khuri (1993, p. 157) ricorda che Gauss ottenne il suo criterio studiando la cosiddetta serie ipergeometrica:

$$p_i = \frac{a(a+1)(a+2)\dots(a+i-1)b(b+1)(b+2)\dots(b+i-1)}{i!c(c+1)(c+2)\dots(c+i-1)};\; i = 1, 2, \dots,$$

dove “n” è intero e a,b,c sono reali non nulli. Il nome di distribuzione ipergeometrica dato al modello incontrato nel paragrafo precedente è dovuto all’uso di un caso particolare di questa serie anche richiamata nella omonima variabile casuale.

Esercizio_VC65: *le probabilità dei valori di un dominio enumerabilmente infinito debbono essere assegnate in base ad uno dei seguenti schemi. Solo uno è corretto. Quale?*

$$a.\; p_i = \frac{k}{i!};\; 0 < k < 1; \quad b)\; p_i = \frac{k}{iLn(i)}; \quad c)\; p_i = \frac{i+2}{i^3 + 2i + 1}$$

Per probabilizzare gli eventi della σ-algebra generata da un universo degli eventi S enumerabilmente infinito è necessaria una funzione di probabilità che verifichi il seguente postulato:

4bis. Se $\{E_1, E_2, E_3, \dots,\}$ è un insieme enumerabile di eventi mutualmente incompatibili inclusi in W, la probabilità del loro evento unione è assegnata in base alla regola:

$$P\left[\bigcup_{i=1}^{\infty} E_i\right] = \sum_{i=1}^{\infty} P(E_i) \quad con \;\; E_i \cap E_j = \emptyset \;\; per \;\; i \neq j$$

Quindi, alla funzione di probabilità si attribuisce la proprietà dell’additivià completa (o σ-additività) che è un ampliamento dell’additività finita utilizzata nella teoria elementare della probabilità. La coerenza con gli altri postulati è illustrata dalle due relazioni seguenti:

$$Se \;\; E_i \uparrow S \;\; cioè \;\; E_i \subset E_{i+1}, \; \bigcup_{i=1}^{\infty} E_i = S \;\; allora \;\; P\left(\bigcup_{i=1}^{\infty} E_i\right) = P(S) = 1$$

$$Se \;\; E_i \downarrow \emptyset \;\; cioè \;\; E_{i+1} \subset E_i, \; \bigcap_{i=1}^{\infty} E_i = \emptyset \;\; allora \;\; P\left(\bigcap_{i=1}^{\infty} E_i\right) = P(\emptyset) = 0$$

La σ-additività permette risultati teorici generali, ma non è indispensabile, neanche per diversi teoremi limite importanti. Del resto, Savage J.L. e B. De Finetti, ritengono questionabile anche l’additività semplice: basterebbe forse una proprietà elementare del tipo $A \subset B \Rightarrow P(A) \leq P(B)$ anche se forse si ridurrebbe il numero di teoremi -collimanti con la realtà osservabile- che potrebbero essere dimostrati rigorosamente in un ambito così angusto.

Esempi:

a) Una funzione di insieme f(.) è additivamente completa se, per ogni partizione di un insieme E in numero “n”, finito o enumerabilmente infinito, di sottoinsiemi mutualmente incompatibili $\{E_i\}$ si ha:

$$f(E) = \sum_{i=1}^{n} f(E_i)$$

b) Prodi (1992, p. 197-198) commentando il numero di colpi che un tiratore spara prima di fare centro vede un insorgere spontaneo, nei problemi di probabilità, dell’uso della σ-additività cioè dell’estensione della additività dal caso discreto finito all’enumerabile. Ha senso pensare che il tiratore continui a sparare all’infinito senza mai colpire il bersaglio soprattutto se dietro il bersaglio ci siete voi.

Teorema della continuità

Per valutare le proprietà della funzione di probabilità nel nuovo tipo di dominio dobbiamo acquisire un importante risultato che consegue dai nuovi postulati. Se $\{E_n\}$ è una successione monotona di eventi allora:

$$P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n)$$

è cioè possibile scambiare l’assegnazione della probabilità al limite della successione con il limite della successione delle probabilità. Ipotizziamo che la successione sia monotona crescente:

$$E_1 \subseteq E_2 \subseteq \; \dots \; \subseteq E_i \subseteq \; \dots \; \Rightarrow \; \underset{n\to\infty}{Lim}\, E_n = \bigcup_{i=1}^{\infty} E_i$$

La successione converge all'evento unione. Poiché E_i è include l'evento che lo precede, si può definire:

$$E_i = E_{i-1} \cup \left(E_i - E_{i-1}\right) \quad i = 1,\; 2, \dots, \; dove \;\; si \;\; conviene \quad E_0 = \varnothing$$

cioè E_i è dato da quello che c'è già in E_{i-1} con aggiunta la parte non compresa in E_{i-1}. Questa relazione consente di riesprimere l'insieme limite come:

$$\bigcup_{i=1}^{\infty} E_i \;\; = \;\; \bigcup_{i=1}^{\infty}\left(E_i - E_{i-1}\right) \qquad con \quad \left(E_i - E_{i-1}\right) \cap \left(E_j - E_{j-1}\right) = \varnothing \;\; se \;\; i \neq j$$

Tenuto conto di quanto si era già dimostrato nel paragrafo 6.1.1 si ha:

$$P\left(E_i - E_{i-1}\right) = P\left(E_i\right) - P\left(E_{i-1}\right) \;\Rightarrow\; \sum_{i=1}^{n} P\left(E_i - E_{i-1}\right) = \sum_{i=1}^{n} P\left(E_i\right) - P\left(E_{i-1}\right) = P\left(E_n\right)$$

dove E_n è l'evento formato dagli eventi elementari ricadenti in almeno un evento della successione. Pertanto:

$$P\left[\underset{n\to\infty}{Lim}\left(E_n\right)\right] \; = \; \underset{n\to\infty}{Lim}\, P\left(E_n\right)$$

Esercizio_VC66: *dimostrare il teorema della continuità per una successione monotona decrescente di eventi.*

Il teorema della continuità è tutt'altro che un ricamo matematico. E' interscambiabile con il postulato della additività completa ed infatti Kolmogorov preferisce adottare come assioma la probabilità del limite come limite delle probabilità e derivare l'additività completa come teorema.

Consideriamo nei numeri reali non negativi $\{p_i, i=1,2,\dots,\}$ con somma convergente all'unità che la funzione di insieme P(.) assegna ai singoletti $E_i=\{e_i\}$ per i=1,2,…, in S e sia C⊂W. Segue dai postulati che:

$$P(C) = \sum_{E_i \subset C} P\left(E_i\right) = \sum_{E_i \subset C} p_i$$

come nella teoria elementare si definisce la funzione di insieme P(.) solo per i singoletti e poi la si estende a tutti i possibili eventi in W grazie all'additività completa. Il fatto nuovo è che ora un evento composto può contenere un'infinità enumerabile di eventi elementari e la sua probabilità è ottenuta da una successione convergente di numeri nell'intervallo unitario.

Esempio:
Per il lancio ripetuto della moneta si può ipotizzare che la probabilità di avere "croce" sia 0.5 e che i lanci siano indipendenti. In questo caso, la probabilità che sia necessario un numero dispari di prove è:

$$P(dispari) = \frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \frac{1}{128} + \frac{1}{512} + \dots = \frac{1}{2}\left(\frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} + \dots\right) = \frac{1}{2}\left(\frac{1}{1 - \frac{1}{4}}\right) = \frac{2}{3}$$

Perché non si ha: P(pari)=P(dispari)=0.5? Perché già al primo lancio P(1)=0.5 e tutti gli altri dispari qualcosa dovranno pur contribuire.

In verità, l'introduzione dei nuovi postulati è tuttaltro che pacifica e sono spesso considerati inutili perché da un esperimento reale non si può ottenere che uno spazio di probabilità finito e come osserva lo stesso Kolmogorov (1933/1995, pp. 32-33): "*... Se gli eventi di S possono aver senso in quanto reali ed osservabili (sia pure approssimativamente), da ciò ancora non segue che gli insiemi della σ-algebra ammettano una stessa ragionevole interpretazione in termini di eventi effettivamente osservabili*".

Esercizio_VC67: *le probabilità che risultano assegnate alle successioni di eventi sono sottosuccessioni convergenti ottenute a partire da quella convergente all'unità assegnate ai singoletti $E_i=\{e_i\}$ dove e_i è un evento elementare di S. Perché non ci dobbiamo preoccupare della convergenza delle sottosuccessioni?*

Eventi quasi certi ed eventi quasi impossibili

Loéve (1977, pp.15-16) osserva "... *Non appena compare il concetto di infinito l'intuizione si perde e il concetto vago, ma familiare di casualità non ci è più d'aiuto*".

Esempio:
Larson (1969, p. 66-67) riflette sul seguente esperimento: scelgo a caso un intero naturale positivo. Qual'è la probabilità di E= "il numero scelto è dispari"? L'universo degli eventi è S={1, 2, 3, 4, …}. Poiché non ci sono ragioni per preferire un numero ad un altro la simmetria dell'esperimento indurrebbe ad assegnare la stessa probabilità "p" a tutti. Questo è però impossibile perché la probabilità di S risulterebbe:

$$P(S) = p + p + p + \ldots = \infty * p \rightarrow \infty$$

con violazione del 4° postulato. Non siamo perciò in grado di descrivere questo esperimento neanche con i nuovi postulati anche se -intuitivamente- riteniamo ragionevole assegnare P(E)=0.5. Non c'è però contraddizione. Con questa assegnazione si opera su un diverso esperimento: inizialmente si pone S={1,2,…,N} con N molto grande e si assegna P(E)=0.5 se N è pari e (0.5-1/N) se N è dispari; successivamente, l'esperimento è allargato a (N+1) numeri, a (N+2), etc. con P(E) che converge a 0.5. Si tratta però solo di un modello astratto dell'esperimento originale.

Scozzafava (1997, cap. 1) nota che il modello di probabilità uniforme entra in profonda contraddizione allorché l'universo degli eventi sia infinito o anche solo enumerabile. In questo caso a tutti i singoletti deve essere assegnata probabilità quasi-zero a prescindere da ogni ipotesi di uguali opportunità di accadimento.

La comparsa dell'infinità, anche solo enumerabile, pone non pochi problemi di coerenza all'impianto di Kolmogorov: siamo di fronte ad esperimenti in cui la probabilità dei singoletti è pressoché nulla, ma la probabilità del complesso degli eventi è uno. L'infinito mette in discussione i punti di ancoraggio più solidi degli assiomi di Kolmogorov: l'evento certo e l'evento impossibile. Con un universo degli eventi finito l'evento certo è S con P(S)=1 senza che si possano avanzare dubbi in proposito. Nella σ-algebra questo è ancora vero, ma l'unità è anche assegnata ad eventi che non sono l'evento certo.

Esempi:
a) Un'ape esce dall'alveare per cercare il fiore che le hanno indicato ricchissimo di nettare. Il numero di fiori simili è infinito e mancano segnali precisi tanto che ognuno di quelli che incontra potrebbe essere quello giusto. Se, nell'astrazione del modello, consentiamo alla brava operaia un' infinità di tentativi per i tanti infiniti fiori, dobbiamo per forza assegnare probabilità uno all'evento "l'ape trova il suo fiore prima o poi". Tuttavia, non si tratta di una certezza logica perché il suo evento contrario: "l'ape non troverà mai il suo fiore" non può avere probabilità nulla (può essere piccolissima, ma non zero) e non è quindi un evento impossibile. D'altra parte se qualcuno togliesse dei fiori diversi da quello giusto, la probabilità di trovare il fiore così accattivante dovrebbe rimanere ancora uno: l'infinito enumerabile non diviene un infinito più piccolo perché è stato ridotto di qualche unità.

b) La serie delle probabilità $\{p_i\}$ deve convergere ad uno per rispettare il 3° postulato. Però, se la serie converge ad uno, ogni sottosuccessione di probabilità di eventi converge pure all'unità, ma questa risulta assegnata ad un evento diverso dall'evento certo.

Probabilità zero implica l'evento impossibile? Non necessariamente: lo zero può essere l'approssimazione di un numero talmente piccolo da non potersi di distinguere dallo zero, ma positivo. Questo tipo di incertezza non sussiste nella teoria elementare in cui la cardinalità di S è finita.

Esempi:
a) Se S={x| x=1,2,…,} e la probabilità ha funzione di distribuzione:

$$p(x) = \frac{0.62867901}{x!(x+1)!}$$

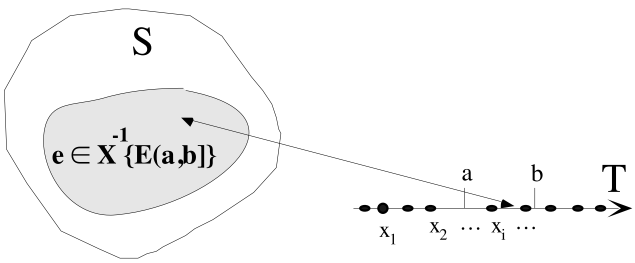
Per modalità "x" grandi la probabilità dell'evento è praticamente zero anche se non è proprio zero perché altrimenti la somma delle probabilità sarebbe inferiore all'unità. Quindi valori tendenti all'infinito hanno probabilità non diversa da zero, ma non sono impossibili.

b) De Finetti (1967) intervenendo ad un convegno evidenziò che un evento può avere probabilità zero, ma non essere l'evento impossibile. Un litro d'acqua può essere considerato puro perché contiene un batterio con probabilità $10^{-100'000'000'000}$, che tuttavia è infinitamente più grande dello zero. Castellano in un intervento nello stesso convegno rileva che è colpa dei matematici la creazione degli eventi di probabilità nulla che sono possibili. La cosa è chiarissima in sé: questi eventi hanno probabilità data da un rapporto di due infiniti di cui quello al denominatore ha un ordine superiore di quello del numeratore e quindi il rapporto è posto pari a zero anche se non è proprio zero. Infatti egli invoca un simbolo diverso dallo zero per indicare una probabilità infinitamente piccola, ma positiva. Matematicamente è irrilevante, ma non dal punto di vista della probabilità. Quanti sarebbero disposti a bere un sorso di quell'acqua sapendo che il batterio è mortale?

Rimane un'altra questione irrisolta. La distribuzione di probabilità P(.) è programmata sui singoletti formatisi intorno agli elementi di S; è possibile assegnare probabilità agli eventi della σ-algebra senza modificare la distribuzione P(.)? A questo risponde il teorema del prolungamento (Wilks, 1962, pp.15-16) secondo il quale una funzione d'insieme additivamente completa può sempre estendersi, e ciò in un unico modo, a tutti gli altri eventi con conservazione di entrambe le proprietà (non negatività e additività). Lo spazio di probabilità acquista così piena libertà di azione.

Definizione della variabile casuale enumerabile

Sia (S, W, P) uno spazio di probabilità per il quale valgano i nuovi postulati di Kolmogorov e definiamo una funzione reale che associ ad ogni "e" dell'universo degli eventi S uno ed uno solo valore X(e) in un intervallo discreto E(a,b) con $a,b \in T\{x_1, x_2,\dots\} \subset R$ e con gli estremi "a" e "b" che possono anche essere infiniti.



Come si è visto nel paragrafo 7.1 la funzione X=X(e) che trasferisce S in T è una variabile casuale se la sua controimmagine $X^{-1}\{E(a,b]\}$ è un evento della σ-algebra W generata da tutti i possibili sottoinsiemi di S e questo per ogni intervallo di modalità in T. Chung (1974, p. 35), Rohatgi (1976, p. 55) evidenziano che qualsiasi funzione a valori sull'asse reale R definita su di un S discreto è una variabile discreta (finita o infinita). Se l'applicazione tra S e T è univoca ed esaustiva, la variabile casuale discreta X con dominio $T\{x_1, x_2,\dots\}$ è associabile ad un insieme di numeri reali non negativi: $\{p_1, p_2, \dots,\}$ tali che P(T)=1. Per accertare le altre caratteristiche dei $\{p_i\}$ ragioniamo su una partizione di T realizzata con degli intervalli disgiunti definiti intorno ad ogni modalità:

$$E(x_i, x_{i+1}] \quad con \quad x_i < x_{i+1}; \; i = 0,1,\dots$$

Sia p_i la probabilità assegnata a $E(x_i,x_{i+1}]$. Ogni intervallo di valori $C(a,b] \subset T$ è esprimibile come unione -finita o infinita- di intervalli del tipo introdotto e quindi:

$$C(a,b] = \bigcup_{E(x_i,x_{i+1}]\subset C(a,b]} E(x_i, x_{i+1}] \Rightarrow P(C) = \sum_i P\{E(x_i, x_{i+1}]\} = \sum_i p_i$$

In definitiva, le probabilità generate dalla funzione di insieme P(.) nel nuovo spazio di probabilità portano a formare una distribuzione di probabilità p(). Non solo, ma ogni insieme di numeri reali non negativi ed a somma convergente all'unità può considerarsi una funzione di distribuzione di probabilità di una qualche variabile casuale discreta senza dover passare per la mediazione di un esperimento casuale e del suo spazio di probabilità.

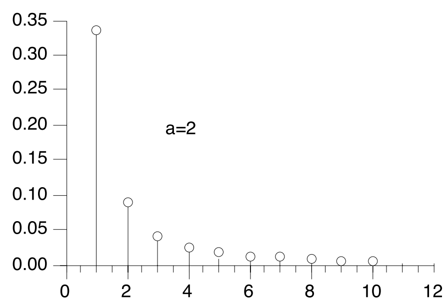
Esempio:
Modello di Yule.

$$p(x) = (x-1)!\left[\frac{\Gamma(a+1)}{\Gamma(x+a+1)}\right]; \quad x = 1,2,\dots,$$

$$dove \quad \Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt \quad è \;\; la \;\; funzione \;\; gamma$$

$$\Gamma(x) = (x-1)! \quad se \quad x \; è \; intero.$$

In figura è rappresentata la funzione di distribuzione per a=2; variando opportunamente "a" si possono descrivere situazioni molto diverse. Il modello di Yule ha avuto successo nell'esprimere la distribuzione dei generi per numero di specie biologiche, ma è scaturito anche nell'analisi della frequenza con cui le parole compaiono in un testo, degli studiosi per numero di pubblicazioni (Simon, 1955).



Esercizio_VC68: *dare la rappresentazione grafica e verificare che l'espressione sia un modello di probabilità:*

$$p(x) = 0.02452539\frac{(x+1)^3}{x!};\ \ x = 0,1,2,\ldots$$

La funzione di ripartizione per variabili casuali enumerabili

Poiché tutti gli insiemi che possono interessare valori della variabile casuale sono esprimibili come unione e intersezione di un numero finito o enumerabile di intervalli del tipo $\{x \leq a\}$ è sufficiente stabilire la probabilità per essi ed a questo fine è utile la funzione di ripartizione.

Consideriamo due successioni di eventi: $H_j=\{e\in S|\ X(e)\in(-\infty,x_j]\}$ con $H_j \subseteq H_{j+1}$ perché $x_j<x_{j+1}$ e $G_j=\{e\in S|\ X(e)\in[x_j,\infty)\}$ con $G_{j+1} \subseteq G_j$. Allora, applicando il teorema della continuità:

$$\underset{j\to\infty}{Lim}\,H_j = \bigcap_{j=1}^{\infty}\left(-\infty,x_j\right] = \varnothing \ \ \Rightarrow\ \ \underset{j\to\infty}{Lim}\,P\left(H_j\right) = P\left[\underset{j\to\infty}{Lim}\left(H_j\right)\right] = P(\varnothing) = 0 = F(-\infty)$$

$$\underset{j\to\infty}{Lim}\,G_j = \bigcup_{j=1}^{\infty}\left[x_j,\infty\right) = T \ \ \Rightarrow\ \ \underset{j\to\infty}{Lim}\,P\left(G_j\right) = P\left[\underset{j\to\infty}{Lim}\left(G_j\right)\right] = P(T) = 1 = F(+\infty)$$

La F(.) è monotona non decrescente; basta infatti considerare che se $x_1 < x_2$ allora $(-\infty,x_1] \subset (-\infty,x_2]$ e quindi:

$$p\left(X \leq x_1\right) = F\left(x_1\right) \leq F\left(x_2\right) = p\left(X \leq x_2\right)$$

Rimane da controllare la continuità a destra. Riprendiamo la definizione del paragrafo 7.1 e teniamo conto della natura ancora discreta del dominio della X:

$$F\left(x_i+\right) = P\left\{\left(-\infty,x_i\right]\right\} - P\left\{\left(-\infty,x_i-\varepsilon\right]\right\} = P\left\{\left(x_i,x_i+\varepsilon\right]\right\} \Rightarrow \underset{\varepsilon\to 0}{Lim}\ F\left(x_i+\right) = \underset{\varepsilon\to 0}{Lim}\ P\left\{\left(x_i,x_i+\varepsilon\right]\right\} = p_i$$

$$F\left(x_i-\right) = P\left\{\left(-\infty,x_i+\varepsilon\right]\right\} - P\left\{\left(-\infty,x_i\right]\right\} = P\left\{\left[x_i-\varepsilon,x_i\right)\right\} \Rightarrow \underset{\varepsilon\to 0}{Lim}\ F\left(x_i-\right) = \underset{\varepsilon\to 0}{Lim}\ P\left\{\left[x_i-\varepsilon,x_i\right)\right\} = p_{i-1}$$

Pertanto, anche qui possiamo adottare la convenzione:

$$F\left(x_i\right) \leq F\left(x_i+\right);\ \ i = 1,2,\ldots$$

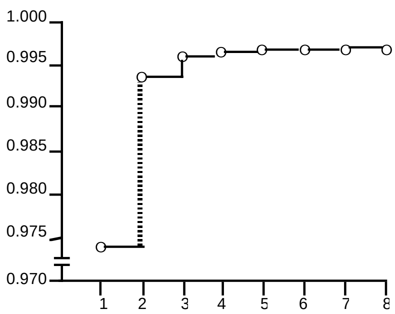
Ne consegue che il passaggio dal discreto finito all'enumerabile non comporta modifica alcuna alla definizione della funzione di ripartizione che rimane una descrizione completa dell'esperimento probabilistico, almeno per quel che concerne l'aspetto rappresentato dalla variabile casuale X.

Esempi:

a) Il numero di tentativi per raggiungere un certo risultato è modellato da un processo sia stabile intorno a X=2. Possono però intervenire perturbazioni che aumentano il numero di tentativi anche all'infinito.

$$p_i = \frac{8.764639}{i\left(4i^2-1\right)^2};\ \ i = 1,2,\ldots,$$

$$F_i = \sum_{j=1}^{i}\frac{8.764639}{j\left(4j^2-1\right)^2};\ \ i = 1,2,\ldots,$$



Spesso, nella simbologia che descrive le variabili casuali discrete si sostituiscono le notazioni "x", "y", etc. per le variabili casuali con i simboli "i", "j", "k" per sottolineare il loro carattere discreto dato che tali lettere sono una notazione classica per i conteggi. Naturalmente, il cambio di simbologia non cambia i significati di ciò che rappresentano.

b) Una progressione più attenuata è il modello:

$$p_i = \left(\frac{1}{e-1}\right)\frac{i^2}{(i+1)!},\quad i = 1, 2, \ldots$$ c

che presenta anche una moda significativa a i=2.

Esercizio_VC69: *le probabilità di una variabile casuale enumerabile sono sintetizzate dall'espressione:*

$$p_i = 96\left[\frac{1}{\pi(2i-1)}\right]^4,\quad i = 1, 2, \ldots,$$

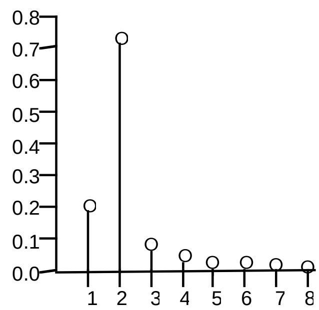
a) Calcolare p(i≤10), p(i>5); b) Calcolare p(i=numero pari);
c) Rappresentare graficamente la funzione di ripartizione.

Se è nota la funzione di ripartizione si ricava -per differenza tra valori successivi- la funzione di distribuzione:

$$p(x_i) = F(x_i) - F(x_{i-1});\quad i = 1, 2, \ldots,\ F(x_0) = p(x_0) = 0$$

Esempio:

$$F_i = 1 - \frac{8}{[\pi(2i-1)]^2};\ i = 1, 2, \ldots,\ p_i = \begin{cases} 1 - \frac{8}{\pi^2} & ; i = 1 \\ \left\{\frac{8}{\pi\left[4(i-1)^2 - 1\right]}\right\}^2 (i-1); & i = 2, 3, \ldots \end{cases}$$



Il modello assegna probabilità poco significative a valori oltre x=5. Presenta una moda spiccata per x=2 a cui è data probabilità P(X=2)=72.5% e 2 è anche la mediana.

Esercizio_VC70: *si consideri il modello:*

$$F_i = \frac{i}{i+\delta}\quad per\ \ i = 0,\ 1,\ 2, \ldots;\ 0 < \delta < \infty$$

a) E' una funzione di ripartizione?
b) Se la risposta in a) è positiva, determinare la distribuzione di probabilità associata al modello.

Trasformazione di variabili casuali enumerabili

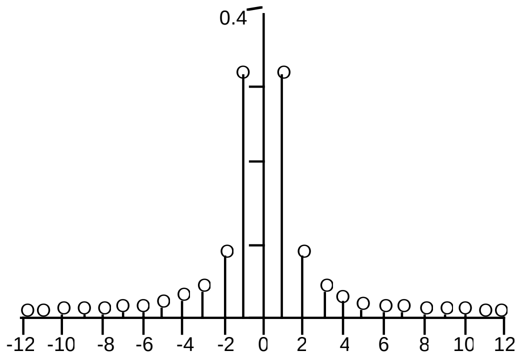
Per le variabili casuali enumerabili possono essere svolte le stesse considerazioni fatte per le discrete finite nel senso che le probabilità non sono coinvolte e risultano assegnate tanto a X quanto a Y=g(X) purché g(.) abbia un numero di inverse finito o enumerabile.

Esempi:
a) Variabile casuale enumerabile con funzione di distribuzione simmetrica:

$$X_i = \pm i,\quad p_i = \frac{3}{(\pi i)^2};\quad i = 1, 2, \ldots,$$

Se la trasformata è $Y=X^2$, la relazione non è 1:1 dato che la Y non può essere negativa e la funzione inversa è doppia: X=±√Y. La probabilità per ogni modalità in Y arriva sia da X che da -X e poiché sono associate alla stessa probabilità possiamo limitarci al semplice raddoppio. Quindi:

$$y_i = i^2,\quad p_i = \frac{6}{(\pi i)^2};\quad i = 1, 2, \ldots,$$



b) Se la distribuzione di probabilità della variabile casuale “i” è p_i che distribuzione avrà j=1/i?

$$p_i = \frac{8i}{\left(4i^2-1\right)^2};\quad i=1,2,\ldots,\quad p_j = \frac{8/j}{\left(\left(2/j\right)^2-1\right)^2};\quad j=1,\frac{1}{2},\frac{1}{3},\ldots,$$

Esercizio_VC71: *per modellare l’eterogeneità delle specie catturate, R.A. Fischer nel 1943 propose il modello logaritmico:*

$$p_i = -\frac{\theta^i}{iLn(1-\theta)};\quad per\ \ i=1,2,\ldots,\quad 0<\theta<1;$$

a) Rappresentare la funzione di ripartizione per θ=0.1;
b) Che applicazioni ha tale modello? c) Che distribuzione di probabilità ha j=1/i?

Sintesi delle variabili casuali enumerabili

La distribuzione di probabilità contiene le informazioni per descrivere il comportamento delle variabili casuali sotto ogni aspetto. In genere non è necessaria una conoscenza così dettagliata e ci si può limitare allo studio delle caratteristiche più importanti, quali la centralità, la variabilità e l’asimmetria espresse con pochi numeri rappresentativi.

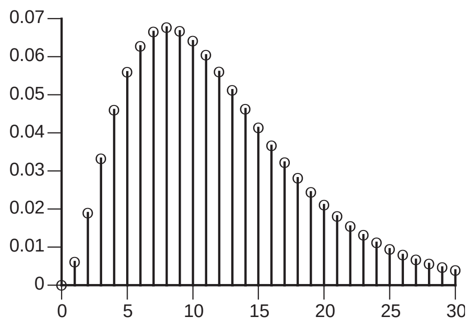
Il passaggio dalle discrete finite alle enumerabili non introduce novità per quanto attiene alla moda, alla mediana ed ai quantili.

Esempi:
a) Calcoliamo la moda del modello:

$$p_i = \frac{i^2 e^{-i/4}}{127.997927};\quad i=0,1,2,\ldots$$

$$\frac{p_{i+1}}{p_i} = \frac{(i+1)^2 e^{-(i+1)/4}}{i^2 e^{-i/4}} = \left(\frac{i+1}{i}\right)^2 e^{-1/4} \le 1 \Rightarrow i > 7.5$$

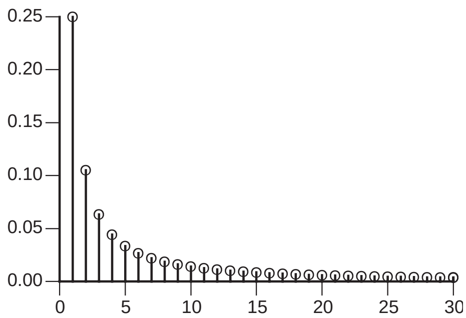
il rapporto tra probabilità successive ha un solo cambio di andamento e scende sotto l’unità a decrescere da i=8 che è appunto il valore più probabile in questo modello.



b) Una distribuzione di probabilità ai margini della convergenza è la seguente:

$$p_i = \frac{0.25}{i^{1.25}},\quad i=1,2,\ldots$$

La massa di probabilità non è divisa equamente: il 50% (circa) è attribuito alle modalità inferiori o uguali a 5 e le infinite altre si dividono il restante 50%.



Esercizio_VC72: *determinare moda e quartili nei due modelli:*

$$a.\ \ p_i = \frac{0.36787}{i!};\ i=0,1,2,\ldots\qquad b.\ \ p_i = \frac{1.629446}{i(2i+1)};\qquad i=1,2,\ldots,$$

Momenti infiniti

I modelli discussi in questo paragrafo presentano un dominio infinito e cioè le modalità tendono all’infinitamente grande. Questo, come afferma Dall’Aglio (1987, p. 137), non crea difficoltà dal punto di vista matematico, ma sul piano intuitivo potrebbe mettere in crisi il concetto di momento:

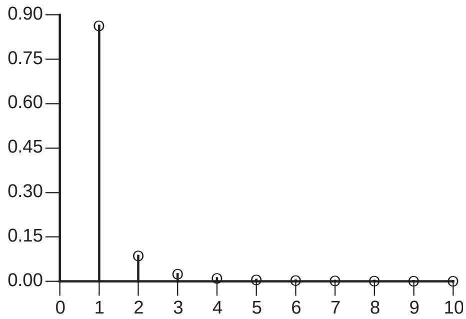
$$\mu_r = E\left(X^r\right) = \sum_{i=0}^{\infty} x_i^r p_i,\quad r=1,2,\ldots,$$

Esempi:

a) Variabile discreta enumerabile con probabilità decrescenti.

$$p(x) = \frac{2.5887}{x\left(4x^2 - 1\right)},\quad x = 1, 2, \ldots,$$

$$\mu = \sum_{i=1}^{\infty} \frac{2.5887}{\left(4i^2 - 1\right)} = 1.2943$$



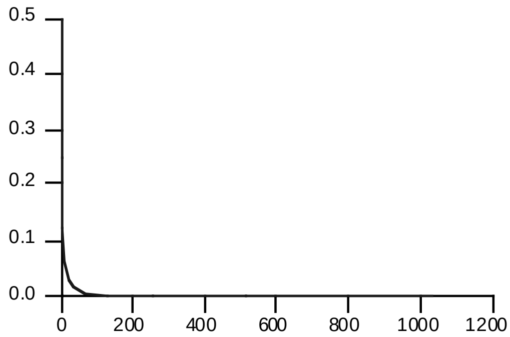
Nonostante l'espansione progressiva dei valori, le probabilità con cui possono verificarsi diminuiscono tanto rapidamente che il contributo i-esimo al valore atteso: "$x_i p_i$" tende subito a zero.

b) Modello con probabilità decrescenti, ma per modalità crescenti in ragione geometrica:

$$x_i = 2^i,\quad p_i = \left(\frac{1}{2}\right)^i,\quad per\quad i = 1, 2, \ldots,$$

Nonostante la caduta repentina delle probabilità il calcolo della media aritmetica è divergente cioè la media è infinita ovvero non esiste finita:

$$\mu = \sum_{i=1}^{\infty} X_i f_i = \sum_{i=1}^{\infty} 2^i \left(\frac{1}{2}\right)^i = \sum_{i=1}^{\infty} 1 = \overbrace{1 + 1 + 1 + \ldots +}^{\text{infiniti "1"}} \to \infty$$



c) Modello con valori in progressione aritmetica e probabilità decrescenti in ragione cubica:

$$x_i = i,\quad p_i = \frac{0.8319}{i^3},\quad i = 1, 2, \ldots,$$

$$\mu_1 = \sum_{i=1}^{\infty} i\left[\frac{0.8319}{i^3}\right] = \sum_{i=1}^{\infty}\left[\frac{0.8319}{i^2}\right] = 0.5057;\quad \mu_2 = \sum_{i=1}^{\infty} i^2\left[\frac{0.8319}{i^3}\right] = \sum_{i=1}^{\infty}\left[\frac{0.8319}{i}\right] = diverge$$

***Esercizio_VC73:** per la distribuzione di probabilità:*

$$p(x) = \begin{cases} \dfrac{6}{(\pi x)^2}; & x = 1, 2, 3, \ldots \\ 0 & altrove \end{cases}$$

a) Verificare che non esiste finito il valore atteso;

b) Quale tipo di fenomeno si può rappresentare con questo modello?

Se per un fissato N si ha $|x_i| \geq 1$ per ogni i>N, l'esistenza finita del momento di ordine "k" assicura la convergenza di μ_r per r=1,2,…,k-1 ovvero esistono finiti i momenti fino al k-esimo. Questo deriva dal fatto che:

$$|x_i|^r p_i \leq |x_i|^k p_i \quad se \quad i > N$$

Fisz (1963, pp. 67-69) rileva che l'esistenza finita del momento r-esimo implica l'esistenza del limite che segue:

$$\lim_{a \to \infty} a^r P\left(|x| > a\right) = 0$$

dove "a" è una costante finita. La relazione lega i momenti alla probabilità delle modalità grandi (in valore assoluto); solo se le probabilità nelle code sono piccole i momenti possono esistere finiti, ma quanto piccole? Se il modello deve avere finito μ_r allora la probabilità negli estremi deve diminuire con il reciproco della potenza r-esima della variabile: $P(|X|>a)=o(x^{-r})$.

Esempi:

a) Supponiamo che il costo annuale di un fattore di produzione dipenda linearmente dal numero di pezzi prodotti e geometricamente dal costo di stoccaggio.

$$x_i = i\left(2^i\right), \quad p_i = \left(\frac{1.71750326}{2^i}\right)\left(\frac{1}{i^2}\right); \;\; i = 1,2,\ldots,$$

La distribuzione di probabilità riflette l'aspettativa di produrre essenzialmente un pezzo l'anno (85%) e già per 4 pezzi annuali la probabilità è 1/500'000. Il valore atteso del costo è:

$$\mu = \sum_{i=1}^{\infty} i\left(2^i\right)\left(\frac{1.71750326}{2^i}\right)\left(\frac{1}{i^2}\right) = 1.71750326\sum_{i=1}^{\infty}\frac{1}{i} \to \infty$$

C'è necessità di frenare gli ordini perché un'apertura eccessiva porterebbe i costi di produzione all'infinito.

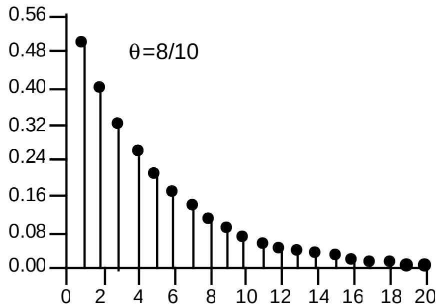
b) Verifichiamo l'esistenza finita di valore atteso e scarto quadratico medio per il modello logaritmico.

$$E(x) = \sum_{x=1}^{\infty} -\frac{x\theta^x}{xLn(1-\theta)} = -\frac{1}{Ln(1-\theta)}\left[-1+\sum_{x=0}^{\infty}\theta^x\right]$$

$$= -\frac{1}{Ln(1-\theta)}\left[-1+\frac{1}{1-\theta}\right] = -\frac{(1-\theta)}{Ln(1-\theta)}$$

$$E\left(x^2\right) = \sum_{x=1}^{\infty} -\frac{x\theta^x}{Ln(1-\theta)} = -\frac{1}{Ln(1-\theta)}\left[\sum_{x=0}^{\infty} x\theta^x\right]$$

$$= -\frac{1}{Ln(1-\theta)}\left[\frac{\theta}{(1-\theta)^2}\right] \Rightarrow \sigma^2 = -\frac{\theta}{(1-\theta)Ln(1-\theta)}$$



c) La componente esponenziale al denominatore del modello:

$$p_i = \frac{12}{2^i i^2\left[\pi^2 - Ln^2(2)\right]}; \;\; i = 1,2,\ldots$$

garantisce l'esistenza finita di tutti i momenti:

$$\mu_r = \sum_{i=1}^{\infty}\frac{i^r 12}{2^i i^2\left[\pi^2 - Ln^2(2)\right]} = \frac{12}{\left[\pi^2 - Ln^2(2)\right]}\sum_{i=1}^{\infty}\frac{i^{r-2}}{2^i} = convergente$$

Esercizio_VC74: *si consideri il seguente modello:*

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

Verificare che esiste finito il momento primo, ma non il momento secondo. Come si interpreta?

Esercizio_VC75: *il modello di variabile discreta Zeta ha distribuzione:*

$$p(x) = \frac{c^{-1}}{x^{1+a}}; \;\; i = 1,2,\ldots; \quad c = \sum_{j=1}^{\infty}\frac{1}{j^{1+a}} \;\; \textit{(funzione di Riemann)}$$

a) Determinare E(X) e σ(x) per a=4;
b) In quali occasioni può essere applicato tale modello?

Il problema della media (o degli altri momenti) infinita non riguarda, come è ovvio, le rilevazioni empiriche in cui i momenti sono finiti per forza di cose: tutto quello che le concerne infatti si mantiene nel limitato. Allora perché un modello dovrebbe recepire una caratteristica che non può essere presente in ciò che si intende rappresentare? Risposta facile. Da un lato è possibile intuire la presenza di un incremento o decremento progressivo verificando la logica assenza di un limite al fenomeno nella o nelle direttrici di espansione. D'altra parte, se è impossibile replicare all'infinito un esperimento anche molto semplice come il lancio di un dado o di una moneta, è anche vero che nulla può impedire di concepire un modello in cui le repliche siano illimitate e studiarne il comportamento in base ad alcune ipotesi senza mai -fisicamente- lanciare un dado o una moneta una sola volta. Tali modelli, a buon diritto, rivendicano un dominio infinito.

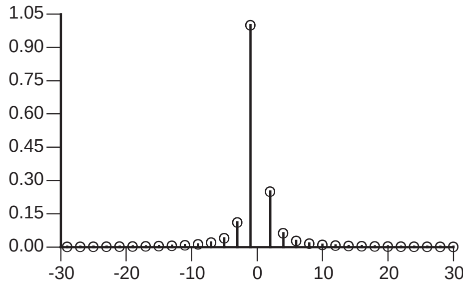
Momenti indeterminati

Nei modelli per variabili enumerabili insorge un ulteriore problema. Ci si aspetta che il calcolo del valore atteso di una qualsiasi funzione della variabile casuale dia lo stesso risultato qualunque sia l'ordine degli addendi. Per le serie finite ciò è scontato, ma per assicurarlo nelle serie infinite sono necessarie condizioni aggiuntive, almeno per quelle coinvolgenti modalità con segni alterni.

