

Richiami sulla Cardinalità.

Due insiemi A e B si dicono *equipotenti* ($A \sim B$) se esiste una corrispondenza biunivoca tra A e B

$$f : A \rightarrow B \quad (\text{biunivoca}).$$

Due insiemi finiti sono equipotenti se e solo se hanno lo stesso numero di elementi.

Esempio. Gli insiemi $A = \{1, 2, 3, 4, 5, 6\}$ e $B = \{a, b, c, d, e, g\}$ sono equipotenti.

Un insieme A si dice *numerabile* (o ha cardinalità numerabile) se è equipotente all'insieme \mathbb{N} dei numeri naturali.

Dato un insieme A infinito, è sempre possibile trovare un sottoinsieme proprio $B \subset A$ tale che $A \sim B$.

Ad esempio, indicando con P l'insieme dei numeri pari, si ha $P \sim \mathbb{N}$, infatti la funzione $f : \mathbb{N} \rightarrow P$ definita nel seguente modo

$$f(n) = 2n, \quad \forall n \in \mathbb{N},$$

è biunivoca.

Sia (a, b) un qualsiasi intervallo aperto nell'insieme dei

numeri reali \mathbb{R} , con $a < b$. Si ha

$$\mathbb{R} \sim (a, b).$$

Ad esempio, per verificare che l'intervallo aperto $(-\frac{\pi}{2}, +\frac{\pi}{2})$ è equipotente ad \mathbb{R} , basta osservare che la funzione $f : (-\frac{\pi}{2}, +\frac{\pi}{2}) \rightarrow \mathbb{R}$, definita da

$$f(x) = \tan(x),$$

è biunivoca.

Un insieme A si dice che ha la *potenza del continuo* se $A \sim \mathbb{R}$.

L'insieme \mathcal{I} dei numeri irrazionali ha la potenza del continuo.

Si può dimostrare che \mathbb{R} non è numerabile.

L'insieme \mathbb{Z} dei numeri relativi e l'insieme \mathbb{Q} dei numeri razionali sono numerabili.

Per gli insiemi numerabili vale il seguente

Teorema 1 L'unione di una successione di insiemi finiti o numerabili (eventualmente vuoti), a due a due disgiunti, è un insieme al più numerabile.

Cioè, data una successione $\{A_1, A_2, \dots, A_n, \dots\}$ di insiemi finiti o numerabili, con $A_i \cap A_j = \emptyset$, l'insieme $A = \bigcup_{n=1}^{\infty} A_n$ è al più numerabile.

(il teorema si dimostra con il procedimento di diagonalizzazione di Cantor)

Data una famiglia \mathcal{A} di cardinalità arbitraria (almeno numerabile) di eventi a due a due incompatibili, sia \mathcal{B} la sottofamiglia di \mathcal{A} formata dagli eventi di probabilità positiva.

$$\mathcal{B} = \{E \in \mathcal{A} : P(E) > 0\} \quad (1)$$

A quanti eventi di \mathcal{A} è possibile attribuire una probabilità positiva?

Indichiamo con \mathcal{A}_n i sottoinsiemi di \mathcal{A} così definiti

$$\mathcal{A}_n = \{E \in \mathcal{A} : \frac{1}{n+1} < P(E) \leq \frac{1}{n}\}$$

si ha

$$\mathcal{B} = \bigcup_{n=1}^{\infty} \mathcal{A}_n$$

Per $n = 1$ si ha $\mathcal{A}_1 = \{E \in \mathcal{A} : \frac{1}{2} < P(E) \leq 1\}$. Inoltre, se esiste un evento $E_1 \in \mathcal{A}_1$, allora \mathcal{A}_1 non

può contenere altri eventi.

Infatti, se esistesse $E_2 \in \mathcal{A}_1$, con $E_1 E_2 = \emptyset$, si avrebbe

$$P(E_1 \vee E_2) > 1/2 + 1/2 = 1 \quad (\textit{assurdo}).$$

Quindi $\textit{card}(A_1) \leq 1$.

Con un ragionamento analogo, per $n = 2$ si ottiene $\textit{card}(A_2) \leq 2$.

Ricapitolando

$$\begin{aligned} n = 1 & \quad \textit{card}(A_1) \leq 1, \\ n = 2 & \quad \textit{card}(A_2) \leq 2, \\ & \quad \vdots \\ n = k & \quad \textit{card}(A_k) \leq k \\ & \quad \vdots \end{aligned}$$

\mathcal{A}_n ha cardinalità finita per ogni n , perciò, per il Teorema 1, $\mathcal{B} = \bigcup_{n=1}^{\infty} \mathcal{A}_n$ è al più numerabile.

Per dirla "alla buona", quando gli eventi sono "troppi" non possiamo attribuire a tutti probabilità positiva.

In alcuni casi siamo costretti ad attribuire probabilità nulla anche ad eventi *non impossibili* e simmetrica-

mente probabilità 1 anche ad eventi *non certi*.

In generale

$$E = \emptyset \Rightarrow P(E) = 0, \quad P(E) = 0 \not\Rightarrow E = \emptyset.$$

$$E = \Omega \Rightarrow P(E) = 1, \quad P(E) = 1 \not\Rightarrow E = \Omega.$$

Esempio. Supponiamo di voler assegnare la probabilità che un dispositivo D si guasti in un preciso istante di tempo t .

Sia T il numero aleatorio che indica il tempo di durata del dispositivo e sia $E_t = (T = t)$, con $t \in \mathbb{R}_0^+$, l'evento " D si guasta all'istante t ".

L'insieme

$$\mathcal{A} = \{E_t : t \in \mathbb{R}_0^+\}$$

è costituito da eventi a 2 a 2 incompatibili ed ha la potenza del continuo. Come sappiamo, possiamo attribuire probabilità positiva al più ad un sottoinsieme numerabile di \mathcal{A} , quindi "*per la quasi totalità*" degli eventi di \mathcal{A} , si ha

$$P(E_t) = 0, \quad P(E_t^c) = 1, \quad \forall t \in \mathbb{R}_0^+.$$

Sia $\mathcal{A} = \{E_1, E_2, \dots, E_n, \dots\}$ una partizione numerabile di Ω , cioè

$$E_i E_j = \emptyset, \quad \forall i \neq j, \quad \bigvee_{n=1}^{\infty} E_n = \Omega.$$

Se anche nel caso numerabile valesse la proprietà additiva, si avrebbe

$$\sum_{n=1}^{\infty} P(E_n) = P\left(\bigvee_{n=1}^{\infty} E_n\right) = P(\Omega) = 1. \quad (2)$$

In effetti, dalla condizione di coerenza segue (soltanto)

$$\sum_{n=1}^{\infty} P(E_n) \leq 1. \quad (3)$$

Infatti

$$\begin{aligned} P(\bigvee_{n=1}^{\infty} E_n) &= P[(\bigvee_{k=1}^n E_k) \vee (\overbrace{\bigvee_{k=n+1}^{\infty} E_k}^{A_{n+1}})] = \\ &= P[(\bigvee_{k=1}^n E_k) \vee A_{n+1}] = P(\bigvee_{k=1}^n E_k) + P(A_{n+1}) \geq \\ &\geq \sum_{k=1}^n P(E_k). \end{aligned}$$

In conclusione

$$1 = P(\Omega) = P(\bigvee_{n=1}^{\infty} E_n) \geq \sum_{k=1}^n P(E_k), \quad \forall n.$$

Esistono quindi situazioni in cui la (2) è valida e situazioni in cui non lo è.

Quando la (2) è valida si parla di σ -*additività* (o *additività completa*).

Esempio (*Scelta a caso di un numero naturale.*)

E_i = "Esce il numero i "

X = numero scelto, $X \in \{1, 2, \dots, n, \dots\}$.

Gli eventi E_i sono tutti equiprobabili, $P(E_i) = P(E_j)$.
Nel caso finito (n numeri) si ha $P(E_i) = 1/n$.
Nel caso infinito invece si ha $P(E_i) = 0$.
Quindi

$$1 = P(\Omega) = P(\bigvee_{n=1}^{\infty} E_n) > \sum_{k=1}^{\infty} P(E_k) = 0.$$

perciò non vale la σ -additività.

Un esempio in cui vale la σ -additività è dato dalla distribuzione geometrica. Infatti, se $X \sim \mathbf{G}(p)$, si ha

$$\begin{aligned} \sum_{n=1}^{\infty} p_n &= \sum_{n=1}^{\infty} p(1-p)^{n-1} = p \sum_{n=1}^{\infty} (1-p)^{n-1} = \\ &= p \cdot \frac{1}{1-(1-p)} = 1. \end{aligned}$$

Esempio: successione di lanci di una moneta;
"successo" in una prova significa "esce testa" nel lancio corrispondente.

Il numero aleatorio di lanci sino al primo successo ha una distribuzione geometrica di parametro $\frac{1}{2}$, ovvero

$X \sim \mathbf{G}(\frac{1}{2})$, e si ha

$$\begin{aligned} \sum_{n=1}^{\infty} p(1-p)^{n-1} &= \sum_{n=1}^{\infty} \frac{1}{2} \left(\frac{1}{2}\right)^{n-1} = \\ &= \frac{1}{2} \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n-1} = \frac{1}{2} \cdot \frac{1}{1-\frac{1}{2}} = 1. \end{aligned}$$

Inoltre, dato l'evento

$$A_{\infty} = (X > n, \forall n \in \mathbb{N}),$$

si ha $A_{\infty} \subset (X > n)$, $\forall n \in \mathbb{N}$, e quindi

$$P(A_{\infty}) \leq P(X > n) = \frac{1}{2^n} \quad \forall n \in \mathbb{N}.$$

Pertanto $P(A_{\infty}) = 0$, cioè l'evento (logicamente possibile)

"non esce mai testa"

ha probabilità nulla.