Esempio:
Consideriamo un modello con modalità di segno alterno in progressione aritmetica:

$$x_i = (-1)^i i, \quad p_i = \frac{0.6079271}{i^2}; \quad i = 1, 2, \dots$$

$$\mu = 0.6079271 \sum_{i=1}^{\infty} \frac{(-1)^i i}{i^2} = 0.6079271 \sum_{i=1}^{\infty} \frac{(-1)^i}{i} = -0.6079271 Ln(2)$$



La media è finita purché la somma avvenga nella sequenza indicata. Se però decidiamo di accorpare i segni negativi ed i segni positivi il risultato è diverso:

$$\mu = 0.6079271\left[\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\dots\right)-\left(1+\frac{1}{3}+\frac{1}{5}+\dots\right)\right] = 0.6079271\left[\frac{1}{2}\left(1+\frac{1}{2}+\frac{1}{3}+\dots\right)-\left(1+\frac{1}{3}+\frac{1}{5}+\dots\right)\right]$$

Il primo addendo è la serie armonica che è divergente ed il secondo è una maggiorante della serie armonica ed è pure divergente. Quindi $\mu=+\infty -\infty$ e la media risulta indeterminata.

Esiste uno strumento semplice per accertare che l'aspettativa della funzione g(x) non dipenda dall'ordine degli addendi: la condizione di convergenza assoluta:

$$\sum_{i=0}^{\infty} \left|g(x_i)p(x_i)\right| = \sum_{i=0}^{\infty} \left|g(x_i)\right| p(x_i) = \left|g(x_1)\right| p(x_1) + \left|g(x_2)\right| p(x_2) + \dots < \infty$$

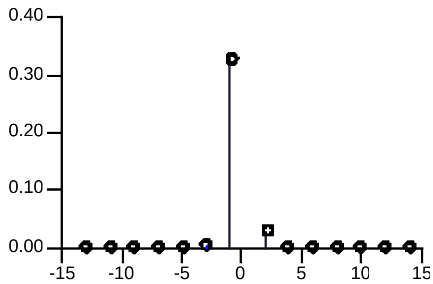
Se esiste finito E{|g(x)|} la serie è assolutamente convergente e converge alla stessa somma quale che sia l'ordine degli addendi (cfr. ad esempio Khuri, 1993, p. 162). Da notare che la convergenza di $\{\sum g(x_i)p_i, i=1,2,\dots\}$ implica la convergenza di $\{\sum |g(x_i)| p_i, i=1,2,\dots\}$, ma il contrario non è sempre vero. Se dovesse succedere che la prima diverge e la seconda converge si dirà che il valore atteso della g(x) non esiste o che è indeterminato.

Esempi:
a) Modello con oscillazioni smorzate.

$$x_i = (-1)^i, \quad p(x) = \frac{0.64805427}{(2x)!}; \quad x = 0, 1, 2, \dots$$

$$\sum_{i=0}^{\infty} \left|(-1)^i\right| \frac{1}{(2i)!} = \sum_{i=0}^{\infty} \left|\frac{1}{(2i)!}\right| = \frac{1}{2}\left(e + \frac{1}{e}\right) = 1.54308$$

$$\mu = 0.64805427 \sum_{i=0}^{1} \frac{(-1)^i}{(2i)!} = 0.35014522$$



b) Modello con oscillazioni esponenziali.

$$p_i = \frac{0.6079271}{i^2}; \quad i = 1, 2, \dots; \quad g(i) = (-1)^i i, \quad \sum_{i=1}^{\infty} \left|\frac{(-1)^i}{i}\right| = \sum_{i=1}^{\infty} \frac{1}{i} = divergente$$

Esercizio_VC76: *la variabile casuale X ha distribuzione:* $p_i = \left[\frac{1}{Ln(2)}\right]\frac{1}{i2^i} \quad per \quad i = 1, 2, \dots$

a) Verificare l'esistenza della media armonica; b) Calcolare il momento secondo μ_2.

7.2.2 Modelli di variabili casuali enumerabili

Alcuni modelli di questo tipo hanno un ruolo di primo piano nella descrizione di aspetti interessanti degli esperimenti ripetuti un numero elevato di volte nelle medesime circostanze, rappresentati con variabili dal domino infinito: numero di cellule tumorali in un campione di tessuto, utenti che richiedono una transazione telematica, realizzazioni necessarie per ottenere un risultato, il numero di disfunzioni in un ciclo di produzione.

La variabile casuale di Poisson

Gli esperimenti che possono dar luogo a questa variabile sono di due tipi: sequenze di prove bernoulliane e ripetizione di eventi in ambiti di tempo o di spazio limitati. Il primo tipo di esperimenti considera una successione di prove bernoulliane tendente all'infinito tale che:

1. Sottosequenze di prove che non si sovrappongono sono indipendenti; ad esempio, la successione degli esiti di posto pari è indipendente da quella degli esiti di posto dispari.
2. La probabilità di rinvenire un successo dipende dalla lunghezza della sequenza e non dal suo punto di inizio. In particolare, la probabilità di successo è proporzionale alla lunghezza della sequenza.
3. E' poco probabile che più successi si verifichino in prove ravvicinate (il successo è un evento raro).

Esempi:

a) Ogni attraversamento a raso della statale 106 jonica calabrese è un punto a rischio che può dar luogo a due esiti: "incidente/non incidente". Se gli incidenti sono più probabili di giorno che di notte la condizione di indipendenza tra l'esito delle prove è violata. Sarebbe pure violata se le probabilità fossero diverse per i due sensi di marcia.

b) Si dice che le disgrazie non vengano mai da sole: nello stesso giorno avete perso il biglietto vincente i 50 milioni di una lotteria, di avere l'influenza, chi vi sta a cuore non chiamerà più e la prova di statistica è andata male. La condizione "3" non è rispettata.

c) E' noto che i bisonti della prateria americana sono delle vere e proprie isole di autosufficienza: uno può essere abbattuto dal cacciatore e l'altro continuare pacificamente a ruminare. Fiutano il nemico, ma restano immobili finché il maschio dominante non decide di caricare. A quel punto diventano una marea inarrestabile: i singoli animali sono prove separate, ma non sono indipendenti visto che seguono tutti la direzione del capobranco. Un fenomeno analogo si riscontra osservando uno stormo di anitre selvatiche in migrazione: le posizioni della formazione a "<" sono dettate dalla migliore efficienza di volo e non sono indipendenti. L'ipotesi è sempre violata in fenomeni in cui le unità tendano naturalmente a raggrupparsi.

Immaginiamo un esperimento in cui rimanga fisso il valore atteso $np_n=\lambda$; cioè si conduce un esperimento che consiste in "n" prove bernoulliane ed in queste "n" prove c'è indipendenza e probabilità costante di successo p_n. Poi si passa a "n+1" e si considerano (n+1) prove dello stesso tipo con probabilità p_{n+1}; si definisce così un'altra binomiale simile dalla prima perché entrambe legate da $np_n= \lambda =(n+1)p_{n+1}$.

$$p_n(X=x)=\binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}=\frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}=\frac{\lambda^x}{x!}\frac{n(n-1)\ldots(n-x+1)}{n^x}\left(1-\frac{\lambda}{n}\right)^n\left(\frac{n}{n-\lambda}\right)^x$$

Al tendere di "n" all'infinito (notare il pedice nella "p") la variabile casuale binomiale perde la sua conformazione originaria e si trasforma nella variabile casuale di Poisson:

$$\lim_{n\to\infty} p_n(X=x)=\frac{\lambda^x}{x!}\lim_{n\to\infty}\frac{n(n-1)\ldots(n-x+1)}{n^x}\left(1-\frac{\lambda}{n}\right)^n\left(\frac{n}{n-\lambda}\right)^x=\frac{\lambda^x}{x!}*1*e^{-\lambda}*1$$

$$p(X=x)=\frac{\lambda^x e^{-\lambda}}{x!},\quad x=0,1,2,\ldots,$$

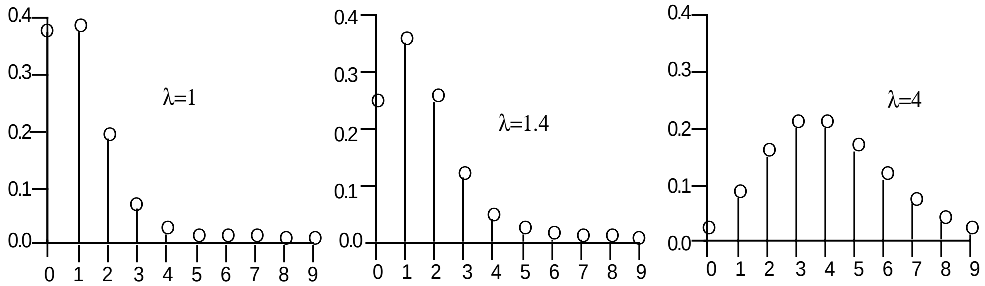
Per verificare che la somma delle probabilità sia unitaria si parte dall'espansione in serie di MacLaurin:

$$\sum_{x=0}^{\infty}\frac{\lambda^x e^{-\lambda}}{x!}=e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2}+\frac{\lambda^3}{3!}+\ldots\right]=e^{-\lambda}e^{\lambda}=1$$

La Poisson dipende da un solo parametro λ che si carica di significati diversi secondo il contesto applicativo ed ha inoltre un'interpretazione articolata perché condensa caratteristiche disparate della distribuzione.

Esempi:

a) Il numero di interruzioni di energie elettrica nel campus è un “evento raro” e quindi modellabile con la variabile casuale di Poisson. Ecco l’andamento della distribuzione di probabilità per tre diversi valori di λ.



La distribuzione è unimodale se λ è frazionario e la moda è X=[λ] cioè la parte intera del suo parametro; se λ è intero si ha:

$$p(X=\lambda)=\frac{\lambda^{\lambda}e^{-\lambda}}{\lambda!}=\frac{\lambda\lambda^{\lambda-1}e^{-\lambda}}{\lambda(\lambda-1)!}=\frac{\lambda^{\lambda-1}e^{-\lambda}}{(\lambda-1)!}=p(X=\lambda-1)$$

con due mode. La mediana cade tra [λ]-1 e [λ] o tra [λ] e [λ]+1 secondo l’entità della parte frazionaria di λ (cfr. Haight, 1967, p.12).

b) Il numero di imperfezioni in un tappeto può essere espresso da una Poisson con media λ=2.2. Effettuiamo alcuni calcoli.

$$p(X=0)=\frac{2.2^{0}e^{-2.2}}{0!}=e^{-2.2}=0.1108;\qquad p(X\le 2)=\sum_{x=0}^{2}\frac{2.2^{x}e^{-2.2}}{x!}=0.1108+0.2438+0.2681=0.6287;$$

0	1	2	3	4	5	6	7	8	9	10
0.11080	0.24377	0.26814	0.19664	0.10815	0.04759	0.01745	0.00548	0.00151	0.00037	0.00008

La moda è ottenuta per X=[2.2]=2; anche la mediana si ottiene per X=2.

Il ruolo del parametro λ è meglio compreso analizzando i primi momenti della variabile casuale

$$E(x)=\sum_{x=0}^{\infty}x\frac{\lambda^{x}e^{-\lambda}}{x!}=0+\sum_{x=1}^{\infty}x\frac{\lambda^{x}e^{-\lambda}}{x!}=\lambda\sum_{x=1}^{\infty}\frac{\lambda^{x-1}e^{-\lambda}}{(x-1)!}=\lambda\sum_{y=0}^{\infty}\frac{\lambda^{y}e^{-\lambda}}{y!}=\lambda*1=\lambda$$

$$\mu_{2!}=\sum_{x=0}^{\infty}x(x-1)\frac{\lambda^{x}e^{-\lambda}}{x!}=0+0+\sum_{x=2}^{\infty}x(x-1)\frac{\lambda^{x}e^{-\lambda}}{x!}=\lambda^{2}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}e^{-\lambda}}{(x-2)!}=\lambda^{2}\sum_{y=0}^{\infty}\frac{\lambda^{y}e^{-\lambda}}{y!}=\lambda^{2}$$

$$\mu_{2}=\mu_{2!}+\mu=\lambda^{2}+\lambda\Rightarrow\sigma^{2}=\mu_{2}-\mu^{2}=\lambda^{2}+\lambda-\lambda^{2}=\lambda;\quad \mu_{3}=\lambda;\ \mu_{4}=\lambda+3\lambda^{2}$$

Nel modello di Poisson il parametro λ esprime anche la varianza. Non solo, compare pure nell’indice di asimmetria: $\gamma_1=\lambda^{-0.5}$ e la distribuzione diventa simmetrica all’aumentare di λ.

Esempi:

a) Un esempio storico pubblicato da von Bortkiewicz nel 1898 e riferito da Fisz (1963) è relativo al numero di cavalleggeri morti per il calcio di cavallo in un periodo di venti anni in dieci reggimenti di cavalleria.

x_i	n_i	f_i	p_i
0	144	0.5143	0.5434
1	91	0.3250	0.3314
2	32	0.1143	0.1011
3	11	0.0393	0.0206
4	2	0.0071	0.0031
≥5	0	0.0000	0.0004
	280	1.0000	1.0000

La media aritmetica è 0.61; inseriamolo nella Poisson come λ=0.61 e calcoliamo le probabilità. L’adattamento è più che soddisfacente.

b) Se in un flusso di dati casuali binari il numero X di lunghette di “r” valori “1” si presentano in ragione costante con il crescere del flusso in modo che [(r+1)!/n]=(2/λ) allora P(X=r)=lre-l/r!

***Esercizio_VC77:** a) Dimostrare che per la Poisson esistono finiti i momenti di ogni ordine;*
b) Dimostrare che nel modello di Poisson p(X=pari)=0.5(1+$e^{-2\lambda}$)

***Esercizio_VC78:** gli studiosi Rutherfod e Geiger hanno osservato il numero di particelle alfa emesse da un barra di radio in 2'606 intervalli di 7.5 secondi. a) Calcolare la media aritmetica per usarla come parametro per le corrispondenti probabilità della Poisson; b) Rappresentare in uno stesso grafico la distribuzione empirica e teorica commentando le differenze.*
c) Ripetere il punto a) stimando λ con la formula di Pillai. Ci sono differenze? quali problemi comporta la presenza di più di una stima per lo stesso parametro?

x	f	x	f
0	0.0219	6	0.1047
1	0.0778	7	0.0533
2	0.1469	8	0.0173
3	0.2013	9	0.0103
4	0.2040	≥10	0.0061
5	0.1564		
			1.0000

$$\hat{\lambda} = \frac{\sum_{i=0}^{k} \frac{n_i n_{i+1}}{i+1}}{\sum_{i=0}^{k} \left(\frac{n_i}{i+1}\right)^2}$$

Approssimazione della binomiale

Per come è stata costruita la Poisson dovrebbe fornire delle buone approssimazioni della Binomiale quando è applicata ad eventi rari. In particolare è richiesto che n≥100 e p≤0.1. Vediamo se l'approssimazione che fornisce è anche comoda dal punto di vista computazionale.

Esempio:

a) In una versione precedente di questo libro ho riscontrato che c'era un errore (di stampa, di data, di formule) più o meno ogni 200'000 parole o formule. Ipotizzando che l'errore in una parola o formula non influenzi l'errore in un'altra (questo non è pacifico dato che la presenza di un errore può innervosire chi scrive e favorire l'insorgenza di altri errori) si può modellare la presenza di refusi con una binomiale in cui p=1/200'000. Se la stessa probabilità è mantenuta per il testo che state leggendo che si compone di circa 4'800'000 termini, qual'è la probabilità di osservare più di 10 errori nell'intero testo?

$$1 - p(X \le 10) = 1 - \sum_{i=0}^{10} \binom{4800000}{i} \left(\frac{1}{200000}\right)^i \left(\frac{199999}{200000}\right)^{4800000-i} \cong \sum_{i=0}^{10} \frac{24^i e^{-24}}{i!} = 0.001$$

Da notare che il calcolo di 4'800'000! che compare nella binomiale ipotizzando che sia sufficiente in media un millesimo di secondo (velocità non certo banale per i numeri coinvolti in ogni prodotto e memorizzazione intermedia) richiederebbe 80 minuti.

b) Una società di assicurazione ha stipulato 20'000 polizze RC auto in una regione. Dalle passate esperienze si può ritenere che la probabilità di un sinistro con danni biologici è p=0.0005 per ognuno dei suoi assicurati. La direttrice dell'agenzia regionale rischia il posto (per aver scelto male i clienti) se gli incidenti anzidetti sono 15 o più. Qual'è la probabilità che si eviti tale evento?

$$p(X < 15) = \sum_{i=0}^{14} \binom{20000}{i} (0.0005)^i (0.9995)^{20000-i} \cong \sum_{i=0}^{14} \frac{10^i e^{-10}}{i!} = 0.9165$$

c) Se la variabile casuale ipergeometrica approssima la binomiale e quest'ultima è ben approssimata dalla Poisson c'è da ritenere che la Poisson possa fornire una approssimazione accettabile anche per l'ipergeometrica. Consideriamo il caso della revisione di una categoria di conti in cui si adotti un campionamento per unità monetaria (Wilburn,1984, cap.12) e che l'importo complessivo sia di 250'000 euri. In base alle informazioni sulla contabilità dell'azienda si è stabilito che la probabilità di un euro in una transazione errata sia $p=N_1/N=0.005$. Il revisore sceglie 3000 eurounità ovviamente senza reimmissione. Se gli errori sono 4 o più l'intero importo è sottoposto a controlli capillari. Qual'è la probabilità di attivare tale procedura?

$$p(X \ge 4) = 1 - p(X \le 3) = 1 - \sum_{x=0}^{3} \frac{\binom{125}{x}\binom{249875}{3000-x}}{\binom{250000}{3000}} = 1 - \sum_{x=0}^{3} \binom{3000}{x} (0.005)^x (0.995)^{3000-x} = 1 - \sum_{x=0}^{3} \frac{1.5^x e^{-1.5}}{x!} = 1 - 0.9344 = 6.6\%$$

d) Amato (1950) suggerisce che, per un "x" fissato, l'approssimazione è valida se:

$$\left| 1 - \frac{(x-\lambda)^2}{np(1-p)} \right| p < \frac{1}{5}$$

***Esercizio_VC79:** a) La demografia segnala che la probabilità di un parto trigemino è 1 su 8'000. Ci si chiede quale sia la probabilità di osservarne almeno 4 su 10'000 parti;*
b) Per quale valore di λ il modello di Poisson verifica p(X=1)=p(X=0);
c) Se, in un modello di Poisson, il primo quartile si ha per X=0, quanto vale λ?

***Esercizio_VC80:** "Videotre" dichiara che nella fascia oraria 22:00-23:00 ha una share del 10% nel suo bacino di utenza che comprende circa 100 mila famiglie. Una società di collocazione pubblicitaria verifica la dichiarazione selezionando casualmente e senza reimmissione 150 famiglie residenti nella zona. Qual'è la probabilità -data per vera l'asserzione della rete locale- che non più di 10 famiglie siano sintonizzate su Videotre?*

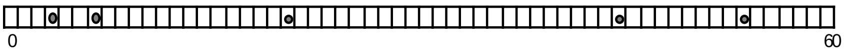
Esercizio_VC81: *la società che gestisce una linea di 12 camion per il trasporto dei rifiuti solidi urbani ha fatto montare su ogni camion 8 pneumatici di cui 2 di marca Pexima. I camion fanno 10 volte il tragitto andata e ritorno e questo per 25 giorni al mese. Il mese scorso sono diventati inservibili 45 pneumatici di cui 27 di marca Pexima. C'è da ritenere che i guasti capitati ai pneumatici Pexima rientrino nella norma oppure vi sembra che rispecchino il loro nome?*

Poisson come eventi in un continuo

La variabile casuale di Poisson è qualcosa di più che la forma limite del modello binomiale. E' infatti la descrizione di un esperimento basilare della Statistica: il monitoraggio per un tempo o in un ambito circoscritto di un fenomeno soggetto alla sorte.

Esempio:

Ripreso da Moskowttz e Wright (1985, p. 178) In un dato istante il telefono di un centralino può squillare (successo) o non squillare (insuccesso). Supponiamo che il telefono suoni, in media, 5 volte in un ora; dividiamo l'ora in 60 intervalli dell'ampiezza di un minuto e consideriamo l'intervallo una prova bernoulliana con probabilità di successo: 5/60.



Se le chiamate sono indipendenti, se in ogni intervallo non può arrivare più di una telefonata e se la probabilità di chiamata è la stessa per ogni intervallo, la situazione è modellabile con una binomiale avente n=60 e p=5/60=1/12

$$p(X=x)=\binom{60}{x}\left(\frac{1}{12}\right)^x\left(\frac{11}{12}\right)^{60-x}$$

L'intervallo di un minuto è forse troppo lungo perché siano valide le condizioni richieste. Raffiniamo la suddivisione e passiamo ai secondi, cioè una successione di n=3'600 prove bernoulliane con p=5/3600=1/720 fermo restando np=5.

$$p(X=x)=\binom{3600}{x}\left(\frac{1}{720}\right)^x\left(\frac{719}{720}\right)^{3600-x}$$

Spingendo oltre il processo di restrizione (ogni decimo di secondo, ogni centesimo, millesimo) si rendono sempre più plausibili le condizioni della sequenza di binomiali e quindi il richiamo della Poisson.

Sia X il numero di ripetizioni nell'intervallo [0,t] e dividiamo tale intervallo in "n" subintervalli di ampiezza costante t/n supponendo che in ognuno di essi il numero di eventi sia zero oppure uno con $p(X=1)=\lambda t/n$ dove λ è il numero medio di accadimenti nell'unità di tempo "t"; stabiliamo infine che gli eventi siano indipendenti e che la probabilità rimanga costante (ipotesi di processo stazionario). Ne consegue:

$$p_n(X=x)=\binom{n}{x}\left(\frac{\lambda t}{n}\right)^x\left(1-\frac{\lambda t}{n}\right)^{n-x} \Rightarrow \lim_{n\to\infty} p_n(X=x)=\frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

Esempi:

a) Gli utenti di un ufficio postale arrivano ad un ritmo di 7 in un'ora. Se gli arrivi sono modellabili con una Poisson qual'è la probabilità che ne arrivino 5? Se lo sportello può sbrigare senza ingolfamenti 10 persone all'ora qual'è la probabilità di più di 10 arrivi?

$$p(X=5)=\frac{(7)^5e^{-7}}{5!}=12.8\%;\quad P(X>10)=1-p(X\le 10)=\sum_{i=0}^{10}\frac{(7)^i e^{-7}}{i!}=9.9\%;$$

Se le condizioni non cambiano da un'ora all'altra si può effettuare -come è già avvenuto per i rapporti statistici- l'estensione proporzionale della media cioè se per un'ora arrivano in media 7 utenti, in due ore ne arriveranno in media 14 e quindi:

$$p(X=5)=\frac{(14)^5e^{-14}}{5!}=0.37\%;\quad p(X>10)=1-p(X\le 10)=\sum_{i=0}^{10}\frac{(7)^i e^{-7}}{i!}=82.4\%;$$

Se la stessa proporzionalità vige anche per i minuti possiamo applicare una Poisson con $\lambda=7*m/60$ dove "m" è il numero di minuti considerati. Per m=12 otteniamo:

$$p(X=5)=\frac{(1.4)^5e^{-1.4}}{5!}=1.11\%;\quad p(X>10)=1-p(X\le 10)=\sum_{i=0}^{10}\frac{(1.4)^i e^{-1.4}}{i!}=0.0002\%;$$

b) L'impostazione del modello di Poisson non è esclusiva del criterio organizzativo temporale, ma si può ragionare anche su serie spaziali. Supponiamo che le saldature lungo una conduttura di acqua presentino una disfunzione ogni 250 metri. La probabilità che in un tratto qualsiasi lungo 2 chilometri si verifichino non più di 3 disfunzioni è:

$$p(X\le 3)=\sum_{i=0}^{3}\frac{(8)^i e^{-8}}{i!}=4.2\%$$

c) Il modello di Poisson è anche adatto a rappresentare eventi che in una un'area limitata suddivisa in sottounità territoriali (maglie, particelle, *plot*, etc.) di dimensione molto ridotta e tali che: 1) eventuali diversità di forma e superficie non incidono sulla probabilità che un evento si verifichi sulla singola maglia e 2) ciò che avviene in un una maglia non sia legato a ciò che avviene in un'altra. Il 12-4-1943 gli aerei americani bombardarono -senza alcuna ragione- Cosenza vecchia.

Immaginando di dividere la zona in quadrati regolari di 10 metri di lato si arriva a circa 30’000 unità. E’ chiaro che non tutte le maglie sono sullo stesso piano: alcune sono vicine a strutture strategiche: ponti, caserme, edifici pubblici, ma l’agglomerato è così fitto che l’equiprobabilità e l’indipendenza rimangono valide, almeno approssimativamente. Gli aerei sgangiarono 70 bombe cosicché il numero medio di bombe atteso nella zona di S. Francesco (circa 15 maglie) è λ=70*15/30000=0.035. La probabilità che la zona fosse colpita da almeno una bomba era:

$$p(X \geq 1) = 1 - p(X = 0) = 1 - \frac{(0.35)^0 e^{-0.35}}{0!} = 1 - 0.9656 = 3.44\%;$$

Quel giorno l’evento raro incluse i bisnonni paterni dell’autore insieme ad altre 41 persone innocenti.

Esercizio_VC82: *il numero di incidenti annuali in una centrale nucleare segue la Poisson con λ=1. a) Qual’è il numero medio di incidenti in due anni? b) Qual’è la probabilità che lo scorso semestre si sia verificato un incidente? c) Qual’è la probabilità che si verifichi almeno un incidente nei prossimi dieci anni?*

Esercizio_VC83: *il numero dei posti occupati giornalmente in un parcheggio di lungo periodo segue il modello di Poisson con media λ=8. Di quanti posti “m” dovrebbe disporre perché siano sufficienti, con probabilità di errore del 95%, per una settimana?*

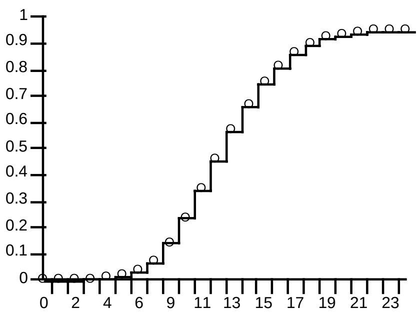
La funzione di ripartizione della variabile casuale di Poisson è: $F(x) = \begin{cases} 0 & se \;\; x < 0 \\ e^{-\lambda} \sum_{j=0}^{i} \frac{\lambda^i}{i!}, & se \;\; i \leq x < i+1 \end{cases}$

Esempio:

Il numero medio di morti per *overdose* in una provincia meridionale è di 1.4 al giorno ed il fenomeno è modellabile con una variabile casuale di Poisson. Indichiamo con X il numero morti in una settimana. Allora X sarà una Poisson con parametro: λ=1.4*7=12.6 con funzione di ripartizione:

$$F(x) = \begin{cases} 0 & se \;\; x < 0 \\ e^{-12.6} \sum_{j=0}^{i} \frac{12.6^i}{i!}, & se \;\; i \leq x < i+1 \end{cases}$$

Il grafico della F(.) corre parallelo all’asse delle ascisse tranne nei punti x=i e per i=0,1,2,… in cui ha un salto di ampiezza data dal valore della distribuzione ad x=i.



Esercizio_VC84*: la presenza nell’aria delle molecole di un composto raro è modellabile con la Poisson. a) Quali condizioni debbono ricorrere perché sia plausibile tale congettura? b) Definire e rappresentare la funzione di ripartizione se, in media, si trovano 15 parti in un metro cubo di aria.*

La variabile casuale binomiale negativa

La sequenza di prove bernoulliane con probabilità costante di successo “p” può anche essere studiata spostando l’attenzione dall’esito delle estrazioni alle prove. Rileviamo con la variabile X il conto delle prove svolte prima di ottenere “r” successi dove “r” è un parametro che con “p” caratterizza il nuovo modello. Il numero minimo di prove è “r” e non c’è un tetto massimo dato che la casualità dell’esperimento non esclude una sequenza senza fine di insuccessi. Il dominio della X è pertanto l’insieme dei naturali a partire da “r”: {r,r+1,r+2, …}. La maturazione di “r” successi alla x-esima prova si concreta allorché nelle (x-1) prove precedenti ne siano stati realizzati (r-1) e l’r-esimo si verifica alla prova numero “x”. Una lunghetta di interesse è pertanto:

$$\underbrace{\overbrace{0 \cap 0 \cap 1 \cap 0 \ldots \cap 0 \cap 1 \cap 0}^{x-1 \;\; prove}}_{r-1 \;\; successi} \cap \overset{x^a \;\; prova}{1} \quad con \quad probabilità \;\; p^{r-1}(1-p)^{x-1-r+1} p = p^r (1-p)^{x-r}$$

Ognuna delle C(x-1,r-1) lunghette di (x-1) prove con (r-1) successi ed un successo alla x-esima è un caso favorevole dell’evento “r successi in x prove” per cui la distribuzione di probabilità è:

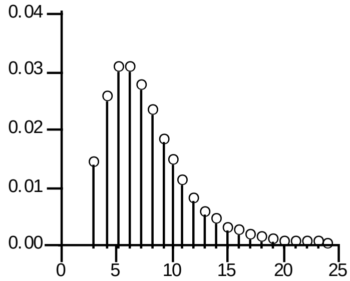
$$p(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, r+2, \ldots$$

nota come variabile casuale binomiale negativa o modello di Pascal.

Esempi:

a) Alla Rebecca informatica s.r.l. la probabilità di una chiamata per il servizio di consulenza è p=0.4 telefonate al minuto (il filtro del centralino non ne consente di più). Tre tecniche aspettano impazienti una telefonata così da poter uscire dall'ufficio. Qual'è la probabilità che arrivino le telefonate di richiesta nei prossimi 3, 4, 5 … minuti?

$$P(X=x)=\binom{x-1}{2}0.40^3(0.60)^{x-3}, \quad x=3,4,5,\ldots$$



La variabile casuale binomiale negativa è adoperata per modellare fenomeni (atmosferici, spaziali, sociali, sanitari) troppo eterogenei per essere compatibili con la Poisson.

b) Il conteggio degli incidenti o delle imperfezioni è l'essenza del monitoraggio delle linee di produzione. Immaginate un articolo in plastica in cui la presenza di un'asperità o di un buco sia un difetto. Un'ipotesi frequente è che sia all'opera un meccanismo che -ripetutamente- aggredisce il prodotto finché non provoca il difetto; ottenuto il risultato si rimette all'opera per causarne un altro. Il modello adatto è quello di Pascal che possiamo anche scrivere come:

$$p_i=\begin{cases} p^k & se \;\; i=0 \\ \dfrac{p^k k(k+1)(k+2)\ldots(k+i-1)}{i!}(1-p)^i & se \;\; i\geq 1 \end{cases}$$

dove "k" e "p" sono dei parametri.

Esercizio_VC85*: i requisiti per l'incarico di direttrice e vicedirettrice sono tali che ogni candidata ha probabilità del 3% di rientrare nel profilo. Determinare la distribuzione di probabilità del numero di candidate da esaminare per individuare le vincitrici.*

La funzione di ripartizione della variabile casuale di Pascal • la seguente:

$$F(x)=\begin{cases} 0 & se \;\; x<r \\ \sum_{j=r}^{i}\binom{i-1}{r-1}p^r(1-p)^{i-r} & se \;\; i\leq x<i; \;\; i\geq r \end{cases}$$

Esempi:

a) Una società sottoscrive polizze contro un incidente che ha probabilità p=0.01 di verificarsi ogni anno potendosi però permettere di rimborsarne 6. Quante polizze può sottoscrivere prima che il rischio di rimborsare più di 6 polizze sia superiore al 10%?

$$\sum_{j=6}^{i}\binom{i-1}{5}(0.01)^6(0.99)^{i-6} \leq 0.1 \;\Rightarrow\; i=317$$

b) La derivazione del modello di Pascal può avvenire con un ragionamento alternativo: X esprime il numero di insuccessi che precedono l'r-esimo successo. In tutto sono necessarie (x+r) prove di cui x+r-1 costellate da "x" insuccessi ed (r-1) successi essendo C(x+r-1,x) le lunghette che producono un tale risultato più l'r-esima che ha dato luogo al successo. Pertanto:

$$p(X=x)=\binom{x+r-1}{x}p^r(1-p)^x, \quad x=0,1,2,\ldots$$

che è lo stesso modello di prima con il solo cambio di variabile: y=x-r. Applichiamolo al caso di una coppia che ha probabilità di avere una figlia femmina p=0.488. Qual'è la probabilità che abbiano "x" figli maschi nel momento in cui nasce la 4ª femmina?

$$p(X=x)=\binom{x+3}{x}0.488^4 0.512^x, \quad x=0,1,2,\ldots$$

c) Se una squadra di calcio avesse probabilità costante di segnare un gol inuna partita, la distribuzione delle reti secondo le partite sarebbe modellabile con la Poisson. Poiché la probabilità cambia sia in ragione delle modifiche nella squadra che degli avversari oltre che di altri fattori (climatici, ad esempio) allora la distribuzione è modellabile con la Pascal (Pollard, 1973).

Controlliamo che la somma delle probabilitˆ del modello di Pascal sia uno. In questo senso basta richiamare il risultato sui coefficienti binomiali con elementi negativi (da cui il nome):

$$p^r\sum_{x=0}^{\infty}\binom{x+r-1}{x}(1-p)^x = p^r\sum_{x=0}^{\infty}\binom{-r}{x}[-p]^x = p^r\left(\frac{1}{p^r}\right)=1$$

Esempi:

a) Le riserve di sangue del gruppo AB nella clinica "Vivi & Sani" debbono essere integrate ed è necessario che arrivino 5 donatori con questo gruppo. La probabilità che un donatore abbia sangue di tipo AB è del 5%. La disponibilità di cassa della clinica permettono di pagare 50 donatori. Basteranno?

$$p(x \le 45) = \sum_{x=0}^{45} \binom{x+4}{x} 0.05^5 0.95^x \quad = 2.5\%$$

La probabilità che si trovino 5 donatori con gruppo AB nei primi 50 (un massimo di 45 donatori non appartenenti al gruppo AB) è solo del 2.5%. Le speranze di colmare la lacuna sono esigue anche se il rischio per la cassa è quasi nullo.

b) La presenza/assenza dal lavoro di un operaio è una variabile binaria con probabilità di assenza p=0.85. Supponiamo che i giorni di assenza consentiti siano r=7. Qual'è la probabilità che vengano saturati in un trimestre? Ogni giorno la probabilità di essere un "successo" è p=0.15; l'evento di interesse è la probabilità che siano necessarie 72 o meno prove per ottenere i 7 successi:

$$p(x \le 72) = \sum_{x=7}^{72} \binom{x+6}{x} 0.15^7 0.95^x \quad = 96.1\%$$

Esercizio_VC86: *un matematico porta sempre due scatole di fiammiferi identiche per accendere la pipa. La S_1 è scelta con probabilità "p" e l'altra (1-p). La probabilità è costante e le scatole contengono entrambe "n" fiammiferi. Calcolare la probabilità che se una delle due è vuota l'altra contenga ancora "r" fiammiferi.*

Esercizio_VC87: *una consulente finanziaria è alla ricerca di tre titoli alto/rischio alto rendimento. Dopo aver individuato un certo numero di titoli con forti perdite nell'ultimo trimestre decide di investire in ciascuno di essi successivamente e di continuare l'investimento finché non ottiene i tre titoli in guadagno. La consulente stima che la probabilità di recupero dei titoli da lei prescelti sia p=0.1; che probabilità ha di fermarsi prima dei 15 investimenti?*

Variabile casuale geometrica

Un caso particolare del modello di Pascal è la cosiddetta variabile casuale geometrica (r=1):

$$p(x) = p(1-p)^{x-1} \qquad x = 1, 2, \ \ldots; \quad F(x) = p\sum_{i=1}^{x}(1-p)^i \quad per \ \ i \le x < i$$

con cui si rileva il numero di prove da effettuare affinché si verifichi il primo successo.

Esempi:

a) Nel gioco del monopoli per uscire dalla prigione senza pagare la cauzione si deve ottenere un doppio (due punti uguali nelle facce superiori con probabilità 6/36=1/6). Qual'è la probabilità che siano necessari meno di 3 giri (cioè due insuccessi) per liberarsi?

$$p(x \le 2) = \left(\frac{1}{6}\right)\sum_{i=0}^{2}\left(\frac{5}{6}\right)^i = 42.13\%$$

b) Nel corso di un esame per il conseguimento di un patentino un candidato risponde a caso a tutte le domande. La probabilità di rispondere correttamente è 0.5 in ogni domanda. Inoltre, le domande non hanno legami tra di loro come prive di legami sono le scelte del concorrente. Se non si risponde bene ad uno dei primi 12 quesiti non si potrà ritentare eventualmente l'esame nella successiva tornata. Qual'è la probabilità che la prima risposta corretta arrivi entro il 12° quiz?

$$p(x \le 12) = \sum_{i=0}^{12} 0.25(0.75)^i = 97.6\%$$

Il concorrente può stare tranquillo: c'è solo una possibilità su cinquanta che sbagli tutti i primi 12 quiz.

Valore atteso e varianza della binomiale negativa si determinano facendo uso dei soliti accorgimenti:

$$E(x) = \sum_{x=0}^{\infty} x\binom{x+r-1}{x}p^r(1-p)^x = \sum_{x=0}^{\infty}\frac{x(x+r-1)!}{x!(r-1)!}p^r(1-p)^x = \frac{r(1-p)}{p}\sum_{x=1}^{\infty}\frac{(x+r-1)!}{(x-1)!r!}p^{r-1}(1-p)^{x-1} = \frac{r(1-p)}{p}$$

$$E[x(x-1)] = \sum_{x=0}^{\infty} x(x-1)\binom{x+r-1}{x}p^r(1-p)^x = \sum_{x=2}^{\infty}\frac{x(x-1)(x+r-1)!}{x!(r-1)!}p^r(1-p)^x =$$

$$= \frac{r(r+1)(1-p)^2}{p^2}\sum_{x=2}^{\infty}\frac{(x+r-1)!}{(x-2)!(r-1)!}p^{r-2}(1-p)^{x-2} = \frac{r(r+1)(1-p)^2}{p^2} \Rightarrow \sigma^2 = \left(\frac{r}{p}\right)\left(\frac{1-p}{p}\right)$$

Nel caso particolare del modello geometrico si ha: $E(x) = \dfrac{1-p}{p}; \qquad \sigma(x) = \dfrac{\sqrt{1-p}}{p}$

Esempi:

a) La strategia di Romilda. Si opera secondo il modello geometrico con probabilità ricavata con il seguente ragionamento: si ottiene un "successo" in caso di decisione conclusiva cioè se escono due teste o due croci. Quindi: P(successo)=P[(C∩C)∪(T∩T)]=0.5 In media Romilda deve effettuare μ=0.5/0.5=1 lancio con scarto quadratico medio σ=√2.

b) Determiniamo la moda della binomiale negativa:

$$_x = \frac{p(X=x)}{p(X=x-1)} = \frac{\binom{x+r-1}{x}p^r(1-p)^x}{\binom{x+r-2}{x-1}p^r(1-p)^{x-1}} = \frac{x+r-1}{x}(1-p)$$

p(X=x) cresce finché x<[a] con a=(r-1)(1-p)/p e poi decresce per x>[a]; la moda è l'x più grande per il quale il rapporto q_x resta superiore all'unità. Se "a" è un intero allora la distribuzione è bimodale con mode in x=[(r-1)(1-p)/p] e x=[(r-1)(1-p)/p]-1.

c) Su di un piano sono state formate "n" buche. Si lancia una biglia in modo che ogni buca abbia probabilità 1/n di ricevere una qualsiasi delle biglie (le buche possono ricevere più di una biglia). Si continua a lanciare biglie indipendentemente l'una dall'altra finché "k" buche qualsiasi contengano almeno una biglia. Sia T_k il numero di biglie da lanciare per ottenere tale risultato. Port (1994, p. 252) osserva che $P(T_1=1)=1$ e che T_k-T_{k-1} - 1 è una variabile casuale geometrica con p=(n-k+1)/n per k=2,3,… ed il cui valore atteso è:

$$1.\ E(T_k) = n\sum_{j=0}^{k-1}\frac{1}{(n-j)};\quad k=1,2,\ldots, \qquad 2.\ E(S_k) = n\sum_{j=1}^{k}\frac{1}{j};\quad k=1,2,\ldots,$$

La seconda formula del valore atteso ipotizza che le buche siano disposte in ordine e che siano riempite le prime "k". L'effetto di un trasferimento neutrale dall'unità più ricca a quella più povera sugli indici di De Vergotttini e di Bonferroni può esprimersi come:

$$1.\ \Delta B = \frac{\varepsilon}{(n-1)\mu}n\sum_{j=1}^{n-1}\frac{1}{j} = \frac{\varepsilon}{(n-1)\mu}E(T_n); \qquad \Delta V = \frac{\varepsilon}{(n-1)\mu}n\sum_{j=0}^{n-1}\frac{1}{(n-j)} = \frac{\varepsilon}{(n-1)\mu}E(S_n)$$

aggiungendo un originale risvolto interpretativo di questi due classici indici di concentrazione.

Esercizio_VC88: *Blom (1989, p. 48) illustra la variabile casuale PPV, (Per la Prima Volta): $P(X=x)=(1-p)^x p$, per x=0,1,2,… che esprime il numero di prove precedenti il primo successo. Calcolare μ e σ;*

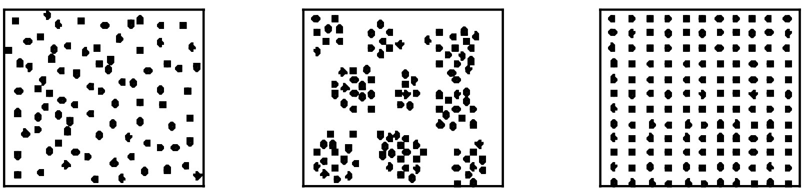
Esercizio_VC89: *Calcolare il momento terzo centrale della variabile casuale di Pascal e verificare che da asimmetrica positiva tende a diventare simmetrica per p=3/4.*

Scelta del modello discreto

Ogni modello esprime un processo astratto di osservazione che però si ravvisa, almeno a grandi linee, in molte rilevazioni reali. I presupposti e le limitazioni dei modelli, tuttavia, debbono trovare puntuale rispondenza nel fenomeno rappresentato in modo da non sembrarne una inutile e pericolosa caricatura.

Esempio:

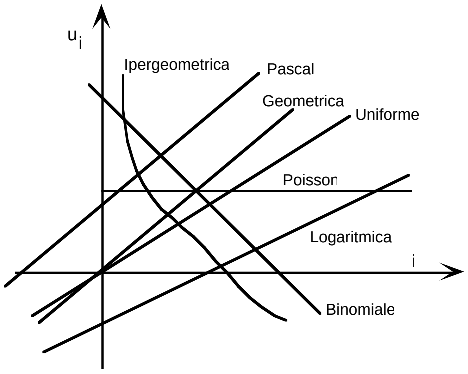
Per modellare le distribuzioni spaziali in ecologia ci si basa sull'indice di dispersione che è dato dal rapporto tra la varianza e la media aritmetica: ID=σ^2/μ.



ID=1 -->diffusione erratica e si usa la Poisson;
ID>1 -->indica la presenza di addensamenti e si usa la Pascal;
D<1--> diffusione equispaziata e si usa la uniforme discreta.

Tentiamo di risolvere il problema di come individuare il modello discreto. Nella vasta gamma di metodi che esistono per verificare se un certo modello descrive adeguatamente i dati empirici spicca il metodo suggerito da Ord (1967) che, sfruttando l'efficacia comunicativa dei grafici, ha proposto un diagramma cartesiano dove in ascissa è posto l'intero naturale "i" e in ordinata il rapporto tra probabilità di modalità consecutive proporzionato alla stessa "i":

$$u_i = i\left(\frac{p_i}{p_{i-1}}\right);\quad i=1,2,\ldots$$



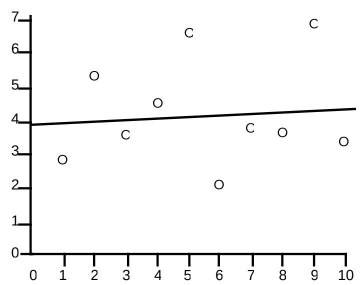
Diversi modelli di variabile casuale discreta possono essere espressi come relazione lineare tra i rapporti di probabilità successive e le modalità:

$$u_i=\beta_0+\beta_1 i,\ i=0,1,2,\ldots$$

Se i punti sembrano allinearsi lungo una retta con inclinazione negativa allora è adatto il modello binomiale; se l'inclinazione è positiva è adatto il Pascal purché l'intercetta sia positiva, altrimenti sarebbe più plausibile il modello logaritmico (intercetta negativa) o quello geometrico (intercetta nulla). Se la retta è parallela all'asse delle ascisse allora quello più appropriato è il modello di Poisson; se invece la retta passa per l'origine ed ha inclinazione unitaria occorre adattare l'uniforme. L'ipergeometrica si riconosce per un andamento discendente quadratico o quasi. La procedura è piuttosto *naive* e dovrebbe essere adoperata solo come indagine preliminare.

Esempio:
Analizziamo i dati di Holmes (1974). I punti non sembrano ben allineati il che lascia spazio alla Poisson (la leggera inclinazione positiva potrebbe però essere un segnale di avvicinamento alla Pascal o alla geometrica).

x_i	n_i	$p_i=n_i/200$	u_i
0	4	0.020	
1	11	0.055	2.7500
2	29	0.145	5.2727
3	34	0.170	3.5172
4	38	0.190	4.4706
5	50	0.250	6.5789
6	17	0.085	2.0400
7	9	0.045	3.7059
8	4	0.020	3.5556
9	3	0.015	6.7500
10	1	0.005	3.3333
	200		



Resterebbe da stimare il parametro l maggiormente compatibile con i dati. Ad esso si potrebbe arrivare attraverso la stima dei parametri b_0 e b_1 della retta la cui determinazione deve però essere rinviata ad un'altra parte del testo. Possiamo però stimare l con la media aritmetica o con la varianza i cui valori non sono troppo lontani: $\mu=4.035$ e $s^2=3.43$ che danno un'ulteriore conferma alla scelta della Poisson (N.B. questa prossimità sarebbe valida anche per il modello logaritmico).

Esercizio_OP90: *il modello di Yule (peraltro molto simile al modello Zeta) è stato impiegato per rappresentare la distribuzione (sebbene con meno successo che in altre applicazioni) degli autori per numero di pubblicazioni. Ecco i dati -riportati in Simon, 1955- relativi ad un ventennio di contributi scientifici ad "Econometrica":*
a) Confrontate le frequenze osservate con quelle teoriche ottenute per a=1;

$$p(x)=(x-1)!\left[\frac{\Gamma(2)}{\Gamma(x+2)}\right]=\frac{1}{(x+1)x};\quad x=1,2,\ldots,$$

Pubblicazioni	Autori		
1	436	7	6
2	107	8	11
3	61	9	1
4	40	10	0
5	14	≥11	22
6	23		721

b) Esiste una peculiarità del grafico di Ord che suggerisca di optare per il modello di Yule?

7.3 Variabili casuali continue

L'universo degli eventi discreto descrive esperimenti su fenomeni di conteggio o di enumerazione, ma è inadatto a trattare la scelta casuale di un punto in un intervallo sull'asse reale quale potrebbe essere richiesto per analizzare distanze, tempi, pesi, altezze, angoli, velocità. L'andamento di queste variabili può essere affrontato con il calcolo delle probabilità riconsiderando i postulati di Kolmogorov all'interno di questo nuovo tipo di esperimenti.

7.3.1 Universo degli eventi continuo

Come si è detto nel paragrafo 1.4.6 non è possibile la rilevazione puntuale di una variabile continua anche se, in via teorica, è del tutto lecito ragionare su suoi valori specifici. Ad esempio si può ritenere che la durata della scheda logica di un elettrodomestico sia di X=3758 giorni, ma questo non è verificabile in pratica. Ciò che una esperienza può provare è che la durata della scheda sia il valore centrale di un intervallo di ampiezza determinata dalla precisione degli strumenti di misura:

$$X_i - \frac{\Delta X_i}{2} \leq X_i \leq X_i + \frac{\Delta X_i}{2}$$

All'aumentare della capacità di dettaglio (ΔX_i), gli estremi dell'intervallo si avvicineranno, rimanendo però sempre distinti. Se invece di considerare un calendario graduato in giorni si considerano le ore, il diametro dell'intorno che circonda il valore si riduce e si ridurrà ancora se consideriamo i secondi e i nanosecondi. Per le variabili continue è impossibile far combaciare, attraverso successivi raffinamenti della misura, i due estremi dell'intervallo in modo che contenga solo un punto: per quanto se ne possa frazionare l'unità, persino a livello subatomico e oltre, l'intervallo $X_i \pm \Delta X_i$ conterrà un numero infinito di valori: questa è l'idea di continuità. Negli esperimenti con variabili continue gli eventi di interesse non sono valori isolati o insiemi di valori separati, ma intervalli di numeri reali dotati della cardinalità del continuo. Non è quindi più realizzabile un meccanismo di assegnazione della probabilità elencando le singole modalità e la rispettiva probabilità che ciascuna si porta dietro, ma occorre sfruttare selettivamente le nuove formulazioni dei postulati discusse nel paragrafo 7.2.

Esempio:
In un *call center* si teme l'arrivo a caso di una chiamata impegnativa proprio negli ultimi dieci minuti di apertura. L'universo degli eventi è l'intervallo di numeri reali [0, 10]. Se non si hanno informazioni sulla distribuzione delle telefonate nel tempo "arrivare a caso" significa che ogni possibile "istante" ha la stessa probabilità di vederla arrivare. Tuttavia, non è eseguibile il conteggio degli istanti perché l'unità di misura del tempo potrebbe essere talmente piccola da non poter più distinguere un istante da quello successivo o da quello precedente. Non è possibile perciò pensare ad "istanti" equiprobabili. Il senso di "arrivare a caso" si può interpretare immaginando, ad esempio, che la probabilità dipenda esclusivamente dalla lunghezza del sottointervallo. La probabilità che arrivi tra il 3° ed il 6° minuto potrebbe essere: (6-3)/10=30%.

La procedura di assegnazione della probabilità per intervalli è interessante: permetterebbe di dare una probabilità ad ogni sottointervallo superando l'ostacolo della impraticabile assegnazione individuale di probabilità. Tuttavia, se l'universo degli eventi S è l'intero asse reale R=(-∞, ∞) o anche solo una sua porzione, la collezione formata da tutti gli eventi definibili su R contiene anche eventi che non sono intervalli e che perciò non sarebbero probabilizzabili (è questa una prima differenza con gli universi discreti).

Esempi:
a) L'intersezione di una successione finita o enumerabile di intervalli è ancora un intervallo, ma l'unione di due intervalli non sempre è un intervallo. Siano : $a=x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$ punti distinti di R. Allora:

$$\bigcup_{i=1}^{n}\left(x_{i-1}, x_i\right] = (a,b] \ \ con \ \left(x_{i-1}, x_i\right] \cap \left(x_i, x_{i+1}\right] = \varnothing; \ \ \bigcup_{i=1}^{n}\left[x_{i-1}, x_i\right] = [a,b] \ \ con \ \left[x_{i-1}, x_i\right] \cap \left[x_i, x_{i+1}\right] \neq \varnothing;$$

$$\bigcup_{i=1}^{n}\left(x_{i-1}, x_i\right) \neq (a,b)$$

Per ottenere l'uguaglianza occorre aggiungere i punti singoli interni che non sono intervalli.

b) I riquadri di una scacchiera di 64 caselle sono variamente colorate: 16 sono bianchi. Se si lancia a caso uno spillo sulla scacchiera con quale probabilità la punta toccherà un riquadro bianco? La risposta intuitiva è 16/64=25% che però non può essere generalizzata in quanto è possibile definire nel piano reale insiemi così irregolari per i quali non è possibile dare una definizione accettabile di area.

c) Gli eventi periodici sono: $E(x_i)=\{x\in R|x=x_i\pm i,\ i=1,2,\ldots,\ x_i\in R\}$. L'insieme formato da tutti gli eventi periodici sull'asse reale è una σ-algebra dato che unione, intersezione e negazione di tali eventi determina eventi dello stesso genere (Feller, 1971, p. 113)

La classe di eventi costituita da tutti i possibili sottoinsiemi dell'asse reale non forma una σ-algebra dato che non può essere chiusa sotto l'operazione di unione (o intersezione o complemento). D'altra parte, tutte le σ-algebre ottenibili da S sono intermedie tra quella più povera $W_*=\{S,\emptyset\}$ e quella più ricca $W^*=\{E|E\subseteq S\}$. Se la cardinalità di S è quella del continuo la W^*, a differenza di quanto è successo nel discreto, è troppo ampia per poter essere probabilizzata nel suo complesso (e W_* rimane troppo semplice per essere utile). C'è da chiedersi allora se esiste una σ-algebra abbastanza grande da contenere gli eventi di un qualche interesse e, nello stesso tempo, abbastanza piccola da non portare a contraddizioni o violazioni quando ad essa si applichino i postulati di Kolmogorov. Fra tutte le σ-algebre di S con le quali si potrebbe ampliare lo schema di Kolmogorov per trattare esperimenti con un S continuo useremo la più piccola cioè l'intersezione di tutte le σ-algebra di S che, opportunamente, è ancora una σ-algebra (cfr. Monfort, 1980). Tale σ-algebra, detta minimale, esiste ed è unica ovvero se W_1 è un'altra σ-algebra di S, esiste una sola σ-algebra minimale W, tale che $W\subseteq W_1$ (Loève, 1977, p. 60).

La sigma-algebra di Borel

In molte prove, l'universo degli eventi è euclideo: la retta, il piano, lo spazio tridimensionale. Limitiamoci a ragionare su questo tipo di universi circoscrivendo l'attenzione all'asse reale R per mantenerci nella statistica univariata. Per costruire sopra R una σ-algebra più operativa di quella massimale adottata nel caso di S enumerabile, dobbiamo specificare quali siano i suoi elementi di base. Anche per lo spazio continuo adottiamo come evento strutturale l'intervallo semiaperto a sinistra e chiuso a destra con la differenza, rispetto agli universi discreti, che le modalità incluse nell'intervallo non sono necessariamente finite né enumerabili:

$$E(a_i,b_i]=\{x\in R|a_i<x\le b_i\};\quad a_1<a_2<\ldots<a_n;\ b_1<b_2<\ldots<b_n\ -\infty<a_1;\ b_n<\infty$$

In questo modo potremo sfruttare l'analogia tra probabilità da dare all'evento e lunghezza dell'intervallo che costituisce una comoda funzione di insieme σ-additiva e non negativa per gli intervalli. Per convenzione, l'intervallo $E(a,\infty)$ verrà inteso come $E(a,\infty]$ cosicché il suo complementare $E(-\infty,a]$ sarà pure un intervallo semiaperto a sinistra e chiuso a destra. In prima approssimazione, costruiamo W inserendo in essa tutti i possibili intervalli $E(a_i,b_i]$ nonché tutte gli eventi che si possono ottenere come negazioni, unioni e intersezioni di collezioni finite di tali intervalli.

$$Se\ \ E(a_i,b_i],E(a_j,b_j]\subset W\ \ con\ \ E(a_i,b_i]\cap E(a_j,b_j]=\emptyset\ \ per\ \ i\ne j\ \Rightarrow\ E_n=\bigcup_{i=1}^{n}E(a_i,b_i]\subset W$$

L'insieme vuoto è un caso particolare con $a_i>b_i$. Verifichiamo che la formulazione appena data abbia i requisiti dell'algebra.

$$1.\ \ Se\ \ E_n\subset W\ \Rightarrow E_n^c\subset W;\quad E_n^c=\left\{\bigcup_{i=1}^{n}E(a_i,b_i]\right\}^c=\bigcap_{i=1}^{n}E(a_i,b_i]^c=E(-\infty,a_1]\cup E(b_n,\infty]\subset W$$

$$2.\ \ Se\ \ E_n=\bigcup_{i=1}^{n}E(a_i,b_i]\ \ ed\ \ F_m=\bigcup_{i=1}^{m}E(c_i,d_i]\subset W\Rightarrow E_n\cup F_m=\bigcup_{i=1}^{n+m}E(\alpha_i,\beta_i]\subset W$$

$$con\ \ \alpha_i=min\{[a_i,c_i]-[\alpha_i,\alpha_2,\ldots,\alpha_{i-1}]\},\quad \beta_i=max\{[b_i,d_i]-[\beta_1,\beta_2,\ldots,\beta_{i-1}]\}$$

Quindi, W risulta chiusa sotto le operazioni di negazione ed unione finita (l'intersezione segue dalle formule di De Morgan). Non si tratta però di una σ-algebra. Infatti:

$$E_i=E\left(0,-\frac{1}{i}\right];\ i=1,2,\ldots\subset W\Rightarrow\bigcup_{i=1}^{\infty}E_i=(0,1)\not\subset W$$

L'intervallo (0,1) non rientra in W perché questa è formata da intervalli semiaperti e quelli aperti non vi sono inclusi. Per ottenere una σ-algebra dobbiamo aggiungere all'algebra W degli intervalli semiaperti a sinistra e chiusi a destra, i limiti delle loro successioni monotone.

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

Burrill (1972, p.43 e Shiryaev (1996, p.140) dimostrano che, condizione necessaria e sufficiente affinché un'algebra diventi una σ-algebra è che essa sia una classe monotona cioè includa anche le successioni crescenti o decrescenti di eventi. Se questo avviene, W diverrà la σ-algebra formata da tutti e solo quegli eventi ottenibili con unione, intersezione, negazione, differenza di un numero finito o enumerabilmente infinito di intervalli di numeri reali semiaperti a sinistra e chiusi a destra. Se poi consideriamo l'intersezione di tutte le σ-algebre contenenti gli intervalli $E(a_i,b_i]$ dell'asse reale si arriva a costituire, sia pure con un processo etereo, la σ-algebra minimale dell'asse reale B. La lettera "B" ricorda che questa classe è anche nota come algebra di Borel dal matematico e probabilista francese (1871-1956).

Esempi:
a) Sebbene la σ-algebra B scaturisca da collezioni di intervalli semiaperti a sinistra e chiusi a destra, l'inserimento delle successioni monotone porta nell'algebra tutti i tipi di intervalli:

$$E(a,b) = \bigcup_{i=1}^{\infty} E\left(a, b-\frac{1}{i}\right]; \quad E[a,b] = \bigcap_{i=1}^{\infty} E\left(a-\frac{1}{n}, b\right]$$

in cui gli estremi "a" e "b" possono anche essere infiniti: E(-∞,b); E(-∞,b], E(a,∞).

b) Il singoletto scaturisce pure dalla intersezione di intervalli semiaperti a sinistra e chiusi a destra:

$$E = \{a\} = \bigcap_{i=1}^{\infty} E\left(a-\frac{1}{n}, a\right]$$

Ogni insieme E{a} formato da un solo punto è in B e vi rientrano anche gli intervalli di interi finiti o enumerabilmente infiniti.

La B contiene tutti i sottoinsiemi di R per i quali abbia senso parlare della loro lunghezza; ma, naturalmente, B non contiene tutti i sottoinsiemi di R in quanto esistono sottoinsiemi di R non riconducibili al concetto di intervallo ed ai quali non potrà essere attribuita la probabilità secondo l'approccio adottato e ci si trova perciò con un impianto teorico in qualche modo inadeguato.

Esempi:
a) Daboni (1992, p. 15) sottolinea che l'evento E={xε[0,1]|x è un irrazionale} non è un intervallo cioè non è un elemento della classe di Borel per cui il sistema assiomatico non è in grado di dare una probabilità al suo verificarsi. Da notare che l'insieme dei numeri irrazionali ha la cardinalità del continuo.

b) Cifarelli (1998, p.22 e p.51) osserva: "*... sottoinsiemi di questo tipo non sono facili da costruire e nelle applicazioni non ricorrono mai*". Zanella, come riportano, Landenna e Marasini,1986, p. 114), è del parere che gli insiemi non probabilizzabili abbiano una mera natura concettuale il che equivale a dire che non possono succedere in pratica e non creano ostacoli.

Il passaggio da un universo discreto ad uno continuo comporta la ridefinizione degli eventi base dell'algebra: intervalli di interi nel primo caso, intervalli di reali nel secondo nonché il restringimento dalla σ-algebra più grande dell'universo discreto S alla σ-algebra minimale B dell'asse reale R. In verità, l'esposizione del modello di Kolmogorov può anche essere svolta presentando in modo unificato le tre diverse conformazioni dell'universo degli eventi: discreto finito, enumerabile, continuo (ad esempio: Billingsley, 1986, cap.1), ma l'esposizione differenziata è sembrata più adatta per i destinatari del presente testo. Inoltre, nessuna innovazione è necessaria per la funzione di probabilità da collegare alla struttura insiemistica sviluppata con la σ-algebra di Borel: una funzione non negativa, normalizzata e completamente additiva analoga a quella costruita per l'universo degli eventi enumerabili è perfettamente idonea allo scopo.

Esercizio_VC91: *un esperimento ricorrente è la ruota della fortuna cioè un cerchio diviso in settori numerati e fatto ruotare con una spinta. Se l'evento base è x= "numero reale in [0,1]" assegnare la probabilità agli eventi:*
a) E_1= l'esito è un numero razionale;
b) E_2= nelle cifre della rappresentazione decimale del numero non compare il "3";
c) E_3= la rappresentazione decimale è costituita da cifre pari.

Probabilità nel continuo

Le variabili casuali sono delle particolari funzioni che hanno come dominio l'universo degli eventi e l'asse reale come codominio. In questo senso useremo la notazione {x≤a} per indicare tutte le determinazioni in S che inducono un valore della variabile casuale inferiore o uguale ad "a":

$$\{x \le a\} = \{e \in S | X(e) \le a\}$$

in cui X(e)≤a è un insieme di risultati possibili nell'esperimento mentre x≤a è un insieme di valori in R. Allo stesso modo, l'intervallo E{a≤x≤b} sarà formato dagli elementi di S tali che

$$E = \{a \le x \le b\} = \{e \in S | a \le X(e) \le b\}$$

Per definire l'impianto probabilistico delle variabili casuali dobbiamo ora trasformare le probabilità definite sulla σ-algebra di Borel in probabilità delle variabili casuali per cui queste saranno collegate ad eventi composti. In altre parole, dobbiamo assicurarci che l'insieme x∈E corrisponda ad un evento di B qualunque sia l'insieme di interesse E.

Esempio:
Melsa e Sage (1973, p. 50) propongono la seguente riflessione. Si consideri l'esperimento di lanciare un dado e con $S=(e_1,e_2,\ldots,e_6)$ dove e_i=i e definiamo la variabile casuale X= "esito del lancio" cosicché $X(e_i)$=i, i=1,2,…,6 . Inoltre, per gestire l'esperimento adottiamo come σ-algebra la collezione W={ ϕ, S, E_1= "esito pari" , E_2= "esito dispari" } e solo a questi eventi forniamo una probabilità. L'insieme di valori della X: $\{x\le 2\}=\{e_1,e_2\}$ non rientra in W e quindi non possiamo assegnargli la probabilità. Se la variabile casuale che definiamo su S è la variabile binaria:

$$X(e_i) = \begin{cases} -1 & se \quad i = 2,4,6 \\ 1 & se \quad i = 1,3,5 \end{cases}$$

In questo caso gli insiemi {x≤a} sono sempre eventi rientranti in W ed ai quali può essere fornita una opportuna probabilità.

Una funzione reale di variabile reale X(.) è detta Borel-misurabile se per ogni "a" l'insieme: E={e∈R|X(e) ≤ a} ricade nella σ-algebra di Borel ovvero è un evento. In verità, il termine "misurabile" non è appropriato dato che non si è ancora delineata alcuna procedura di misurazione. Deve essere inteso nel senso che la probabilità è assegnata solo agli eventi in B e che il meccanismo di assegnazione X(e) è Borel-misurabile cioè prescrive valori o intervalli di valori ancora in B. La restrizione però non deve preoccupare dato che le applicazioni realistiche coinvolgono eventi di questo tipo.

Funzione di ripartizione delle variabili casuali continue

La variabile casuale associata con lo spazio di probabilità (S,P,B) è una funzione X(.) che trasferisce l'universo degli eventi S nell'asse reale R cioè assegna un numero reale x=X(e) ad e∈S in modo che l'insieme {x≤a} sia un evento in B per ogni "a". Si può dimostrare (Shiryaev, 1996, pp. 153-154; Cifarelli, 1998, pp. 51-55) che esiste una corrispondenza biunivoca tra una funzione di insieme σ-additiva: P(.) dello spazio di probabilità definito sull'asse reale attraverso la σ-algebra di Borel ed una funzione puntuale sull'asse reale F(.) dotata delle proprietà già viste per la funzione di ripartizione delle variabili casuali discrete:

1. $0 \le F(x) \le 1, \;\; x \in R;$
2. $F(-\infty) = 0; \;\; F(\infty) = 1;$
3. $F(x) = F(x+);$
4. $F(x_1) \le F(x_2) \;\; se \;\; x_1 < x_2$

L'assegnazione della probabilità agli intervalli della variabile casuale continua, avviene in base alla regola:

$$P\{a < x \le b,\} = P\{E(a,b]\} = P\left\{\bigcup_{i=1}^{n} E(a_i, b_i]\right\} \quad a \le a_1 < b_1 \le a_2 < b_2 < \ldots < b_n \le b$$

$$= \sum_{i=1}^{n}\left[F(b_i) - F(a_i)\right] = F(b) - F(a) \quad per \quad E(a,b] \subset B$$

Se le modalità di una variabile casuale discreta sono a livello nominale, possiamo -con un po' di forzatura- interpretare la loro $F(x_i)$ come un'aggregazione di indici ovvero, come nell'ortogramma paretiano, come un accorpamento sequenziale di categorie, emerge un risultato fondamentale: la funzione di ripartizione è la stessa qualunque sia l'aspetto sperimentale descritto dalla variabile casuale: discreto, enumerabile, continuo; non solo, rimane lo stesso sia per le rilevazioni empiriche che per i modelli teorici. In altre parole, scrivere F(x) non richiede di specificare quale tipo di variabile sia la X, né se si tratta di un modello oppure di una osservazione concreta. L'incidenza notevole di questo risultato diventerà più rilevante nell'inferenza statistica.

Evento certo ed evento impossibile nel continuo

Nel paragrafo precedente si è discusso di come l'esclusività degli abbinamenti "evento certo/probabilità uno" e "evento impossibile /probabilità zero" si rompa nel passaggio dal discreto finito all'enumerabile. Lo stesso accade nella σ-algebra di Borel.

Esempi:

a) Stoka (19, p. 30): *"Consideriamo S={xεR|0≤ x ≤4} cioè l'universo degli eventi è formato da tutti i numeri compresi tra zero e quattro. Se ora poniamo E={xεS|xè irrazionale} ad E dobbiamo per forza assegnare probabilità uno dato che nell'intervallo vi sono sicuramente √2 e π. Però l'evento E è diverso da S"*

b) La probabilità che la variabile casuale X assuma un valore specifico è:

$$P(X=c)=F(c)-\underset{x\to c-}{Lim}\,F(x)=F(c)-F(x-)$$

dove F(.) è la funzione di ripartizione. Il risultato significa che la funzione di ripartizione ha un salto se e solo se P(X=c)>0 e che, se X è continua allora P(X=c)=0. Inoltre, si ha: P{X∈ (R-(c})}=1 che ripropone la tematica di eventi con probabilità uno diversi dall'evento certo. Si dà inoltre conto della influenza nulla del fatto che la derivata h(.) della F(.) sia discontinua in qualche punto del dominio.

Esercizio_VC92: *a partire dalla funzione di ripartizione:*

$$F(x)=\begin{cases}0 & se\;\; x<-1\\ \dfrac{1+x}{3} & se\;\; -1\le x<0\\ \dfrac{2+x^2}{6} & se\;\; 0\le x<2\\ 1 & se\;\; x\ge 2\end{cases}$$

Quale probabilità sarà assegnata agli intervalli: a)(2], b)[-0.5,3), c)(-1,0]∪(1,2), d)E={x |x²>0.5}

Tipologia della funzione di ripartizione

La funzione di ripartizione di una qualsiasi variabile casuale o è continua su R o è una funzione a gradini oppure si compone di due parti una delle quali è continua e l'altra è a gradini. L'insieme dei punti di discontinuità della F(.) ha la cardinalità dell'enumerabile: se vi sono punti di discontinuità, essi possono contati uno ad uno (Chung, 1974, p. 4). La funzione di ripartizione a gradini è legata a domini discreti e quindi non interessa in questo frangente.

Le funzioni di ripartizione continue possono sempre essere scritte come:

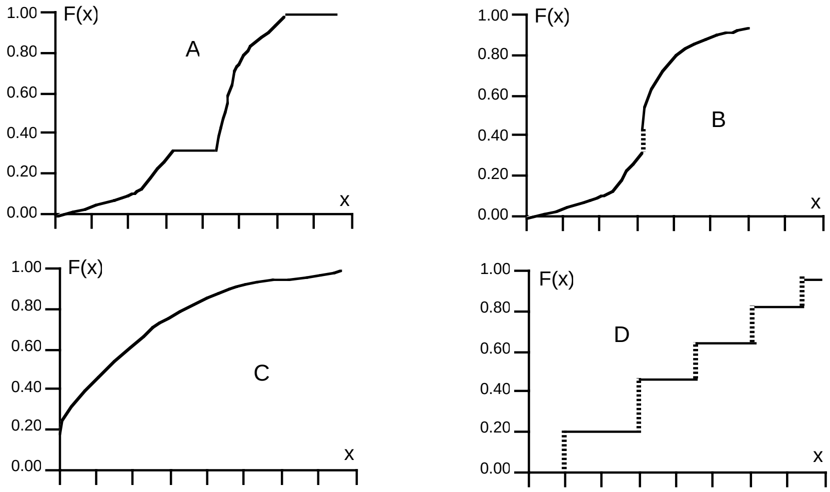
$$F_c(x)=\alpha F_{ac}+(1-\alpha)F_s;\quad 0\le\alpha\le 1$$

La prima componente è una funzione assolutamente continua (ne parleremo tra poco), la seconda è una funzione singolare. Quest'ultimo tipo è particolarmente ostico dato che pur essendo continua e non costante ha derivata nulla in ogni punto. Tali funzioni non sembrano avere grande rilevanza pratica. Feller (1971, p. 141) tuttavia, considera solo un *cliché* il fatto che le funzioni di ripartizione singolari "in pratica" non si incontrino perché diverse procedure statistiche dipendono dalla loro esistenza. per momento decidiamo di escludere le variabili casuali singolari e di discutere le combinazioni:

$$F(x)=\alpha F_{ac}(x)+(1-\alpha)F_d(x);\qquad 0\le\alpha\le 1$$

cioè discrete, miste o continue. Le miste hanno un andamento continuo salvo che in un numero finito o enumerabile di punti in cui presentano un salto senza però far riscontrare i tipici gradini delle variabili casuali discrete.

Esempi:



La A è continua, la D è discreta, B e C sono miste.

Lo studio delle variabili casuali si prospetta in modo differenziato secondo le diverse tipologie della F().

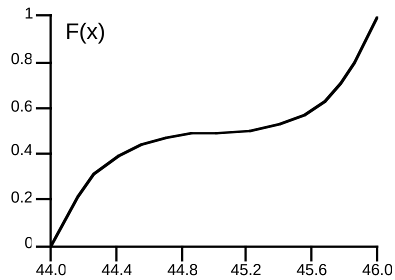
Esempi:
a) Nella funzione di ripartizione

$$F(x)=\begin{cases}0 & se \;\; x<0\\ x & se \;\; 0\le x\le 1\\ 1 & se \;\; x>1\end{cases} \qquad p\left(\frac{1}{4}\le x\le\frac{3}{5}\right)=F\left(\frac{3}{5}\right)-F\left(\frac{1}{4}\right)=\frac{3}{5}-\frac{1}{4}=\frac{7}{20}$$

La probabilità che la X ricada in un intervallo è uguale alla lunghezza dell'intervallo.

b) Il concorrente di una gara a cronometro vuole battere il record della pista (31 chilometri di lunghezza) fissato a 45 minuti. La durata del percorso è una variabile continua: fra 44.0000 e 44.9999 minuti sono presenti infiniti valori (tutti graditi all'aspirante *recordman*) come infiniti se ne trovano tra 45.0001 e 45.0002 (per un battito di ciglia il record è fallito). Poiché in ogni gara c'è almeno un poco di casualità, ciascuna lunghezza può essere associata alla probabilità di raggiungerla. Il modello potrebbe essere:

$$44<x<46; \quad F(x)=0.5\left[1+(x-45)^3\right]$$



I tempi rilevabili ricadono in (44,46). Ciò non è in contrasto con l'assegnazione di probabilità su tutto R dato che agli intervalli esterni a (44,46) hanno probabilità zero.

c) Un gruppo finanziario stabilisce il ricarico sulle fidejussioni in base al mercato rimanendo però ancorato ad un suo intervallo: [1+a,z+a7b con F(x)=Ln(x-a). Gli eventi che interessano ricadono nella classe di Borel che è comune a tutti gli esperimenti rivolti a caratteristiche continue. Tuttavia, la funzione di probabilità di tale esperimento assegna probabilità positiva solo ai sottointervalli di [1+a,e+a].

Esercizio_VC93: *la variabile casuale X ha funzione di ripartizione:*

$$F(x)=\begin{cases}0 \;\; se \;\; x<0\\ \frac{1}{2} \;\; se \;\; x=0\\ \frac{1}{2}+\frac{x}{4} \;\; 0<x<2\\ 1 \;\; se \;\; x\ge 2\end{cases}$$

a) Esprimere la F(.) come somma ponderata di una funzione assolutamente continua e di una funzione a gradini.
b) Quali modifiche occorrerebbe apportare se F(0)=1/4?

La densità di probabilità

Kolmogorov ha concepito i suoi postulati pensando alle probabilità come frequenze relative e interpretando queste ultime come aree sottese ad una curva. La stessa logica può essere adottata per le probabilità assegnandole con il calcolo integrale. Il collegamento più appropriato tra la probabilità di una variabile casuale e il calcolo delle aree è fornito da un concetto che ricalca la densità di frequenza negli istogrammi e cioè la densità di probabilità.

La funzione di densità di probabilità h(x) di una variabile casuale avente funzione di ripartizione F(x) è una funzione non negativa h(.) tale che:

$$F(x) = \int_{-\infty}^{x} h(t)dt, \quad h(x) \geq 0 \quad per \quad ogni \quad x \in R \quad e \quad \int_{-\infty}^{+\infty} h(t)dt = 1$$

Esempio:
Sia S=(0,∞), B la σ-algebra di Borel di S e $E_1, E_2 \in$ B; inoltre, definiamo P(.) come:

$$P(E_1 \in B) = \int_{E_1} e^{-x}dx; \;\; P(E_2 \in B) = \int_{E_2} e^{-x}dx; \;\; P(E_1 \cup E_2) = \int_{E_1 \cup E_2} e^{-x}dx = \int_{E_1} e^{-x}dx; + \int_{E_2} e^{-x}dx$$

E' evidente che P(E)≥0; P(S)=1 e che P(.) è σ-additiva rispondendo in pieno ai postulati di Kolmogorov. Ha anche il non trascurabile merito di riportare una funzione di insieme ad una funzione puntuale. Inoltre, se al posto di e^{-x} si pone una costante positiva si ripropone la probabilità come misura della lunghezza di un intervallo confermando la naturalezza dell'approccio.

Non tutte le funzioni di distribuzioni attivabili per la coppia (R,B) hanno una funzione di densità. La definizione data implica che la funzione di ripartizione sia continua e non tutte le F(.) lo sono ed infatti non lo sono le funzioni singolari (cfr. Jacod e Protter, 1991, p. 74).

Esercizio_VC94: *consideriamo la funzione di distribuzione degenere nel punto a∈R con p(x)=1 se x=a e p(x)=0 altrimenti. Verificare che la sua funzione di ripartizione è discontinua e quindi non dotata di funzione di densità.*

Il problema con la funzione di ripartizione è che se l'integrale è inteso nel senso di Riemann questo non è definito per tutti i tipi di intervalli a meno di limitarsi alla unione e intersezione finita di intervalli semiaperti a sinistra e chiusi a destra. Per trattare il caso più generale si deve ricorrere un altro concetto di integrazione, quello di Lebesgue-Stieltijes, che colma le lacune:

$$F(x) = \int_{-\infty}^{x} dF(t) \quad per \quad ogni \quad x \in R$$

Questa impostazione, per quanto rigorosa ed elegante (ad esempio consente la trattazione unificata delle variabili casuali discrete, miste e continue) crea qualche problema espositivo e non serve nelle nostre applicazioni in cui le miste sono rare e le variabili discrete o continue hanno trattazioni separate formalmente ineccepibili. Nel prosieguo l'integrale sarà perciò inteso nel senso di Riemann a meno di indicazioni in senso contrario.

In generale la funzione di densità non è semplicemente la derivata della funzione di ripartizione sebbene ne sia uno sbocco naturale. Ad esempio, nelle variabili casuali discrete la derivata della F(.) esiste ed è uguale a zero ovunque tranne che nei punti di discontinuità in cui la probabilità è positiva. Nelle variabili casuali miste la derivata esiste su tutto il dominio ad eccezione dei punti di salto. Ne consegue che, in aggiunta alle due descrizioni dell'esperimento casuale già introdotte per universi discreti finiti e ed enumerabili: lo spazio di probabilità e la funzione di ripartizione, in tantissimi casi ne esiste anche una terza, quella della funzione di densità, che può vantare un ruolo alternativo ed autonomo per lo sviluppo degli aspetti probabilistici degli esperimenti.

La densità di probabilità serve per il calcolo delle probabilità, ma non è una probabilità; infatti, può anche succedere che h(x)>1 o addirittura che h(x) tenda all'infinito, pur in presenza di un'area sottesa unitaria.

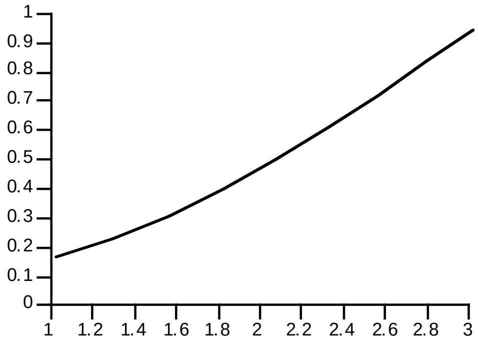
Esempi:
a) Riprendiamo il teorema del valor medio per gli integrali ed applichiamolo alla funzione di ripartizione:

$$P(x \leq X \leq x + dx) = \int_{x}^{x+dx} h(t)dt = h(c)dx \qquad x \leq c \leq dx$$

se "dx" è molto piccolo h(c)dx approssima la probabilità di [c,c+dx], ma h(c) da sola non è una probabilità come l'altezza di un rettangolo non è l'area del rettangolo; peraltro, h(c) può anche essere grande (quindi superiore all'unità) se "dx" è prossimo allo zero.

b) Verifichiamo che la seguente espressione descriva una funzione di densità:

$$h(x)=\begin{cases}\dfrac{3x^2+2x+1}{36} & 1\le x\le 3\\ 0 & altrove\end{cases}$$



L'intervallo (1,3) dove la funzione di densità è positiva è detto supporto della variabile casuale. Al di fuori del supporto la funzione di densità è identicamente nulla. Notiamo subito che la funzione è positiva nel supporto. Controlliamo l'area sottesa:

$$\int_1^3 \frac{3x^2+2x+1}{36}dx=\frac{1}{36}\left[\int_1^3 3x^2dx+\int_1^3 2xdx+\int_1^3 1dx\right]=\frac{1}{36}\left[x^3\Big|_1^3+x^2\Big|_1^3+x\Big|_1^3\right]$$
$$=\frac{1}{36}[26+8+2]=1$$

L'integrale si calcola limitatamente al supporto dato che, al di fuori di questo, il contributo della h(.) all'area è nullo.

c) Effettuiamo lo stesso controllo per la funzione:

$$h(x)=\begin{cases}\text{sen}(x) & 0\le x\le \pi\\ 0 & altrove\end{cases}$$

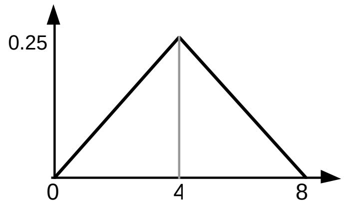
La h(.) è non negativa, ma l'area sottesa non è pari ad uno:

$$\int_0^\pi \text{sen}(x)dx=-\cos(x)\Big|_0^\pi=\cos(0)-\cos(\pi)=1+1=2$$

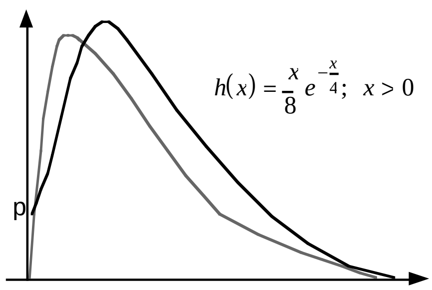
Lo è invece l'area sottesa a h(x)=sen(x)/2 che può essere considerata una funzione di densità (la variabile casuale è da trovarsi).

d) Determinare il valore della costante "b" di modo che la funzione seguente diventi una funzione di densità:

$$h(x)=\begin{cases}bx \quad per & 0\le x<4\\ b(8-x) \quad per & 4\le x\le 8\\ 0 & altrove\end{cases}\Rightarrow \int_0^4 bxdx+\int_4^8 b(8-x)dx=b16\Rightarrow b=\frac{1}{16}$$



e) Sia T il tempo trascorso prima che il computer usato appena acquistato mostri un problema serio. La variabile è di tipo continuo e richiede una densità di probabilità crescente. Un 'espressione adatta potrebbe essere il modello proposto nel 1951 da W. Weibull la cui densità è espressa dalla linea più chiara della figura.