Numeri Aleatori Continui

Sia \mathcal{A} una partizione dell'evento certo Ω , con $\text{card}(\mathcal{A}) > \text{card}(\mathbb{N})$. Si può dimostrare che in questo caso non è possibile attribuire a tutti gli eventi di \mathcal{A} probabilità positiva, ma al più ad un sottoinsieme $\mathcal{B} \subset \mathcal{A}$, con $\text{card}(\mathcal{B}) = \text{card}(\mathbb{N})$.

In queste situazioni si possono presentare numeri aleatori che assumono valori in un insieme con cardinalità del continuo. Un n. a. X di questo tipo può essere visto come una funzione

$$X : \Omega \rightarrow \mathbb{R}$$

che ad ogni caso possibile $\omega \in \Omega$ assegna un valore $X(\omega) = x \in \mathbb{R}$.

Ad esempio: X può essere l'istante di tempo in cui si guasta un determinato dispositivo.

In questo genere di applicazioni tipicamente i valori possibili di X sono (per ragioni di carattere matematico) tutti di probabilità nulla, cioè

$$P(X = x) = 0, \quad \forall x.$$

Distribuzioni (assolutamente) continue. Un numero aleatorio X si dice continuo se

- (i) $P(X = x) = 0, \forall x \in \mathbb{R}$ (*non esistono probabilità concentrate, tutti gli eventi hanno probabilità nulla*);
- (ii) esiste una funzione reale $f \geq 0$ (non negativa) integrabile secondo Riemann, tale che per ogni sottoinsieme $A \subseteq \mathbb{R}$, misurabile secondo Peano-Jordan, si ha

$$P(A) = P(X \in A) = \int_A f(x) dx .$$

La funzione f si chiama *densità di probabilità* del n.a. X .

Osservazioni. Se A è un generico intervallo $[a, b]$ (limitato o non), la (ii) diventa

$$P(X \in [a, b]) = \int_a^b f(x) dx ,$$

cioè $P(X \in [a, b])$ coincide con l'area sottesa al diagramma di $f(x)$ nell'intervallo (a, b) .

In particolare se $A = (-\infty, +\infty)$ otteniamo

$$P(\Omega) = P(X \in (-\infty, +\infty)) = \int_{-\infty}^{+\infty} f(x)dx = 1. \quad (4)$$

Cioè l'area totale sotto la curva è 1.

Da un punto di vista meccanico la densità di probabilità si può interpretare come la densità di massa con cui una massa unitaria è diffusa sull'asse reale.

Vediamo il legame che sussiste tra la densità di probabilità e la funzione di ripartizione. Data una densità di probabilità $f(x)$, ricordando che

$$F(x) = P(X \leq x) = P(X \in] - \infty, x]),$$

si ha

$$F(x) = \int_{-\infty}^x f(t)dt$$

cioè la f.d.r. calcolata in x rappresenta l'area sotto la curva da $-\infty$ a x .

Viceversa data una funzione di ripartizione $F(x)$, se $F(x)$ è derivabile in x , si ha

$$F'(x) = f(x).$$

Quindi una f.d.r. nel caso di distribuzioni continue soddisfa le seguenti proprietà

1. $0 \leq F(x) \leq 1, \forall x \in \mathbb{R}$;
2. $F'(x) \geq 0$ ($F(x)$ è non decrescente).

Se abbiamo la f.d.r. $F(x)$ e si vuole calcolare la probabilità di $A = [a, b]$, cioè $P(a \leq X \leq b)$, basta osservare che

$$P(A) = \int_a^b f(x)dx = F(b) - F(a).$$

Sia $g(x)$ una funzione reale di variabile reale integrabile secondo Riemann, per vedere se essa rappresenta una densità di probabilità occorre provare che

$$1. g(x) \geq 0, \quad \forall x \in \mathbb{R};$$

$$2. \int_{-\infty}^{+\infty} g(x) dx = 1.$$

Se $g(x) \geq 0, \forall x \in \mathbb{R}$ e risulta

$$\int_{-\infty}^{+\infty} g(x) dx = k > 0,$$

la seguente funzione

$$f(x) = \frac{1}{k} \cdot g(x), \quad \forall x \in \mathbb{R},$$

ottenuta normalizzando la funzione g , è una densità di probabilità.

Distribuzione Uniforme

Un n.a. continuo X ha una distribuzione uniforme in un intervallo $[a, b]$, in simboli

$$X \sim \mathbf{U}([a, b]),$$

se ha la seguente densità di probabilità

$$f(x) = \begin{cases} k > 0 & \text{se } x \in [a, b], \\ 0 & \text{altrove.} \end{cases}$$

La costante k si determina osservando che

$$\int_{-\infty}^{+\infty} f(x) dx = \int_a^b k dx = k(b - a) = 1,$$

da cui segue

$$k = \frac{1}{b - a}.$$

Quindi, se $X \sim \mathbf{U}([a, b])$, si ha

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b], \\ 0 & \text{altrove.} \end{cases}$$

Dato un intervallo $I = (c, d)$ contenuto in $[a, b]$ si prova che la $P(X \in I)$ dipende solo dall'ampiezza dell'intervallo. Infatti se $l = d - c$ è l'ampiezza di I , si ottiene

$$P(X \in I) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a} = \frac{l}{b-a}.$$

Pertanto dati due intervalli I, J contenuti in $[a, b]$, rispettivamente di ampiezza l e λ , si ha

$$P(X \in I) = P(X \in J) \Leftrightarrow l = \lambda.$$

La funzione di ripartizione di $X \sim \mathbf{U}([a, b])$ è definita da

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & \text{se } x < a, \\ \frac{x-a}{b-a}, & \text{se } x \in [a, b], \\ 1, & \text{se } x > b. \end{cases}$$

Esempio 1 Se $X \sim \mathbf{U}([0, 1])$, la densità di probabilità è data da

$$f(x) = \begin{cases} 1, & \text{se } x \in [0, 1], \\ 0, & \text{altrove,} \end{cases}$$

e la f.d.r. diventa

$$F(x) = \begin{cases} 0, & \text{se } x < 0, \\ x, & \text{se } x \in [0, 1], \\ 1, & \text{se } x > 1. \end{cases}$$

Se consideriamo l'intervallo $I = [\frac{1}{2}, \frac{2}{3}]$, è facile verificare che

$$P(I) = P(X \in I) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}.$$

Se inoltre consideriamo l'intervallo $J = [\frac{1}{2}, \frac{3}{2}]$ si ha

$$P(J) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Previsione. Dato un n.a. continuo X con densità di prob. $f(x)$, la previsione di X si definisce nel modo seguente

$$\mathbb{P}(X) = \int_{-\infty}^{+\infty} x f(x) dx ,$$

sotto la condizione

$$\int_{-\infty}^{+\infty} |x| f(x) dx < \infty ,$$

cioè che l'integrale sopra esista e sia finito.

Se $X \sim \mathbf{U}([a, b])$ si ha

$$\mathbb{P}(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} .$$

Varianza. Dato un n.a. continuo X con densità di prob. $f(x)$, ponendo $\mathbb{P}(X) = m$, la varianza di X è definita da

$$\text{Var}(X) = \mathbb{P}[(X - m)^2] = \int_{-\infty}^{+\infty} (x - m)^2 f(x) dx ,$$

sotto la condizione che l'integrale a secondo membro esista e sia finito.

Deviazione standard. Ricordiamo che lo scarto quadratico medio (o scarto standard o deviazione standard) è dato da: $\sigma_X = \sqrt{Var(X)}$.

Se $X \sim \mathbf{U}([a, b])$, si ha

$$\begin{aligned} Var(X) &= \sigma_X^2 = \int_{-\infty}^{+\infty} (x - m)^2 f(x) dx = \\ &= \int_{-\infty}^{+\infty} x^2 f(x) dx - m^2 = \int_a^b x^2 \cdot \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \\ &= \dots = \frac{(b-a)^2}{12}, \end{aligned}$$

e quindi: $\sigma_X = \frac{b-a}{2\sqrt{3}}$.

Proprietà della previsione. Sia X un n.a. continuo ed Y un n.a. definito da

$$Y = h(X),$$

dove h è una funzione definita nel codominio di X . Allora si può provare, sotto l'ipotesi di esistenza di

$\mathbb{P}(Y)$, che risulta

$$\mathbb{P}(Y) = \mathbb{P}[h(X)] = \int_{-\infty}^{+\infty} h(x)f(x)dx.$$

Ad esempio, se $h(X) = X^2$, si ha $\mathbb{P}[h(X)] = \mathbb{P}(X^2)$.
Se $h(X) = (X - m)^2$, si ha

$$\mathbb{P}[h(X)] = \mathbb{P}[(X - m)^2] = \sigma_X^2.$$

Linearità della previsione. Sia X un n.a. continuo. Consideriamo $Y = cX + d$. Allora

$$\begin{aligned} \mathbb{P}(Y) &= \mathbb{P}(cX + d) = \int_{-\infty}^{+\infty} (cx + d)f(x)dx = \\ &= c \int_{-\infty}^{+\infty} xf(x)dx + d \underbrace{\int_{-\infty}^{+\infty} f(x)dx}_{=1} = c\mathbb{P}(X) + d. \end{aligned}$$

Osservazione. Le disuguaglianze di Markov e Cebicev valgono anche per n.a. continui. Inoltre

- $\min(X) \leq \mathbb{P}(X) \leq \max(X)$.

Distribuzione Esponenziale.

Un n.a. continuo X con densità di probabilità

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0, \\ 0 & \text{se } x < 0, \end{cases} \quad \lambda \in \mathbb{R}^+ \quad (5)$$

si dice che ha *distribuzione esponenziale* di parametro λ e si indica con $X \sim \text{Exp}(\lambda)$.