Per rendere più realistico il modello si deve inserire una probabilità positiva P(T=0)=p>0 che il computer non funzioni fin dal primo momento (può succedere di fare un cattivo affare) e quindi la variabile casuale è di tipo misto. La nuova densità è rappresentata con la linea scura che comincia da h(0)=p e l'area sottesa è stata riscalata per renderla pari a (1-p).

Esercizio_VC95: *verificate che la seguente espressione sia una funzione di densità:*

$$h(x)=\begin{cases}\dfrac{e^x}{\left(1+e^x\right)^2} & x\ge 0\\ 0 \quad altrove\end{cases}$$

Esercizio_VC96: *Per quale valore di "c" l'espressione che segue è una funzione di densità?*

$$h(x)=\begin{cases}\dfrac{c}{\sqrt{1-x^2}} & 0\le x<1\\ 0 \quad altrove\end{cases}$$

7.3.2 Variabili casuali assolutamente continue

In un caso particolare l'approccio della funzione di ripartizione e della funzione di densità coincidono: se vige la continuità assoluta. Una variabile casuale (o la sua funzione di ripartizione) è assolutamente continua se:

$$x_1 < x_2 \Rightarrow F(x_2) - F(x_1) = \int_{x_1}^{x_2} h(t)dt$$

cioè se esiste una funzione di densità tale che la probabilità dell'intervallo (x_1,x_2) sia data proprio dall'area ad essa sottesa. Una funzione F(.) assolutamente continua è anche continua; il contrario non è necessariamente vero (cfr. Baldi, 1998, pp. 73-75). Per le variabili casuali assolutamente continue la funzione di densità può essere ottenuta con la relazione:

$$h(x) = \frac{dF(x)}{dx}; \quad -\infty < x < \infty$$

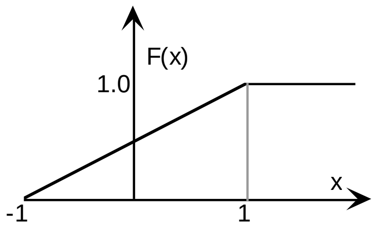
in cui la derivata esiste quasi ovunque. La locuzione "quasi ovunque" non è limitativa in quanto dovuta al fatto che la funzione di ripartizione non cambia se si cambiano o si sopprimono un numero finito di modalità (od anche un loro insieme enumerabile) alle quali sia però associata probabilità zero.

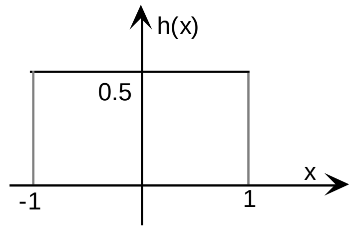
Esempi:

a) La probabilità su R è assegnata tramite la funzione di ripartizione:

$$F(x) = \begin{cases} 0 & se \;\; x \leq -1 \\ \frac{x+1}{2} & se \;\; -1 < x \leq 1 \\ 1 & se \;\; x > 1 \end{cases}$$

$$h(x) = F'(x) = \begin{cases} 0 & se \;\; x < -1, \;\; x > 1 \\ \frac{1}{2} & se \;\; -1 < x < 1 \end{cases};$$

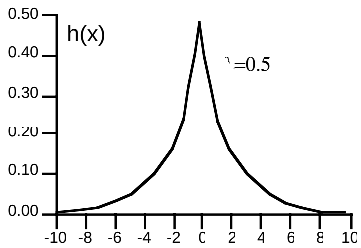




La derivata non è definita per x=-1 e x=1. Convenzionalmente si è posto h(-1)=h(1)=0, ma ogni altra scelta sarebbe stata valida.

b) Consideriamo il modello di Laplace (doppio esponenziale):

$$F(x) = \begin{cases} 0.5e^{\frac{x}{\lambda}} & se \;\; x < 0 \\ 0.5\left(1 - e^{-\frac{x}{\lambda}}\right) & se \;\; x \geq 0 \end{cases} \Rightarrow h(x) = \frac{e^{-\frac{|x|}{\lambda}}}{2\lambda}$$



che potrebbe rappresentare l'errore di approssimazione in una misura. Infatti, si adatta bene alla distribuzione del confronto -in un tempo fissato- con il parametro da misurare λ secondo il logaritmo della distanza da λ. La variabile casuale di Laplace è assolutamente continua.

Esercizio_VC97: *si dimostri che la F(.) qui riportata è continua, ma non assolutamente continua.*

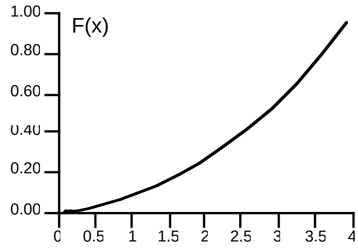
$$F(x) = \begin{cases} 0 & se \;\; x \leq 0 \\ x & se \;\; 0 < x \leq 1 \\ 1 & se \;\; x > 1 \end{cases}$$

La funzione di ripartizione F(.) di una variabile casuale assolutamente continua è l'integrale definito della sua funzione di densità:

$$F(x) = \int_{-\infty}^{x} h(t)dt \;\; con \;\; h(x) \geq 0; \quad \int_{-\infty}^{+\infty} h(x)dx = 1$$

Esempi:

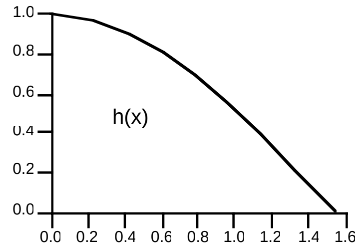
a) Le richieste di spedizioni di una categoria di pacchi si distribuiscono proporzionalmente al peso:



$$h(x)=\begin{cases}\dfrac{x}{8} & per \quad 0\le x\le 4\\ 0 & altrove\end{cases} \quad\Rightarrow\quad F(x)=\begin{cases}0 & se \quad x<0\\ \dfrac{x^2}{16} & per \quad 0\le x\le 4\\ 1 & x>4\end{cases}$$

La funzione di ripartizione risulta quindi quadratica. Una volta nota la funzione di densità si ottiene, se necessario, la funzione di ripartizione con il calcolo integrale.

b) Dalla funzione di ripartizione della variabile casuale X legata ad una misura angolare:



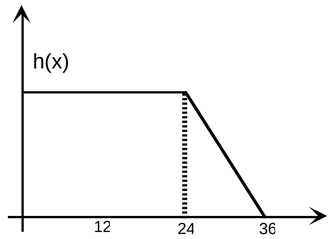
$$F(x)=\begin{cases}0 & se \quad x<0\\ \text{sen}(x) & se \quad 0\le x\le \dfrac{\pi}{2}\\ 1 & se \quad x>\dfrac{\pi}{2}\end{cases} \Rightarrow h(x)=\begin{cases}\cos(x) & se \quad 0\le x\le \dfrac{\pi}{2}\\ 0 & altrove\end{cases}$$

si arriva ad una funzione di densità dello stesso tipo.

c) Per quale valore non negativo del parametro θ la seguente funzione diventa di densità?

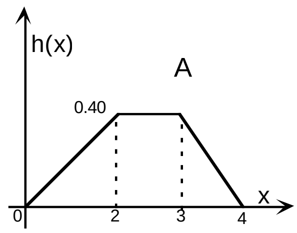
$$f(x)=\frac{\theta x+1}{2};\quad -1\le x\le 1;\quad 1.\, f(x)\ge 0\Rightarrow 0\le\theta\le 1;\ \ 2.\ \int_{-1}^{1} f(x)dx=1\Rightarrow \left.\frac{\theta x^2+2x}{2}\right|_{-1}^{1}=1 \quad per \quad ogni \quad \theta$$

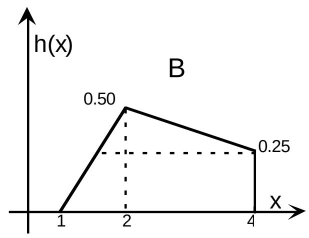
Esercizio_VC98: *la durata di una batteria ha come modello di comportamento la funzione descritta nel grafico a destra.*
Determinare la sua funzione di densità e di ripartizione.



Esercizio_VC99: *per le funzioni indicate nelle due figure effettuare i seguenti calcoli*:

$$a)\ \int_{0}^{3} h(x)dx;\quad b)\int_{-\infty}^{2.5} h(x)dx;\quad c)\int_{1.5}^{\infty} h(x)dx;$$





Verifichiamo che la funzione di densità delle variabili casuali assolutamente continue abbia i requisiti di non negatività ed area sottesa unitaria:

$$h(x)=\frac{dF(x)}{dx}=\underset{x_2\to x}{Lim}\frac{F(x_2)-F(x)}{x_2-x};\ \ con \quad x<x_2$$

Poiché F(.) è monotona non decrescente si ha h(x)≥0. La continuità della h(.) comporta:

$$F(x)-F(x_1)=\int_{x_1}^{x} h(t)dt \quad per \quad x_1\le x \qquad \underset{x\to-\infty}{Lim}\int_{-\infty}^{x} h(t)dt=F(-\infty)=0;\ \ \underset{x\to\infty}{Lim}\int_{-\infty}^{x} h(t)dt=F(\infty)=1$$

perciò, per le variabili assolutamente continue (ma solo per queste) è indifferente che la probabilità sia descritta dalla funzione di ripartizione oppure dalla funzione di densità.

Esempi:

a) La densità h(.) sottostante una variabile casuale assolutamente continua non è unica, anzi ne esistono infinite, ma differiscono tra di loro solo in alcuni punti per i quali però è prescritta probabilità zero. D'altra parte, se per definire h(x) si adopera $F_1(x)=F(x)+a$ dove "a" è una costante, il risultato è lo stesso dato che la derivata della costante è nulla.

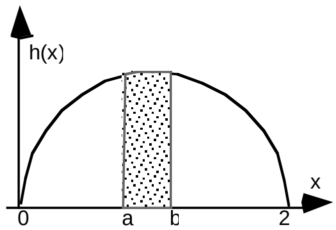
b) Caso particolare del modello di Pareto.

$$h(x)=\begin{cases}\dfrac{3}{2x^4} & se\ \ |x|\geq 1\\ 0 & se\ \ |x|<1\end{cases}\Rightarrow P(-1.5<x<1.5)=\int_1^{1.5}\frac{3}{x^4}dx=0.7037=P(-1.5\leq x\leq 1.5)$$

c) Nella seguente funzione di densità, la probabilità che la variabile casuale ricada nell'intervallo [a,b] è:

$$h(x)=\begin{cases}\dfrac{3}{4}x(2-x) & 0<x<2\\ 0\ \ altrove\end{cases}$$

$$p(a\leq x\leq b)=\int_0^b\frac{3}{4}x(2-x)dx-\int_0^a\frac{3}{4}x(2-x)dx=\frac{\left(a^3-b^3\right)+3\left(b^2-a^2\right)}{4}$$



Sono evidenti le indiscutibili facilitazioni offerte dalla conoscenza del modello di densità e dal calcolo integrale.

Nel calcolare la probabilità da assegnare ad un intervallo continuo è irrilevante considerare gli estremi inclusi o esclusi.

$$\begin{aligned}P\{x\in(a,b]\}&=P\{x\in(a,b)\cup[b,b]\}=P\{x\in(a,b)\}+P\{x\in[b,b]\}\\&=\int_a^b h(x)dx+\int_b^b h(x)dx=F(b)-F(a)+0=F(b)-F(a)\end{aligned}$$

Ragionamenti analoghi valgono per gli altri intervalli per cui trattando con variabili continue non sarà necessario precisare se gli estremi ricadono o no nel supporto.

Esercizio_VC100: *per le funzioni indicate determinare la costante che le rende funzioni di densità:*

$$1)\ \ h(x)=\frac{x^3}{4},\quad 0<x<c;\qquad 2)\ \ h(x)=\frac{c}{x^3};\quad x>2\qquad 3)\ \ h(x)=\begin{cases}c+x; & -c<x<0\\ c-x & 0\leq x<c\end{cases}$$

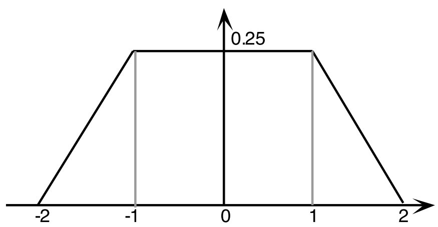
Esercizio_VC101: *dimostrare che è una funzione di densità l'espressione:*

$$\sum_{i=1}^{k}\alpha_i h_i(x)\ \ con\ \ \alpha_i\geq 0,\quad \sum_{i=1}^{k}\alpha_i=1$$

se sono densità le singole $\{h_i\}$. Che conseguenza ha l'internalità di Cauchy?

Esercizio_VC102: *per il modello trapezoidale*

$$h(x)=\begin{cases}0\ \ se\ \ x<-2\\ \dfrac{1}{2}+\dfrac{x}{4}\ \ se\ -2\leq x<-1\\ \dfrac{1}{4}\ \ se\ -1\leq x<1\\ \dfrac{1}{2}-\dfrac{x}{4}\ \ se\ \ 1\leq x<2\\ 0\ \ se\ \ x\geq 2\end{cases}$$



Calcolare: a) p(X<-1); b) p(-0.5, 2.5); c) p(|x+0.5|≤1)

Modelli per la densità

Esistono diversi modelli di funzione di densità tra i quali il noto sistema di curve di K. Pearson di cui si riportano alcuni casi ed il sistema di Johnson.

Denominazione	*Densità*	*Dominio*
Uniforme	$h(x)=\frac{1}{b-a}$	$a<x<b$
Burr – Mielke	$h(x)=abcx^{b(c-1)}\left[\frac{x^b}{1+x^b}\right]^{1+c}$	$x>0;\ \ a,b,c>0$
Pareto	$h(x)=\frac{b}{a}\left(\frac{a}{x}\right)^{b+1}$	$x>a;\ \ a,b>0$
Weibull	$h(x)=\frac{a}{b}\left(\frac{x}{b}\right)^{a-1}e^{-\left(\frac{x}{b}\right)^a}$	$x>0;\ \ a,b>0$
Normale	$h(x)=\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$	$\sigma>0$
Log – Normale	$h(x)=\frac{e^{-\frac{[Ln(x)-\mu]^2}{2\sigma^2}}}{x\sigma\sqrt{2\pi}}$	$\sigma>0$
Beta	$h(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(a)}x^{a-1}(1-x)^{b-1}$	$0<x<1;\ \ a,b>0$ $\Gamma(\alpha)=\int_0^\infty t^{\alpha-1}e^{-t}dt$
Gamma	$h(x)=\frac{\left(\frac{x}{b}\right)^{a-1}e^{-\frac{x}{b}}}{\Gamma(a)}$	$x>0;\ \ a,b>0$
Cauchy	$h(x)=\frac{\sigma}{\pi\left[1+\sigma^2(x-a)^2\right]}$	$\sigma>0$
Gumbel	$h(x)=\frac{e^{-\left[\left(\frac{x-a}{\sigma}\right)+e^{-\left(\frac{x-a}{\sigma}\right)}\right]}}{\sigma}$	$\sigma>0$
Laplace	$h(x)=\frac{e^{-\left\lvert\frac{x-a}{\sigma}\right\rvert}}{2\sigma}$	$\sigma>0$
Normale inv.	$h(x)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)\left(\frac{\mu}{x}\right)^{1.5}e^{-\frac{1}{2}\left[\left(\frac{\mu}{x}\right)\left(\frac{x-\mu}{\sigma}\right)^2\right]}$	$x,\mu,\sigma>0$

L' interesse per questi modelli è duplice: da un lato per il loro carattere descrittivo che raggiunge l'obiettivo con l'introduzione della comoda scorciatoia dell'interpolazione del poligono delle frequenze. Ciò consente ragionamenti più approfonditi e generalizzabili sul fenomeno in quanto espressi nei loro elementi essenziali, senza dettagli inutili; inoltre, i calcoli sono più rapidi, meglio approssimati e più resistenti agli errori accidentali (Costanzo 1968, cap. 9; Leti, 1983, cap.8; Girone e Salvemini 1991, cap. 9; Di Ciaccio e Borra, 1996, cap. 7; Vajani 1997, cap.11).

Esempio:

Leti (1983, p. 555) osserva: "*... Talvolta nello studio dei collettivi reali si ottengono distribuzioni i cui grafici presentano andamenti così regolari da ricordare quelli di funzioni matematiche ben note. Secondo alcuni ciò significa che il fenomeno segue una legge matematica, espressa da una funzione, e che gli scarti da questa, che vengono osservati, sono dovuti soltanto a cause accidentali. Ciò può essere talvolta vero per i fenomeni studiati dalle scienze sperimentali, ma certamente non è vero per i fenomeni biologici, sociali ed economici, che sono il principale oggetto della statistica.]...[E' erroneo quindi cercare di individuare la legge matematica che regoli un fenomeno oggetto della statistica: in realtà bisogna soltanto determinare un modello esplicativo che si adatti soddisfacentemente alla distribuzione osservata e che possa dare informazioni aggiuntive rispetto a quelle fornite dalle sintesi della distribuzione empirica.*"

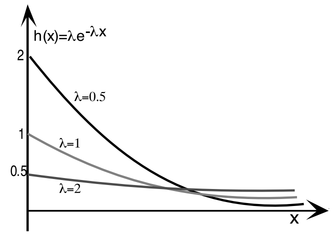
Questa linea di studi, tuttavia, ha perduto gran parte dell'importanza che le era stata attribuita in passato perché:
1) La simulazione con il computer è ora in grado di estendere artificialmente, ma con grande realismo, le rilevazioni permettendo di definire accuratamente le distribuzioni empiriche e di calcolarne le caratteristiche più rilevanti senza il ricorso a modelli fittizi.

2) I modelli proposti mancano di una spiegazione convincente che porti a scegliere tra curve simili dal punto di vista del mero compito dell'interpolazione, ma diverse nella base teorica. In aggiunta, accade spesso che i poligoni di frequenza riscontrati su di un fenomeno siano fortemente somiglianti alla curva di un modello teorico anche se le condizioni di base di tale modello sono palesemente violate.

Esempi:

a) Molti esperimenti sono centrati sulle durate. Se una sequenza di eventi Y si manifesta nel tempo secondo i modello di Poisson, quanto bisogna aspettare perché si verifichi il primo? Fissiamo a λ il numero medio di "eventi" e definiamo la variabile casuale X = "Tempo antecedente la prima manifestazione" (coincide con il tempo intercorrente tra due manifestazioni consecutive) per cui X è una variabile continua con valori tra zero ed infinito. L'evento X>x cioè il primo evento si verifica dopo il tempo "x" si realizza se nel lasso temporale [0,x] il modello di Poisson non ha prodotto successi, cioè Y=0. Quindi:

$$p(X > x) = p(Y = 0) = \frac{(\lambda x)^0 e^{-\lambda x}}{0!} = e^{-\lambda x} \Rightarrow p(X \le x) = F(x) = 1 - e^{-\lambda x}$$

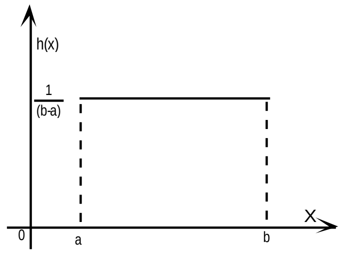


Questo modello è noto come variabile casuale esponenziale ed ha numerose applicazioni nei tempi di attesa in coda, nella durata delle conversazioni telefoniche, nel decadimento delle particelle radioattive, nel tempo di funzionamento prima di una disfunzione per la componente in un sistema, etc. La funzione di densità si ricava con il calcolo differenziale:

$$F'(x) = h(x) = \begin{cases} \lambda e^{-\lambda x} & per \quad x > 0 \\ 0 & altrove \end{cases}; \quad \lambda > 0$$

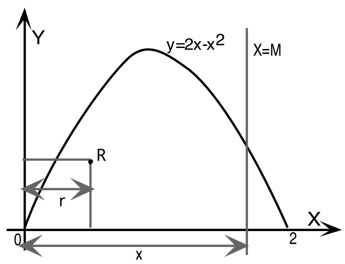
b) La variabile casuale uniforme è un modello elementare che, oltre a rappresentare diverse situazioni pratiche (è richiamato spesso quando non si hanno informazioni sulla densità di probabilità più appropriata in una situazione) serve da base per tante altre variabili casuali. Il modello si fonda su due presupposti: campo di variazione limitato (a,b) e densità di probabilità costante.

$$h(x) = \frac{1}{b-a}, per \quad a \le x \le b; \quad F(x) = \frac{x-a}{b-a}$$



L'aspetto che assume la funzione di densità spiega anche il nome di modello rettangolare.

Esercizio_VC103: *un punto R è scelto a caso nella regione delimitata dalla parabola riportata in figura.*



Determinate il modello della variabile casuale X="distanza dall'asse delle ordinate del punto".

Costanzo (1968, p. 267) riporta una considerazione di B. Barberi: "*Se si tentasse il bilancio di oltre mezzo secolo di studi e di applicazioni non sarebbe da escludere un saldo passivo tra i contributi e gli svantaggi recati al progresso della matematica statistica delle celebri curve di frequenza di K. Pearson*". Anche Vajani (1997, p. 279) trova superato lo studio delle curve di densità. Di parere opposto sono Kotz e Johnson (1973) per i quali lo studio dei sistemi di distribuzione ha alle spalle una lunga ed onorata carriera, sia pure tra abbandoni e ritorni di fiamma e che desta sempre grande interesse.

Un altro filone di studi, meno descrittivo e più investigativo secondo la distinzione di Girone e Salvemini (1991, p. 248) vede nei modelli uno strumento utile per lo studio simulato del comportamento delle statistiche calcolate quando la rilevazione è ripetuta virtualmente con il computer (Tarsitano, 1986).

Esempi:
a) Se il logaritmo della frequenza relativa di modalità che distano dal centro una distanza "d" è inferiore alla frequenza relativa al centro di una quantità proporzionale a "d^2", allora il modello adatto per rappresentare il poligono delle frequenze è quello Normale.

b) Frosini (1995, P. 215) osserva che nelle situazioni in cui una massa omogenea subisce rotture o spezzamenti in dipendenza dell'effetto congiunto di molti fattori, il poligono di frequenza è ben approssimato dal modello lognormale.

c) La densità di Cauchy è raccomandata per descrivere variabili legate a misurazioni angolari quali ad esempio la direzione del vento (Essenwanger, 1976). Le sue peculiarità sono la simmetria rispetto alla mediana e le code molto spesse.

e) La densità Burr-Mielke, caratterizzata da forte asimmetria a sinistra e da una coda a destra particolarmente spessa, è applicata in campi quali lo studio della distribuzione dei redditi, delle precipitazioni atmosferiche, dei rischi assicurativi.

Secondo questo approccio, il fatto che la curva di densità abbia un andamento simile al poligono di frequenza non è sufficiente per giustificare l'uso di un modello di variabile casuale (si pensi al problema della non identificabilità), ma la sua genesi deve essere interpretabile come traccia ideale del comportamento del fenomeno.

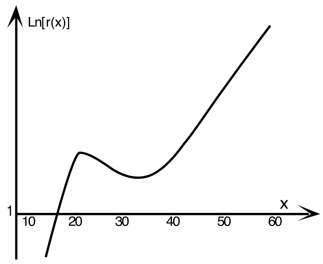
Esempi:
a) Il modello logistico ha funzione di ripartizione:

$$F(x) = \frac{1}{1+e^{-(ax+b)}}; \quad x > 0, \quad a > 0$$

Dice Feller (1971, pp. 52-53): "*... Una cospicua mole di pubblicazioni ha tentato di stabilire la legge di crescita logistica applicata praticamente a tutti i processi di sviluppo: popolazioni di batteri, imprese, ferrovie. Ci sono però due problemi: 1) La stessa legge si applica ugualmente bene al peso ed all'altezza di piante ed animali sebbene sia noto che le due variabili non possono essere soggette alla stessa distribuzione. 2) Se è vero che il modello logistico si adatta bene ai dati osservati, è vero anche che esistono altri modelli che hanno un adattamento altrettanto soddisfacente se non migliore*" . Il modello logistico è perciò in competizione con altri e senza ulteriori indicazioni ci si trova di fronte ad un problema non identificabile.

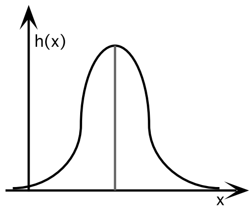
b) Riprendiamo da Berry e Lindgren (1990, p. 292) un esempio su di una delle applicazioni classiche della Statistica: la costruzione di tavole attuariali ed il problema dei premi di assicurazione sulla vita. Sia X l'età di una persona all'atto della morte cioè una variabile continua e positiva (sono esclusi i nati morti). Alle compagnie interessa un indicatore dell'evento: muore entro l'anno in età X dato che ha già raggiunto l'età X cioè il rischio di morte che è ottenuto dividendo la densità di frequenza h(X) dei morti in età X per la G(X) che esprime i sopravviventi in età maggiore o uguale ad X.

$$r(X) = \frac{h(X)}{G(X)}$$



Nella figura è riportato il rischio di morte per età (in logaritmi). Dopo i 40 anni si nota un andamento lineare (quindi esponenziale in scala naturale) che aumenta con l'età. Non sarebbe difficile trovare un modello che rappresenti tale evoluzione. La ricerca è però complicata dal picco -anomalo- che si riscontra intorno ai 20 anni e che coincide in modo eloquente con l'età in cui si comincia a guidare.

Esercizio_VC104: *diversi fenomeni quali le caratteristiche fisiche di piante o animali e i tratti somatici delle persone producono un poligono di frequenza di tipo campanulare.*



Effettuate una prova, ad esempio rilevando le altezze di una trentina di persone dello stesso sesso e costruitene il poligono di frequenza. Si ottiene un andamento campanulare?

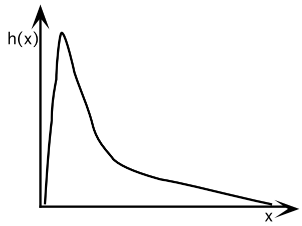
Altre funzioni legate a quella di densità

Richiamiamo alcune delle funzioni già introdotte per le distribuzioni di frequenza che anche nell'ambito delle variabili casuali possono avere un ruolo importante. La funzione di sopravvivenza lega la variabile casuale alle probabilità retrocumulate:

Esempi:

a) Modello fortemente unimodale:

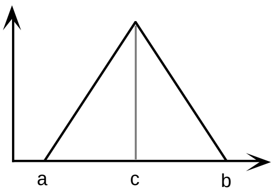
$$h(x)=\frac{6x}{(1+x)^4},\quad x>0$$
$$F(x)=1-\frac{3}{(1+x)^2}+\frac{2}{(1+x)^3}$$
$$G(x)=\frac{3}{(1+x)^2}-\frac{2}{(1+x)^3}$$



I modelli di questo tipo sono utili per presentare i concetti dal punto di vista didattico, ma poco pratici perché “spogli” cioè privi di parametri che diano loro la flessibilità per modellare situazioni reali.

b) Modello triangolare o di Simpson:

$$h(x)=\begin{cases}\dfrac{2(x-a)}{(b-a)(c-a)} & a<x\le c\\ \dfrac{2(b-x)}{(b-a)(c-a)} & c<x\le b\end{cases}$$



è proposto spesso per esprimere la somma di due variabili aventi ciascuna densità uniforme: la prima in (a,c) e la seconda in (c,b). Le funzioni di ripartizione e di sopravvivenza sono:

$$F(x)=\begin{cases}\dfrac{(x-a)^2}{(b-a)(c-a)} & a<x\le c\\ 1-\dfrac{(b-x)^2}{(b-a)(c-a)} & c<x\le b\end{cases};\quad G(x)=\begin{cases}1-\dfrac{(x-a)^2}{(b-a)(c-a)} & a<x\le c\\ \dfrac{(b-x)^2}{(b-a)(c-a)} & c<x\le b\end{cases}$$

Esercizio_VC105: *determinare la funzione di sopravvivenza per il modello paretiano del 3° tipo*

$$h(x)=(a+bx)\frac{e^{-bx}}{x^{a+1}},\quad x>c\quad dove\quad e^{-bc}=c^a$$

Esercizio_VC106: *il modello di Rayleigh, invocato per rappresentare misure di grandezze fisiche, ha funzione di densità:*

$$h(x)=\begin{cases}a^2xe^{-a^2\frac{x^2}{2}} & per\quad x>0\\ 0 & altrove\end{cases}$$

Individuare la funzione di ripartizione e di sopravvivenza.

La funzione di graduazione di una variabile casuale continua è ottenuta considerando:

$$X_p=F^{-1}(p)=Min\{x|F(x)=p\}$$

dove X_p è il quantile di ordine “p”.

Esempi:

a) Il modello di Benini ha funzione di ripartizione e di sopravvivenza:

$$F(x)=1-ae^{-b[Ln(x)]^3};\quad X>0;\quad G(x)=ae^{-b[Ln(x)]^3};\quad X_p=e^{\left[-\frac{Ln\left(\frac{1-p}{a}\right)}{b}\right]^{\frac{1}{3}}}$$

b) La funzione di graduazione di due modelli classici quali il modello esponenziale e quello uniforme sono:

$$F(x)=1-e^{\lambda x}\ \Rightarrow\ X_p=\frac{-Ln(1-p)}{\lambda};\quad F(x)=\frac{x-a}{b-a}\ \Rightarrow\ X_p=a+p(b-a);$$

c) Un punto è distribuito uniformemente sulla circonferenza di equazione x^2-2Rx +y^2=0. La sua funzione di densità è:

$$h(x)=\frac{1}{\pi\sqrt{2Rx-x^2}},\quad 0\le x\le 2R;\quad F(x)=\frac{arcos\left(\frac{R-x}{R}\right)}{\pi}\Rightarrow X_p=\frac{R}{1+cos(\pi p)}$$

d) Modello di Burr-Mielke (ad un solo parametro).

$$h(x)=\frac{1}{\left(1+\frac{x}{\sigma}\right)^{\sigma+1}},\quad x>0;\quad F(x)=1-\frac{1}{\left(1+\frac{x}{\sigma}\right)^{\sigma}}\Rightarrow X_p=\sigma\left[\left(\frac{1}{1-p}\right)^{\frac{1}{\sigma}}-1\right];\quad X_{1-p}=\sigma\left(p^{-\frac{1}{\sigma}}-1\right)$$

e) Modello di Cauchy. Un punto, partito dall'origine del piano cartesiano viaggia in linea retta nel semipiano in cui x>0. Sia U l'angolo formato tra il verso positivo dell'asse delle ascisse e la linea su cui si muove il punto e si supponga che l'angolo sia una variabile casuale uniforme nell'intervallo ($-\pi/2,\pi/2$). Se X è l'ordinata del punto in cui la verticale x=1 incrocia la linea di movimento del punto allora Y ha la funzione di densità di Cauchy:

$$h(x)=\frac{1}{\pi\left(1+x^2\right)},\quad -\infty<x<\infty;\ F(x)=\frac{1}{\pi}\left[\frac{\pi}{2}+arctan(x)\right];\quad X_p=tan[\pi(p-0.5)]$$

Esercizio_VC107: *si consideri la seguente funzione di ripartizione:*

$$F(x)=\begin{cases}0 & se\ \ x<1\\ \dfrac{x-1}{3} & se\ \ 1\le x\le 4\\ 1 & se\ \ x>4\end{cases}$$

Determinare: a) La funzione di densità; b) La funzione di ripartizione complementare;c) La funzione di graduazione diretta e complementare.

Esercizio_VC108: *una delle versioni della variabile casuale logistica è:*

$$p(X\le x)=F(x)=\begin{cases}\dfrac{1}{1+e^{-\frac{x-\mu}{\sigma}}} & per\ \ x>0\\ 0 & altrove\end{cases}$$

*Determinare: a) La funzione di densità per μ=0 e σ=1.b)La funzione di sopravvivenza e di graduazion*e.

Affidabilità e funzione di rischio

L'affidabilità di una attrezzatura, di una componente, di un organo, di un materiale, di un apparato, di una persona ha come indicatore naturale la frequenza con la quale si succedono nel tempo le disfunzioni o gli errori o le inadempienze che ne rendono inaccettabili le prestazioni. I modelli di variabile casuale assolutamente continua tentano di descrivere il funzionamento di un sistema partendo dalla osservazione dei tempi precedenti la sua *defaillance* e tenendo conto sia della asimmetria che delle rarefazioni dei dati empirici nelle modalità più elevate (le ampiezze campionarie su cui si opera in queste applicazioni sono in genere esigue). La scelta del modello è facilitata dalla funzione di rischio *(hazard ratio)* o tasso istantaneo di mortalità od anche tasso di guasto che è definita come il negativo della derivata della funzione di log-sopravvivenza :

$$r(x)=-\frac{d}{dx}Log[1-F(x)]=\frac{h(x)}{1-F(x)}=\ \lim_{\Delta t\to 0}\frac{P(t\le x\le t+\Delta t)}{\Delta t}$$

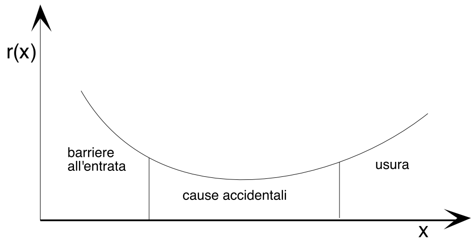
se il limite esiste finito. La r(x) esprime la probabilità di errore al tempo x+Δx dopo che l'elemento sotto esame sia rimasto operativo fino al tempo "x". L'azione della sorte si esplica provocando rotture rapide in sistemi con tassi di rischio molto basso e con durate inusitate per apparati sulla cui affidabilità pochi avrebbero scommesso. L'andamento della funzione di rischio è governato dalla derivata:

$$\frac{dLn[r(x)]}{dLn[h(x)]}=[1+r(x)]$$

e quindi il tasso di incremento relativo (o elasticità) della funzione di rischio rispetto alla funzione di densità rimane costante, aumenta o diminuisce in ragione del comportamento della funzione di rischio stessa.

Esempi:

a) Mann ed al. (1974, p.143) propongono una funzione di rischio a forma di U.

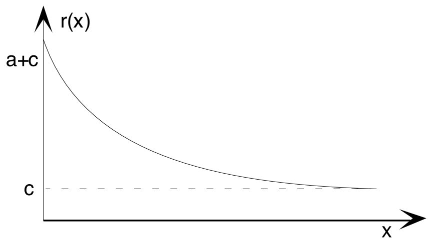


Nel primo tratto il funzionamento corre pericoli soprattutto a causa delle barriere all'entrata: cattive condizioni iniziali, difetti all'origine, eventi eccezionali; nella fase centrale dominano i fattori accidentali e nella terza fase le cause più verosimili sono l'usura e il deterioramento delle condizioni di operatività. L'ottenimento di una funzione di rischio di questa forma ha dato luogo a diversi studi. Ad esempio, Wang (2000) propone di utilizzare il modello di Burr:

$$h(x)=\frac{abx^{b-1}}{1+x^b} \Rightarrow r(x)=\frac{1}{\left(1+x^b\right)^c}$$

b) Dobson (1985) descrive la seguente funzione di rischio per la sopravvivenza di una specie di uccelli:

$$r(x)=ae^{-bx}+c; \quad x\geq 0; \; a,b,c>0$$



detto *competing risk model*. Nell'istante x=0 il tasso è al suo massimo r(0)=a+c; al crescere dell'individuo il tasso di mortalità tende a diminuire per stabilizzarsi al livello r(x)=c che è configurazione per l'età adulta.

Se è nota la funzione di rischio è possibile ricavare da questa la funzione di densità e di ripartizione. Infatti:

$$\int_0^x r(t)dt=-Ln\left[1-F(x)\right] \Rightarrow F(x)=1-e^{-\int_0^x r(t)dt}, \quad h(x)=r(x)e^{-\int_0^x r(t)dt}$$

Esempi:

a) Esponenziale: $r(x)=\lambda \Rightarrow h(x)=\lambda e^{-\lambda x}$; b) Weibull: $r(x)=x^\beta \Rightarrow h(x)=\beta x^{\beta-1}e^{-x^\beta}$;

c) Gompertz: $r(x)=ab^x \Rightarrow h(x)=ab^{x-ab^x}$

b) Iuculano (1996, p. 130) osserva che nella prassi si introduce un tempo t_0 al di sotto del quale non si verifica, con assoluta certezza, alcun guasto o avaria. In tal caso la funzione di ripartizione e la densità di probabilità risultano nulle per $t<t_0$ e corrispondentemente la funzione di rischio è uguale ad uno:

$$F(x)=0;\; h(x)=0, \qquad r(x)=0 \;\; per \;\; t<t_0$$

$$F(x)=1-e^{-\int_{t_0}^x r(t)dt};\; h(x)=r(x)e^{-\int_{t_0}^x r(t)dt}, \quad r(x)>0 \;\; per \;\; t<t_0$$

Esercizio_VC109: *determinare la funzione di rischio sottostante il modello di Gumbel*

$$F(x)=1-e^{e^{-\left(\frac{x-\mu}{\sigma}\right)}}$$

Esercizio_VC110: *Determinare il modello di funzione di densità associata alla funzione di rischio proposta da Dobson. Come potrebbe essere interpretata se applicata a delle aziende?*

7.3.3 Trasformazione di variabili casuali continue

La trasformazione Y=g(X) è un sistema in cui la variabile casuale X agisce da input trasformato in output Y dal meccanismo g(.); tale esigenza si pone ad esempio allorché si debbano modificare le unità di misura o di conto di una rilevazione oppure quando si misura il diametro di una sfera X di un materiale di densità g(.) e si opera con il volume $Y=(g/6)\pi X^3$; la trasformazione logaritmica è un classico nelle serie storiche e l'estrazione di radice o l'elevamento al cubo sono spesso richiamate per modificare la forma della curva di densità. Noto il comportamento della variabile casuale X (che parte dallo spazio di probabilità e arriva sull'asse reale) e definita la trasformazione g(.): R→R, la Y è del tutto specificata. Resta da stabilire se è una variabile casuale , e, in caso affermativo, quale ne siano la funzione di ripartizione e di densità.

Il primo problema è legato alla natura della g(.) ed in particolare al fatto che {g(X)≤y} sia o no Borel-misurabile cioè ricadente nella σ-algebra di Borel. A questo fine è necessario che la trasformazione g(.) sia monotona crescente o decrescente cioè la g(.) non deve modificare la sua linea di tendenza e non deve rimanere costante nell'intervallo considerato. La stretta monotonia implica che l'inversa di g(.) esiste ed è unica: $X=g^{-1}(Y)$. Se g(.) è strettamente crescente l'evento {g(X)≤y} è equivalente all'evento $\{X\leq g^{-1}(y)\}$

$$F_y = p(Y \leq y) = p[g(X) \leq y] = p\left[X \leq g^{-1}(y)\right] = F_x\left[g^{-1}(y)\right]$$

Se g(.) fosse strettamente decrescente {g(X)≤y} corrisponderebbe a $\{X>g^{-1}(y)\}$ per cui:

$$F_y = p(Y \leq y) = p[g(X) \leq y] = p\left[X > g^{-1}(y)\right] = 1 - F_x\left[g^{-1}(y)\right]$$

Pertanto, la funzione di ripartizione della trasformata coincide con quella della variabile originaria, valutata però in un intervallo che termina o inizia nel punto immagine $g^{-1}(y)$.

Esempi:

a) Distribuzione uniforme (0,1):

$$F_x(x) = \begin{cases} 0 & se \;\; x \leq 0 \\ x & se \;\; 0 < x < 1; \\ 1 & se \;\; x \geq 1 \end{cases} \qquad Y = 2x + 4 \Rightarrow \left[\begin{array}{l} x = 0 \rightarrow y = 4 \\ x = 1 \rightarrow y = 6 \end{array}\right. \Rightarrow F_x(x) = \begin{cases} 0 & se \;\; x \leq 4 \\ \dfrac{y-4}{2} & se \;\; 4 < x < 6 \\ 1 & se \geq 6 \end{cases}$$

b) Se X ha densità uniforme sull'intervallo [a,b] qual'è la densità di Y=1/X?

$$h(x) = \frac{1}{b-a}; \;\; a \leq x \leq b \Rightarrow F_X(x) = \begin{cases} \dfrac{x-a}{b-a} & se \;\; a \leq x \leq b \\ 0 & altrove \end{cases}$$

$$F_Y(y) = p(Y \leq y) = p\left(\frac{1}{X} \leq y\right) = p\left(X \geq \frac{1}{y}\right) = 1 - F_X\left(\frac{1}{y}\right); \quad \Rightarrow \quad F_Y(y) = \begin{cases} 1 & se \;\; \dfrac{1}{y} \leq a \\ 1 - \dfrac{\frac{1}{y} - a}{b-a} & se \;\; a \leq \dfrac{1}{y} \leq b \\ 0 & se \;\; \dfrac{1}{y} > b \end{cases}$$

c) Trasformazione invertibile a tratti:

$$F_X(x) = \begin{cases} 0 & se \;\; x \leq -1 \\ \dfrac{x+1}{2} & se \;\; -1 < x < 1; \\ 1 \;\; \geq & se \;\; y \geq 1 \end{cases} \quad \Rightarrow \quad y = x^2; \;\; 0 \leq y \leq 1$$

Come è già successo per la trasformazione delle variabili casuali discrete occorre tenere conto che gli eventi di interesse possono essere più di uno per ogni "x" in input.

$$F_Y(y) = P(Y \leq y) = P\left(X^2 \leq y\right) = P\left(-\sqrt{y} \leq X \leq \sqrt{y}\right) = F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right) = \frac{1+\sqrt{y}}{2} - \frac{1-\sqrt{y}}{2} = \sqrt{y}$$

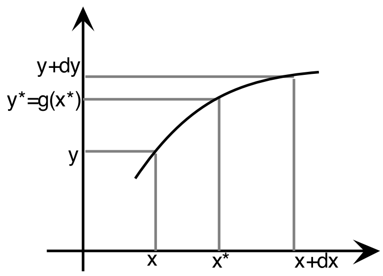
Ne consegue:

$$F_Y(y) = \begin{cases} 0 & se \;\; y \leq 0 \\ \sqrt{y} & se \;\; 0 < y < 1; \\ 1 & y \geq 1 \end{cases}$$

Esercizio_VC111:
a) Se X è uniforme sull'intervallo (-c,c) determinare la funzione di ripartizione per $Y=1/X^2$;
b) Se X è Normale con media $\mu=0$ e varianza $\sigma^2=0.5$ qual'è la funzione di ripartizione di $Y=|X|$?

Se è nota la funzione di densità si può prima ottenere la funzione di ripartizione e poi da questa ricavare la F(.) della trasformata ed eventualmente - per differenziazione- la funzione di densità. Esiste però un metodo più diretto. Sia X una variabile casuale continua con densità h(.) e sia y=g(x) una funzione strettamente monotona, differenziabile quasi ovunque e tale da non essere mai nulla. Per il momento ragioniamo con g'(x)>0. Dato che y*=g(x*) si ha y≤y*≤y+dy se e solo se x≤x*≤x+dx:

$$\int_y^{y+dy} h_y(t)dt = \int_x^{x+dy} h_x(t)dt \Rightarrow h_y(y)dy \cong h_x(x)dx$$



Per valori piccoli di "dx" e "dy" sarà valida, approssimativamente, l'ultima relazione. Inoltre, al tendere di "dx" a zero, il rapporto dy/dx tenderà alla derivata g'(x) cosicché:

$$h_y(y) = \frac{1}{g'(x)} h_x(x)\Big|_{x=g^{-1}y} \qquad g(-\infty) < y < g(\infty)$$

La stretta monotonicità di g(.) non necessariamente implica che g'(x)>0 per ogni "x" (ad esempio, la funzione $g(x)=x^3$ è strettamente monotona, ma ha derivata zero nell'origine) e quindi il risultato rimane valido solo per quei valori che non annullino la derivata g'(x). Se il numero di tali punti forma un insieme finito o enumerabilmente infinito l'eccezione è irrilevante dato che la definizione della funzione di densità consente di dare probabilità nulla a tali valori.

Per superare la limitazione della derivata monotona crescente ed ampliare il ragionamento a funzioni differenziabili decrescenti basterà ricorrere al valore assoluto:

$$h_y(y) = \frac{1}{|g'(x)|} h_x(x)\Big|_{x=g^{-1}(y)} = h_x\left[g^{-1}(y)\right]\left|\frac{dg^{-1}(x)}{dx}\right|$$

determinando l'inversa $x=g^{-1}(y)$ per poi sostituirla nell'espressione della funzione di densità.

Esempio:
Modello esponenziale e trasformazione quadratica

$$h_x(x) = e^{-x}, \quad x>0; \quad y=\sqrt{x} \Rightarrow x=y^2; \ \frac{dy}{dx}=\frac{1}{2\sqrt{x}}; \quad h_y(y)=\frac{e^{-y^2}}{\left|\frac{1}{2\sqrt{y^2}}\right|}=\frac{e^{-y^2}}{2y} \quad y>0$$

La forma del modello cambia notevolmente anche per una trasformazione piuttosto semplice come la radice quadrata.

$$b. \ h_x(x)=4x^3, \quad 0<x<1 \ \ y=\frac{1}{x} \Rightarrow x=\frac{1}{y}; \ \left|\frac{dy}{dx}\right|=\frac{1}{x^2}; \quad h_y(y)=\frac{\frac{4}{y^3}}{\frac{1}{y^{-2}}}=\frac{4}{y^5} \quad y>1$$

Se la densità della variabile originaria si annulla al di fuori dell'intervallo (a,b) le condizioni di differenziabilità e monotonia possono valere solo all'interno di (a,b) considerando poi come estremi della variabile trasformata: min{g(a),g(b)} e max{g(a),g(b)}.

Esercizio_VC112: *per le funzioni di densità e relative trasformazioni:*

$$1.\ h(x)=\frac{1}{b-a};\quad a<x<b;\quad W=Ln(x);\qquad 2.\ h(y)=6y(1-y)\quad 0<y<1;\quad Z=e^{-y};$$

$$3.\ h(u)=\frac{3(u-1)^2}{8};\ -1\le u\le 1;\ V=u^2$$

determinare la densità delle variabili trasformate;

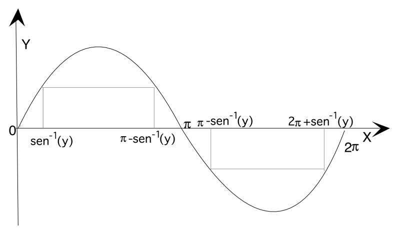
Se la trasformazione è invertibile a tratti il metodo appena delineato può essere adoperato anche per funzioni non monotone. Basterà infatti applicare separatamente la procedura ai singoli tratti:

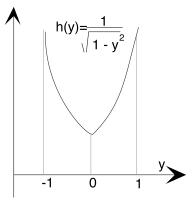
$$h_y(y)=\sum_{i=1}^{m} h_x\left[g_i^{-1'}(y)\right]\left|g_i^{-1'}(y)\right| \quad con \quad h_x\left[g_i^{-1'}(y)\right]\left|g_i^{-1'}(y)\right|=0 \quad se \quad y\notin J_i$$

in cui J_i è l'intervallo in cui incide il tratto i-esimo della trasformazione. E' però richiesto che l'equazione Y=g(X) abbia un numero di soluzioni finito o enumerabilmente infinito nel segmento considerato. Non è necessario che g'(x) esista ovunque; la derivata può, ad esempio, non essere definita nei punti terminali.

Esempi:

a) Sia X la misura dell'angolo formato dalla freccia ruotante su di un disco con la posizione d'avvio. Se foggia e materiale -insieme al tipo di impulso imposto- giustificano l'ipotesi di simmetria tra i vari settori del disco, il modello di densità che può descrivere l'angolo è la distribuzione uniforme: : h(x)=1/(2π) per 0≤x≤2π. Supponiamo ora di essere interessati al seno formato dall'angolo: Y=sen(X)



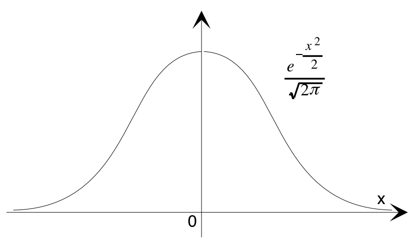


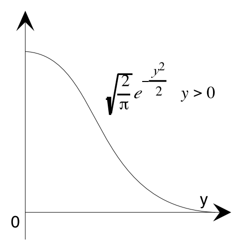
$$h_y(y)=h\left[arcosen(y)\right]\left|\frac{d\ \ arcosen(y)}{dy}\right|+h\left[\pi - arcosen(y)\right]\left|\frac{d\ \left[\pi - arcosen(y)\right]}{dy}\right|=\frac{1}{\pi\sqrt{1-y^2}};\quad 0<y<1;$$

$$h_y(y)=h\left[arcosen(y)\right]\left|\frac{d\left[\pi - arcosen(y)\right]}{dy}\right|+h\left[\pi - arcosen(y)\right]\left|\frac{d\ \left[2\pi + arcosen(y)\right]}{dy}\right|=\frac{1}{\pi\sqrt{1-y^2}};\quad -1<y<0$$

b) Consideriamo la trasformazione in valore assoluto Y= |X | ed ipotizziamo che la variabile di riferimento X abbia come modello la normale standardizzata:

$$h_x(x)=\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\qquad -\infty<x<\infty$$





Per la funzione di ripartizione si ottiene

$$F_y(y)=p(Y\le y)=p(|X|\le y)=p(-y\le x\le y)=\int_{-y}^{y}h(x)dx=F_x(y)-F_x(-y)$$

La funzione di densità di Y si ottiene agevolmente con il calcolo differenziale

$$h_y(y)=\frac{d}{dy}\left[F_x(y)-F_x(-y)\right]=h_x(y)+h_x(-y)=\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}+\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}=\sqrt{\frac{2}{\pi}}e^{-\frac{y^2}{2}}\quad y>0$$

Esercizio_VC113:
a) Se X è una variabile casuale continua ed Y=X^2 allora la densità della Y è data da:

$$h_y(y)=\begin{cases}\dfrac{h_x\left(-\sqrt{y}\right)+h_x\left(\sqrt{y}\right)}{2\sqrt{y}} & se\;\; y>0\\ 0 & se\;\; y\le 0\end{cases}$$

Applicare il risultato alla variabile casuale uniforme (-2,3).

b) Applicare il risultato alla variabile casuale X avente funzione di densità: $h(x)=\dfrac{2\sqrt{R^2-x^2}}{\pi R^2};\quad -R<x<R$

Discretizzazione di variabili casuali continue
Nel paragrafo 1.4.6 è stata segnalata l'eventualità di discretizzare una variabile continua per meglio affrontare diversi casi di incertezza di misurazione e di classificazione.

$$\begin{cases}y_1 \;\; se \;\; x\in\left(-\infty,x_1\right],\;\; F\left(x_1\right)-F(-\infty)=p_1\\ y_2 \;\; se \;\; x\in\left(x_1,x_2\right],\;\; F\left(x_2\right)-F\left(x_1\right)=p_2\\ y_i \;\; se \;\; x\in\left(x_{i-1},x_i\right],\;\; F\left(x_i\right)-F\left(x_{i-1}\right)=p_i\\ \dots\\ y_k \;\; se \;\; x\in\left(x_{k-1},\infty\right],\;\; F(\infty)-F\left(x_{i-1}\right)=p_k\end{cases}$$

Il meccanismo di trasformazione non è dissimile da quello già visto in opera nella trasformazione delle variabili casuali continue: $P[g(x)=y_i]=p_i$.

Esempi:
a) Sia $F(x)=3x^2$ per $0\le x\le 1$. Tale modello potrebbe esprimere la funzione di ripartizione della quota di capitale posseduta dalla società che lancia un'offerta pubblica di acquisto su di un'impresa. Si suddivide il supporto in tre sezioni di pari ampiezza assegnando i livelli di ostilità: $Y\in\{-1, 0, 1\}$; vediamo quale ne sia la distribuzione di probabilità.

$$\begin{cases}-1 \;\; x\in\left[0,\dfrac{1}{3}\right],\;\; P(Y=-1)=\dfrac{1}{3^3}\\ \;\;0 \;\; x\in\left(\dfrac{1}{3},\dfrac{2}{3}\right],\;\; P(Y=0)=\left(\dfrac{2}{3}\right)^3-\left(\dfrac{1}{3}\right)^3 \Rightarrow F_y(y)=\left(\dfrac{2+y}{3}\right)^3;\;\; y=-1,0,1\\ +1 \;\; x\in\left(\dfrac{2}{3},1\right],\;\; P(Y=+1)=1-\left(\dfrac{2}{3}\right)^3\end{cases}$$

b) Dougherty (1990, p.105) segnala che il modo più elementare di quantificare un segnale X è un sensore Y tale che Y=1 se X≥a e Y=0 se X<a dove "a" è una soglia prefissata. Ipotizziamo a=0.5 e che il segnale sia una variabile casuale con:

$$h(x)=\begin{cases}x & se\;\; 0\le x\le 1\\ -x+2 & se\;\; 1<x\le 2\\ 0 & altrove\end{cases} \Rightarrow F(x)=\begin{cases}\dfrac{x^2}{2} & se\;\; x\le 1\\ -\dfrac{x^2}{2}+2x & se\;\; 1<x\le 2\\ 1 & se\;\; x>2\end{cases}$$

La Y è una variabile binaria la cui funzione di distribuzione può essere ottenuta adoperando la procedura di discretizzazione illustrata in precedenza:

$$f_y(0)=P(Y=0)=P(X<0.5)=F(0.5)=\frac{0.5^2}{2}=\frac{1}{8};\;\; f_y(1)=P(Y=1)=P(X\ge 0.5)=1-F(0.5)=\frac{7}{8};$$

Esercizio_VC114: *la X ha la densità di Cauchy h(x)=1/[π(1+x^2)], F(x)=0.5+arctan(x)/π e si pone:*

$$Y=g(X)=\begin{cases}X-1 & se\;\; X\ge 1\\ 0 & se\;\; |X|<1\\ X+1 & se\;\; X\le -1\end{cases}$$

a) Di che tipo è la variabile Y? b) Determinare la funzione di densità della Y.

La trasformazione dell'integrale di probabilità

Se X è una variabile casuale con funzione di ripartizione F(.) la trasformazione y=F(x) produce una variabile casuale uniforme nell'intervallo [0,1]. Tale relazione è detta trasformazione dell'integrale di probabilità perché se X ha una funzione di densità h(.) la F(.) ne è l'integrale indefinito. Riprendiamo la funzione di graduazione: $g(y)=F^{-1}(y)$. Se la F(.) è monotona, ad ogni $x\in T$ corrisponde una sola $y\in[0,1]$ tale che F(x)=y e all'intervallo semiaperto $-\infty<X\leq x$ corrisponde l'intervallo $0\leq Y\leq y=F(x)$. Il primo tipo è probabilizzabile dato che X è una variabile casuale; dobbiamo accertare che lo sia anche l'altro. Notiamo l'uguaglianza tra: $E_1=\{F^{-1}(y)\leq x\}$ e $E_2=\{y\leq F(x)\}$; infatti, $F^{-1}(y)\leq x$ implica che per ogni $\varepsilon>0$, $y<F(x+\varepsilon)$ perché F(.) è continua a destra e, data l'arbitrarietà di "ε", si ha anche $y\leq F(x)$; ne consegue che $F^{-1}(y)\leq x$ e $E_1=E_2$ ed inoltre:

$$P(E_1)=P(E_2)\Rightarrow p\left[F^{-1}(y)\leq x\right]=p\left[y\leq F(x)\right]$$

In definitiva, Y=F(X) è una variabile casuale; vediamo come determinarne la funzione di densità.

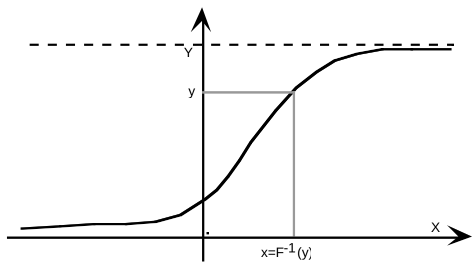
$$p(Y\leq y)=p\left[F(X)\leq y\right]=p\left\{F^{-1}\left[F(X)\right]\leq F^{-1}(y)\right\}=p\left[X\leq F^{-1}(y)\right]=F\left[F^{-1}(y)\right]=y$$

in cui la relazione $F[F^{-1}(X)]=X$ scaturisce dalla univocità della funzione di graduazione $F^{-1}()$. Se con $F_y(y)$ indichiamo la funzione di ripartizione di Y allora, in caso di variabile casuale assolutamente continua, si otterrà:

$$F_y(y)=\begin{cases}0 & se\ \ y\leq 0\\ y & se\ \ 0<y<1\\ 1 & se\ \ y\geq 1\end{cases}\Rightarrow\ h(y)=\begin{cases}1 & se\quad 0\leq x\leq 1\\ 0 & altrove\end{cases}$$

cioè il modello di densità uniforme sull'intervallo unitario. L'aspetto più interessante di questa trasformazione è però la sua inversa. Partiamo dalla variabile casuale uniforme [0,1] Y e definiamo una nuova variabile casuale con la trasformazione $X=F^{-1}(Y)$ dove $F^{-1}(.)$ è la funzione di graduazione della X. Ne consegue:

$$p(X\leq x)=p\left[F^{-1}(Y)\leq x\right]=p\left[Y\leq F(x)\right]=F(x)$$



Quindi, per ottenere una variabile casuale che abbia la funzione di ripartizione F(.) basta disporre della sua inversa in cui inserire come argomento la variabile casuale uniforme.

Esempi:

a) Se X è uniforme sull'intervallo [0,1], la trasformazione per arrivare alla ripartizione di Gumbel è:

$$F(x)=1-e^{-e^{-\left(\frac{x-a}{b}\right)}};\ \ x>0;\quad p=F(x)\Rightarrow x_p=a+bLn\left[-Ln(p)\right];\quad 0<p<1$$

b) Per ottenere una variabile casuale con la densità di Weibull si adotta la trasformazione: $F(x)=1-e^{-\left[\frac{x-\mu}{\sigma}\right]^{\gamma}}\ \Rightarrow X=\mu+\sigma\left[-Ln(U)\right]^{\frac{1}{\gamma}}$

c) Per arrivare ad una variabile casuale con funzione di densità esponenziale (Weibull con $\gamma=1$) si può utilizzare: $X=\mu+\sigma[-Ln(U)]$, ma anche $X=\mu+\sigma[-Ln(1-U)]$ dato che se U è uniforme [0,1], lo è anche Y=1-U.

Esercizio_VC115: *determinate le trasformazioni per ottenere variabili casuali:*

$$a.\ \ Champernowne: F(x)=1-\frac{2\,\mathrm{arctg}\left[\frac{a}{x}\right]}{\pi};\ \ b.Dagum: F(x)=\alpha+\frac{1-\alpha}{\left[1+x^{-a}\right]^{b}};\ \ c.h(x)=\frac{1}{2}(2-x),\ \ 0<x<2$$

7.3.4 Sintesi delle variabili casuali continue

La descrizione completa della variabile casuale continua attraverso la sua funzione di ripartizione o di densità è considerata eccessiva rispetto alle esigenze informative di molti problemi e, soprattutto, rispetto alle possibilità operative che lasciano le rilevazioni empiriche a cui sarà poi accostato lo studio delle variabili casuali. Ci si limita perciò a valutare pochi tratti essenziali, peculiari del modello, ed il resto deve o può essere tralasciato.

Centralità

La moda di una variabile casuale continua è il valore più probabile ovvero il punto in cui la densità di probabilità, se esiste, raggiunge il suo massimo globale. Se la h(x) è dotata di derivata prima per ogni punto nell'intervallo aperto [a,b], la presenza di un massimo in X_0 implicherà $h'(X_0)=0$.

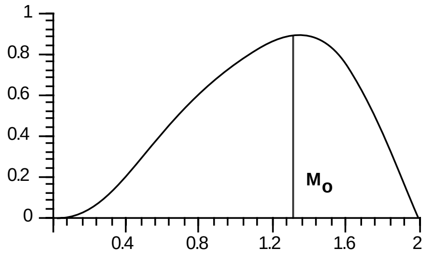
Esempi:

a) Modello polinomiale.

$$h(x)=\frac{3x^2(2-x)}{4};\quad 0<x<2;$$

$$h'(x)=3x-\frac{9}{4}x^2=0\Rightarrow x=0 \;\; o \;\; x=\frac{4}{3}$$

Il segno negativo di h''(4/3)=-3 conferma il massimo. Nel caso di supporto limitato è necessario controllare gli estremi h(a) e h(b) prima di concludere sulla moda.

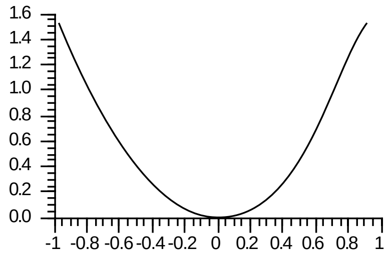


b) L'azzeramento di h'(x) è una condizione necessaria, ma non sufficiente:

$$h(x)=\frac{3}{2}x^2;\quad -1<x<1;$$

$$h'(x)=3x=0\Rightarrow x=0$$

il punto stazionario, cioè l'ascissa in cui si annulla la h'(x), individua un minimo (detto antimoda). In effetti, la derivata seconda è positiva: h''(x)=3.



c) La funzione di densità di Laplace è:

$$h(x)=\frac{e^{-\left|\frac{x-a}{b}\right|}}{2b};\quad b>0$$

che nel punto x=a non è derivabile, ma ha qui un evidente massimo globale e quindi la moda è $M_o=a$.

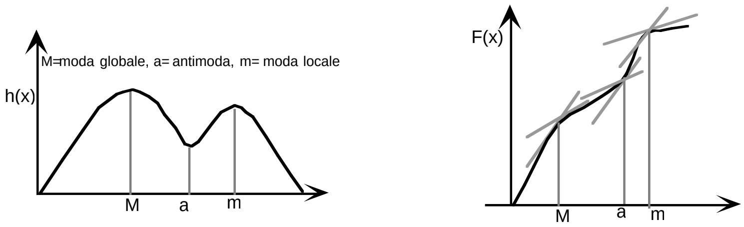
Esericzio_VC116*: per le funzioni di densità*

$$1.\;\; h(x)=\left[\frac{2}{Ln(101)}\right]\frac{x}{\left(1+x^2\right)};\quad 0<x<10;\qquad 2.\;\; h(x)=\frac{xe^{-\frac{x}{4}}}{16};\;\; x\geq 0$$

a) Determinare la moda; b) Si tratta di un valore effettivamente tipico?

Moda e funzione di ripartizione

L'individuazione grafica della moda può avvenire anche con la funzione di ripartizione. Riprendiamo il grafico con cui C.E. Bonferroni (1940-41) elaborò l'idea:



La pendenza della F(x) coincide con la densità di probabilità e la tangente alla curva di densità è la derivata seconda della funzione di ripartizione. Ad un arco crescente di h(.) corrisponde un arco convesso della F():

$$se \;\; x_1 < x_2 < M \Rightarrow F\left[\frac{(x_1+x_2)}{2}\right] \le \frac{1}{2}F(X_1) + \frac{1}{2}F(X_2)$$

e ad un arco di densità decrescente corrisponde un arco concavo:

$$se \;\; x_4 > x_3 > M \Rightarrow F\left[\frac{(x_3+x_4)}{2}\right] \ge \frac{1}{2}F(X_3) + \frac{1}{2}F(X_4)$$

In corrispondenza di X=M la h(x) ha un punto di massimo e la funzione di ripartizione un punto di flesso. Un comportamento analogo si verificherà in prossimità della submoda "m" (nel punto "a" c'è un antimoda). Quindi la densità è unimodale se la funzione di ripartizione ha un solo punto di flesso ovvero cambia curvatura una sola volta; se la curvatura non si inverte mai la moda è in uno degli estremi se il dominio è limitato oppure ne esistono infinite. Se la densità è unimodale con moda M allora:

$$\left\{F\left(\frac{x_1+x_2}{2}\right) - \left[\frac{1}{2}F(X_1) + \frac{1}{2}F(X_2)\right]\right\} * \left\{F\left(\frac{x_3+x_4}{2}\right) - \left[\frac{1}{2}F(X_3) + \frac{1}{2}F(X_4)\right]\right\}$$

è negativo per ogni quaterna di modalità tali che: $x_1<x_2<M<x_3<x_4$.

Esercizio_VC117: *per il modello:*

$$h(x) = 4xe^{-2x},\;\; x>0;\;\; F(x) = 1 - e^{-2x}(2x+1)$$

Individuare graficamente la moda usando la funzione di ripartizione.

La mediana della variabile casuale continua è la soluzione dell'equazione:

$$0.5 = F(M_e) = \int_{-\infty}^{M_e} h(x)dx \;\; ovvero \;\; M_e = F^{-1}(0.5)$$

Esempi:

a) Un noto modello di densità, usato nello studio della distribuzione dei redditi, è quello di Pareto (cfr. Arnold, 1983) la cui funzione di ripartizione è: $F(x)=1-x^{-\theta}$ con $x>\theta>1$. La mediana si ottiene con comodi passaggi algebrici:

$$0.5 = 1 - M_e^{-\theta} \Rightarrow M_e^{-\theta} = 0.5 \Rightarrow -\theta Ln(M_e) = -Ln(2) \Rightarrow M_e = e^{\frac{Ln(2)}{\theta}}$$

che dà conto delle semplificazioni promesse nella trattazione dei dati quando se ne conosca il modello.

b) Consideriamo una funzione di densità moderatamente crescente nell'intervallo [0, 0.5] che potrebbe esprimere la percentuale di astensionismo potenziale in un *referendum*.

$$h(x) = \frac{1}{Ln(2)(1-x)};\quad 0<x<0.5;\qquad F(x) = -\frac{Ln(1-x)}{Ln(2)} \Rightarrow M_e = 1 - 0.5^{0.5} = 0.2929$$

Esercizio_VC118: *la densità lineare discendente fornisce il quadro di un fenomeno che si presenta con densità in progressione regolare inversa rispetto alle modalità:*

$$h(x) = \frac{c_2 - x}{c_2 - c_1};\quad c_1 < \;\; x \;\; < c_2$$

a) Calcolare la mediana; b) Che significato assume in questo contesto?

Più in generale, per i quantili vale la formula:

$$\int_{-\infty}^{X_p} h(x)dx = p$$

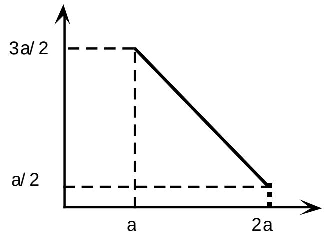
Esempi:

a) Modello di Laplace:

$$h(x)=\begin{cases}\dfrac{3e^{3x}}{2} & se \;\; x<0\\ \dfrac{3e^{-3x}}{2} & se \;\; x\geq 0\end{cases} \Rightarrow F(x)=\begin{cases}\dfrac{e^{3x}}{2} & se \;\; x<0\\ \dfrac{1-e^{-3x}}{2} & se \;\; x\geq 0\end{cases} \Rightarrow X_p=\begin{cases}\dfrac{Ln(2p)}{3} & se \;\; p\leq 0.5\\ \dfrac{-Ln(1-2p)}{3} & se \;\; p>0.5\end{cases}$$

La mediana è M_e=0 per cui i quantili inferiori a 0.50 si ottengono invertendo il primo ramo e quelli superiori l'altro cioè risolvendo $F(x_p)$=1-p rispetto ad x_p.

b) La retribuzione dei soci fondatori di una cooperativa ha probabilità inversamente proporzionale alla retribuzione stessa:



$$h(x)=\frac{1}{a^2}\left(\frac{5}{2}a\; - \; x\right); \quad a<x<2a \; ; \quad F(x)=\frac{-x^2+5ax-4a^2}{2a^2}; \;\; X_p=\frac{a}{2}\left[5-\sqrt{9-8p}\right]$$

Come verifica si può accertare che $X_{0.0}$=a e $X_{1.0}$=2a.

c) Determiniamo la funzione di graduazione per il modello di Pareto:

$$h(x)=\frac{\alpha}{x^{\alpha+1}} \quad x>1 \Rightarrow F(x)=1-\frac{1}{x^{\alpha}} \Rightarrow X_p=(1-p)^{-\frac{1}{\alpha}}$$

Se α=3 allora M_e=1.5874

Esercizio_VC119: *l'andamento iperbolico è una caratteristica riportata spesso nei modelli di variabili economiche. Ecco una funzione con tale peculiarità:*

$$F(x)=1-\frac{1}{x^2} \;\; se \;\; x\geq 1$$

a) Definire la funzione di graduazione;

b)Calcolare il quantile di ordine 0.20.

Esercizio_VC120: *a) Quale situazione potrebbe modellare la funzione di densità?*

$$h(x)=\frac{3\sqrt{x}}{2}; \quad 0\leq x\leq 1$$

b) Calcolare i tre quartili

Valore atteso di una funzione di variabile casuale continua

Le caratteristiche salienti di una variabile casuale sono esprimibili come valore atteso di una sua trasformata: y=g(x) dove g(.) è una funzione continua e differenziabile:

$$E(y)=\int_a^b y h_y(y)dy=\int_a^b y h_x\left[g^{-1}(y)\right]\frac{d}{dy}g^{-1}(y)dy;$$

Se poniamo $x=g^{-1}(y)$ ed ipotizziamo che g(.) sia monotona crescente allora dy=g'(x)dx con g'(x)>0. Sostituendo la variabile all'interno dell'integrale si ottiene:

$$E[g(x)]=\int_{-\infty}^{\infty} g(x)h_x(x)\frac{d}{dy}g^{-1}(y)g'(x)dx=\int_{-\infty}^{\infty} g(x)h_x(x)dx \quad dove \quad \frac{d}{dy}g^{-1}(y)=\frac{1}{g'\left[g^{-1}(y)\right]}=\frac{1}{g'(x)}$$

Analoga dimostrazione può essere svolta per g(.) monotona decrescente. La funzione più semplice di cui capita di dover calcolare il valore atteso è quella identità cioè g(x)=x.

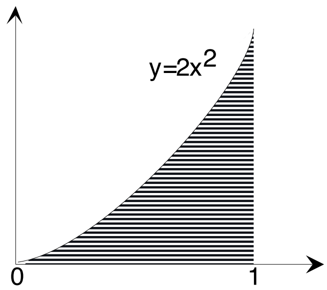
$$\mu=\int_{-\infty}^{\infty} x h(x)dx$$

che coincide con l'area sottesa alla funzione di densità moltiplicata per la variabile.

Esempi:

a) Consideriamo la funzione di densità lineare crescente:

$$h(x)=\begin{cases}2x \quad per \quad 0\le x\le 1\\ 0 \quad altrove\end{cases};\quad \mu=\int_0^1 2x^2dx=2\left(\frac{x^3}{3}\right)\Bigg|_0^1=2\left(\frac{1}{3}-0\right)=\frac{2}{3}$$



Il valore atteso μ corrisponde all'area della figura delimitata da x=0, x=1 e $y=2x^2$.

b) Calcolo del valore atteso della trasformata y=5x+3 della densità lineare: h(x)=2x per 0<x<1.

$$y=5x+3;\quad x=g(y)=\frac{y-3}{5};\ 3<y<8;\ |g'(y)|=\frac{1}{5};\ h(y)=\frac{2}{5}(y-3)$$

$$\mu_y=\int_3^8\frac{y2}{25}(y-3)dy=\frac{2}{25}\frac{y^3}{3}-\frac{3y}{2}\Bigg|_3^8=\frac{19}{3}=5\frac{2}{3}+3$$

c) Ripreso da Wadsworth e Bryan (1960, p.188). Le entrate dall'uso di un impianto di generazione di energia elettrica nel mese di novembre dipendono, tra l'altro, della caduta di pioggia in quel mese. Ipotizziamo che la funzione di guadagno sia $g(x)=5(1-e^{-x})$ miliardi dove "x" rappresenta i centimetri di pioggia caduti durante il mese. Se la funzione di densità della X è $h(x)=e^{-x}$ per x>0, quale sarà il guadagno atteso?

$$E[g(x)]=\int_0^\infty 5\left(1-e^{-x}\right)e^{-x}dx=5\int_0^\infty\left(e^{-x}-e^{-2x}\right)dx=5\left[1-\frac{1}{2}\int_0^\infty 2e^{-2x}dx\right]=5\left[1-\frac{1}{2}\right]=2.5$$

Esercizio_VC121: *calcolare il valore atteso nei modelli di densità:*

$$a.\ \ h(x)=\frac{x}{\sigma^2}e^{-\frac{x}{2\sigma}},\quad x>0;\quad b.\ \ h(x)=\frac{2}{5}(2+x),\quad 0<x<1$$

Il valore atteso ha le stesse proprietà della media aritmetica. In particolare, se la variabile casuale continua ha un supporto limitato (a,b) la E(X) non ricadrà mai all'esterno di tale intervallo:

$$a\le x\le b\Rightarrow\int_a^b ah(x)dx\le\int_a^b xh(x)dx\le\int_a^b bh(x)dx\Rightarrow a\le\mu\le b$$

Inoltre, se la variabile è trasformata linearmente, lo sarà anche il suo valore atteso:

$$E(a+bX)=\int_{-\infty}^{\infty}(a+bx)h(x)dx=\int_{-\infty}^{\infty}ah(x)dx+\int_{-\infty}^{\infty}bxh(x)dx=a+b\mu$$

Esempio:

Un pacchetto turistico sulla Calabria -disponibile per M posti- comporta un guadagno di "m" euri se collocato ed una perdita di "n" euri se rimane invenduto; il numero "x" di pacchetti collocati è una variabile casuale con densità h(x). Qual'è il numero di pacchetti "y" che un'agenzia dovrebbe prenotare? La funzione di guadagno è a due rami secondo che rimangano o no posti vuoti:

$$g(y)=\begin{cases}mx-n(y-x) & se\ \ x\le y\\ mx & se\ \ x>y\end{cases}$$

$$E[g(y)]=\int_0^y[(m+n)x-ny]h(x)dx+\int_y^M nyh(x)dx=\int_0^y[(m+n)x-ny]h(x)dx+my\left[1-\int_0^y yh(x)dx\right]=my+(m+n)\int_0^y(x-y)h(x)dx$$

La derivata dell'aspettativa è:

$$\frac{dE[g(y)]}{dy} = m + (m+n)(x-x)h(x) - (m+n)\int_0^y h(x)dx = m - (m+n)F(y) = 0 \Rightarrow F(y) = \frac{m}{m+n}$$

in cui si è fatto uso della formula di derivazione sotto il segno di integrale:

$$Se\quad \lambda(y) = \int_{r(y)}^{s(y)} f(x;y)dx \Rightarrow \frac{d\lambda(y)}{dy} = \int_{r(y)}^{s(y)} \frac{\partial}{\partial y}\left[f(x;y)\right]dx + f(x;y)\frac{ds(y)}{dy} - f(x;y)\frac{dr(y)}{dy}$$

dove r(y)=0, s(y)=y, f(x;y)=(x-y)h(x).In caso di densità uniforme si ha:

$$h(x) = \frac{1}{2b}, \quad a-b \le x \le a+b; \quad F(x) = \frac{x-(a-b)}{2b} \Rightarrow y = (a-b) + \frac{2bm}{m+n}$$

prevedendo di collocare “a” pacchetti, ma con un errore equiprobabile di sopravvalutare o sottovalutare la domanda fino ad un massimo di “b” pacchetti. Il numero ottimale di pacchetti sarà il quantile della uniforme corrispondente al rapporto m/(m+n): se m=3, n=1, a=10, b=2 la scelta ottimale è y=11.

Esercizio_VC122: *Wadsworth e Bryan (1960, pp. 196-197) introducono la possibilità di una perdita nel caso in cui la domanda superi la disponibilità dell’agenzia:*

$$g(y) = \begin{cases} mx - n(y-x) & se\ \ x \le y \\ mx - k(x-y) & se\ \ x > y \end{cases}$$

Qual’è ora il numero ottimale se k=2?

Esercizio_VC123: *il fisico-matematico Maxwell ha dimostrato che ad una fissata temperatura, la velocità X di una molecola di gas ha funzione di densità:*

$$h(x) = kx^2 e^{-\frac{3x^2}{2\sigma^2}}, \quad x > 0$$

dove σ è un parametro di scala e “k” una costante che rende l’area sottesa alla curva pari ad uno.
a) Determinare “k”; b) Calcolare E(mx^2) dove “m” è un parametro che indica la massa di una singola molecola.

Esistenza del valore atteso

L’integrale del valore atteso non sempre esiste finito; affinché ciò avvenga bisogna e basta che:

$$\lim_{\substack{a\to-\infty \\ b\to\infty}} \int_a^b |g(x)h(x)|dx = \lim_{\substack{a\to-\infty \\ b\to\infty}} \int_a^b |g(x)|h(x)dx = E\big[|g(x)|\big] < \infty$$

Se questo non si verifica allora il valore atteso E[g(x)] non esiste finito e non è tale nemmeno nel caso in cui sia convergente l’integrale di g(x), ma non quello di |g(x)|.

Esempi:
a) Per la funzione di densità:

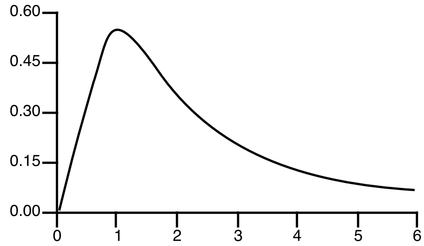
$$h(x) = \frac{x}{\left[\frac{1}{2b^2} + bx\right]^3}, \quad x > 0, \ \ b > 0$$

La media non esiste finita, come del resto si intuisce dallo spessore della coda.

b) Nel modello esponenziale il valore atteso esiste finito.

$$h(x) = \lambda e^{-\lambda x}, \quad x > 0 \Rightarrow \mu = \lambda\int_0^\infty xe^{-\lambda x} = \frac{1}{\lambda}$$

Anche qui sono possibili valori infinitamente grandi cioè in un campione descritto da tale modello ci si aspetta di tanto in tanto di rinvenire qualche valore fuori dall’ordinario. Come mai questo non porta ad un valore atteso infinito? Perché qui le *chances* dei valori grandi sono relativamente più piccole che non nel modello precedente.



c) Densità iperbolica.

$$h(x)=\frac{a}{x^2};\ \ x>a;\ \ \mu=\int_a^\infty \frac{xa}{x^2}dx=\int_a^\infty \frac{a}{x}dx=aLn(x)\Big|_a^\infty=-aLn(a)+a\left[\lim_{x\to\infty} Ln(x)\right]=\infty$$

L'area sottesa all'iperbole non tende ad un limite finito dato che la modalità cresce più rapidamente di quanto non decresca la densità di probabilità ad essa associata. Come si interpreta μ? Non certo come valore ripartitorio dato che l'ammontare da suddividere è infinito ed ogni quota potrebbe essere infinita e nemmeno come speranza matematica di una partita in cui tutti sarebbero disposti a giocare e nessuno a fare da banco. In senso fisico si può dire che il fenomeno rappresentato non possiede un punto di equilibrio stabile in quanto il suo fulcro è trascinato all'infinito dalla presenza di valori grandissimi che agiscono con pesi piccolissimi, ma ancora rilevanti. La centralità di queste distribuzioni dovrà essere misurata con indici immuni da questo difetto (i quantili o le medie troncate ad esempio).

Esercizio_VC124: *per la funzione di ripartizione*

$$F(x)=1-x^{-2};\ \ x\geq 1$$

Verificare che esiste finito E(x), ma non E(x^2).