La distribuzione esponenziale viene utilizzata ad esempio quando X rappresenta

- il tempo di durata di un dispositivo (non soggetto ad usura);
- il tempo di attesa del verificarsi di un certo evento (arrivo di un cliente in una coda, arrivo di una telefonata).

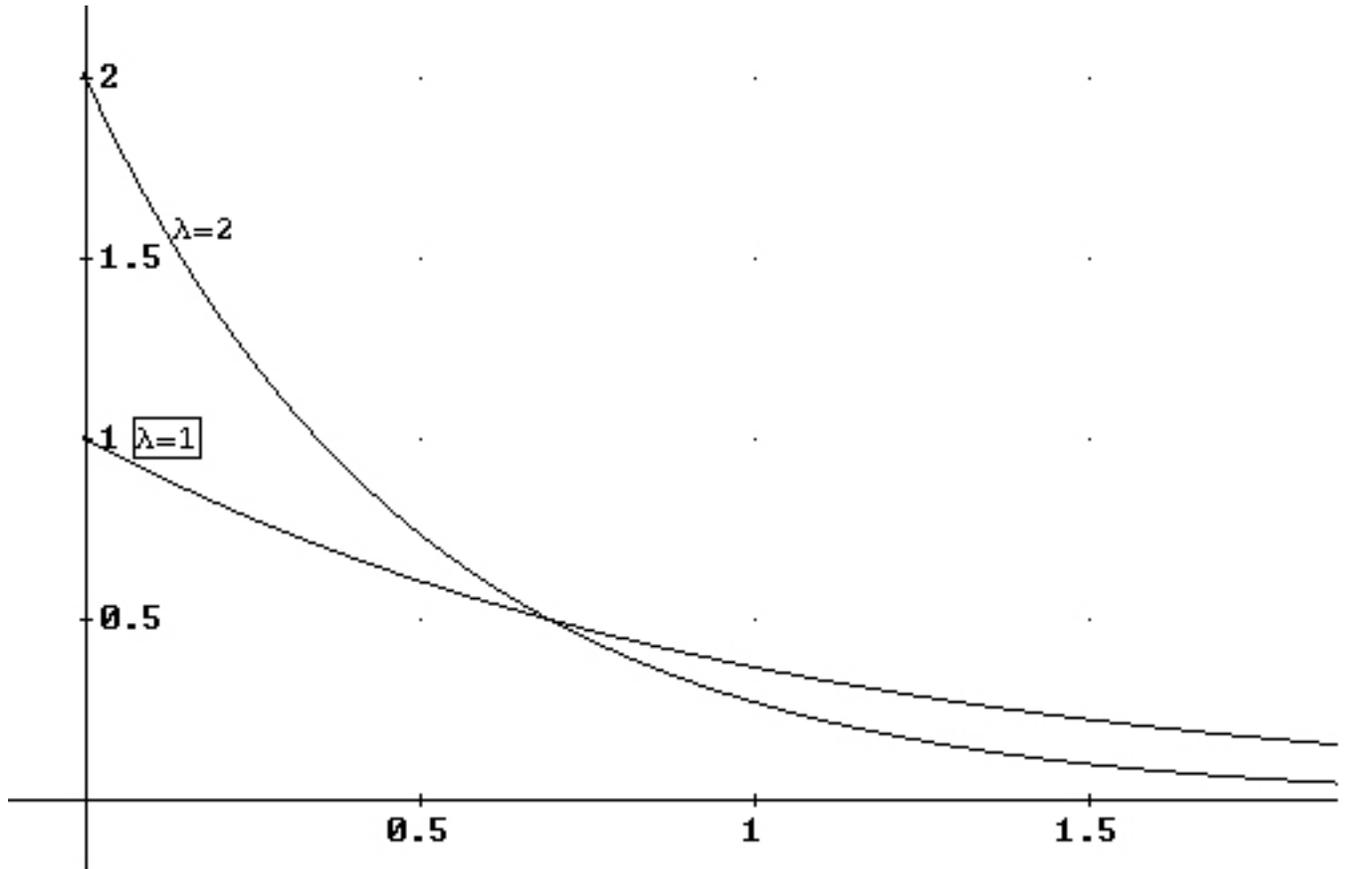


Figura 1: Esponenziale

L'area sotto la curva $y = f(x)$ al crescere del parametro λ si concentra sempre più verso l'origine.

Ricordiamo che l'area totale sotto la curva è uguale a 1. Infatti

$$\int_{-\infty}^{+\infty} f(x) = \int_0^{+\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^{+\infty} = 1.$$

La distribuzione esponenziale è l'analogo nel continuo della distribuzione geometrica. Infatti nel discreto il tempo di attesa può essere visto come il numero di prove necessarie per il verificarsi di un evento (numero di lanci di una moneta fino a quando per la prima volta esce testa).

La funzione di ripartizione è data da

$$F(x) = \begin{cases} \int_0^x \lambda e^{-\lambda t} dt, & \text{se } x > 0, \\ 0, & \text{se } x \leq 0. \end{cases}$$

Osservando che $\int_0^x \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^x = 1 - e^{-\lambda x}$, si ottiene

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{se } x > 0, \\ 0, & \text{se } x \leq 0. \end{cases}$$

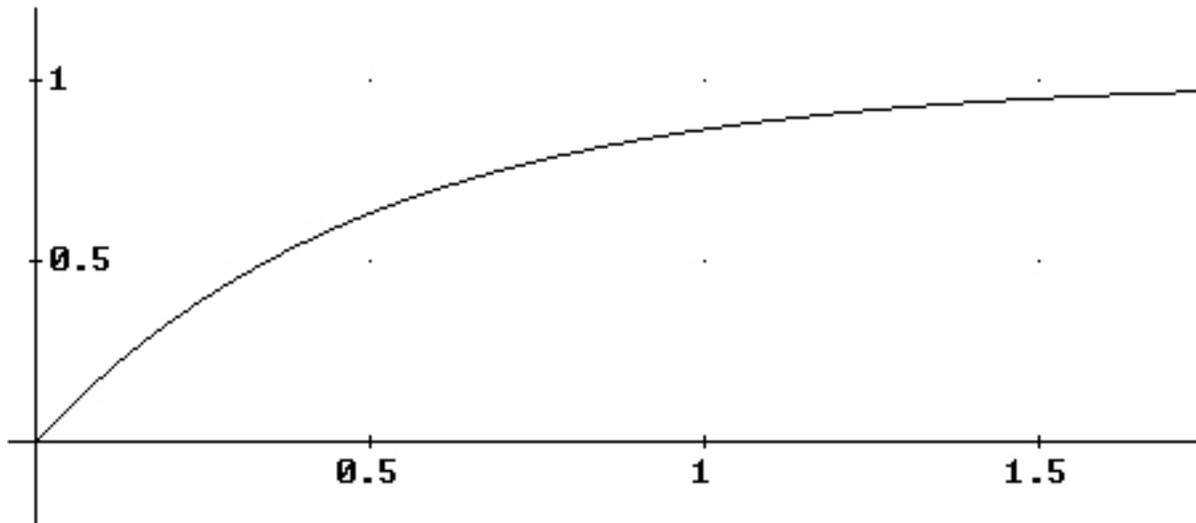


Figura 2: f.d.r. Exp

La funzione $S(x) = 1 - F(x) = P(X > x)$, detta funzione di sopravvivenza, è data da

$$S(x) = \begin{cases} e^{-\lambda x}, & \text{se } x > 0, \\ 1, & \text{se } x \leq 0, \end{cases}$$

La previsione è

$$\mathbb{P}(X) = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \dots = \frac{1}{\lambda},$$

mentre

$$\mathbb{P}(X^2) = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = \dots = \frac{2}{\lambda^2}.$$

Quindi la varianza e lo scarto sono rispettivamente

$$\begin{aligned} \text{Var}(X) &= \mathbb{P}(X^2) - [\mathbb{P}(X)]^2 = \frac{1}{\lambda^2}, \\ \sigma_X &= \frac{1}{\lambda}. \end{aligned}$$

Lo scarto quadratico medio coincide con la previsione.

Nota: il calcolo diretto della previsione e della varianza di X si può evitare utilizzando la funzione Gamma (vedi distribuzione beta).

Proprietà di Assenza di memoria. Un numero aleatorio continuo e non negativo X ha distribuzione esponenziale se e solo se vale la seguente proprietà (detta di assenza di memoria)

$$P(X > x_0 + x | X > x_0) = P(X > x), \quad \forall x_0, x \in \mathbb{R}_0^+. \quad (6)$$

Se X rappresenta il tempo (aleatorio) fino al guasto di un dispositivo, la proprietà di assenza di memoria

ha il seguente significato: supposto che il dispositivo non si guasti sino al tempo x_0 , la probabilità che non si guasti per un ulteriore tempo x è la stessa che il dispositivo non si guasti nell'intervallo $[0, x]$.

Tale proprietà è valida per le apparecchiature che, durante il loro funzionamento, non sono soggette ad usura (o, più realisticamente, quando l'usura è trascurabile).

dim. (\Rightarrow) *Hp*) $X \sim \text{Exp}(\lambda)$; *Th*) vale la (6).

$$\begin{aligned} P(X > x_0 + x | X > x_0) &= \frac{P(X > x_0 + x, X > x_0)}{P(X > x_0)} = \\ &= \frac{P(X > x_0 + x)}{P(X > x_0)} = \frac{S(x_0 + x)}{S(x_0)} = \frac{e^{-\lambda(x_0 + x)}}{e^{-\lambda x_0}} = e^{-\lambda x} = \\ &= S(x) = P(X > x). \end{aligned}$$

(\Leftarrow) *Hp*) vale la (6); *Th*) $X \sim \text{Exp}(\lambda)$.

Da quanto visto nella precedente dimostrazione la proprietà di assenza di memoria si può scrivere anche come:

$$\frac{S(x_0 + x)}{S(x_0)} = S(x),$$

cioè

$$S(x + x_0) = S(x)S(x_0).$$

Essendo la funzione di sopravvivenza definita come $1 - F(x)$, con $F(x)$ crescente, allora $S(x)$ è positiva e decrescente e quindi

$$S(x) > 0, \quad S'(x) < 0, \quad \forall x \in R.$$

Osserviamo che

$$\frac{S'(x+x_0)}{S(x+x_0)} = \frac{S(x_0)S'(x)}{S(x_0)S(x)} = \frac{S'(x)}{S(x)} = -\lambda, \quad \lambda > 0,$$

quindi

$$\begin{aligned} D[\ln(S(x))] &= \frac{S'(x)}{S(x)} = -\lambda \Rightarrow \\ \ln(S(x)) &= -\lambda x + k, \end{aligned}$$

allora

$$S(x) = e^{-\lambda x} e^k.$$

Essendo X un n.a. non negativo, si ha $S(0) = 1$, per cui $e^k = 1$. Allora

$$S(x) = e^{-\lambda x},$$

ovvero $X \sim Exp(\lambda)$.

Distribuzione Beta.

Dati due parametri r, s entrambi positivi, un n.a. continuo X con densità di probabilità data da

$$B_{r,s}(x) = \begin{cases} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1} (1-x)^{s-1}, & \text{se } x \in (0, 1), \\ 0, & \text{altrimenti} \end{cases} \quad (7)$$

si dice che ha distribuzione beta, di parametri r ed s , e si indica nel seguente modo: $X \sim B_{r,s}(x)$.

La funzione $\Gamma(\cdot)$ è così definita

$$\Gamma(c) = \int_0^{+\infty} x^{c-1} e^{-x} dx, \quad \forall x \in R^+. \quad (8)$$

Applicando l'integrazione per parti a $\Gamma(c+1)$ si ottiene

$$\Gamma(c+1) = c\Gamma(c)$$

infatti posto

$$\begin{aligned} h(x) &= x^c & h'(x) &= cx^{c-1} \\ g(x) &= -e^{-x} & g'(x) &= e^{-x} \end{aligned}$$

si ha

$$\begin{aligned}\Gamma(c+1) &= \int_0^{+\infty} x^c e^{-x} dx = \int_0^{+\infty} h(x)g'(x) dx = \\ &= [h(x)g(x)]_0^{+\infty} - \int_0^{+\infty} h'(x)g(x) dx = \\ &= \underbrace{[-x^c e^{-x}]_0^{+\infty}}_{=0} + c \int_0^{+\infty} x^{c-1} e^{-x} dx = c\Gamma(c).\end{aligned}$$

In particolare

$$\Gamma(1) = \int_0^{+\infty} x^{1-1} e^{-x} dx = \int_0^{+\infty} e^{-x} dx = 1$$

e quindi se consideriamo solo i valori interi di c si ha

$$\Gamma(n+1) = n\Gamma(n) = \dots = n!\Gamma(1) = n!$$

La funzione Γ applicata al numero intero n restituisce il fattoriale di $n - 1$.

La distribuzione $B_{r,s}(x)$ con $r, s = 1$ diviene la distribuzione $\mathbf{U}(0, 1)$, infatti

$$\begin{aligned}\frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1} (1-x)^{s-1} &= \\ = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} x^0 (1-x)^0 &= \frac{1!}{0!0!} = 1.\end{aligned}$$

Vediamo i grafici della funzione densità al variare dei parametri r, s . (File betagrafici.dvi o .pdf)

Si può dimostrare che

$$\int_0^1 x^{r-1}(1-x)^{s-1}dx = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)},$$

pertanto

$$\begin{aligned} \int_{-\infty}^{+\infty} B_{r,s}(x)dx &= \int_0^1 \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1}dx = \\ &= \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} = 1. \end{aligned}$$

La previsione di X è data da

$$\begin{aligned} \mathbb{P}(X) &= \int_{-\infty}^{+\infty} xB_{r,s}(x)dx = \int_0^1 \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^r(1-x)^{s-1}dx = \\ &= \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \frac{\Gamma(r+1)\Gamma(s)}{\Gamma(r+s+1)} = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \frac{(r)\Gamma(r)\Gamma(s)}{(r+s)\Gamma(r+s)} = \frac{r}{r+s}. \end{aligned}$$

In modo analogo si prova che

$$\mathbb{P}(X^2) = \frac{r(r+1)}{(r+s)(r+s+1)},$$

e quindi

$$\text{Var}(X) = \mathbb{P}(X^2) - [\mathbb{P}(X)]^2 = \frac{rs}{(r+s)^2(r+s+1)}.$$

Distribuzione normale standard

Un n.a. continuo X , con densità di probabilità

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}, \quad (9)$$

si dice che ha *distribuzione normale standard* (di parametri 0,1) e si indica con $X \sim N_{0,1} = N$. La densità $f(x)$ si indica con $N(x)$, mentre la funzione di ripartizione $F(x)$ si indica con $\Phi(x)$. Di tale funzione non è possibile dare un'espressione, ma si possono cercare soltanto alcuni valori riportati su apposite tavole.

Alcune proprietà:

1. il diagramma della densità ha un andamento a forma di campana (con il massimo nell'origine e due flessi in $x = -1$, $x = 1$) ed è simmetrico rispetto all'asse y , cioè $N(x)$ è una funzione pari ($N(-x) = N(x)$);
2. dalla simmetria di $N(x)$, per ogni $x \in \mathbb{R}$ si ha

$\Phi(-x) = 1 - \Phi(x)$, e quindi

$$\begin{aligned} P(|X| \leq x) &= P(-x \leq X \leq x) = \int_{-x}^x N(t) dt = \\ &= \Phi(x) - \Phi(-x) = 2\Phi(x) - 1; \end{aligned}$$

$$P(|X| > x) = 1 - P(|X| \leq x) = 2[1 - \Phi(x)];$$

3. in particolare

$$\Phi(1) \simeq 0.8413, \quad \Phi(2) \simeq 0.9772, \quad \Phi(3) \simeq 0.9987,$$

e quindi

$$P(|X| \leq 1) = 2\Phi(1) - 1 \simeq 0.6826;$$

$$P(|X| \leq 2) = 2\Phi(2) - 1 \simeq 0.9544;$$

$$P(|X| \leq 3) = 2\Phi(3) - 1 \simeq 0.9974.$$

Si può verificare che

$$\mathbb{P}(X) = \int_{-\infty}^{+\infty} xN(x)dx = \dots = 0,$$

$$\mathbb{P}(X^2) = \int_{-\infty}^{+\infty} x^2 N(x) dx = \dots = 1,$$

e quindi:

$$\text{Var}(X) = \mathbb{P}(X^2) = 1.$$

In generale, si dice che X ha una distribuzione normale di parametri m, σ , con $m \in \mathbb{R}, \sigma > 0$, se la densità di X ha la seguente forma:

$$f(x) = N_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \quad (10)$$

In simboli, si scrive: $X \sim N_{m,\sigma}$. La funzione di ripartizione si indica con $\Phi_{m,\sigma}(x)$.

Il diagramma della densità ha un andamento a forma di campana (con il massimo in $x = m$ e due flessi in $x = m - \sigma, x = m + \sigma$) ed è simmetrico rispetto alla retta $x = m$.

Se $Y = aX + b$, con $X \sim N_{m,\sigma}, a > 0$, indicando con G la funzione di ripartizione di Y e g la sua densità, si

ha:

$$G(y) = P(Y \leq y) = P(X \leq \frac{y-b}{a}) = \Phi_{m,\sigma}(\frac{y-b}{a}),$$

e quindi

$$g(y) = G'(y) = \Phi'_{m,\sigma}(\frac{y-b}{a}) \cdot \frac{1}{a} = \dots = N_{m_Y,\sigma_Y}(y),$$

dove

$$m_Y = am + b, \quad \sigma_Y = a\sigma. \quad (11)$$

Si può dimostrare che, se $a < 0$, risulta $Y \sim N_{m_Y,\sigma_Y}$, con

$$m_Y = am + b, \quad \sigma_Y = -a\sigma. \quad (12)$$

In altri termini, se dal n.a. X , con distribuzione normale, si passa al n.a. $Y = aX + b$, con $a \neq 0$, la distribuzione rimane di tipo normale, con i parametri che cambiano come indicato nella (11), oppure (12). In particolare, se $Y = \frac{X-m}{\sigma}$, si ha

$$m_Y = 0, \quad \sigma_Y = 1, \quad (13)$$

cioè la distribuzione diventa normale standard. Allora, tenendo conto che, se $Y = \frac{X-m}{\sigma}$, si ha $\mathbb{P}(Y) = 0$, $\sigma_Y = 1$, e che $X = \sigma Y + m$, si ottiene

$$\mathbb{P}(X) = \mathbb{P}(\sigma Y + m) = m, \quad \sigma_X^2 = \text{Var}(\sigma Y + m) = \sigma^2.$$

Pertanto i parametri m, σ sono rispettivamente la previsione e lo scarto quadratico medio. Lo stesso risultato si può ottenere con calcoli diretti, verificando che

$$\mathbb{P}(X) = \int_{-\infty}^{+\infty} x N_{m,\sigma}(x) dx = \dots = m,$$

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - m)^2 N_{m,\sigma}(x) dx = \dots = \sigma^2.$$

Se $X \sim N_{m,\sigma}$, osservando che

$$(X \leq x) \iff \left(\frac{X - m}{\sigma} \leq \frac{x - m}{\sigma} \right),$$

e che

$$\frac{X - m}{\sigma} \sim N,$$

si ottiene

$$\Phi_{m,\sigma}(x) = P(X \leq x) = P\left(\frac{X-m}{\sigma} \leq \frac{x-m}{\sigma}\right) = \Phi\left(\frac{x-m}{\sigma}\right).$$

Pertanto, utilizzando le tavole della distribuzione normale standard è possibile calcolare i valori di una distribuzione normale con parametri m, σ arbitrari.