Valore atteso indeterminato

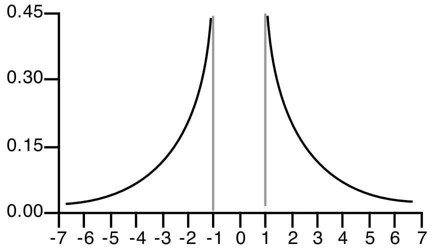
Ash (1970, p.105-106), e Sinai (1991, pp. 102-103) definiscono il valore atteso di una variabile casuale come la differenza tra i valori attesi di due particolari trasformazioni:

$$\begin{cases} X^+ = Max\{X,0\} \\ X^- = Max\{-X,0\}\end{cases} \Rightarrow E(X)=E\left(X^+\right)-E\left(X^-\right)$$

Se il valore atteso assume la forma $+\infty$ **-**∞ allora E(X) è indeterminata (questo può succedere solo per variabili casuali con supporto negativo e positivo). Se le due aspettative sono entrambe finite allora la media esiste finita; se solo una delle due è finita allora la media è infinita ovvero non esiste finita.

Esempio:

a) Densità con vuoto al centro.



$$h(x)=\begin{cases}0 \ \ se\ \ -1<x<1\\ \dfrac{1}{2x^2}\ \ altrove\end{cases}$$

$$E\left(X^+\right)=\int_{-\infty}^{\infty}\frac{x^+}{2x^2}dx=\int_1^\infty\frac{x}{2x^2}dx=\frac{Ln(x)}{2}\Bigg|_1^\infty=\infty$$

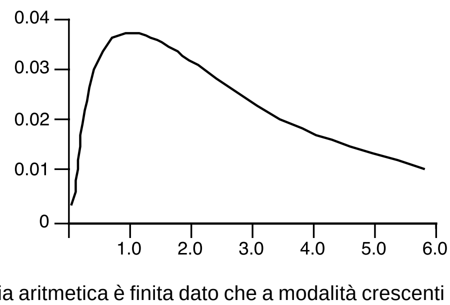
$$E\left(X^-\right)=\int_{-\infty}^{\infty}\frac{x^-}{2x^2}dx=\int_{-\infty}^{-1}\frac{-x}{2x^2}dx=\frac{Ln(x)}{2}\Bigg|_1^\infty=\infty$$

$$E(X)=\infty-\infty=indeterminata$$

b) Modello di Erlang (versione continua del modello di Pascal).

$$h(x)=\frac{x^n e^{-x}}{n!},\ \ x>0$$

$$\mu=\int_0^\infty\frac{x^{n+1}e^{-x}}{n!}dx=\frac{(n+1)!}{n!}=n+1$$



Per n=2 la media aritmetica è pari a 3. nonostante $x\to\infty$ la media aritmetica è finita dato che a modalità crescenti il modello assegna probabilità decrescenti a ritmo più rapido dell'aumento delle prime.

c) La legge di Ohm I=V/R stabilisce una relazione tra l'intensità di corrente I che attraversa un dipolo, la tensione V misurata ai capi e la sua resistenza R. Supponiamo che V sia costante, ma che la resistenza sia una variabile casuale. Anche I è una variabile casuale, ma ha media infinita:

$$h(R)=\mathrm{Re}^{-R},\ \ R>0;\ \ R=\frac{V}{I}\ \ \Rightarrow\ \ h(I)=\frac{V}{I}e^{-\frac{V}{I}}\left|\frac{dR}{dI}\right|=\frac{V^2}{I^3}e^{-\frac{V}{I}}\ \ \Rightarrow\ \ E(I)=\int_0^\infty\frac{V^2}{I^2}e^{-\frac{V}{I}}dI=\infty$$

I modelli con media infinita sono efficaci per addomesticare fenomeni a coda pesante cioè situazioni in cui certi valori o intervalli di valori, per quanto remoti ed estremi rispetto al centro della distribuzione, mantengono possibilità concrete di manifestazione: la distribuzione della ricchezza rientra in questa tipologia e vi rientrano anche le precipitazioni atmosferiche.

Esercizio_VC125: *si supponga per un fenomeno sia adatto il modello di Singh e Maddala (1976):*

$$F(x)=1-\left(\frac{1}{1+ax^b}\right)^c;\quad x>0,\ \ a,b,c>0$$

utile per la previsione dell'importo da liquidare per danni provocati da pioggia, vento, grandine. Per quali valori dei parametri la media esiste finita?

I momenti

I momenti riassumono gli aspetti principali del modello di una variabile casuale: il valore atteso, lo scarto quadratico medio, l'asimmetria. Tale capacità specificativa si è dimostrata utile sia con le distribuzioni di frequenza che per le distribuzioni di probabilità delle variabili casuali discrete. Un ruolo analogo è loro riservato con le continue.

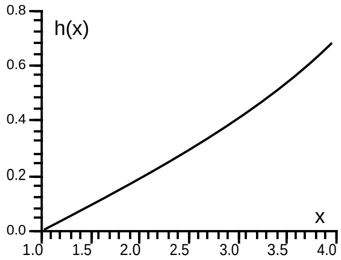
$$\textit{dall' origine}:\ \mu_\alpha=\int_{-\infty}^{\infty}x^\alpha h(x)dx;\quad \textit{dalla media}:\ \mu'_\alpha=\int_{-\infty}^{\infty}(x-\mu)^\alpha h(x)dx;$$

Esempi:

a) Densità crescente.

$$h(x)=\frac{6\left(x-\sqrt{x}\right)}{17},\ \ 1\le x\le 4;\ \ \mu_1=\int_1^4 x\,\frac{6\left(x-\sqrt{x}\right)}{17}dx=2.8;$$

$$\mu_2=\int_1^4 x^2\,\frac{6\left(x-\sqrt{x}\right)}{17}dx=9.693;\ \ \sigma=\sqrt{9.693-2.8^2}=1.36$$

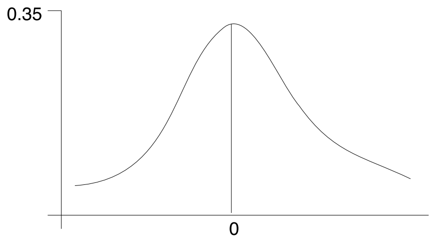


Sono sempre vigenti le relazioni tra momenti già viste per le discrete. Tra le altre, il calcolo della varianza: $\sigma^2=\mu_2-\mu^2$.

b) Il modello di Cauchy ha andamento campanulare simmetrico intorno allo zero.

$$h(x)=\frac{1}{\pi\left(1+x^2\right)},\ -\infty<x<\infty;\ \ E(x)=\int_{-\infty}^{\infty}\frac{x}{\pi\left(1+x^2\right)}dx=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{2x}{\left(1+x^2\right)}dx$$

$$E(x)=\frac{1}{2\pi}Ln\left(1+x^2\right)\Big|_{-\infty}^{\infty}=+\infty-\infty\ \ \textit{non esiste}$$



Il disegno delle code è rimasto sospeso per rimarcarne lo spessore. L'entità delle code può essere saggiata rappresentando graficamente la funzione xh(x) rispetto alla "x".

Esercizio_VC126: *calcolare i primi tre momenti all'origine per i modelli:*

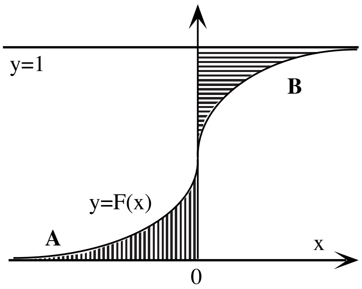
$$h(x)=\left[\frac{b}{Ln(1+b)}\right]\frac{1}{1+bx};\quad 0<x<1,\ \ b>0\ \textit{1.}\qquad \textit{2.}\ \ h(x)=\lambda e^{-\lambda x},\ \ x>0,\ \ \lambda>0$$

Se il valore atteso esiste finito se ne può dare un'efficace interpretazione adoperando la funzione di ripartizione e quella di sopravvivenza e l'area loro sottesa. Infatti, E(x) è uguale alla differenza delle aree tratteggiate e marcate con A e B.

$$"B":\ \int_0^{\infty}[1-F(x)]dx = x[1-F(x)]\Big|_0^{\infty} + \int_0^{\infty} xh(x)dx = \lim_{x\to\infty} x[1-F(x)] + \int_0^{\infty} xh(x)dx$$

$$"A":\ \int_{-\infty}^{0} F(x)dx = xF(x)\Big|_{-\infty}^{0} - \int_{-\infty}^{0} xh(x)dx = \lim_{x\to-\infty} xF(x) - \int_{-\infty}^{0} xh(x)dx$$

$$"B" - "A":\ \int_0^{\infty} xh(x)dx + \int_{-\infty}^{0} xh(x)dx = \int_{-\infty}^{\infty} xh(x)dx = \mu$$



Il tendere di μ all'infinito significa che una o entrambe le aree diventano illimitate.

Esercizio_VC127: *dimostrare la disuguaglianza di Markov per una variabile casuale continua non negativa*

$$P(x \ge c) \le \frac{E(x)}{c} \quad se \quad x > 0$$

Momenti assoluti e momenti parziali

I momenti hanno avuto grande attenzione cosicché su di essi si è accumulato molto materiale che è difficile esporre in breve. Ricordiamo i momenti assoluti che si definiscono considerando in modulo la variabile:

$$\nu_r = \int_{-\infty}^{\infty} |x| h(x)dx$$

I momenti assoluti coincidono con i quelli ordinari per variabili positive e coincidono comunque per i momenti di ordine pari. Peraltro, vale la disuguaglianza: $\mu_r \le |\mu_r| \le \nu_r$. I momenti parziali o incompleti sono i momenti dall'origine, ma calcolati solo per valori inferiori ad un dato limite oppure eccedenti una soglia prefissata:

$$\mu_k^+(M) = \begin{cases} Discrete: \dfrac{\sum_{X_{(i)} \le M} X_{(i)}^k f_i}{F(M)} \\ Continue: \dfrac{\int_{-\infty}^{M} x^k h(x)dx}{F(M)} \end{cases}; \quad \mu_k^-(M) = \begin{cases} Discrete: \dfrac{\sum_{X_{(i)} > M} X_{(i)}^k f_i}{1-F(M)} \\ Continue: \dfrac{\int_{M}^{\infty} x^k h(x)dx}{1-F(M)} \end{cases}$$

dove M<∞ è la porzione di supporto su cui incide il calcolo e F(M) ed G(M) sono le probabilità cumulate e retrocumulate delle modalità fino ad M e da M in poi.

Esempio:

$$h(x) = \frac{15x^2(1+x^2)}{272},\ -2 < x < 2; \quad F(x) = \frac{5x^3 + 3x^5 + 136}{272}$$

$$\nu_1 = \frac{15}{272}\int_{-2}^{2} |x| x^2(1+x^2)dx = \frac{15}{272}\left[-\int_{-2}^{0} x^3(1+x^2)dx + \int_0^2 x^3(1+x^2)dx\right] = \frac{30}{272}\int_0^2 x^3(1+x^2)dx = \frac{55}{4}$$

$$\mu_1^+(0) = \frac{\frac{15}{272}\int_{-2}^{0} x^3(1+x^2)dx}{F(0)} = \frac{-\frac{15}{272}\frac{352}{24}}{\frac{136}{272}} = -\frac{55}{34}; \quad \mu_1^-(1) = \frac{\frac{15}{272}\int_0^2 x^3(1+x^2)dx}{G(1)} = \frac{\frac{15}{272}\frac{342}{24}}{\frac{128}{272}} = \frac{55}{32}$$

Esercizio_VC128: *per la densità:*

$$f(x) = \begin{cases} \dfrac{3+2x}{28} & per \quad 0 < x < 4 \\ 0 & altrove \end{cases}$$

a) Calcolare i primi due momenti assoluti. Perché coincidono con i momenti ordinari?
b) Calcolare il valore atteso per i valori inferiori e per i valori superiori alla mediana.

Momenti finiti e probabilità nelle code

L'esistenza dei momenti è legata al comportamento della variabile nelle code. Infatti, g(x) è una trasformazione non negativa monotona, almeno a tratti, e se E[g(x)] esiste finito, allora si dimostra (Rohatgi, 1976, p.100) che la probabilità nelle code per la trasformazione g(x) è soggetta al vincolo:

$$P\left[g(x) \geq a\right] \leq \frac{E\left[g(x)\right]}{a}, \quad con \quad a > 0$$

Se ora si pone $g(x)=|x|^k$ con k>0 e a>0 si ottiene la disuguaglianza di Markov:

$$aP\left[|x|^k \geq a\right] \leq E\left[|x|^k\right]$$

Se il k-esimo momento esiste finito, allora

$$\mu_k = \int_{-\infty}^{+\infty} |x|^k h(x)dx < \infty \Rightarrow \underset{a\to\infty}{Lim} \int_{|x|\leq a} |x|^k h(x)dx < \infty \quad e \quad \underset{a\to\infty}{Lim} \int_{|x|>a} |x|^k h(x)dx \to 0$$

Poiché risulta:

$$aP\left(|x|^k > a\right) = a \int_{|x|^k > a} h(x)dx = \int_{|x|^k > a} ah(x)dx \leq \int_{|x|^k > a} |x|^k h(x)dx$$

con la conseguenza che

$$\underset{a\to\infty}{Lim} \int_{|x|>a} |x|^k h(x)dx \to 0 \Rightarrow \underset{a\to\infty}{Lim} \; aP\left(|x|^k > a\right) = 0$$

La probabilità nelle code tende a zero con lo stesso ordine di grandezza di $(1/x^k)$: maggiore è l'esponente "k" più rapido sarà il decadimento.

Esempio:
Rohatgi (1976, p. 83) osserva che il converso di tale risultato non è necessariamente vero. Per illustrare questo punto considera una funzione di densità che si incontra nello studio dei numeri primi:

$$h(x) = \frac{1}{2|x|\left(Ln|x|\right)^2}; \quad |x| > e$$

Per tale funzione non esiste finito alcun momento di ordine k>0, ma aP(|x|>a) tende a zero. Se la probabilità nelle code si riduce rapidamente questo può assicurare forse l'esistenza dei momenti di ordine più basso, ma non c'è garanzia che ciò succeda anche per potenze più grandi.

Se, per un certo "r", i momenti sono infiniti allora sono infiniti tutti i momenti di ordine superiore. Infatti, dalla disuguaglianza di Liapunov, si può dedurre che:

$$\upsilon_k^{s-t} \leq \upsilon_t^{s-k} \upsilon_s^{k-t} \Rightarrow \upsilon_{k+1} \leq \frac{\upsilon_k^2}{\upsilon_{k-1}}$$

Peraltro, se esiste finito il momento k-esimo, esisteranno anche tutti gli altri momenti di ordine inferiore, sia rispetto all'origine che rispetto ad una modalità (qualsiasi purché finita) di riferimento e se è infinito il momento k-esimo lo sono anche quelli di ordine superiore. L'esistenza finita del momento k-esimo nulla consente di aggiungere sui momenti di ordine superiore.

Esempio:
Se esiste finito il momento secondo all'origine, esisterà finito anche il valore atteso dato che $E(x)<E(x^2)$ e quindi è finita anche la varianza poiché $\sigma^2(x) \leq E(x^2)$. Nulla si può dire sul momento 3° e quindi sulle misure di asimmetria ad esso legate.

Esercizio_VC129: *il volume delle vendite di un prodotto è modellato con la densità:*

$$h(x)=\frac{2\sigma}{x^2}e^{-\frac{2\sigma}{x}},\ \ x>0$$

a) Verificare l'esistenza dei momenti di qualsiasi ordine;
b) Cosa di può dire se le vendite ammontano -per contratto- ad almeno una unità: x>1?

Indici di variabilità per le variabili casuali continue

La natura di "media" di molti indici descrittivi ne consente l'espressione come area sottesa alla funzione di densità moltiplicata per una opportuna trasformazione della X:

$$\textit{Deviazione standard}:\ \sigma=\sqrt{E[X-E(X)]}=\sqrt{\int_{-\infty}^{\infty}(x-\mu)^2h(x)dx}=\sqrt{\int_{-\infty}^{\infty}x^2h(x)dx-\mu^2}$$

$$\textit{Scarto ass. mediano}:\ S_{Me}=\int_{-\infty}^{+\infty}|x-M_e|h(x)dx=\int_{-\infty}^{M_e}(M_e-x)h(x)dx+\int_{-\infty}^{M_e}(x-M_e)h(x)dx$$

$$\textit{Deviazione media}:S_\mu=\int_{-\infty}^{+\infty}|x-\mu|h(x)dx=2\mu F(\mu)-2\int_{-\infty}^{\mu}xh(x)dx$$

$$\textit{Differenza media}:\Delta=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x-y|h(x)h(y)dxdy=2\int_{-\infty}^{\infty}F(x)[1-F(x)]dx=4\int_{-\infty}^{+\infty}x[F(x)-0.5]h(x)dx=2F(\mu)[\mu-\mu_1^+(\mu)];$$

Esempi:
a) Modello di densità uniforme: h(x)=1/(b-a) per a≤x≤b.

$$\int_a^b\frac{x^2}{(b-a)}dx=\frac{x^3}{3(b-a)}\Bigg|_a^b=\frac{b^3-a^3}{3(b-a)};\quad \int_a^b\frac{x}{(b-a)}dx=\frac{x^2}{2(b-a)}\Bigg|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{(b+a)(b-a)}{2(b-a)}=\frac{b+a}{2}$$

$$\sigma^2=\frac{b^3-a^3}{3(b-a)}-\left[\frac{b+a}{2}\right]^2=\frac{(b-a)^3}{12(b-a)}=\frac{(b-a)^2}{12}\Rightarrow\sigma=\frac{b-a}{\sqrt{12}}$$

Se la X varia uniformemente tra 0 e 10, lo scarto quadratico medio sarà: 10/√12=2.89. A tale risultato si arriva direttamente conoscendo i parametri, senza procedere alla escussione di alcun valore rilevato empiricamente: è questa la forza dei modelli.

b) I tempi di produzione di un *item* in due diversi processi X ed Y sono descritti dalle variabili casuali aventi densità:

$$h(x)=\begin{cases}x-1 & se\ \ 1\le x<2\\ 3-x & se\ \ 2\le x\le 3\end{cases};\qquad h(y)=\begin{cases}4y-6 & se\ \ 1.5\le y<2\\ -4y+10 & se\ \ 2\le y\le 2.5\end{cases};$$

Entrambi i modelli prevedono una media μ=2 ore. Dove si riscontra maggiore variabilità? Nel processo X.

$$E(x^2)=\int_1^2(x^3-x^2)dx+\int_2^3(3x^2-x^3)dx=\frac{50}{12}\Rightarrow\sigma_x^2=\frac{50}{12}-4=0.1667$$

$$E(y^2)=\int_{1.5}^2(4y^3-6y^2)dy+\int_2^{2.5}(-4y^3+10y^2)dy=\frac{97}{24}\Rightarrow\sigma_y^2=\frac{97}{24}-4=0.0417$$

c) Se si pensa ad una variabile a valori in [0,1] con densità decrescente a ritmi decrescenti potrebbe essere utile il modello:

$$h(x)=\frac{3}{2}(1-x^2)\quad 0\le x\le 1;\qquad F(x)=\frac{3x-x^3}{2};\quad M_e=0.3473$$

La mediana è definita da F(x)=0.5 che qui implica la soluzione di un'equazione cubica. Il calcolo può essere condotto con il metodo Newton-Raphson anche su foglio elettronico.

$$S_{Me}=\int_0^{Me}(M_e-x)\frac{3}{2}(1-x^2)dx+\int_{Me}^1(x-M_e)\frac{3}{2}(1-x^2)dx=M_eF(M_e)-\frac{3}{8}(2M_e^2-M_e^4)+\frac{3}{8}\left[1-(2M_e^2-M_e^4)\right]-M_e[1-F(M_e)]$$

$$=M_e0.5+\frac{3}{8}-\frac{6}{8}(2M_e^2-M_e^4)-M_e0.5=\frac{3}{8}-\frac{6}{8}(2M_e^2-M_e^4)=0.205$$

d) Calcolo della deviazione media.

$$h(x)=\frac{\left(x^2-1\right)^{-\frac{1}{2}}}{Ln\left(2+\sqrt{5}\right)},\quad 1\ \le x\le\sqrt{5};\quad F(x)=\frac{Ln\left(x+\sqrt{x^2-1}\right)}{Ln\left(2+\sqrt{5}\right)};$$

$$\mu=\int_1^{\sqrt{5}}x\frac{\left(x^2-1\right)^{-\frac{1}{2}}}{Ln\left(2+\sqrt{5}\right)}dx=\left.\frac{\sqrt{x^2-1}}{Ln\left(2+\sqrt{5}\right)}\right|_1^{\sqrt{5}}\cong 1.39;\ S_\mu=\int_1^{\mu}(\mu-x)h(x)dx+\int_\mu^{\sqrt{5}}(x-\mu)h(x)dx=0.31$$

e) Calcolo della differenza media.

$$h(x)=2x,\quad 0\le x\le 1;$$

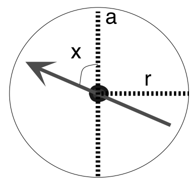
$$\Delta=\int_0^1\int_0^1|x-y|4xydxdy=2\int_0^1 2x\int_0^x(x-t)2tdtdx=2\int_0^1 2x\left[\left.xt^2-\frac{2}{3}t^3\right|_0^x\right]dx=2\int_0^1 2x\left(\frac{x^3}{3}\right)dx=\frac{4}{3}\int_0^1 x^4dx=\frac{4}{15}$$

Scegliendo a caso due valori dalla densità lineare essi differiranno in valore assoluto e in media di 4/15.

f) Nel caso della funzione di densità proporzionale si ha:

$$h(x)=\frac{3x^2}{a^3},\quad 0\le\ x\ \le a;\ \ F(x)=\left(\frac{x}{a}\right)^3\ \Rightarrow\ \Delta=2\int_0^a\left(\frac{x}{a}\right)^3\left[1-\left(\frac{x}{a}\right)^3\right]dx=\frac{3a}{14}$$

g) Su di un disco di raggio “r” e centro “c” si fa ruotare una freccia con un impulso casuale. Se il sistema freccia/perno è ben equilibrato l’angolo (in radianti) formato con la posizione “a” di avvio è una variabile casuale uniforme con supporto in [0,2πr]. Valore atteso e scarto quadratico medio sono, rispettivamente:



$$\mu=\int_0^{2\pi r}\frac{x}{2\pi r}dx=\pi r;\ \ \sigma=\sqrt{\int_0^{2\pi r}\frac{x^2}{2\pi r}dx-(\pi r)^2}=\frac{\pi r}{\sqrt{3}}$$

Se la densità ha code spesse può succedere che esista finito il valore atteso, ma non lo scarto quadratico medio. In questi casi conviene misurare la variabilità con la deviazione media o con la differenza media che hanno come precondizione l’esistenza finita del 1° momento e non del 2°.

Esercizio_CV130:

a) Per il modello di densità: $h(x)=2\alpha e^{-\alpha x}\left(1-e^{-\alpha x}\right),\quad x>0$. *Determinare la varianza.*

b) Per il modello di densità lineare: $h(x)=\frac{x}{4}+\frac{1}{2};\ \ -1\le x\le 1$. *Determinare lo scarto assoluto mediano.*

Esercizio_VC131:

a) Calcolare deviazione media e differenza media per i due modelli:

$$1)\ \ h(x)=\begin{cases}6x(1-x) & per\ \ 0<x<1\\ 0 & altrove\end{cases};\qquad 2)\ \ h(x)=\begin{cases}1+x & se\ \ -1<x<0\\ 1-x & se\ \ 0\le x<1\end{cases}$$

b) Calcolare σ *e* S_{Me} *per la variabile casuale con densità di tipo quadratico:*

$$h(x)=\left(\frac{3}{10}\right)\left(2-x^2\right);\quad -1\le x\le 1$$

c) Calcolare la deviazione media per il modello: $h(x)=\frac{|2-x|}{16};\ \ -4\le x\le 4$

Esercizio_VC132: *qual’è l’effetto di una trasformazione lineare Y=aX+b sulle diverse misure di variabilità?*

La disuguaglianza di Tchebycheff nel continuo

Parliamo ora della disuguaglianza di Tchebycheff (estendibile alle variabili casuali continue purché esista finito il momento secondo). Le soglie che essa fornisce definiscono un minimo assoluto alla probabilità degli intervalli e quindi sono troppo prudenziali (conservative) per essere utilizzabili in pratica, anche se danno un’idea abbastanza chiara della distribuzione quando si sa poco della variabile. Le carenze della Tchebycheff si possono verificare allorché sia noto il modello di densità; essa ha infatti una efficacia differenziata secondo la forma della curva di densità.

Esempio:
Per le funzioni densità seguenti:

$$h_1(x) = 3(x-1)^2;\ 1 \le x \le 2; \qquad h_2(x) = 1;\ \ 1 \le x \le 2$$

L'intervallo x=μ±1.25σ include, rispettivamente l'84.6% ed il 58.74% laddove la Tchebycheff propone, per entrambe, la soglia del 36%.

Se h(x) è nota, il ricorso alla Tchebycheff è sbagliato, a meno che non abbia un'espressione troppo complessa.

Esercizio_VC133: *in alcuni modelli è possibile migliorare la disuguaglianza di Tchebycheff cioè renderla più aderente alle soglie vere. Se esiste finito il momento 4° e si ipotizza che μ=0 allora vale la disuguaglianza:*

$$P(|x| > \varepsilon\sigma) \ge \frac{\mu_4 - \sigma^4}{\mu_4 + \left(\varepsilon^4 - 2\varepsilon^2\right)\sigma^4},\ \ per\ \ \varepsilon > 1$$

che sorpassa la Tchebycheff se $\varepsilon^2 > \mu_4/\sigma^4$ (Rohatgi, 1976, p. 102). Verificate che tale risultato per:

$$h(x) = \frac{8 - x^3}{12};\ \ 0 \le x \le 2$$

La curva di Lorenz per variabili assolutamente continue

E' possibile dare una definizione parametrica della curva di Lorenz di una variabile casuale continua per x>a con valore atteso positivo.

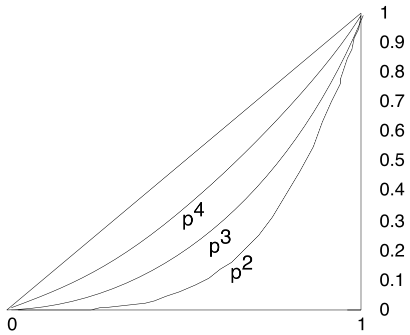
$$p = F(x);\quad L(p) = \frac{1}{\mu}\int_a^{g(p)}(x-a)h(x)dx;\ \ \mu = \int_a^{\infty} xh(x)dx;\ \ g(p) = F^{-1}(p),\ \ x \ge a$$

Esempi:
a) La funzione di concentrazione coincide con la retta di equidistribuzione L(p)=p se e solo se la distribuzione della X è la distribuzione degenere. Seguiamo l'illustrazione di Moothatu (1983):

$$F(x) = \begin{cases} 0 & se\ \ x < \mu \\ 1 & se\ \ x \ge \mu \end{cases} \Rightarrow g(p) = \mu\ \ per\ \ 0 < p \le 1$$

Quindi, la funzione di concentrazione sarà: $L(p) = \frac{1}{\mu}\int_0^p \mu dp = p;\ \ 0 \le p \le 1$

b) C'è un preciso legame tra la derivata seconda della curva di Lorenz e la funzione di densità. Infatti:



$$L'(p) = \frac{g(p)}{\mu} = \frac{F^{-1}(x_p)}{\mu} \Rightarrow L''(p) = \frac{1}{\mu}\left[\frac{1}{\frac{dF}{dx}}\right]_{x=x_p} = \frac{1}{\mu h(x_p)} > 0$$

La convessità è confermata dal fatto che μ>0 e che la densità è positiva per x>a.

c) A quale modello di variabile casuale è associata una curva di Lorenz potenziata? E' sufficiente ragionare sulla derivata prima della curva di Lorenz:

$$L(p) = p^b,\ \ b > 1 \Rightarrow L'(p) = bp^{b-1} = \frac{F^{-1}(x_p)}{\mu} \Rightarrow x_p = \mu bp^{b-1}$$

$$F(x) = \left[\frac{x}{b\mu}\right]^{\frac{1}{b-1}}\ \ per\ \ 0 < x < b\mu$$

cioè il modello proporzionale. Ragioniamo all'inverso e partiamo dal modello di variabile casuale proporzionale.

$$h(x) = ax^{a-1};\ \ 0 \le x \le 1; \qquad F(x) = x^a$$

$$\mu = \int_0^1 xax^{a-1}dx = \frac{a}{a+1};\ \ g(p) = F^{-1}(p) = p^{\frac{1}{a}} \Rightarrow L(p) = \frac{a+1}{a}\int_0^{p^{\frac{1}{a}}} xax^{a-1}dx = p^{\frac{a+1}{a}}$$

Le curve si annidano in ragione dell'esponente. La retta di equidistribuzione è raggiunta per "a" tendente ad infinito e la curva di massima concentrazione si ottiene per "a" tendente a zero.

La curvatura della funzione di concentrazione

Nel linguaggio comune si dice che un tratto di strada è pericoloso se la direzione cambia bruscamente rispetto alla lunghezza del tratto. Questo concetto si traduce nella curvatura che è una generalizzazione dell'idea di tangente. Per la curva di Lorenz si ha:

$$K(p)=\frac{L''(p)}{\left\{1+\left[L'(p)\right]^2\right\}^{\frac{3}{2}}}=\frac{\frac{1}{\mu h(x)}}{\left[1+\left(\frac{x}{\mu}\right)^2\right]^{\frac{3}{2}}}=\frac{\mu^2}{h(x)\left[\mu^2+x^2\right]^{\frac{3}{2}}}$$

Taguchi (1968) ha proposto alcuni indici per caratterizzare la curva di Lorenz basati sul concetto di curvatura.

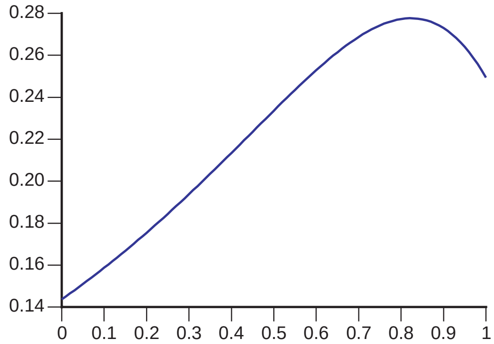
Esempio:

Riprendiamo una delle curve proposte nel quinto capitolo:

$$L(p)=\frac{Ln(1-\beta p)}{Ln(1-\beta)}\quad per\quad 0<\beta<1$$

$$L'(p)=\frac{-\beta(1-\beta p)^{-1}}{Ln(1-\beta)};\quad L''(p)=\frac{-\beta^2(1-\beta p)^{-2}}{Ln(1-\beta)};$$

$$K(p)=\frac{\beta^2(1-\beta p)^{-2}}{\left[Ln(1-\beta)^2+\beta^2(1-\beta p)^{-2}\right]^{\frac{3}{2}}}$$



La curvatura di questa funzione di concentrazione per $\beta=0.75$ ha un massimo a p=0.85. Tale punto, noto come punto di saturazione, se esiste ed è unico può essere utilizzato per specificare la curva di Lorenz.

Esercizio_VC134: *a partire dalla curva di Lorenz:*

$$L(p)=1-(1-p)^{0.5}$$

Determinare la corrispondente funzione di densità e la sua curvatura. Esiste un punto di saturazione?

Esercizio_VC135:

a) Ad una funzione di densità simmetrica corrisponde necessariamente una curva di Lorenz simmetrica?

b) Se la curva di Lorenz è simmetrica se ne può dedurre che la densità è simmetrica?

Asimmetria delle variabili casuali continue

E' riferita alle caratteristica di simmetria intorno alla retta $x=M_e$ della loro densità di probabilità: $h(M_e-x)=h(M_e+x)$ per ogni "x" ovvero la densità di probabilità è la stessa per ascisse equidistanti dalla mediana. In tal caso se $E(|X|)<\infty$, mediana e valore atteso sono uguali (se c'è unimodalità coincide pure la moda).

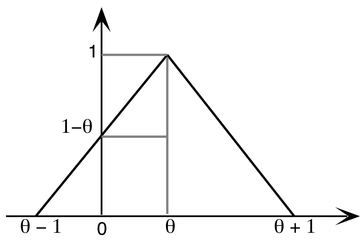
Esempi:

a) Consideriamo il modello di densità triangolare:

$$h(x)=1-|x-\theta|,\qquad \theta-1<x<\theta+1;\ \ \theta>0$$

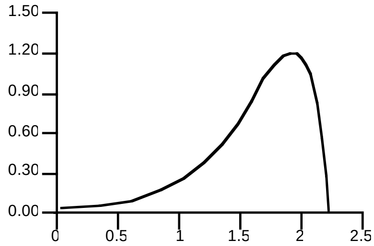
E' semplice accertare che $M_e=\theta$. Per valutare la simmetria basta considerare le relazioni:

$$h(\theta-x)=1-|\theta-x-\theta|=1-|x|;$$
$$h(\theta+x)=1-|\theta+x-\theta|=1-|x|;$$



b) La funzione di densità che segue:

$$h(x)=\begin{cases}0.06261\left[x^4\left(5-x^2\right)\right] & per \quad 0<x<\sqrt{5}\\ 0 & altrove\end{cases}$$



non è simmetrica in quanto, come mostra il grafico, si distende verso i valori bassi e rimane alta per le modalità più grandi del supporto.

Esercizio_VC136: *tenuto conto che per variabili casuali simmetriche (continue o discrete) si ha*

$$F(M_e - \varepsilon) = 1 - F(M_e + \varepsilon) \text{ per ogni } \varepsilon$$

Verificare la simmetria del modello logistico:

$$F(x)=\frac{1}{1+e^{-\left(\frac{x-\mu}{\sigma}\right)}}$$

Per misurare l'asimmetria delle funzioni di densità si possono richiamare gli indici già adoperati per le distribuzioni empiriche delle frequenze relative, quali ad esempio, l'indice α_1 ed il γ_1.

Esempi:

a) Calcoliamo l'indice semplice di asimmetria per un modello quadratico:

$$h(x)=\begin{cases}\dfrac{2x^2-x}{34.5}; & per \quad 1\le x\le 4\\ 0 \quad altrove\end{cases} \qquad M_e=3.2395,\ \mu=3.0869,\ \sigma=0.69547,\ \alpha_1=-0.2194$$

b) Uno dei tanti modelli proposti per migliorare l'adattamento della Pareto alla distribuzione dei redditi è il modello di Lomax:

$$h(x)=\frac{ba}{(1+bx)^{a+1}},\quad x>0;\quad F(x)=1-\frac{1}{(1+bx)^a};$$

Per valori di "a" compresi tra 2 e 3 l'asimmetria è positiva. Se a=3+ε, il numeratore dell'indice di Fisher va all'infinito per ε→0 descrivendo la tendenza alla forma "L" della funzione di densità.

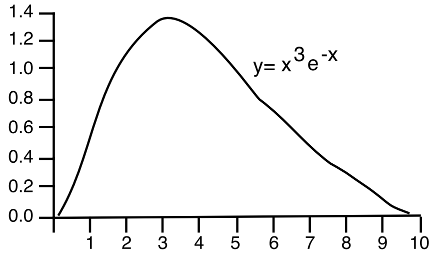
c) Calcolo dell'indice di Pearson per il modello di Erlang

$$h(x)=x^n e^{-bx}\quad x>0;\quad \mu=\int_0^\infty x^{n+1}e^{-bx}dx=\frac{(n+1)!}{b^{n+2}};$$

$$h'(x)=x^{n-1}e^{-bx}(n-bx)\Rightarrow M_o=\frac{n}{b}$$

$$\sigma=\int_0^\infty x^{n+2}e^{-bx}dx-\mu^2=\frac{(n+2)!}{b^{n+3}}-\left[\frac{(n+1)!}{b^{n+2}}\right]^2;$$

$$\alpha_2=\frac{\dfrac{(n+1)!}{b^{n+2}}-\dfrac{n}{b}}{\sigma};\ nullo\ \ se\ \ b=\left[\frac{(n+1)!}{n}\right]^{\frac{1}{n+1}}$$



Per n=3 e b=1 si ottiene a_2=-0.0461 che segnala una leggera asimmetria negativa non troppo in accordo con la figura. L'indice di Pearson non è in effetti considerato affidabile.

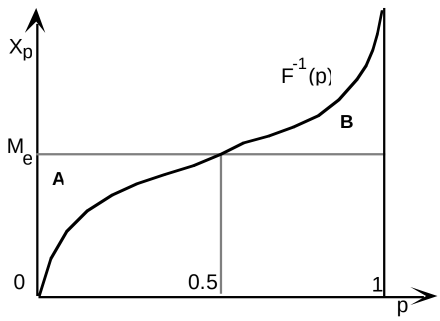
Esercizio_VC137: *calcolare l'indice* α_1 *per il modello esponenziale:*

$$h(x)=\begin{cases}\dfrac{e^{-\frac{x}{5}}}{5} & x>0\ ;\quad F(x)=1-e^{-\frac{x}{5}}\\ 0 & altrove\end{cases}$$

Le misure di asimmetria per variabili continue hanno le stesse finalità e le stesse carenze incontrate nel 3° capitolo. Ad esempio, α_1 è giustamente nullo quando la funzione di densità è simmetrica:

$$
\begin{aligned}
&F(M_e-\mu)+F(M_e+\mu)-1=0 \Rightarrow \int_0^{\infty}\left\{F(M_e-\mu)+F(M_e+\mu)-1\right\}=0\\
&\Rightarrow \int_{-\infty}^{M_e} F(x)dx-\int_{M_e}^{\infty}[1-F(x)]dx \Rightarrow \int_{-\infty}^{M_e}(M_e-x)h(x)dx-\int_{M_e}^{\infty}(M_e-x)h(x)dx=M_e-\mu=0
\end{aligned}
$$

(cfr. van Zwet, 1979). Il contrario non è sempre vero. Il difetto più evidente di α_1, comune ad altri indici di asimmetria, è la non univocità del valore nullo.



L'indice semplice di asimmetria corrisponde infatti alla differenza tra l'area B e l'area A delimitate dalla retta $X_p=M_e$ e dalla curva di graduazione, rapportata alla loro somma (A+B).

$$
\alpha_1=\frac{\int_{0.5}^{1} F^{-1}(p)dp-\frac{M_e}{2}-\left(\frac{M_e}{2}-\int_0^{0.5}F^{-1}(p)dp\right)}{\int_{0.5}^{1} F^{-1}(p)dp-\frac{M_e}{2}+\left(\frac{M_e}{2}-\int_0^{0.5}F^{-1}(p)dp\right)}=\frac{\int_{0.5}^{1} F^{-1}(p)dp+\int_0^{0.5}F^{-1}(p)dp-M_e}{\int_{0.5}^{1} F^{-1}(p)dp-\int_0^{0.5}F^{-1}(p)dp}=\frac{\mu-M_e}{S_{M_e}}
$$

Le due aree possono essere uguali, ma racchiuse da curve differenti (cfr. Groeneveld e Meeden, 1984). Quindi l'indice α_1 è in grado di cogliere solo certi aspetti dell'asimmetria, in particolare quello legato ad uno squilibrio tra centro mediano e baricentro, ma altri sbilanciamenti potrebbero rimanere nascosti.