Inoltre, per ogni $k > 0$, si ha

$$P(|X - m| \leq k\sigma) = \dots = 2\Phi(k) - 1.$$

Distribuzione Gamma

Un numero aleatorio X , continuo e non negativo, si dice che ha una distribuzione Gamma con parametri (positivi) c, λ , che indichiamo con il simbolo $X \sim G_{c,\lambda}$, se la densità di X è

$$f(x) = G_{c,\lambda}(x) = \begin{cases} \frac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Si ha:

$$IP(X) = \int_0^{+\infty} x \frac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x} dx = \dots = \frac{c}{\lambda}.$$

Inoltre

$$IP(X^2) = \int_0^{+\infty} x^2 \frac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x} dx = \dots = \frac{c}{\lambda^2} + \frac{c^2}{\lambda^2};$$

pertanto

$$Var(X) = IP(X^2) - [IP(X)]^2 = \frac{c}{\lambda^2}.$$

Osservazioni:

(i) se $c = 1$ si ha

$$G_{1,\lambda}(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0; \end{cases}$$

ovvero, la distribuzione Gamma di parametri $c = 1, \lambda$ è una distribuzione esponenziale di parametro λ ;

(ii) se $c < 1$ c'è un asintoto verticale in $x = 0$;

(iii) dati due numeri aleatori X, Y , con

$$X \sim G_{c_1, \lambda}, Y \sim G_{c_2, \lambda},$$

sotto opportune ipotesi (di indipendenza stocastica) si può verificare che

$$X + Y \sim G_{c_1 + c_2, \lambda}.$$

(iv) in particolare, dati n numeri aleatori T_1, \dots, T_n , indipendenti e con distribuzione esponenziale $G_{1, \lambda}$, posto $X = T_1 + \dots + T_n$, si ha

$$X \sim G_{1 + \dots + 1, \lambda} = G_{n, \lambda}.$$

Affidabilità

Ricordiamo che, dato un n. a. X non negativo con distribuzione esponenziale di parametro λ , vale la seguente proprietà di *assenza di memoria*

$$\begin{aligned} P(X > x + y | X > y) &= P(X > x) = \dots = \\ &= e^{-\lambda x}, \quad \forall x > 0, y > 0. \end{aligned} \tag{14}$$

Dalla (14), considerando l'evento contrario, si ottiene

$$\begin{aligned} P(X \leq x + y | X > y) &= P(X \leq x) = \\ &= 1 - e^{-\lambda x}, \quad \forall x > 0, y > 0, \end{aligned} \tag{15}$$

e, più in generale,

$$\begin{aligned}
 P(x + y < X \leq x + y + \Delta x \mid X > y) &= \\
 &= P(x < X \leq x + \Delta x) = F(x + \Delta x) - F(x) = \\
 &= (1 - e^{-\lambda(x+\Delta x)}) - (1 - e^{-\lambda x}) = \\
 &= e^{-\lambda x}(1 - e^{-\lambda \Delta x}), \quad \forall x > 0, y > 0.
 \end{aligned}
 \tag{16}$$

Se la distribuzione di X non è esponenziale le formule precedenti non valgono e, per fissati valori x, y , potrà risultare

$$P(X > x + y \mid X > y) < P(X > x), \tag{17}$$

oppure

$$P(X > x + y \mid X > y) > P(X > x), \tag{18}$$

o in casi particolari

$$P(X > x + y \mid X > y) = P(X > x). \tag{19}$$

Se X rappresenta il tempo aleatorio fino al guasto di una data apparecchiatura, il fatto che vale la (14) corrisponde all'assenza di usura, mentre la (17) e la (18) corrispondono rispettivamente al caso di usura positiva (invecchiamento dell'apparecchiatura) e di usura negativa (ringiovanimento dell'apparecchiatura).

Indicando con $f(x)$ la densità di probabilità e con $S(x)$ la funzione di sopravvivenza, se consideriamo l'evento condizionato $(x < X \leq x + \Delta x | X > x)$, con Δx abbastanza piccolo, si ha (sotto opportune condizioni)

$$\begin{aligned} P(x < X \leq x + \Delta x | X > x) &= \frac{P(x < X \leq x + \Delta x)}{P(X > x)} = \\ &= \frac{\int_x^{x+\Delta x} f(x) dx}{S(x)} \simeq \frac{f(x)\Delta x}{S(x)} = h(x)\Delta x. \end{aligned} \tag{20}$$

La funzione non negativa $h(x) = \frac{f(x)}{S(x)}$ si chiama funzione di rischio (o intensità, o tasso di avaria) di X e, come abbiamo visto, permette di approssimare $P(x < X \leq x + \Delta x | X > x)$ con $h(x)\Delta x$.

Assegnare $f(x)$ è equivalente ad assegnare $h(x)$.

Infatti, data la densità $f(x)$, si ha

$$S(x) = \int_x^{+\infty} f(t)dt, \quad h(x) = \frac{f(x)}{\int_x^{+\infty} f(t)dt}.$$

Viceversa, data la funzione di rischio $h(x)$, si ha

$$h(x) = \frac{f(x)}{S(x)} = -\frac{S'(x)}{S(x)},$$

e quindi

$$\frac{S'(x)}{S(x)} = D \ln S(x) = -h(x).$$

Allora

$$\ln S(x) = -\int_0^x h(t)dt + c,$$

dove c è una costante arbitraria. Ricordando che per un n.a. non negativo è $S(0) = 1$, si ha $\ln S(0) = c = 0$ e quindi

$$S(x) = e^{-\int_0^x h(t)dt}, \quad (21)$$

da cui segue

$$f(x) = h(x)S(x) = h(x)e^{-\int_0^x h(t)dt}. \quad (22)$$

La funzione di rischio, oltre ad essere non negativa, soddisfa la seguente proprietà

$$\int_0^{+\infty} h(x) dx = +\infty .$$

Infatti tale condizione segue dalla (21), osservando che

$$\lim_{x \rightarrow +\infty} S(x) = \lim_{x \rightarrow +\infty} \int_x^{+\infty} f(t) dt = 0 .$$

Osserviamo anche che, come appare dalla (20), se la funzione di rischio $h(x)$ è crescente l'apparecchiatura subisce un'usura positiva (invecchiamento). Infatti, si ha

$$\begin{aligned} P(x < X \leq x + \Delta x \mid X > x) &= \frac{S(x) - S(x + \Delta x)}{S(x)} = \\ &= \dots = 1 - e^{-\int_x^{x+\Delta x} h(t) dt} , \end{aligned}$$

da cui, se $h(x)$ è crescente, per $x_1 < x_2$ si ha

$$\int_{x_1}^{x_1 + \Delta x} h(t) dt < \int_{x_2}^{x_2 + \Delta x} h(t) dt .$$

Allora

$$1 - e^{-\int_{x_1}^{x_1+\Delta x} h(t)dt} < 1 - e^{-\int_{x_2}^{x_2+\Delta x} h(t)dt},$$

e quindi

$$\begin{aligned} P(x_1 < X \leq x_1 + \Delta x \mid X > x_1) < \\ < P(x_2 < X \leq x_2 + \Delta x \mid X > x_2). \end{aligned}$$

Con lo stesso ragionamento, si dimostra che se $h(x)$ è decrescente c'è usura negativa (ringiovanimento).

Infine, il caso in cui $h(x)$ è costante (assenza di usura) è caratteristico della distribuzione esponenziale. Infatti, se

$$f(x) = \lambda e^{-\lambda x}, \quad \forall x \geq 0,$$

allora

$$h(x) = \frac{f(x)}{S(x)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda.$$

Viceversa, se $h(x) = \text{cost} = \lambda > 0$, allora

$$f(x) = h(x)S(x) = \lambda e^{-\int_0^x \lambda dt} = \lambda e^{-\lambda x}, \quad \forall x \geq 0.$$

Alcuni modelli particolari di funzioni di rischio sono:

$$(a) \quad h(x) = \alpha + \beta x; \quad (b) \quad h(x) = cx^\beta .$$

Nel caso (a) (modello lineare), essendo

$$h(x) \geq 0, \quad \int_{-\infty}^{+\infty} h(x)dx = +\infty ,$$

segue che le costanti α e β devono essere non negative ed almeno una positiva, cioè devono soddisfare le condizioni

$$\alpha \geq 0, \quad \beta \geq 0, \quad \alpha + \beta > 0 .$$

Pertanto, nel caso $\beta > 0$, $h(x)$ è crescente, mentre nel caso $\beta = 0$, $h(x)$ è costante e la corrispondente distribuzione è esponenziale di parametro α . Con il modello lineare, quindi, non si può rappresentare la situazione di usura negativa.

Nel caso (b), dalle proprietà di $h(x)$ segue intanto che dev'essere $c > 0$. Inoltre, non può essere $\beta \leq -1$,

altrimenti, per ogni fissato $x > 0$, si avrebbe

$$\int_0^x ct^\beta dt = +\infty,$$

e quindi risulterebbe

$$S(x) = e^{-\int_0^x ct^\beta dt} = 0, \quad \forall x > 0.$$

Pertanto, dev'essere $\beta > -1$ e possiamo distinguere tre casi:

$$(i) \quad -1 < \beta < 0; \quad (ii) \quad \beta > 0; \quad (iii) \quad \beta = 0.$$

Nel primo caso $h(x)$ è decrescente e quindi siamo in presenza di usura negativa; nel secondo caso $h(x)$ è crescente (usura positiva); nel terzo caso $h(x)$ è costante (assenza di usura) e la distribuzione è esponenziale di parametro c .

La distribuzione di probabilità corrispondente alla funzione di rischio $h(x) = cx^\beta$ è detta distribuzione di Weibull ed ha la seguente densità

$$f(x) = cx^\beta e^{-\int_0^x ct^\beta dt} = cx^\beta e^{-\frac{c}{\beta+1}x^{\beta+1}}.$$

Esercizio. Un sistema S è costituito da due dispositivi A e B in parallelo funzionanti simultaneamente (e quindi S funziona finchè almeno uno dei due dispositivi funziona). Siano X e Y i tempi aleatori di durata di A e B , rispettivamente, e supponiamo che le loro densità siano:

$$f_1(x) = e^{-x}, \quad x \geq 0, \quad f_2(y) = 2e^{-2y}, \quad y \geq 0,$$

con $f_1(x) = f_2(y) = 0$ per $x < 0, y < 0$. Si supponga inoltre che valga la condizione

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall (x, y).$$

- (i) calcolare la funzione di rischio $h(z)$ del tempo aleatorio Z fino al guasto di S .
- (ii) è possibile calcolare, con i dati del problema, la probabilità che A si guasti prima di B ?

Soluzione.

- (i) ricordando che, per ogni $\lambda > 0, x \geq 0$, si ha

$$\int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x},$$

e osservando che $(Z \leq z) = (X \leq z, Y \leq z)$, segue

$$\begin{aligned} P(Z \leq z) &= G(z) = P(X \leq z, Y \leq z) = \\ &= P(X \leq z)P(Y \leq z) = (1 - e^{-z})(1 - e^{-2z}) = \\ &= 1 - e^{-z} - e^{-2z} + e^{-3z}. \end{aligned}$$

Allora, indicando con $g(z)$ la densità di Z , si ha

$$\begin{aligned} h(z) &= \frac{g(z)}{S(z)} = \frac{G'(z)}{1-G(z)} = \frac{e^{-z} + 2e^{-2z} - 3e^{-3z}}{e^{-z} + e^{-2z} - e^{-3z}} = \\ &= \frac{e^{2z} + 2e^z - 3}{e^{2z} + e^z - 1} = 1 + \frac{e^z - 2}{e^{2z} + e^z - 1}. \end{aligned}$$

La funzione di rischio $h(z)$ è crescente in $[0, 2 + \sqrt{5})$ ed è decrescente in $(2 + \sqrt{5}, +\infty)$.

Quindi, per i singoli dispositivi A e B non c'è usura, mentre per il sistema S è come se ci fosse usura *positiva* in $[0, 2 + \sqrt{5})$ ed usura *negativa* in $(2 + \sqrt{5}, +\infty)$. Quindi, il termine usura va anche inteso in senso convenzionale.

In ogni caso, l'affidabilità di S è superiore rispetto ai

singoli dispositivi. Infatti:

$$\begin{aligned} P(Z > z) = S(z) &= e^{-z} + e^{-2z} - e^{-3z} = \\ &= e^{-z} + (e^{-2z} - e^{-3z}) = e^{-2z} + (e^{-z} - e^{-3z}), \end{aligned}$$

e quindi

$$S_1(z) = P(X > z) = e^{-z} < S(z);$$

$$S_2(z) = P(Y > z) = e^{-2z} < S(z).$$

In particolare:

$$0 < h(z) < 1, \text{ per } 0 < z < \log 2,$$

$$1 < h(z) < 2, \text{ per } z > \log 2,$$

e osservando che $h_1(z) = 1$, $h_2(z) = 2$, $\forall z \geq 0$, si ha:

$$h(z) < h_1(z), \forall z < \log 2;$$

$$h_1(z) < h(z) < h_2(z), \forall z > \log 2.$$

(ii) per quanto riguarda la seconda domanda, la risposta è sì e no. Si ha

$$p = P(X < Y) = P[(X, Y) \in \{(x, y) : x < y\}] = \dots = \frac{1}{3}.$$

I dettagli saranno esaminati ampliando il discorso al caso di distribuzioni multiple (per vettori aleatori).

Vettori Aleatori

In molti esperimenti aleatori, indicando con Ω l'insieme dei possibili risultati, al generico risultato dell'esperimento, $\omega \in \Omega$, sono associati n numeri reali x_1, \dots, x_n , con $n \geq 2$, che costituiscono i valori di n numeri aleatori X_1, \dots, X_n . Tali n. a. sono le componenti di un vettore aleatorio $X = (X_1, \dots, X_n)$, che può essere visto come una funzione definita su Ω a valori in \mathbb{R}^n , cioè

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R}^n \\ \omega &\longrightarrow x = X(\omega). \end{aligned}$$

Due casi importanti da considerare sono i v. a. discreti e v.a. continui.

Vettori aleatori discreti. Un vettore aleatorio $X = (X_1, X_2, \dots, X_n)$ si dice discreto se esiste un insieme finito o numerabile $C \subset \mathbb{R}^n$ tale che

- $P(X = x) > 0, \quad \forall x \in C,$
- $P(X = x) = 0, \quad \forall x \notin C,$

dove, ponendo $x = (x_1, \dots, x_n)$, l'evento $(X = x)$ rappresenta l'evento

$$(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n).$$

Vettori aleatori continui. Un vettore aleatorio $X = (X_1, X_2, \dots, X_n)$ si dice continuo se

- $P(X = x) = 0, \quad \forall x \in \mathbb{R}^n,$

- $\exists f : \mathbb{R}^n \rightarrow \mathbb{R}$ tale che

(i) $f(x) \geq 0, \quad \forall x \in \mathbb{R}^n;$

(ii) $\forall A \subseteq \mathbb{R}^n$, misurabile secondo Peano - Jordan, si ha

$$\begin{aligned} P(X \in A) &= P(A) = \int_A f(x) dx = \\ &= \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n. \end{aligned}$$

La funzione $f(x)$ si chiama densità di probabilità congiunta del v. a. X .

Proprietà di normalizzazione:

$$\begin{aligned} \int_{\mathbb{R}^n} f(x) dx &= \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_n = 1. \end{aligned}$$

La funzione di ripartizione congiunta di $X = (X_1, \dots, X_n)$ è definita nel seguente modo:

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Nel caso continuo si ha:

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n.$$

Analizziamo in dettaglio il caso discreto, con $n = 2$. Per semplicità di notazione, indichiamo con (X, Y) il v.a. (X_1, X_2) .

(tale notazione può essere utilizzata anche nel caso $n > 2$, indicando con X ed Y due sottovettori del vettore aleatorio (X_1, \dots, X_n)).

Distribuzioni marginali. Sia

$$X \in C_x, Y \in C_y, (X, Y) \in C \subseteq C_x \times C_y.$$

Ricordiamo che C è l'insieme (al più numerabile) dei punti di \mathbb{R}^2 che hanno probabilità positiva. Quindi per ogni coppia $(x_h, y_k) \in C$ si ha

$$P(X = x_h, Y = y_k) = p_{x_h, y_k} > 0.$$

Fissato un punto $x_h \in C_x$, osservando che

$$\Omega = \bigvee_{y_k \in C_y} (Y = y_k),$$

possiamo decomporre l'evento $(X = x_h)$ nel seguente modo

$$\begin{aligned} (X = x_h) &= (X = x_h) \wedge \Omega = \\ &= (X = x_h) \wedge \left[\bigvee_{y_k \in C_y} (Y = y_k) \right] = \\ &= \bigvee_{y_k \in C_y} (X = x_h, Y = y_k). \end{aligned}$$

Quindi, $\forall x_h \in C_x$, si ha

$$\begin{aligned} P(X = x_h) &= p_{x_h} = \sum_{y_k \in C_y} P(X = x_h, Y = y_k) = \\ &= \sum_{y_k} p_{x_h, y_k}; \quad (\text{distribuzione marginale di } X). \end{aligned}$$

In modo analogo si ottiene

$$\begin{aligned} P(Y = y_k) &= p_{y_k} = \sum_{x_h \in C_x} P(X = x_h, Y = y_k) = \\ &= \sum_{x_h} p_{x_h, y_k}; \quad (\text{distribuzione marginale di } Y). \end{aligned}$$

Distribuzioni marginali condizionate.

$$\begin{aligned} p_{x_h|y_k} &= P(X = x_h | Y = y_k) = \\ &= \frac{P(X=x_h, Y=y_k)}{P(Y=y_k)} = \frac{\overbrace{p_{x_h, y_k}}^{\text{congiunta}}}{\underbrace{p_{y_k}}_{\text{marginale}}}. \end{aligned}$$

La distribuzione $\{p_{x_h|y_k}, x_h \in C_X\}$ si chiama distribuzione marginale di X condizionata all'evento $(Y = y_k)$.

In maniera analoga, la distribuzione $\{p_{y_k|x_h}, y_k \in C_Y\}$ si chiama distribuzione marginale di Y condizionata all'evento $(X = x_h)$, ovvero

$$\begin{aligned} p_{y_k|x_h} &= P(Y = y_k | X = x_h) = \\ &= \frac{P(X=x_h, Y=y_k)}{P(X=x_h)} = \frac{\overbrace{p_{x_h, y_k}}^{\text{congiunta}}}{\underbrace{p_{x_h}}_{\text{marginale}}} \end{aligned}$$

Dalle ultime relazioni, per il teorema delle probabilità composte, si ottiene:

$$\begin{aligned} P(X = x_h, Y = y_k) &= P(Y = y_k | X = x_h)P(X = x_h) = \\ &= P(X = x_h | Y = y_k)P(Y = y_k), \end{aligned}$$

ovvero

$$p_{x_h, y_k} = p_{x_h|y_k} \cdot p_{y_k} = p_{y_k|x_h} p_{x_h}.$$

Osserviamo che, in generale, risulta

$$p_{x_h, y_k} \neq p_{x_h} p_{y_k}.$$

Indipendenza stocastica. I numeri aleatori X, Y si dicono stocasticamente indipendenti (in breve, indipendenti) se, $\forall (x_h, y_k)$, vale

$$P(X = x_h, Y = y_k) = P(X = x_h)P(Y = y_k),$$

ovvero la distribuzione congiunta è data dal prodotto delle marginali

$$p_{x_h, y_k} = p_{x_h} p_{y_k}, \quad \forall (x_h, y_k).$$

Quindi, se X, Y sono indipendenti, le distribuzioni condizionate coincidono con le marginali

$$p_{x_h|y_k} = p_{x_h}, \quad p_{y_k|x_h} = p_{y_k}.$$

Esempio. Si lancia due volte un dado, definendo

X = risultato del primo lancio;

Y = risultato del secondo lancio.