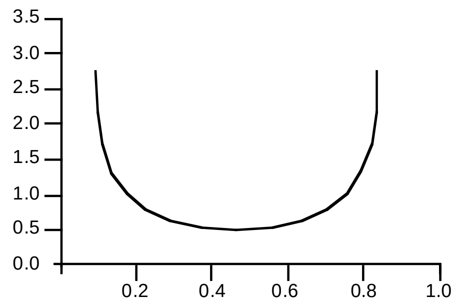
Esempio:

Feller (1950, pp.78-81) propone il modello arcoseno per situazioni in cui si verificano frequenze elevate nei due estremi della distribuzione e basse al centro quali per esempio le durate dei soggiorni turistici. Tale modello ha anche un'interessante interpretazione fisica (Shiryaev, 1996, pp.100-102).

$$h(x)=\frac{1}{\pi\sqrt{x(1-x)}}\quad 0<x<1;\quad F(x)=\left(\frac{2}{\pi}\right)arcoseno\left(\sqrt{x}\right)$$

$$X_p=sen^2\left(\frac{\pi p}{2}\right)\Rightarrow M_e=sen^2\left(\frac{\pi}{4}\right)=\frac{1}{2}$$

$$\mu=\int_0^1 sen^2\left(\frac{\pi p}{2}\right)dp=\frac{2}{\pi}\left\{\frac{x}{4}-\frac{\text{sen}(2x)}{4}\Bigg|_0^{\frac{\pi}{2}}\right\}=\frac{1}{2};\quad S_{M_e}=\frac{1}{\pi};\quad \alpha_1=\pi\left(\frac{1}{2}-\frac{1}{2}\right)=0$$



Esercizio_VC138: *calcolate l'indice semplice di asimmetria per il modello:*

$$F(x)=\sqrt{\frac{x-a}{b-a}}\quad a<x<b$$

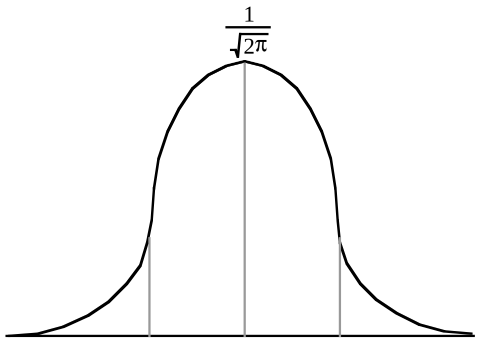
7.3.5 La densità Normale

L'esperienza ha dimostrato che un gran numero di fenomeni fisici, antropologici, biologici, genetici, socioeconomici, geografici, finanziari presenta un poligono delle frequenze di forma campanulare. Fra i modelli in grado di riprodurre questo comportamento ha trovato largo impiego il modello Normale o funzione di Gauss-Laplace:

$$X \sim N(\mu,\sigma) \rightarrow h(x) = \frac{e^{-\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}; \quad -\infty < \mu < \infty; \ \sigma > 0$$

$$\textit{Media aritmetica}\ \mu = \int_{-\infty}^{\infty} x h(x) dx;$$

$$\textit{Deviazione standard}\ \ \sigma = \sqrt{\int_{-\infty}^{\infty} (x-\mu)^2 h(x) dx}$$



L'andamento campanulare e simmetrico della curva Normale sta ad indicare che:

1) Gli scostamenti negativi dal centro sono altrettanto probabili di quelli positivi;
2) I valori sono addensati intorno al centro;
3) Gli scostamenti si verificano con probabilità decrescente man mano che diventano grandi in valore assoluto.

L'importanza di questo modello non risiede solo nel fatto che moltissimi fenomeni possono esservi rappresentati, ma anche in alcuni vantaggi formali che lo rendono il modello più ricorrente in Statistica. Infatti, la cosiddetta densità gaussiana, oltre a presentare caratteristiche convenienti di per sé, serve anche da efficace approssimazione di molte altre funzioni di densità, continue e discrete, note e meno note. In particolare, la media aritmetica di "n" valori presi a caso da una non meglio specificata popolazione di valori tende ad avere distribuzione Normale all'aumentare dell'ampiezza campionaria.

Esempio:
M. Boldrini (1968, p. 731) osserva che ogni volta che un fenomeno appare distribuito normalmente, dall'interpolazione si ricava tutta una catena di conseguenze. Tale l'affermazione dell'uguaglianza presumibile tra lo scarto quadratico medio ottenuto con una rilevazione campionaria e quello ottenibile analizzando -in via teorica- dall'intera popolazione; tale ancora il significato tipico, nel senso queteletiano, della media aritmetica; tale, infine, la conclusione circa la natura puramente accidentale della divergenza dalla media da parte delle singole rilevazioni. Assimilare la curva effettiva del fenomeno alla curva gaussiana equivale a penetrare nel suo intimo determinismo.

Già nel 17° secolo, Galileo discusse il comportamento delle misurazioni delle distanze astronomiche avendo in mente il modello Normale della compensazione tra errori di segno opposto. La sua scoperta è però da attribuirsi al matematico britannico De Moivre che nel 1733 pubblicò una memoria in cui la presentava come forma limite della binomiale. Laplace, un altro nome legato al modello, non mostra di conoscerla prima del 1774. Il matematico svizzero Gauss incontrò il modello affrontando il problema della ricerca del valore di una grandezza di cui si posseggono più misure ugualmente attendibili, ma discordanti e la indicò come la funzione degli errori accidentali ed è conosciuta anche come densità gaussiana o modello gaussiano, sebbene il primo apparire ufficiale della Normale negli scritti di Gauss è del 1809.

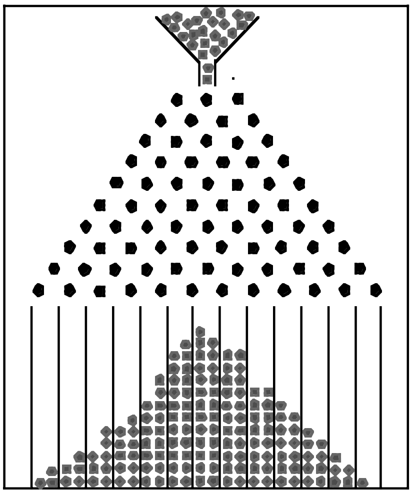
La spiegazione teorica della diffusione di poligonali di forma campanulare poggia in gran parte su di un teorema rientrante nel gruppo noto come teoremi dei limiti centrali, che qui introduciamo solo informalmente:

Lo scostamento, dalla propria media, di un fenomeno che risulti dalla somma di un gran numero di altri fenomeni tra di loro indipendenti e dello stesso ordine di grandezza ha densità Normale.

Si intuisce subito la portata del teorema: interpretando un fenomeno come sovrapposizione di un gran numero di cause prive di legami, ciascuna troppo piccola perché se ne possa seguire l'influenza separatamente dalle altre, è possibile prevedere quale sarà il risultato della loro interazione nelle rilevazioni che saranno effettuate sul fenomeno che mostrerà un poligono di frequenza campanulare.

Esempi:

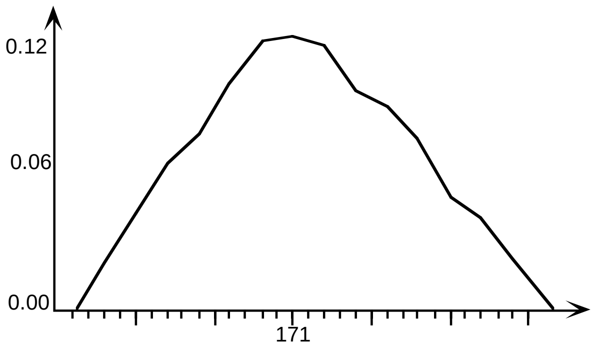
a) Il *quincunx* è una disposizione di cinque punti su si un quadrato o rettangolo di cui quattro si trovano agli angoli ed uno al centro (come il cinque in un dado). In Statistica indica un apparato ideato da Galton alla fine dell'800 che dà una rappresentazione tangibile del teorema del limite centrale. Immaginate un contenitore di legno coperto da *plexiglas* costituito da più parti: un collettore ad imbuto chiuso da una valvola che viene aperta per lasciar passare delle biglie di uguale foggia e materiale nella seconda zona. Le biglie passano una ad una o comunque in modo che le loro traiettorie non si influenzino.



Le biglie, guidate dalla forza di gravità, tentano di raggiungere gli scomparti posti al fondo del *quincunx.* Lungo il tragitto incontrano file regolari di pioli circolari di uguale dimensione che le fanno deviare -con la stessa forza- a sinistra o a destra. Se non ci fossero i pioli, le biglie cadrebbero nelle scanalature centrali fino a colmarle; invece, l'effetto dei pioli distoglie le palline dalla linea retta dirottandole in ogni punto, ma finendo e col raccogliersi secondo una funzione di densità normale.

b) Le stature sono un caso classico di tendenza alla normalità (l'approssimazione delle variabili antropometriche risale al Quetelet cioè alla prima metà dell'800). Ecco i dati ripresi da Frosini 1995, p. 224. L'andamento della distribuzione mostra unimodalità e simmetria richiesti per l'approssimazione normale. Peraltro, pur essendo istintivamente improponibile un dominio infinito per la variabile "altezza di una persona", si potrebbe comunque usare l'asse reale nella certezza che oltre un certo livello inferiore o superiore le frequenze saranno nulle o quasi tali.

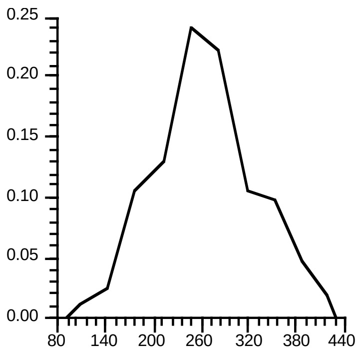
Altezza		Adulti			
0	≤157	51	171	173	225
157	159	41	173	175	188
159	161	84	175	177	174
161	163	126	177	179	148
163	165	152	179	181	96
165	167	194	181	183	79
167	169	230	183	185	45
169	171	234		≥185	65
					2132



c) Il Superenalotto coinvolge sei primi estratti del gioco del lotto. Si è rilevata la somma degli estratti per n=171 concorsi.

380	98	132	136	149	156	161	168	169	170	173
323	193	193	193	199	202	206	206	207	208	208
165	200	282	282	283	284	284	285	285	287	287
194	228	298	298	299	304	306	307	309	311	317
172	216	334	336	340	342	342	345	349	349	350
193	227	225	227	230	230	232	232	233	234	237
169	209	249	249	250	250	251	253	254	255	255
192	212	262	262	263	265	265	266	266	269	270
191	215	379	379	380	386	386	388	408	409	413
176	199	290	325	354	238	258	273	182	214	351
184	219	293	325	359	240	258	275	257	273	353
188	214	294	327	360	241	258	275	288	324	
179	208	295	331	366	242	258	278	256	272	
189	224	296	332	368	243	260	278	179	208	
340	198	296	332	369	246	261	279	323	351	
401	121	255	272	210	325	181	238	237	288	

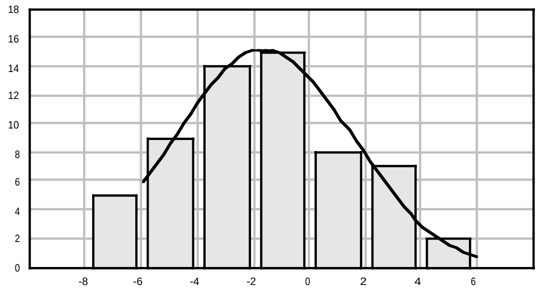
Totale		Estr.	f
90	125	2	0.012
126	160	4	0.023
161	195	18	0.105
196	230	22	0.129
231	265	41	0.240
266	300	38	0.222
301	335	18	0.105
336	370	17	0.099
371	405	8	0.047
406	430	3	0.018
		171	1.000



Sebbene le condizioni del teorema limite centrale non siano tutte rispettate (l'indipendenza c'è, ma 6 "cause" sono poche), il poligono di frequenza mostra i segni della normalità.

Le condizioni di applicabilità della Normale sono generali, ma non elastiche al punto da coprire ogni situazione di indagine. Ad esempio, se l'errore di misurazione, cresce con il numero delle misurazioni effettuate, allora è più adatto il modello lognormale. Williams (1984, P. 92) sostiene: *"... il termine "normale" è male applicato perché fuorviante. Le distribuzioni Normali non sono affatto normali in Statistica"*.

Esercizio_VC139: *si è rilevato il peso di 60 lotti di prodotti in cui ogni lotto conteneva 35 item. Le misure sono state arrotondate con la regola del 5. In tabella sono riportati gli errori di approssimazione complessivi per lotto. E' inoltre riportato l'esito dell'adattamento della Normale ai dati.*



-7.2	-4.8	-1.7	-3.1	-0.7	1.6
-6.7	-4.8	-1.6	-3.1	-0.5	2.2
-6.7	-4.8	-1.5	-2.9	-0.1	2.3
-6.6	-4.6	-1.4	-2.8	0.3	3.4
-6.5	-3.8	-1.4	-2.7	0.3	3.4
-5.6	-3.7	-1.3	-2.7	0.6	3.5
-5.5	-3.6	-1.3	-2.2	0.8	3.5
-5.3	-3.5	-1.2	-2.2	1.3	3.6
-5.1	-3.5	-1.0	-1.9	1.5	4.2
-5.0	-3.3	-0.8	-1.8	1.5	5.5

a) Vi sembra che ricorrano le condizioni per l'approssimazione normale?
b) Che cosa implica il passaggio dalle frequenze relative alle probabilità?

Morfologia della normale

La normale, indicata con N(μ,σ), è continua sull'intero asse reale ed è dotata da due parametri: μ e σ che le danno notevole flessibilità di rappresentazione e capacità di adattamento. Analizziamone le caratteristiche salienti.

1. La non negatività della funzione di densità è evidente dato che si tratta di un esponenziale e che si è ipotizzato σ>0. Inoltre, l'area sottesa alla curva è pari ad uno. Per verificarlo standardizziamo la variabile:

$$\int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} d\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}dz$$

Accertiamo preliminarmente che il quadrato dell'ultimo integrale valga uno.

$$A = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}dz \Rightarrow A^2 = \left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}dz\right]\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{w^2}{2}}dw\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{z^2+w^2}{2}}dzdw$$

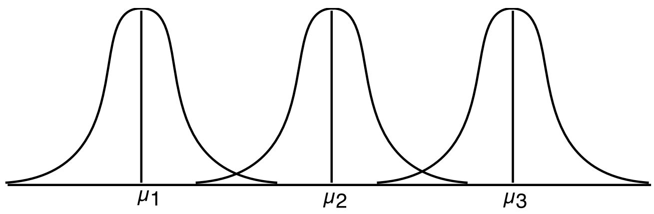
Se ora si passa alle coordinate polari z=ycos(t), w=ysen(t), si ottiene: $y^2=z^2+w^2$ con y∈[0,∞) e t∈[0,2π]

$$\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{-\frac{y^2}{2}}ydydt = \frac{1}{2\pi}2\pi\int_0^{\infty} e^{-\left(\frac{y^2}{2}\right)}d\left(\frac{y^2}{2}\right) = -e^{-\left(\frac{y^2}{2}\right)}\Bigg|_0^{\infty} = 1 \Rightarrow A = 1$$

2. L'asse delle ascisse è un evidente asintoto sia positivo che negativo.
3. La funzione di densità è unimodale con massimo raggiunto nel punto x=μ dato che, per x che cresce da -∞ a μ, la densità aumenta da zero a 1/√2π, per poi diminuire da tale picco fino a zero al variare di x da μ ad ∞.
4. La funzione di densità è simmetrica:

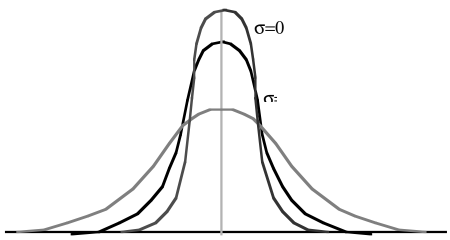
$$h(\mu+\varepsilon) = \frac{e^{-\frac{1}{2}\left(\frac{\mu+\varepsilon-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}};\; h(\mu-\varepsilon) = \frac{e^{-\frac{1}{2}\left(\frac{\mu-\varepsilon-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \Rightarrow h(\mu+\varepsilon) = h(\mu-\varepsilon)$$

La natura di σ è altrettanto esplicita dato che costituisce lo scarto quadratico medio del modello. La semplicità ed univocità di interpretazione dei parametri è uno dei pregi del modello normale.



Al variare di μ il grafico resta inalterato nella sua forma complessiva. Si modifica solo la sua localizzazione: più a destra se μ aumenta, più a sinistra se μ diminuisce. Si deve ricordare che, essendo la densità simmetrica, μ, oltre che di moda e di media aritmetica, ha anche il ruolo anche di mediana.

Al diminuire di σ e a parità di media i due punti di flesso (in corrispondenza delle ascisse: $X=\mu \pm \sigma$) tendono ad accentrarsi e l'ordinata massima aumenta: l'effetto è un graduale assottigliarsi ed appuntirsi della funzione a causa del maggiore addensamento al centro della distribuzione.



L'esiguità delle code ha inoltre come conseguenza l'esistenza dei momenti di ogni ordine che è coerente con la l'idea di sorte "benigna" nel senso di Mandelbrot che accompagna questo modello.

Esercizio_VC140: *dimostrare che se X ha funzione di densità normale N(μ,σ) allora Y=a+bX è pure normale con E(Y)=a+bμ e Var(y)=$(b\sigma)^2$.*

La funzione di ripartizione della Normale

La funzione di ripartizione corrispondente alla densità normale è:

$$F(x;\mu;\sigma)=\int_{-\infty}^{x}\frac{e^{-\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}dx=\phi(x)=\int_{-\infty}^{x}\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}dz \quad dove \quad z=\frac{x-\mu}{\sigma}$$

La F(.) dipende da μ e da σ ed il calcolo dovrebbe essere ripetuto per ogni combinazione dei due parametri. In realtà, come mostra il cambio di variabile, si può operare con la sola normale standardizzata: $Z=(x-\mu)/\sigma$ che ha $\mu=0$ e $\sigma^2=1$. L'integrale, anche nella versione standardizzata, non ha una soluzione esplicita ed il calcolo delle aree avviene con tecniche di integrazione numerica (in verità, piuttosto precise). Lo stesso vale per la funzione di graduazione $Z_p=\phi^{-1}(z)$. Nel formulario allegato al testo è dato uno stralcio della tavola delle aree sottese alla normale, ma le formule per il loro calcolo, come quelle per i quantili, si trovano inserite anche nelle calcolatrici tascabili oltreché sui fogli elettronici più diffusi.

Esempi:

a) Area sottesa su $(-\infty,1.83)$ cioè $\phi(1.83)$. Cerchiamo la riga che corrisponde alla parte intera ed alla prima cifra significativa: 1.8; scorriamo da sinistra verso destra fino a trovare la colonna intestata con il secondo decimale: 0.03. All'incrocio troviamo la misura dell'area: 0.9664.

b) Calcolo dell'area a destra di 0.72. Poiché l'area totale è uno si ha l'identità: area a sinistra=1-area a destra e quindi: area a destra di 0.72=1-ϕ(0.72)=1-0.7642=0.2358.

c) Calcolo dell'area per ascisse negative: ϕ(-1.44). La curva normale è simmetrica per cui l'area a sinistra di -1.44 è uguale all'area destra: ϕ(-1.44)=1-ϕ(1.44)=1-0.9251=0.0749.

d) Area compresa tra due valori: (-0.8,1.4): ϕ(1.4)-ϕ(-0.8)=0.9192-[1-ϕ(0.8)]=0.9192-0.2119=0.7073.

e) Area esterna a due valori:$(-\infty,-1.91)\cup(2.12,\infty)$: ϕ(-1.91)+[1–ϕ(2.12)]=1-0.9719+1-0.9826=0.0455.

Esercizio_VC141: *calcolare l'area sottesa alla curva normale per gli intervalli:*
a) $|Z - 0.5| \geq 1$; b) $Z \geq -2$; c) $(1-Z) \leq 1.5$; d) $|Z| \leq 2.38$

Approssimazione aree della normale
Per estendere le tavole o per calcolare i valori non presenti nella tabella sono disponibili varie approssimazioni:

$$\textit{Formula di Page:}\ \varphi(z) \cong \left[1+e^{-2y}\right]^{-1},\ y = \sqrt{2/\pi}\left(1+0.044715z^2\right)z$$

con un errore inferiore a 0.0002 per $Z \in (0,4)$ (Patel e Read, 1996, Cap. 3). Un'approssimazione accurata (errore in valore assoluto inferiore a 7.5x10^{-8}) è la formula polinomiale data in Abramowitz e Stegun (1971, p. 932).

$$\phi(z) = 1 - \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\left[t\left(b_1 + t\left(b_2 + t\left(b_3 + t\left(b_4 + b_5 t\right)\right)\right)\right)\right],\ t = \frac{1}{1+0.2316419z}$$

$$b_1 = 0.3193\,8153,\ b_2 = -0.3565\,63782,\ b_3 = 1.7814\,77937,\ b_4 = -1.8212\,555918,\ b_5 = 1.3302\,74429$$

Esempi:
a) Ipotizziamo che la spesa delle famiglie segua un modello gaussiano con N(1'940,1'155). Calcoliamo le frazione di famiglie che ha una spesa compresa tra 1'250 e 2'250.

$$\textit{Tavole}:\ \phi\left(\frac{2250-1940}{1155}\right) - \phi\left(\frac{1250-1940}{1155}\right) = \varphi(0.27) - \phi(-0.60) = 0.6064 - 1 + 0.7257 = 0.3321$$

$$\textit{Approssimazione}: \phi(0.27) = 0.6064198,\ \phi(0.60) = 0.7257469$$

con valori pressoché coincidenti.

b) Il peso alla nascita dei neonati è distribuito normalmente con media 3200g e deviazione standard 800g. Se le nascite sottopeso sono quelle inferiori a 2300g, qual'è la probabilità che tale evento si verifichi? Circa il 13% secondo la formula di Page:

$$\phi\left(\frac{2300-3200}{800}\right) = \phi(-1.125) = 1 - \phi(1.125) = 1 - 0.869533 = 0.13047$$

c) Tra le donne di età compresa nell'intervallo (30-35) ed alle quali è stato diagnosticato un cancro ai polmoni la distribuzione del periodo di fumo di sigarette è normale con media μ=8.35 anni e deviazione standard di 2.71 anni. Qual'è la probabilità (approssimazione polinomiale) che una donna in questo gruppo abbia fumato per più di 15 anni?

$$1 - \phi\left(\frac{15-8.35}{2.71}\right) = 1 - \phi(2.454) = 1 - 0.992936 \cong 0.007$$

Esercizio_VC142: *approssimare con la formula polinomiale l'area sottesa alla curva normale standardizzata in corrispondenza di: a) Z=1.234, b) Z=4.012, c) Z=-1.319.*

Esercizio_VC143: *anche per i quantili della gaussiana esistono buone approssimazioni. Verificate quella di Abramowitz e Stegun (1971, p. 933) determinando il 1° quartile e l'ultimo ventile della variabile standardizzata.*

$$Z_p = \frac{c_0 + t\left(c_1 + c_2 t\right)}{1 + t\left(d_1 + t\left(d_2 + d_3 t\right)\right)};\ t = \sqrt{-2Ln(p)};\quad per\ \ 0 < p \leq 0.5$$

$$c_0 = 2.515517,\ c_1 = 0.802853;\ c_2 = 0.010328;\ d_1 = 1.432788,\ d_2 = 0.189269,\ d_3 = 0.001308$$

In taluni casi è necessario determinare il quantile che delimita un ammontare prefissato di probabilità nella Normale N(μ,σ). E' facile verificare che:

$$p(|x-\mu|>r)=\alpha \Rightarrow r=\sigma\phi^{-1}\left(1-\frac{\alpha}{2}\right);\ \ p(X>X_\alpha)=\alpha \Rightarrow X_\alpha=\mu+\sigma\phi^{-1}(1-\alpha);\ \ p(X<X_\alpha)=\alpha \Rightarrow X_\alpha=\mu+\sigma\phi^{-1}(\alpha);$$

Esempi:

a) La misura di un *item* ha distribuzione normale con media $\mu=5$ e deviazione standard $\sigma=0.005$ se si decide di scartare il 3% dei valori più piccoli ed il 3% di quelli più grandi, quali sono le soglie a cui ancorare le scelte?

$$p(|x-5|>r)=0.94 \Rightarrow r=0.005*\phi^{-1}\left(1-\frac{0.06}{2}\right)=0.005*\phi^{-1}(0.97)=0.005*1.881=0.0094;\ \ Soglie: 5\pm 0.0094$$

b) Gli assegni emessi per pagare le forniture della MacBartel hanno media 9 mila euri e deviazione standard 2.5 mila euri. Si sospetta che i pagamenti maggiori servano a costituire fondi neri. Il Fisco decide di controllare le fatture con importi superiori ad una certa soglia. Determinare tale soglia tenendo conto che non si può operare su più dell'8.5% delle fatture.

$$p(X>r)=0.915 \Rightarrow r=9+2.5*\phi^{-1}(0.915)=9+2.5*1.372=12.43$$

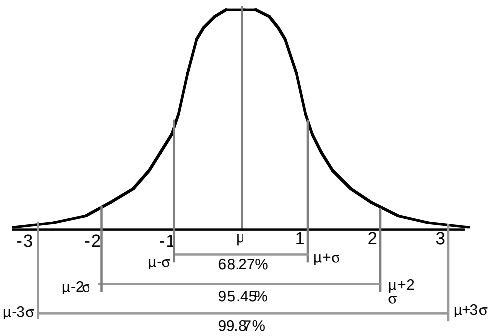
Esercizio_VC144:

a) Il contenuto di principio attivo in un farmaco è N(0.2,0.01). Quali sono le soglie (simmetriche) minime e massime che racchiudono il 96.8% delle confezioni?

b) *Il tempo in minuti del volo Sibari-Milano è ben descritto dal modello N(65,7). Qual'è il tempo da indicare negli avvisi pubblicitari se si intende assicurare che il 99% dei voli arriverà in orario.*

Criterio degli intervalli tipici

La prossimità di una distribuzione -empirica o teorica- a quella normale può essere valutata grazie alla frequenza relativa o alla probabilità che ricade in intervalli simmetrici del tipo: [μ - kσ; μ + kσ] con k=1,2,3. Nella versione standardizzata si riscontrano le percentuali indicate in figura.
La disuguaglianza di Tchebycheff stabilisce che entro i tre intervalli tipici siano incluse le frazioni minime: 0%,75% e 89% che sono molto più lasche. L'approssimazione con la Normale è accettabile se c'è unimodalità, simmetria e se nella distribuzione circa il 68%, il 95% ed il 99% delle modalità è compreso, rispettivamente, entro 1, 2 e 3 volte σ.

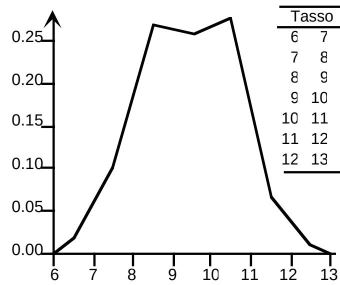


Esempi:

a) Tasso di crescita 1987-1991 del valore aggiunto netto per dipendente (banche del Centro-Nord).

6.06	6.31	7.30	7.38	7.43	7.53	7.68	7.69	7.70	7.89	7.93	7.94	7.99	8.06	8.11	8.15	8.19	8.21
8.28	8.35	8.40	8.45	8.47	8.50	8.56	8.59	8.62	8.71	8.71	8.71	8.72	8.74	8.76	8.80	8.81	8.85
8.87	8.91	8.91	8.93	8.94	8.97	9.04	9.09	9.11	9.13	9.17	9.19	9.19	9.22	9.23	9.31	9.46	9.49
9.53	9.53	9.57	9.58	9.60	9.63	9.77	9.83	9.85	9.88	9.90	9.93	9.94	9.96	9.96	9.97	10.01	10.03
10.03	10.04	10.06	10.07	10.09	10.14	10.18	10.20	10.24	10.28	10.35	10.35	10.40	10.41	10.44	10.51	10.53	10.54
10.58	10.65	10.68	10.80	10.83	10.84	10.84	10.85	10.93	10.96	11.26	11.30	11.36	11.61	11.66	11.82	11.99	12.50

Intervallo	*Modalità corrispondenti*	*frequenza relativa*	*Normale*
$\mu \pm \sigma \Rightarrow$	$8.24 \le X \le 10.64$	$73/108 = 67.59\%$	68.27%
$\mu \pm 2\sigma \Rightarrow$	$7.04 \le X \le 11.84$	$104/108 = 96.30\%$	95.45%
$\mu \pm 3\sigma \Rightarrow$	$5.84 \le X \le 13.04$	$108/108 = 100.00\%$	99.73%



Tasso		Banche
6	7	2
7	8	11
8	9	29
9	10	28
10	11	30
11	12	7
12	13	1
		108

Le percentuali dei tassi di crescita sono simili a quelle previste dalla Normale. Il poligono di frequenza non sembra smentire la normalità anche se, una migliore scelta delle classi centrali, avrebbe potuto eliminare la sgradevole doppia cuspide in mezzo al grafico.

b) La caduta di neve in una località sciistica è una variabile casuale con valore atteso μ=55 cm e varianza σ^2=2100. Se fosse approssimabile con la normale, quale sarebbe la probabilità che nella prossima stagione invernale la neve superi i 120 cm?

$$p\left(\frac{X-55}{\sqrt{2100}}>\frac{120-55}{\sqrt{2100}}\right)=1-P(Z\le 1.42)\cong 1-\Phi(1.42)=7.78\%$$

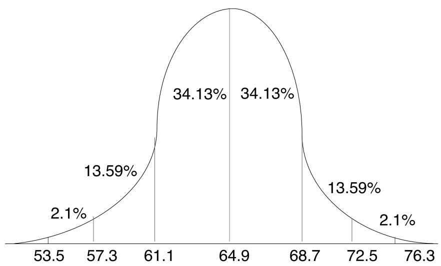
Tale evento (che significherebbe blocco delle strade e isolamento) nelle ultime 45 stagioni non si è mai verificato. Se ne può dedurre che il modello normale è inadeguato? Un evento che ha probabilità costante del 7% di verificarsi su 45 prove genera una media di 3.15 successi con scarto quadratico medio di 1.71. Si può ragionevolmente ritenere che il 95% dei casi è compreso tra 0 e 6 prove e quindi lo zero (assenza di successi) non è un fatto anomalo.

c) L'esito valido del lancio del peso nelle gare ufficiali è ben approssimato da una variabile casuale normale con valore atteso 18m e scarto quadratico medio 2m. Le prove realmente eccezionali sono quelle che realizzano misure collocate nell1% inferiore e nell'1% superiore. Quali sono le soglie?

$$p\left(\frac{X-18}{2} > U\right) = 0.01 \Rightarrow 1 - \Phi(U) = 0.01 \Rightarrow \Phi(U) = 0.99 \Rightarrow U = 22.653;\quad p\left(\frac{X-18}{2} < L\right) = 0.01 \Rightarrow \Phi(L) = 0.01 \Rightarrow L = 13.347$$

Il criterio degli intervalli tipici non è solo un riscontro della normalità di una densità od i un poligono delle frequenze, ma è anche una regola empirica per integrare le rilevazioni empiriche. Se abbiamo già calcolato media aritmetica e scarto quadratico medio e, a nostro parere, ricorrono le condizioni del teorema limite centrale, allora le due informazioni: μ e σ, possono essere ulteriormente valorizzate.

Esempio:
Da indagini molto estese risulta che la distanza interpupillare di persone di sesso maschile di età superiore ai 18 anni ha media μ=64.9mm e scarto quadratico medio σ=3.8mm. Un settimanale di larga diffusione vuole inserire un paio di occhiali come omaggio in un numero speciale della rivista. La direzione ignora i dati di dettaglio, ma ritiene che la distribuzione sia di tipo normale. Verso quali misure dovrà orientarsi?



Il criterio degli intervalli tipici pone la direzione di fronte al quadro riportato in figura. Sono individuate sei classi di distanza interpupillare che possono servire a confezionare la rivista in modo da scontentare meno clienti possibili. Quindi, combinando dei dati empirici ed una congettura si ricostruisce -con pochissimo sforzo- un'intera rilevazione ovvero si proietta su di un'indagine i risultanti ottenuti in un'altra indagine più piccola.

Esercizio_VC145: *importo complessivo incassato da una linea di ristoranti per giorni di apertura negli ultimi tre anni. Dati in migliaia di euri.*

X		n
0	60	13
60	90	39
90	120	69
120	150	156
150	200	238
200	250	129
250	300	107
300	350	64
350	400	11
		826

a) Costruire il poligono di frequenza; b) Verificare l'approssimazione gaussiana confrontando le frequenze assegnate dal modello con quelle effettivamente rilevate negli intervalli tipici.

Esercizio_VC146: *diversi autori hanno cercato dei modelli analitici simili alla gaussiana, ma con una funzione di ripartizione più trattabile. Verificate la qualità di quello proposto da Burr (1967) con gli intervalli tipici.*

$$\phi(x) \cong F(x) = \begin{cases} 1 - \left[1 + (0.644693 + 0.161984x)^{4.874}\right]^{-6.158} & se \quad x > -3.97998 \\ 0 & se \quad x < -3.97998 \end{cases}$$

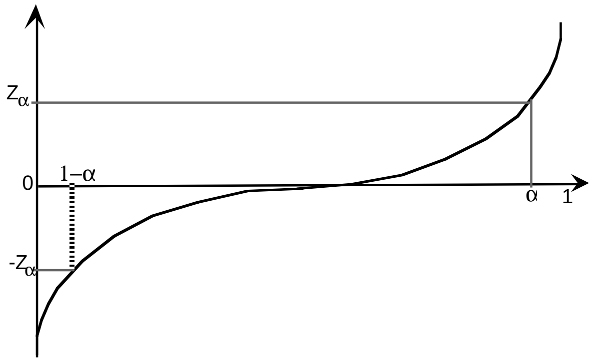
Suddivisione in classi

Una interessante applicazione del modello normale è la suddivisione in classi di una rilevazione. L'idea è che esista un quadro di riferimento ben definito che permetta di stabilire a priori le percentuali di unità incluse in ciascuna classe e che fra una classe e l'altra esista una demarcazione netta, valida in ogni occasione. Il problema consiste nel calcolare le soglie della suddivisione in intervalli contenenti la frazione prestabilita di unità.

Esempio.
Un campione di studenti è suddiviso in *cinque* gruppi:

Gruppi	**A**	**B**	**C**	**D**	**E**
Contenuto	5%	20%	50%	20%	5%
Giudizio	Distratti	Scarsi	Sufficienti	Buoni	Ottimi
Percentile	$X_{0.05}$	$X_{0.25}$	$X_{0.50}$	$X_{0.75}$	$X_{0.95}$

Occorre quindi determinare il primo e l'ultimo ventile nonché il primo ed il terzo quartile. Si supponga che il voto medio della classe sia stato 24.5 con scarto quadratico medio 2.5. Il calcolo dei quantili del modello gaussiano non è semplice in quanto la funzione di ripartizione non può essere espressa in forma esplicita (l'approssimazione in VC143 è comune ottima). Di seguito è riportata uno stralcio per i percentili più ricorrenti.



α	Z_α	α	Z_α	α	Z_α	α	Z_α
0.50	0.00	**0.910**	1.34	**0.960**	1.75	**0.975**	1.96
0.55	0.13	**0.920**	1.41	**0.962**	1.77	**0.976**	1.98
0.60	0.25	**0.930**	1.48	**0.964**	1.80	**0.977**	2.00
0.65	0.39	**0.940**	1.56	**0.966**	1.83	**0.978**	2.01
0.70	0.52	**0.950**	1.64	**0.970**	1.88	**0.980**	2.05
0.75	0.67	**0.952**	1.66	**0.971**	1.90	**0.990**	2.33
0.80	0.84	**0.954**	1.68	**0.972**	1.91	**0.995**	2.58
0.85	1.04	**0.956**	1.71	**0.973**	1.93	**0.999**	3.09
0.90	1.28	**0.958**	1.73	**0.974**	1.94		

Per i livelli indicati di α la tabella fornisce il quantile Z_α normale standardizzata. Ad esempio, la modalità che qui supera il 95% ed è superata dal 5% delle altre modalità è 1.64. I quantili inferiori al 50% non sono riportati perché ricavabili da: $Z_{1-\alpha} = -Z_\alpha$ (per la simmetria della funzione di densità). Per il quantile X_α di una normale qualsiasi occorre effettuare la trasformazione: $X_\alpha = \mu + \sigma Z_\alpha$.

$Z_{0.05} = -Z_{0.95} = -1.64$; $X_{0.05} = 24.5 - 1.64*2.5 = 20.4$
$Z_{0.25} = -Z_{0.75} = -0.67$; $X_{0.25} = 24.5 - 0.67*2.5 = 22.8$
$Z_{0.75} = 0.67$; $X_{0.75} = 24.5 + 0.67*2.5 = 26.2$
$Z_{0.95} = 1.64$; $X_{0.95} = 24.5 + 1.64*2.5 = 28.6$

La risultante suddivisione è dunque:

Voti	Giudizi
< 20.4	Distratti
da 20.4 a 22.8	Scarsi
da 22.8 a 26.2	Sufficienti
da 26.2 a 28.6	Buoni
> 28.6	Ottimi

Esercizio_VC147: *il consumo di elettricità dovuto ad un nuovo tipo di lampada da 100 watt in un'ora di accensione ininterrotta ha andamento gaussiano con media μ=300 e deviazione standard 24. E' considerato difettoso il primo 5% di lampade e di qualità bassa il successivo 10%. Quali sono le durate massime che individuano tali soglie? Sugg. Adoperate l'EXCEL per il calcolo (Funzioni statistiche, INV.NORM oppure INV.NORM.ST).*

Una carenza grave di questo metodo è che uno stesso giudizio può richiedere risultati diversi in prove analoghe con la sensazione, per gli esclusi, non di essersi prodotti in una prova scadente, ma di essere capitati nel gruppo sbagliato. Tuttavia, se i parametri sono stimati con una rilevazione numerosa, l'applicazione è corretta.

Esempio:
Cox (1957) dimostra che la scelta più efficace delle soglie di suddivisione del campo di variazione della normale sono quelle riportate in tabella e che sono che sono diverse da quelle fornite dagli intervalli tipici.

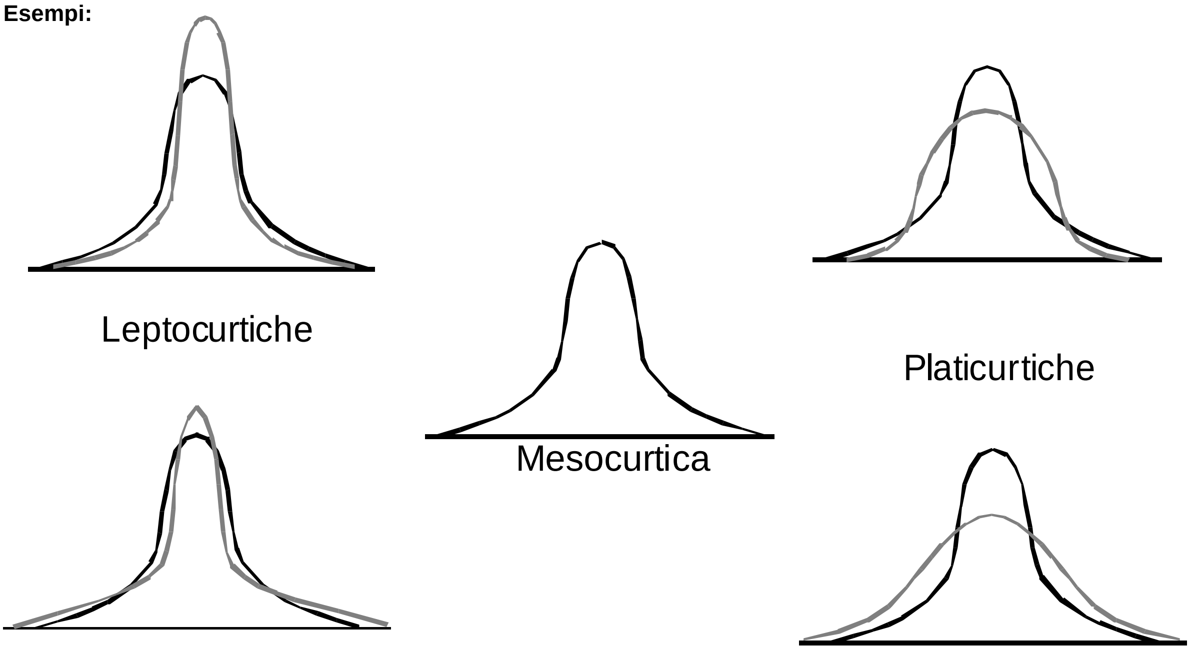
Classi	Soglie			Percentuali					
2	0			50.0	50.0				
3	±0.612			27.0	45.9	27.0			
4	±0.980	0		16.4	33.6	33.6	16.4		
5	±1.230	±0.395		10.9	23.7	30.7	23.7	10.9	
6	±1.449	±0.660	0	7.4	18.1	24.5	24.5	18.1	7.4

Esercizio_VC148: *nel valutare un programma di premi-fedeltà a sorteggio, la direzione di un supermercato ha raccolto i dati sui clienti per importi acquistati nell'ultimo mese. La loro distribuzione è risultata di forma quasi gaussiana. Il 10% dei clienti che acquista meno o il 20% che più spende non hanno bisogno di rinforzi psicologici cosicché i premi fedeltà sono dati ai clienti "tiepidi". Quali sono le soglie?*

La curtosi

Un altro aspetto che a volte compare nello studio della morfologia delle distribuzioni è la curtosi cioè la deformazione cui deve essere sottoposta una densità simmetrica e unimodale per sovrapporsi ad una curva gaussiana di pari media e pari scarto quadratico. La ragione di questo concetto e della relativa misura è la grande importanza attribuita alla somiglianza con la curva gaussiana (Balanda e MacGillivray, 1988).