Ovviamente, X, Y sono indipendenti e quindi, per ogni coppia $(m, n) \in \{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\}$, si ha

$$P(X = m, Y = n) = P(X = m)P(Y = n) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

$Y \setminus X$	1	2	3	4	5	6	$P(Y = n)$
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
$P(X = m)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	

Esempio. Due estrazioni senza restituzione da un'urna contenente cinque palline numerate da 1 a 5. I numeri

aleatori

X = risultato della prima estrazione,
 Y = risultato della seconda estrazione,

non sono indipendenti. Infatti, ad esempio

$$P(X = 2, Y = 1) = \frac{1}{20},$$

$$P(X = 2)P(Y = 1) = \frac{1}{5} \cdot \frac{1}{5} = \frac{1}{25} \neq \frac{1}{20}.$$

Esempio. Siano X, Y due n. a. indipendenti con distribuzione di Poisson, rispettivamente di parametri λ_1 e λ_2 , ovvero

$$X \sim \mathcal{P}(\lambda_1), \quad Y \sim \mathcal{P}(\lambda_2).$$

Calcoliamo la distribuzione di probabilità del n. a.

$Z = X + Y$. Osserviamo che, fissato $n \in \mathbb{N}$, si ha

$$\begin{aligned} P(Z = n) &= P[\bigvee_{i=0}^n (X = i, Y = n - i)] = \\ &= \sum_{i=0}^n P(X = i, Y = n - i) = \\ &= \sum_{i=0}^n e^{-\lambda_1} \frac{\lambda_1^i}{i!} \cdot e^{-\lambda_2} \frac{\lambda_2^{n-i}}{(n-i)!} = \dots = \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!}. \end{aligned}$$

Pertanto: $Z \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

Inoltre, si può verificare che

$$X|(Z = n) \sim \mathcal{B}(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}),$$

$$Y|(Z = n) \sim \mathcal{B}(n, \frac{\lambda_2}{\lambda_1 + \lambda_2}).$$

Teorema. Se X ed Y sono indipendenti, si ha:
 $Cov(X, Y) = 0$.

Dim.: Supponiamo che, $\forall (x_h, y_k) \in C$, sia

$$P(X = x_h, Y = y_k) = P(X = x_h)P(Y = y_k).$$

Allora, segue

$$\begin{aligned}\mathbb{P}(XY) &= \sum_{x_h} \sum_{y_k} x_h y_k p_{x_h, y_k} = \sum_{x_h} \sum_{y_k} x_h y_k p_{x_h} p_{y_k} = \\ &= \left(\sum_{x_h} x_h p_{x_h} \right) \left(\sum_{y_k} y_k p_{y_k} \right) = \mathbb{P}(X) \mathbb{P}(Y),\end{aligned}$$

e quindi: $Cov(X, Y) = 0$.

Osserviamo che il viceversa non vale, come mostra il seguente controesempio.

Esempio. Si consideri il seguente vettore aleatorio (X, Y) , con la distribuzione congiunta riportata nella tabella:

$Y \setminus X$	-1	0	1
-1	a	/	a
0	/	b	/
1	a	/	a

Si ha

$$C = \{(-1, -1), (-1, 1), (0, 0), (1, -1), (1, 1)\},$$

con $P(X = 0, Y = 0) = b$ e $P(X = x, Y = y) = a$ negli altri casi. Ovviamente, deve essere:

$$4a + b = 1, \quad a \geq 0, \quad b \geq 0.$$

Come si può verificare, si ha

$$X \in \{-1, 0, 1\}, \quad Y \in \{-1, 0, 1\}, \quad XY \in \{-1, 0, 1\},$$

con

$$P(X = -1) = P(Y = -1) = P(XY = -1) = 2a,$$

$$P(X = 0) = P(Y = 0) = P(XY = 0) = b,$$

$$P(X = 1) = P(Y = 1) = P(XY = 1) = 2a.$$

Pertanto X, Y ed XY hanno la stessa distribuzione di probabilità. Inoltre

$$\mathbb{P}(X) = \mathbb{P}(Y) = \mathbb{P}(XY) = 0,$$

e quindi $Cov(X, Y) = 0$, ovvero X, Y sono incorrelati. Però X ed Y non sono indipendenti, in quanto risulta

$$P(X = x, Y = y) \neq P(X = x)P(Y = y),$$

ad esempio:

$$P(X = 0, Y = 0) = b \neq P(X = 0)P(Y = 0) = b \cdot b = b^2.$$

Vettori aleatori continui: distribuzioni marginali e condizionate.

Dato un vettore aleatorio continuo (X_1, \dots, X_n) , sia $f(x_1, \dots, x_n)$ la sua densità congiunta. Le densità marginali $f_1(x_1), \dots, f_n(x_n)$ dei n. a. X_1, \dots, X_n sono date dalle seguenti formule:

$$\begin{aligned} f_i(x_i) &= \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n, \end{aligned}$$

$$i = 1, \dots, n.$$

Ovvero, per calcolare $f_i(x_i)$ si integra $f(x_1, \dots, x_n)$ rispetto alle variabili $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$.

Consideriamo in particolare il caso $n = 2$, indicando con (X, Y) il v. a. (X_1, X_2) e con $f(x, y)$ la densità congiunta. Si ha

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y) dy, \quad f_2(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

Le densità condizionate (di $Y|x$ ed $X|y$), per fissati valori x, y ed assumendo $f_1(x) > 0$, $f_2(y) > 0$, sono definite nel seguente modo

$$f_2(y|x) = \frac{f(x, y)}{f_1(x)}, \quad f_1(x|y) = \frac{f(x, y)}{f_2(y)}.$$

Pertanto

$$f(x, y) = f_1(x)f_2(y|x) = f_2(y)f_1(x|y).$$

Se risulta

$$f(x, y) = f_1(x)f_2(y), \quad \forall (x, y)$$

i n. a. si dicono stocasticamente indipendenti e in questo caso si ha

$$f_2(y|x) = f_2(y), \quad \forall y; \quad f_1(x|y) = f_1(x), \quad \forall x,$$

cioè le densità condizionate coincidono con le densità marginali.

Osserviamo che la relazione di indipendenza tra X e Y può essere definita anche richiedendo che valga

$$F(x, y) = F_1(x)F_2(y), \quad \forall (x, y),$$

cioè

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall (x, y).$$

Come già visto nel caso discreto, si può dimostrare che, se X e Y sono indipendenti, segue che sono incorrelati, mentre il viceversa non vale. Infatti, assumendo

$f(x, y) = f_1(x)f_2(y)$, $\forall (x, y)$, si ottiene

$$\begin{aligned}\mathbb{P}(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dxdy = \dots \\ &= \left(\int_{-\infty}^{+\infty} xf_1(x)dx\right)\left(\int_{-\infty}^{+\infty} yf_2(y)dy\right) = \mathbb{P}(X)\mathbb{P}(Y),\end{aligned}$$

e quindi $Cov(X, Y) = 0$. Per mostrare attraverso un controesempio che il viceversa non vale, introduciamo la distribuzione uniforme su un insieme $A \subset \mathbb{R}^2$, limitato e misurabile.

Si dice che (X, Y) ha distribuzione uniforme su A , in simboli

$$(X, Y) \sim U(A),$$

se la densità congiunta assume un valore costante $k > 0$ su A ed è nulla altrove. Imponendo la condizione

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)dxdy = 1,$$

ovvero

$$\int \int_A f(x, y)dxdy = 1,$$

si ottiene $k = \frac{1}{\mu(A)}$, dove $\mu(A)$ è l'area di A .

Esempio. Supponiamo che $(X, Y) \sim U(C)$, dove C è il cerchio di raggio 1 e centro nell'origine. Allora

$$f(x, y) = \frac{1}{\pi}, \quad (x, y) \in C,$$

con $f(x, y) = 0$ altrove. Si dimostra che

$$f_1(x) = \frac{2}{\pi} \sqrt{1 - x^2}, \quad x \in [-1, 1],$$

con $f_1(x) = 0$ altrove. Inoltre

$$f_2(y) = \frac{2}{\pi} \sqrt{1 - y^2}, \quad y \in [-1, 1],$$

con $f_2(y) = 0$ altrove. Allora $\mathbb{P}(X) = \mathbb{P}(Y) = 0$. Inoltre

$$\mathbb{P}(XY) = \int \int_C xy f(x, y) dx dy = \dots = 0,$$

pertanto X e Y sono incorrelati. D'altra parte $f(x, y) \neq f_1(x)f_2(y)$, pertanto X e Y non sono indipendenti.