Esempi:



Una densità mesocurtica ha picco centrale e code simili alla normale; la leptocurtica (o iponormale) presenta ordinate più alte al centro (la sommità è più acuminata: *leptòs* significa "stretto"); la platicurtica (o ipernormale) hanno ordinate più basse al centro (sommità più appiattita: *platùs* significa "ampio"). In entrambi i casi le code possono essere sia più spesse e allungate che più sottili e brevi rispetto alla normale.

Per accertare l'entità dello scostamento dalla normalità esistono diversi indicatori. La misura più nota è:

$$\gamma_2 = \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^4 h(x)dx - 3; \quad con \ \ \gamma_2 \begin{cases} = 0; & mesocurtica \\ < 0; & platicurtica \\ > 0; & leptocurtica \end{cases}$$

proposto da Fisher. Il γ_2 non varia rispetto a trasformazioni di scala; non è però normalizzato. Per distribuzioni platicurtiche il γ_2 è tendenzialmente negativo ed è tendenzialmente positivo per distribuzioni leptocurtiche (è prossimo allo zero per curve mesocurtiche). L'avverbio "tendenzialmente" avverte che il segno di γ_2 non sempre dà indicazioni corrette sull'aspetto della densità e ciò è dovuto alla sua intrinseca natura di media che può far scomparire anche consistenti difformità qualora queste si bilancino nell'arco della distribuzione. Infine, l'indice presuppone finito il 4° momento ed è quindi applicabile solo a modelli con code sottili laddove la presenza di code spesse e dei connessi valori remoti è proprio una delle cause più frequenti di curtosi.

Esempi:

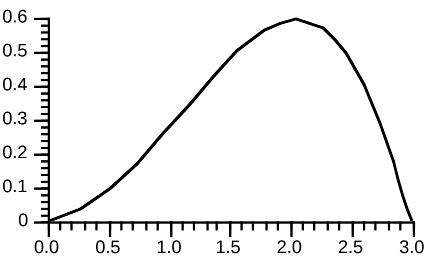
a) Densità su supporto limitato.

$$h(x) = \frac{4(3-x)x^2}{27}, \quad 0 \le x \le 3$$

$$\mu_1 = 1.8,\ \mu_2 = 3.6,\ \mu_3 = 7.71,\ \mu_4 = 17.6 \quad \mu'_2 = 3.6 - 1.8^2 = 0.36$$

$$\mu'_4 = 17.6 - 4*1.8*7.71 + 6*1.8^2*3.6 - 3*1.8^4 = 22.112$$

$$\gamma_2 = \frac{\mu'_4}{\sigma} - 3 = \frac{22.12}{\sqrt{0.36}} - 3 = 167.62$$



il modello risulta fortemente leptocurtico anche se dal grafico non si direbbe. Il valore sorprendente dell'indice è forse dovuto alla ridotta variabilità quale risulta dalla deviazione standard.

b) La conoscenza di γ_2 rende più stringente la disuguaglianza di Tchebycheff. David e Barton (1962, pp.54-55) segnalano che:

$$P\left(|x-\mu|\geq k\sigma\right)\leq\frac{\mu_4-1}{\mu_4-1+\left(k^2-1\right)^2}$$

Se poniamo $\mu_4=\sigma(3+\gamma_2)=3$, all'esterno dell'intervallo (μ-2σ,μ+2σ) si troverà alpiù il 18.2%. Tale indicazione prescinde dalla simmetria/asimmetria della funzione di densità.

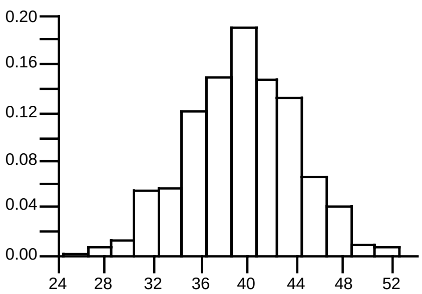
***Esercizio_VC149:** determinare il coefficiente di curtosi per la densità uniforme sull'intervallo unitario:*

$$h(x)=\begin{cases}1 & 0\leq x\leq 1\\ 0 & altrove\end{cases}$$

Lo studio della curtosi è applicabile anche ai poligoni di frequenza purché abbiano forma campanulare e sostanziale simmetria. La finalità più usuale è di accertare l'adattamento del modello gaussiano ai fini del trattamento teorico delle statistiche campionarie. E' noto che media e scarto quadratico medio sono stime appropriate dei corrispondenti valori della popolazione se la distribuzione è di tipo normale o moderatamente platicurtica, ma danno stime inadeguate se è leptocurtica (Cleveland, 1993, p. 72).

Esempio:
Un campione di 300 esemplari di una specie sono stati sottoposti ad un test di sopravvivenza. I giorni di vita sono riportati in tabella.

L_i	U_i	n_i
24.5	26.5	1
26.5	28.5	2
28.5	30.5	4
30.5	32.5	16
32.5	34.5	17
34.5	36.5	37
36.5	38.5	45
38.5	40.5	57
40.5	42.5	44
42.5	44.5	40
44.5	46.5	20
46.5	48.5	12
48.5	50.5	3
50.5	52.5	2
		300



L'istogramma mostra una buona vicinanza alla normale. Calcoliamo l'indice di Fisher e l'indice di curtosi per trovare una ulteriore conferma: γ_1=-0.115, γ_2=-0.067.

Per aggirare il problema dello scollamento tra i valori dell'indice di curtosi e la forma della distribuzione Hogg (1974) ha proposto un ulteriore indice basato sul confronto di medie parziali:

$$KH=\frac{\mu_1^-\left(x_{0.2}\right)-\mu_1^+\left(x_{0.2}\right)}{\mu_1^-\left(M_e\right)-\mu_1^+\left(M_e\right)}$$

E' facile verificare che la misura è invariante rispetto a trasformazioni lineari, ma molto sensibile ai valori estremi (questo non è necessariamente un male); comunque lo è meno del γ_2 (Ruppert, 1987).

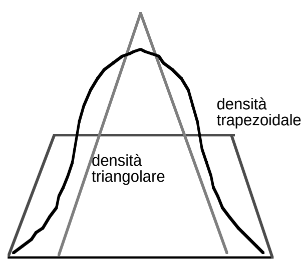
Esempio:

Per il modello esponenziale $F(x)=1-e^{-\lambda x}$, l'indice vale: $KH=2.5\dfrac{1-2e^{-\lambda x_{0.2}}\left(1-x_{0.2}\right)}{1-2e^{-\lambda M_e}\left(1-M_e\right)}=2.16\lambda-0.8047$

risultando crescente rispetto a λ cioè all'aumentare dello spessore della coda.

Problemi di definizione

Il problema con la curtosi è che con essa si tenta di descrivere un effetto composto da due elementi distinti: il grado di appuntimento al centro e lo spessore delle code. Per raggiungere lo stesso grado di appuntimento della densità triangolare, la normale dovrà essere tirata verso l'alto e assottigliata nei fianchi con solo leggere modifiche nelle code. Per portare le code della normale allo spessore della densità trapezoidale si possono allargare i fianchi e rigonfiare le code senza praticamente intervenire sulla sommità. In breve, la coda e la cupola dovrebbero essere trattate separatamente. (Cfr. Balanda K.P. MacGillivray H.L. 988).



L'effetto congiunto di questi due fattori può portare a valori di γ_2 incoerenti con l'aspetto della distribuzione. Per superare tale carenza Groeneveld e Meeden (1984) e Groeneveld (1998) hanno proposto degli indici basati sull'idea che se X ha distribuzione simmetrica con mediana M_e allora l'asimmetria della variabile fittizia creata dagli scarti assoluti dalla mediana $|X-M_e|$ può servire come misura della curtosi di X. In particolare si può utilizzare un indice tipo Yule-Bowley limitatamente alla metà superiore della distribuzione:

$$GM_p = \frac{\left(X_{1-\frac{p}{2}} - Q_3\right) - \left(Q_3 - X_{\frac{1+p}{2}}\right)}{\left(X_{1-\frac{p}{2}} - Q_3\right) + \left(Q_3 - X_{\frac{1+p}{2}}\right)} = \frac{X_{1-\frac{p}{2}} + X_{\frac{1+p}{2}} - 2Q_3}{X_{1-\frac{p}{2}} + X_{\frac{1+p}{2}}};\;\; 0 < p < 0.5$$

GM_p assume valori nell'intervallo [-1,1] ed è invariante rispetto a trasformazioni lineari. Valori grandi e positivi sono associati a curve leptocurtiche; le curve platicurtiche danno luogo a valori positivi, ma piccoli. Valori nulli di GM si verificano solo per distribuzioni uniformi ed i negativi indicheranno curve a forma di U. L'indice GM può ancora essere calcolato in caso di asimmetria soprattutto per le zone più estreme della distribuzione. Si modifica però l'interpretazione: non è più un indicatore di allontanamento dalla densità normale, ma una misura dello spessore della coda se "p" è prossimo allo zero e dell'importanza del centro se "p" è prossimo a 0.5

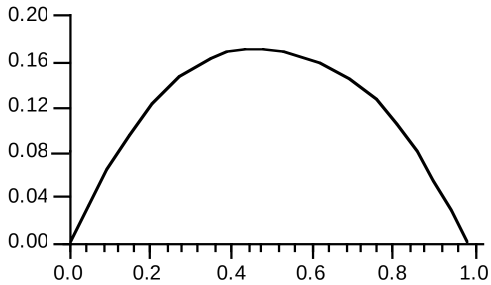
Esempi:

a) Parabola sull'intervallo unitario. Le misure confermano la iponormalità della densità che ha infatti code corte e spesse.

$$h(x) = \frac{3x(1-x)}{4};\quad 0 \le x \le 1$$

$$\mu = 1,\;\; \sigma = 0.141,\;\; \gamma_2 = -1.279,$$

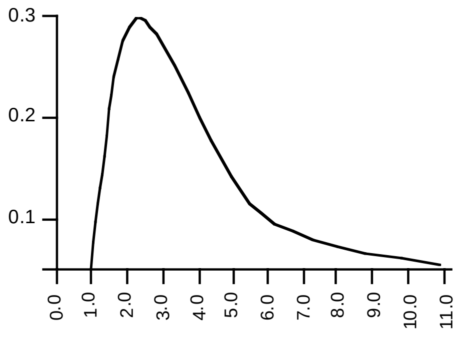
$$GM_{0.3} = \frac{X_{0.85} + X_{0.65} - 2 * X_{0.75}}{X_{0.85} - X_{0.65}} = 0.067$$



b) Modello privo di media finita (il γ_2 non sarebbe calcolabile).

$$h(x) = \frac{2\left(x^2-1\right)}{\left(x^2+1\right)}\quad x > 1;\quad F(x) = 1 - \frac{2x}{1+x^2}$$

$$X_p = \frac{1 + \sqrt{1-(1-p)^2}}{(1-p)};\;\; GM_{0.05} = 0.897$$



E' da ritenere che lo studio dell'intero arco dei valori di GM (funzione di curtosi) possa essere utile per caratterizzare la morfologia di una distribuzione di probabilità o di frequenza.

Esercizio_VC150:

a) Determinare la formula dell'indice di Groeneveld e Meeden per il modello logistico:

$$F(x) = \frac{1}{1+e^{-(ax+b)}};\;\; a > 0,\qquad X_p = \frac{1}{a} Ln\left(\frac{p}{1-p}\right) - b$$

b) Rappresentare la funzione di curtosi del modello per b=1/2.

Approssimazione di variabili discrete

Nonostante la palese contraddizione, le distribuzioni delle variabili discrete possono essere approssimate efficacemente dal modello normale. Occorre però tenere conto che i valori possibili delle discrete sono assimilabili ad un insieme di numeri interi naturali: 0, 1, 2, ..., n laddove la normale può assumere ogni valore reale per cui si renderanno necessarie alcune convenzioni.

Esempio:
Per le variabili discrete si ha: $p(X<a)+p(X\geq a+1)=1$ dato che nessuna modalità può trovarsi fra "a" ed "a+1"; per le variabili continue invece:

$$p(x<a)+p(x\geq a)\neq 1$$

in quanto la densità sull'intervallo (a,a+1) non è necessariamente nulla.

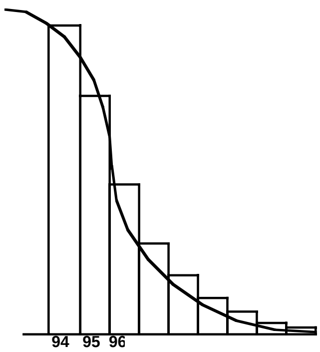
Per ovviare all'inconveniente si usa un fattore di correzione additivo pari a 0.5. Se X è discreta e la sua distribuzione deve essere approssimata con quella gaussiana, si definisce:

$$p(a)=\phi(a+0.5)-\phi(a-0.5)$$

dove p(a) è la probabilità della variabile discreta nel punto X=a ed $\phi(x)$ è la funzione di ripartizione della normale nell'intervallo $(-\infty,x)$. Poiché includere o escludere l'estremo è rilevante nelle variabili discrete si conviene che:

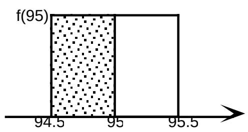
$$p(X<a)=\phi(a-0.5); \qquad p(X\leq a)=\phi(a+0.5);$$
$$p(X>a)=1-\phi(a+0.5); \qquad p(X\geq a)=1-\phi(a-0.5)$$

Esempi:
a) X è il numero di punti di attacco di un parassita sulle foglie di un albero in cui $\mu=80$ e $\sigma=10$. Approssimiamo con la normale la frazione di foglie con punti di attacco inferiori a 95.



X ≥ 95 significa in realtà ≥ 94.5

X ≤ 95 significa in realtà ≤ 95.5



$$p(X<95)\cong\phi(94.5)=\phi\left(\frac{94.5-80}{10}\right)=\phi(1.45)=0.9265;$$

I rettangoli dell'istogramma sono "centrati" sul valore intero e coprono un intervallo ottenuto sommando e sottraendo la metà della distanza tra due valori discreti consecutivi

b) La situazione più tipica di approssimazione è quella della binomiale che, grazie al teorema del limite centrale, converge alla normale con media $\mu=np$ e varianza $\sigma^2=np(1-p)$ all'aumentare del numero delle prove. L'approssimazione è considerata soddisfacente se $np\geq 5$. Proviamo a calcolare la probabilità che su 250 prove bernoulliane con probabilità di successo p=0.2 si abbia un numero di successi maggiore di 35, inferiore o uguale a 70, compreso tra 40 e 60. Abbiamo: $\mu=np=50$, $\sigma^2=np(1-p)=40$ per cui:

$$p(X>35)\cong 1-\phi\left(\frac{35-50+0.5}{\sqrt{40}}\right)=1-\phi(-2.29)=\phi(2.29)=0.9890;\quad p(X<70)\cong\phi\left(\frac{70-50-0.5}{\sqrt{40}}\right)=\phi(3.08)=0.99856$$

$$p(40\leq x\leq 60)=\phi\left(\frac{60-50+0.5}{\sqrt{40}}\right)-\phi\left(\frac{40-50+0.5}{\sqrt{40}}\right)=\phi(1.66)$$

La qualità dell'approssimazione migliora man mano che "p" si avvicina a 0.5. Dudewicz ed al. (1989, p.16) osservano che la standardizzazione è un passo obbligato per l'approssimazione delle variabili. Infatti, la sostituzione di $p(X\leq 10)$ con $\phi(10.5)$ è legittima sia per $n\leq 20$ che per $n>20$, ma non è altrettanto accurata: $\phi(10.5)$ è indistinguibile dall'unità, ma $p(X\leq 10)$ diventa sempre più piccola all'aumentare di "n". La standardizzazione, comunque, fa rientrare nel quadro dell'approssimazione sia "n" che "p".

c) Le elezioni provinciali hanno fatto registrare 10'000 mancate partecipazioni al voto. La Destra ha vinto per uno scarto di soli 100 voti. Nell'ipotesi che i non votanti, oltre ad essere in vita e a non aver cambiato cittadinanza, avessero votato lanciando una moneta senza trucchi, qual'è la probabilità che la Destra perdesse le elezioni? Il voto X dei non partecipanti è una binomiale con $\mu=5000$ e $\sigma=50$; si perderebbero le elezioni se $X<4900$. Calcoliamone la probabilità con l'approssimazione normale:

$$p(X < 4900) \cong \phi\left(\frac{4900 - 0.5 - 5000}{50}\right) = \phi(-2.01) = 1 - \phi(2.01) = 2.22\%$$

C'è un certo margine di sicurezza: la Destra sarebbe sconfitta solo una volta su 45 (1/0.0222).

d) Anche la Poisson è approssimabile dalla Normale ponendo $\mu=\lambda$ e $\sigma=\sqrt{\lambda}$ purché λ sia elevato (ricordiamo che $\gamma_1=1/\sqrt{\lambda}$ e $\gamma_2=1/\lambda$) diciamo, $\lambda\geq 10$. Supponiamo che il numero annuale di contatti sbagliati ad un numero telefonico sia approssimabile da una Poisson con media $\lambda=14$. Qual'è la probabilità che in un anno si abbiamo meno di 23 chiamate per sbagliato numero?

$$p(0 \leq x \leq 23) \cong \phi\left(\frac{23 + 0.5 - 14}{3.7417}\right) - \phi\left(\frac{0 - 0.5 - 14}{3.7417}\right) = \phi(2.54) - \phi(-3.88) = 0.9945$$

e) Il numero X di lavoratori che in un mese subiscono un infortunio nella SCA Spa (che ha 4870 operai) è modellabile con la binomiale p=0.003. Qual'è la probabilità che gli infortuni siano meno di 10?

$$\textit{Binomiale}: p(x \leq 10) = 0.1383; \quad \textit{Normale per la Binomiale}: \phi\left(\frac{10 + 0.5 - 14.61}{3.8166}\right) = \phi(-1.0769) = 0.1408$$

$$\textit{Poisson}\ (l = 14.61): 0.1387; \quad \textit{Normale per la Poisson}: \phi\left(\frac{10 + 0.5 - 14.61}{3.8223}\right) = \phi(-1.0753) = 0.1411$$

f) All'aumentare del numero di prove la differenza tra scelte con reimmissione e scelte senza reimmissione diventa irrilevante per cui l'approssimazione normale si può estendere anche al modello ipergeometrico. Un campione di n=30 operai -scelti senza reimmissione- in uno stabilimento di 900 dipendenti è sottoposto a controllo periodico per l'insorgenza di patologie legate all'ambiente di lavoro. Ipotizzando che gli operai con inizio di malattia siano il 5% calcolare la probabilità che nel campione se ne trovi uno.

$$\textit{Ipergeometrica}: p(x=1) = \frac{\binom{45}{1}\binom{855}{29}}{\binom{900}{30}} = 0.3418; \qquad \textit{Binomiale}: \binom{45}{1}\left(\frac{30}{900}\right)\left(\frac{870}{900}\right)^{44} = 0.3375$$

$$\textit{Normale}: \phi\left(\frac{1 + 0.5 - 1.5}{1.114}\right) - \phi\left(\frac{1 - 0.5 - 1.5}{1.114}\right) = \phi(0) - \phi(-0.9) = 0.5 - 1 + \phi(0.9) = 0.3159$$

Esercizio_VC151: *il modello di Pascal. Il coefficiente binomiale diventa intrattabile dal punto di vista numerico quando le ampiezze diventano grandi ed è difficile ottenere valori plausibili per x>100 e qui interviene il modello Normale. Supponiamo che r=20. Qual'è la probabilità da attribuire alle modalità inferiori a 50?*

Esercizio_VC152: *l'ammissione ad un corso avviene con il superamento di una prova basata su 50 quiz a scelta multipla. Ogni quiz ha 4 possibili risposte di cui una sola è esatta (le risposte errate non comportano penalità).*
a) Qual'è la probabilità che rispondendo a caso si ottengano 15 risposte esatte?
b) Se la commissione ritiene di poter tollerare una probabilità dell'uno per diecimila o inferiore che un rispondente a caso sia ammesso, quante risposte esatte deve richiedere? c) Se Ciccillo ha una preparazione tale che ogni volta che risponde ha probabilità del 50% di dare la risposta giusta qual'è la probabilità che sia ammesso al corso? d) Nel punto precedente insorge un problema. Quale? Come si risolve?

Grafico di normalità

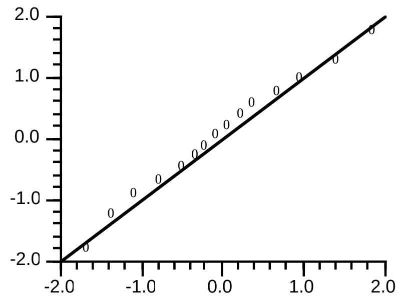
Molte procedure statistiche presuppongono il modello gaussiano. Un metodo rapido per verificare tale congettura è il grafico di normalità (caso particolare del grafico Q-Q, quantile/quantile presentato da Gnanadesikan e Wilk (1968). Il grafico è ottenuto rappresentando in un sistema di assi cartesiano le modalità standardizzate ed i quantili della Normale corrispondenti alle frequenze relative cumulate p_i della X_i:

$$\hat{Z}_i = \frac{X_{(i)} - \mu}{\sigma}, \ Z_{p_i} \quad i = 1, 2, \ldots, n \ \ dove \ \ p_i = \frac{i - 0.5}{n}$$

Per "n" modalità distinte (la modifica per modalità ripetute è facile) ordinate in senso ascendente: $X_{(i)}$. Se il grafico somiglia molto ad una retta che passa per l'origine e con inclinazione unitaria (la bisettrice del quadrante per intenderci), l'approssimazione normale è ottima.

Esempio:
Analizziamo i valori: {4, 6, 8, 10, 12, 13, 14, 15, 16, 17, 18, 20, 22, 25, 28}

i	X_i	$\hat{Z}_i$	p_i	Z_i
1	4	-1.71	0.03	-1.83
2	6	-1.40	0.10	-1.28
3	8	-1.10	0.17	-0.97
4	10	-0.79	0.23	-0.73
5	12	-0.49	0.30	-0.52
6	13	-0.34	0.37	-0.34
7	14	-0.18	0.43	-0.17
8	15	-0.03	0.50	0.00
9	16	0.12	0.57	0.66
10	17	0.27	0.63	0.86
11	18	0.43	0.70	1.06
12	20	0.73	0.77	1.28
13	22	1.04	0.83	1.53
14	25	1.49	0.90	1.81
15	28	1.95	0.97	2.17



I quantili della normale standardizzata sono stati ottenuti mediante l'approssimazione proposta nell'esercizio VC143. Se le osservazioni provenissero dalla normale i punti dovrebbero allinearsi lungo la bisettrice del grafico. Si nota invece un certo allontanamento per i valori medio-alti.

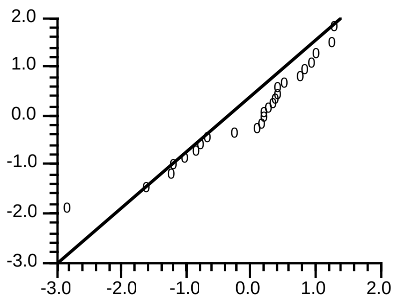
Grafico Q-Q

E' chiaro che il grafico di normalità non è limitato all'accertamento della prossimità tra i quantili empirici e quelli della normale, ma anche al confronto tra i quantili empirici ed i quantili di un qualsiasi altro modello diverso dalla normale ovvero tra quantili teorici di un modello e i quantili della normale. La scelta dei quantili cui far corrispondere $x_{(i)}$ è controversa. Scartata la soluzione p_i=i/n che è inadatta per un supporto infinito (raggiunge subito l'unità per i=n), ci si è orientati sulla opzione dell'esempio cioè il valore centrale dell'intervallo [(i-1)/n,i/n] e che è anche la più diffusa. L'alternativa, p_i=i/(n+1) detta di van der Waerden gode pure di una certa popolarità soprattutto per il confronto tra la Normale ed il modello di Weibull. Ecco delle altre possibili scelte:

$$\textit{Gerson}\ (1974):\quad p_i=\frac{i-0.375}{n+0.25};\qquad \textit{Tukey}\ (1962):\quad p_i=\frac{3i-1}{3n+1};\qquad \textit{Barnett}(1975):\quad p_i=\frac{i-0.3}{n+0.4}$$

Esempio:
Variazioni percentuali del PIL in alcuni Paesi emergenti. Dati 1998. Posizioni dei quantili date da p_i=(3i-1)/(3n+1).

5.9	7.1	5.5	1.5	4.6
-2.1	3.1	-0.5	4.8	-6.6
0.1	4.8	5.3	5.5	0.8
5.4	-0.7	5.0	6.9	7.5
8.6	8.8	7.8	1.0	4.4



Il grafico evidenzia la presenza di un valore anomalo o che appare tale sotto la lente del modello normale. D'altra parte, l'allineamento è troppo scarso per poter proporre tale modello. Meglio optare per altri modelli.

Il grafico Q-Q è un semplice ed utile ausilio qualitativo per accertare l'aderenza dei dati al modello gaussiano ovvero a qualsiasi altra distribuzione di cui siano noti -in forma esplicita o ben approssimata- i quantili. E' inoltre utile per accertare la situazione negli estremi della distribuzione: a code molto spesse, diciamo per quantili nelle zone $p_i<0.05$ e $p_i>0.95$, corrisponderà una deviazione consistente dalla retta che si formerebbe se i dati fossero originati da una variabile casuale diversa dalla normale. La sua realizzazione è semplice con il foglio elettronico ed è comunque disponibile in quasi tutti i pacchetti applicativi statistici.

Esercizio_VC153: *percentuale di abilitati all'esercizio della professione (dati 1995 e 1996).*

Categoria	% Abilitati				
Commercialisti	34.5	Biologi	93.2	Psicologi	65.6
Attuari	71.8	Agronomi	68.2	Veterinari	98.7
Ingegneri	90.0	Forestali	65.5	Notai	3.1
Architetti	40.5	Farmacist	94.4	Avvocati	40.0
Geologi	51.3	Medici	98.4	Ragioneri	30.8
Chimici	94.9	Odontoiatri	99.8	Cons. Lavoro	65.0

Dopo aver programmato (su foglio elettronico o in un linguaggio di programmazione) il calcolo dei quantili (VC143) verificare l'ipotesi di normalità con l'apposito diagramma.

Trasformazioni per la normalità.

Le indagini statistiche coinvolgono solitamente moltissime variabili e non sempre si ha la possibilità di indagare ognuna per tutti gli aspetti finora considerati: centralità, variabilità e asimmetria, appuntimento al centro, spessore delle code. In genere, ci si limita alle prime due caratteristiche e per le altre, visto che per determinarle occorrerebbe ogni volta un accurato esame del grafico, si presuppone -spesso senza verifica- che la distribuzione sia normale o quasi. Talvolta l'ipotesi regge e tante procedure, pur richiedendo la normalità, valgono lo stesso anche in presenza di una moderata disnormalità (è questo che si intende per robustezza). In molti casi però l'ipotesi di normalità è poco difendibile: ad esempio nelle distribuzioni ad "U" oppure a "J".

Esempio:
Rosenzweig e Fuller, 1970) avvertono: nelle distribuzioni fortemente asimmetriche si riscontra una elevata variabilità dovuta a poche osservazioni remote (qui legittimate a presentarsi) che possono pregiudicare i risultati inferenziali sulla media ed è quindi opportuno intervenire per attenuare l'effetto di code troppo accentuate.

Per rendere più accettabile l'ipotesi di normalità si adottano delle trasformazioni che rendono il poligono di frequenza più simile alla curva gaussiana.

Trasformazione Box-Cox
Fra le diverse trasformazioni in uso ricordiamo la più famosa, la Box-Cox per variabili strettamente positive:

$$Y_i(\alpha) = \begin{cases} \dfrac{X_i^\alpha - 1}{\alpha} & per \quad \alpha \neq 0 \\ Ln(X) & per \quad \alpha = 0 \end{cases}$$

Tale trasformazione, a differenza di quelle lineari, non si limita a correggere la centralità e la variabilità, ma interviene anche sulla forma della curva di densità.

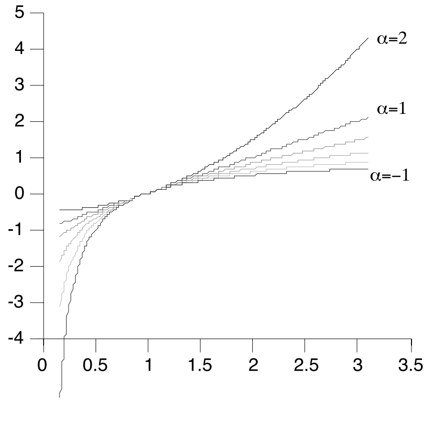
Esempi:
a) Casi particolari della Box-Cox

$\alpha = 1$, *trasformazione lineare*: $Y_i = X_i - 1$;

$\alpha = 0.5$, *radice*: $Y_i = 2\left(\sqrt{X_i} - 1\right)$;

$\alpha = 2$, *trasformazione quadratica*: $Y_i = \dfrac{X_i^2 - 1}{2}$;

$\alpha = -1$, *reciproca*: $Y_i = 1 - \dfrac{1}{X_i}$;



b) Bhattacharyya e Johnson (1977, p. 4) suggeriscono le seguenti trasformazioni:
1) $y=x^3$, $y=x^2$ per rendere i valori grandi ancora più grandi; 2) $y=x^{0.5}$, $y=x^{0.25}$, y=Log(x), $y=x^{-1}$ per rendere i valori grandi più piccoli.

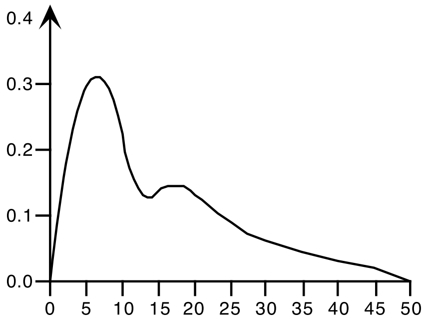
c) Per variabili discrete x=0, 1, 2,… Anscombe ha proposto: $y_i = \sqrt{\frac{3}{8} + x_i}$

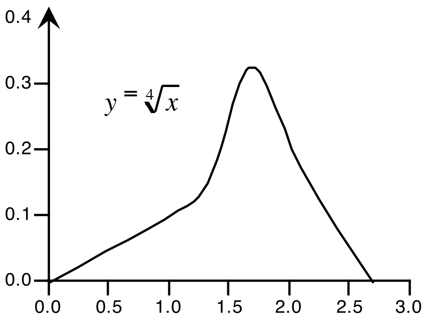
d) Se la X può assumere valori negativi si può applicare la Box-Cox alla variabile : $U=(X-X_{min})+0.001$

La trasformazione ha diverse proprietà utili come ricordano Emerson e Stoto (1993).
1) E' monotona cosicché i quantili di Y(α) corrispondono a quelli di Y.
2) E' continua di modo che modalità ravvicinate nella variabile originale rimangono vicine nella trasformata.
3) E' smussata (non presenta cioè punti angolosi).
4) Le curve hanno in comune il punto (1,0) e sono molto vicine nell'intorno di tale punto.
5) Con α>1 la concavità è rivolta verso l'alto e se α<1 la concavità è rivolta verso il basso.

Esempi:
a) Bhattacharyya e Johnson (1977, p. 224) presentano dei dati relativi alla produzione di legname di 49 sezioni regolari di un bosco.

0.7	0.9	1.0	1.3	1.9	2.7	3.2	3.4	3.4	3.5	3.5	4.3	5.2
5.9	6.0	6.3	6.5	6.6	7.1	7.4	7.6	7.9	8.3	8.3	8.3	8.3
8.7	10.0	10.0	10.3	12.0	13.4	14.1	14.8	16.7	16.8	17.1	17.7	18.9
19.0	19.4	19.7	24.3	26.2	26.2	28.3	31.7	39.3	44.8			





Per i dati originali il cui poligono (smussato) è presentato sulla sinistra la somiglianza alla normale sembrerebbe da scartare. La radice quartica ravvicinando i valori grandi al centro sembra invece produrre un poligono ragionevolmente approssimabile dalla normale.

b) Pearce (1965, p.57) nota che l'area della lesione procurata da un'infezione fungina viola alcune ipotesi precludendo la normalità. La radice quadrata delle misure rimuove l'ostacolo. La ragionevolezza di questo stratagemma nasce dal fatto che le lesioni sono misurate con una grandezza proporzionale alla distanza dalla spora originale; l'estrazione di radice ha quindi senso poiché potrebbe essere questa la variabile cui applicare il modello gaussiano.

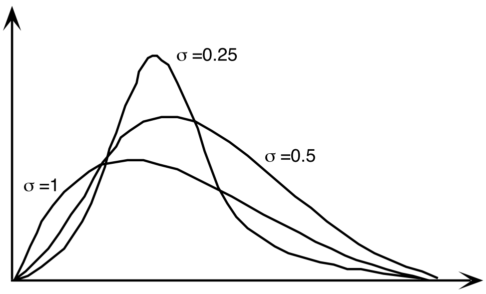
Il caso forse più emblematico della trasformazione Box-Cox è il modello lognormale:

$$h(x) = \frac{e^{-0.5\left(\frac{Ln(x)-\mu}{\sigma}\right)^2}}{x\sigma\sqrt{2\pi}}; \quad x > 0; \qquad E(x) = e^{\mu+\frac{\sigma^2}{2}} = \alpha$$

$$\sigma^2(x) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$$

$$CV = \sqrt{\left(e^{\sigma^2} - 1\right)} = \beta$$

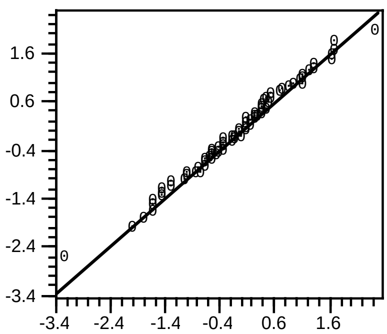
$$\gamma_1 = \alpha^3\left(\beta^6 + 3\beta^4\right); \quad \gamma_2 = \alpha^4\left(\beta^{12} + 6\beta^{10} + 15\beta^8 + 16\beta^6 + 3\beta^4\right)$$



La curva di densità è unimodale, asimmetrica a sinistra (infatti γ_1 è positivo) e con una coda consistente, ma mai tanto spessa da impedire l'esistenza dei momenti di qualsiasi ordine. Se si considera il logaritmo della variabile la sua funzione di densità diventa normale.

Esempio:
Riprendiamo i dati di Williams (1984, p. 108) sulla densità dell'area di drenaggio in 75 maglie regolari in una zona del Devon (GB).

0.7	1.2	1.3	1.4	1.4	1.4	1.5	1.5	1.5	1.6	1.6	1.8	1.8
1.8	1.9	2.0	2.0	2.1	2.1	2.1	2.2	2.2	2.2	2.2	2.2	2.2
2.3	2.3	2.4	2.4	2.4	2.4	2.5	2.5	2.6	2.6	2.7	2.7	2.7
2.8	2.8	2.8	2.8	2.8	2.9	3.0	3.0	3.0	3.0	3.1	3.1	3.1
3.1	3.1	3.2	3.3	3.3	3.4	3.4	3.6	3.7	3.9	4.2	4.2	4.3
4.3	4.0	4.5	4.7	4.7	5.3	5.3	5.4	5.4	7.4			



La trasformazione logaritmica rende abbastanza ragionevole il modello normale come mostra il grafico di normalità, anche se sembra mancare un poco nelle due code.

Determinazione del parametro

L'idea della trasformazione Box-Cox è che sia necessario qualcosa di più che non un semplice slittamento del centro o una modifica dell'unità di misura per rendere comprensibili i dati e che la giusta modifica nella forma della curva di distribuzione rientri in uno dei casi particolari di questa trasformazione. Un modo efficace di determinare il valore di "α" più appropriato è di calcolare una quantità basata su di un confronto per alcuni caratteristiche della curva quale:

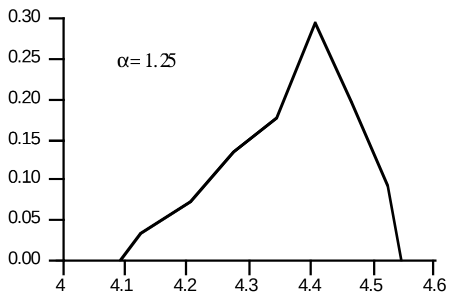
$$T(\alpha) = \sqrt{\left[\gamma_1(\alpha)\right]^2 + \left[\gamma_2(\alpha)\right]^2}$$

valutata per valori prefissati di a: {-2.5,-2.0,- 1.5,-1,-0.5, 0, 0.5, 1.0, 1.5, 2.0,2.5} cercando il minimo di Q(α). Questo, ovviamente, non garantisce che il valore determinato sia poi in effetti piccolo nel senso che la distribuzione potrebbe essere così selvaggia che non esiste una trasformazione Box-Cox in grado di "normalizzarla". Inoltre, non sono poche le distribuzioni unimodali, simmetriche, con momento quarto pari a tre, però dissimili dalla normale per cui il criterio non può dare troppe garanzie.

Esempi:

a) Distribuzione di un campione di richiedenti un'assicurazione sulla vita per pressione sanguigna diastolica (in mm/hg).

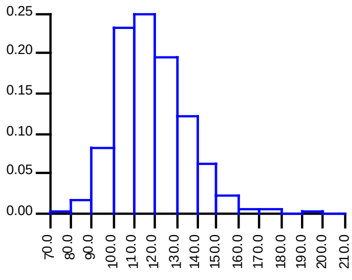
α	γ_1	γ_2	T(α)
-1.00	-3.8365	0.2373	3.8438
-0.75	-3.7817	0.1254	3.7838
-0.50	-3.7275	0.0219	3.7276
-0.25	-3.6739	-0.0735	3.6746
0.00	-3.6207	-0.1609	3.6243
0.25	-3.5682	-0.2406	3.5763
0.50	-3.5161	-0.3128	3.5300
0.75	-3.4646	-0.3777	3.4852
1.00	-3.4137	-0.4356	3.4413
1.25	-3.3632	-0.4866	3.3982
1.50	-3.6207	-0.1609	3.6243

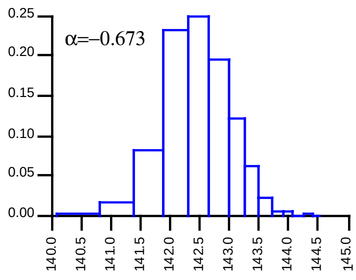


Cerchiamo di individuare la trasformazioni Box-Cox adatta per normalizzarla. In questo caso c'è incompatibilità tra riduzione della asimmetria ed eliminazione della iponormalità e per ottenere la normalità c'è bisogno di una trasformazione ancora più flessibile della Box-Cox.

b) Dati sul peso -in libbre- di 1000 studentesse (ripresi da Cohen, 1991, p.183). In questo caso la trasformazione Box-Cox è applicata ai valori centrali delle classi.

L_i	U_i	n_i			
70	79.9	2	150	159.9	23
80	89.9	16	160	169.9	5
90	99.9	82	170	179.9	7
100	109.9	231	180	189.9	1
110	119.9	248	190	199.9	2
120	129.9	196	200	209.9	1
130	139.9	122	210	219.9	1
140	149.9	63			1000





Esercizio_VC154: *tasso di disoccupazione delle province del Centro-Nord in un anno recente.*

9.27	9.95	11.19	13.41	11.89	14.54	16.88	25.77
9.27	10.03	11.25	13.69	12.08	14.84	17.50	28.11
9.71	10.08	11.26	13.73	12.40	15.23	19.48	29.16
9.71	10.17	11.39	13.75	12.41	15.31	19.58	
9.73	10.71	11.42	13.79	13.00	16.34	19.86	
9.91	10.75	11.47	14.05	13.26	16.37	24.64	

Determinare il parametro della trasformazione Box-Cox che più avvicini la rilevazione al modello normale.