Distribuzione multinomiale

Si considerino n ripetizioni di un esperimento aleatorio, con $m + 1$ possibili risultati in ciascuna ripetizione. Ad esempio, si consideri un'urna contenente N palline, delle quali p_0N sono segnate con il numero 0, p_1N sono segnate con il numero 1, ..., p_mN sono segnate con il numero m , dove

$$p_0 + p_1 + \cdots + p_m = 1, \quad p_k \geq 0, \quad \forall k \in \{0, 1, \dots, m\}.$$

Supposto di effettuare n estrazioni con restituzione da tale urna, definiamo i seguenti eventi e numeri aleatori: $E_k^{(i)}$ = *nell' i -ma prova viene estratta una pallina segnata con il numero k , $i = 1, \dots, n$,*

$$X_k = |E_k^{(1)}| + \cdots + |E_k^{(n)}|, \quad k = 0, 1, \dots, m.$$

Ovviamente: $X_0 = n - (X_1 + \cdots + X_m)$. Inoltre, per ogni $i = 1, \dots, n$, la famiglia $\{E_0^{(i)}, \dots, E_m^{(i)}\}$ forma una partizione di Ω e si ha:

1. $P(E_k^{(i)}) = p_k, \quad \forall k \in \{0, 1, \dots, m\};$

2. gli eventi relativi a partizioni distinte, cioè associati a prove distinte, sono stocasticamente indipendenti.

Proponiamoci di calcolare la distribuzione di probabilità del vettore aleatorio discreto (X_1, \dots, X_m) . Come si può verificare, l'evento $(X_1 = x_1, \dots, X_m = x_m)$ è possibile se e solo se x_1, \dots, x_m sono dei valori interi non negativi tali che: $x_1 + \dots + x_m \leq n$. Posto $x_0 = n - (x_1 + \dots + x_m)$, si può verificare che:

a) il numero di costituenti favorevoli all'evento $(X_1 = x_1, \dots, X_m = x_m)$ è pari al coefficiente multinomiale $\frac{n!}{x_0!x_1!\dots x_m!}$;

b) ognuno di tali costituenti ha probabilità $p_0^{x_0} p_1^{x_1} \dots p_m^{x_m}$; pertanto:

$$P(X_1 = x_1, \dots, X_m = x_m) = \frac{n!}{x_0!x_1!\dots x_m!} p_0^{x_0} p_1^{x_1} \dots p_m^{x_m}.$$

La distribuzione di (X_1, \dots, X_m) si dice *multinomiale* di parametri n, p_1, \dots, p_m .

Per $m = 1$ si ottiene in particolare la distribuzione binomiale di parametri n, p_1 .

Rette di regressione

Dato un vettore aleatorio (X, Y) , cerchiamo una retta di equazione $y = a + bx$ che meglio si adatti alla distribuzione di probabilità congiunta di (X, Y) , ovvero che risulti più vicina possibile a tale distribuzione. Da un certo punto di vista, si potrebbe pensare di voler stimare Y mediante una funzione lineare $a + bX$, con i coefficienti a, b da determinare sulla base di un opportuno criterio. Un criterio ben noto in statistica è il metodo dei minimi quadrati che consiste nel cercare i valori a, b che rendono minima la previsione del numero aleatorio $(Y - a - bX)^2$. La retta che si ottiene si chiama retta di regressione di Y su X . Considerando il caso continuo e ponendo $\mathbb{P}[(Y - a - bX)^2] = g(a, b)$, se la densità congiunta è $f(x, y)$, si ha (applicando la linearità della previsione)

$$g(a, b) = \int \int_{\mathbb{R}^2} (y - a - bx)^2 f(x, y) dx dy =$$

$$\mathbb{P}(Y^2) + a^2 + b^2\mathbb{P}(X^2) - 2a\mathbb{P}(Y) - 2b\mathbb{P}(XY) + 2ab\mathbb{P}(X).$$

Uguagliando a zero le derivate parziali di $g(a, b)$ rispetto ad a, b (indicando con $m_1, m_2, \sigma_1, \sigma_2$ le previsioni e

gli scarti standard di X e Y , e con ρ il coefficiente di correlazione) si ha

$$\begin{cases} \frac{\partial g}{\partial a} = 2a - 2m_2 + 2bm_1 = 0, \\ \frac{\partial g}{\partial b} = 2b(m_1^2 + \sigma_1^2) - 2(m_1m_2 + \rho\sigma_1\sigma_2) + 2am_1 = 0. \end{cases}$$

Ricavando a dalla prima equazione ($a = m_2 - bm_1$) e risolvendo rispetto a b la seconda, si ottiene

$$a = m_2 - \rho \frac{\sigma_2}{\sigma_1} m_1, \quad b = \rho \frac{\sigma_2}{\sigma_1}.$$

Pertanto, l'equazione della retta di regressione di Y su X è data da

$$y = m_2 + \rho \frac{\sigma_2}{\sigma_1} (x - m_1).$$

Simmetricamente, l'equazione della retta di regressione di X su Y è

$$x = m_1 + \rho \frac{\sigma_1}{\sigma_2} (y - m_2),$$

che si può scrivere

$$y = m_2 + \frac{1}{\rho} \cdot \frac{\sigma_2}{\sigma_1}(x - m_1).$$

Le due rette si incontrano nel punto di coordinate (m_1, m_2) e, nel caso $\rho = 0$, sono perpendicolari e di equazioni: $y = m_2$, $x = m_1$.

Se $|\rho| = 1$, le due rette coincidono ed hanno equazione (a seconda che sia $\rho = 1$ oppure $\rho = -1$)

$$y = m_2 \pm \frac{\sigma_2}{\sigma_1}(x - m_1).$$

Vett. aleatori con distribuzione normale

Un vettore aleatorio continuo (X, Y) ha una distribuzione normale bidimensionale, di parametri $m_1, m_2, \sigma_1, \sigma_2, \rho$, con $\sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1$, se ha la seguente densità di probabilità

$$f(x, y) = k e^{-\frac{1}{2(1-\rho^2)} \cdot Q(x, y)},$$

dove $k = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$ e $Q(x, y)$ è la seguente espressione

$$\left(\frac{x - m_1}{\sigma_1}\right)^2 - 2\rho \left(\frac{x - m_1}{\sigma_1}\right) \left(\frac{y - m_2}{\sigma_2}\right) + \left(\frac{y - m_2}{\sigma_2}\right)^2.$$

Tale distribuzione gode delle seguenti proprietà:

- $f_1(x) = N_{m_1, \sigma_1}(x)$, $f_2(y) = N_{m_2, \sigma_2}(y)$, pertanto le previsioni e gli scarti quadratici medi di X e Y sono rispettivamente m_1, m_2 e σ_1, σ_2 ;

- $f_1(x|y) = N_{\bar{m}_1, \bar{\sigma}_1}(x)$, con

$$\bar{m}_1 = m_1 + \rho \frac{\sigma_1}{\sigma_2} (y - m_2), \quad \bar{\sigma}_1 = \sigma_1 \sqrt{1 - \rho^2};$$

- $f_2(y|x) = N_{\bar{m}_2, \bar{\sigma}_2}(y)$, con

$$\bar{m}_2 = m_2 + \rho \frac{\sigma_2}{\sigma_1} (x - m_1), \quad \bar{\sigma}_2 = \sigma_2 \sqrt{1 - \rho^2};$$

- $\mathbb{P}(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dx dy = \dots = m_1 m_2 + \rho \sigma_1 \sigma_2$, pertanto ρ rappresenta il coefficiente di correlazione di X e Y ;
- se $\rho = 0$ risulta $f(x, y) = f_1(x)f_2(y)$, pertanto se X e Y sono incorrelati, segue che sono indipendenti;
- infine, se i parametri $m_1, m_2, \sigma_1, \sigma_2$ sono fissati, al variare di ρ si ottengono infinite distribuzioni normali bidimensionali con le stesse marginali $N_{m_1, \sigma_1}(x), N_{m_2, \sigma_2}(y)$; il che significa che date le distribuzioni marginali non è possibile determinare la distribuzione congiunta.

Osserviamo che la matrice delle varianze-covarianze del vettore (X, Y) è data da

$$C_2 = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

e si ha

$$\det C_2 = \dots = \sigma_1^2 \sigma_2^2 (1 - \rho^2),$$

$$C_2^{-1} = \frac{1}{\det C_2} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

Allora, com'è possibile verificare, la densità congiunta si può rappresentare nella forma matriciale seguente

$$f(x, y) = \frac{1}{2\pi\sqrt{\det C_2}} e^{-\frac{1}{2}A(x-m_1, y-m_2)},$$

dove

$$A(x-m_1, y-m_2) = (x-m_1, y-m_2) \cdot C_2^{-1} \cdot \begin{pmatrix} x - m_1 \\ y - m_2 \end{pmatrix}.$$

In generale, dato un vettore aleatorio continuo $X = (X_1, \dots, X_n)$, sia C_n la matrice delle varianze-covarianze di X . Si dice che X ha una distribuzione

normale n -dimensionale se la densità congiunta è data da

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det C_n}} e^{-\frac{1}{2}A(x_1 - m_1, \dots, x_n - m_n)},$$

dove

$$\begin{aligned} A(x_1 - m_1, \dots, x_n - m_n) &= \\ &= (x_1 - m_1, \dots, x_n - m_n) \cdot C_n^{-1} \cdot \begin{pmatrix} x_1 - m_1 \\ \dots \\ x_n - m_n \end{pmatrix}. \end{aligned}$$

La distribuzione normale n -dimensionale gode di proprietà simili a quella bidimensionale; in particolare

$$X_i \sim N_{m_i, \sigma_i}, \quad i = 1, \dots, n.$$

Inoltre, se per ogni $i \neq j$ si ha $\sigma_{ij} = 0$, la matrice delle varianze-covarianze diventa diagonale e la densità congiunta coincide con il prodotto delle densità marginali, ovvero i numeri aleatori X_1, \dots, X_n sono stocasticamente indipendenti.

Funzione caratteristica

La funzione caratteristica è uno strumento teorico utile sotto diversi aspetti per studiare la distribuzione di probabilità di numeri aleatori discreti e continui.

Dato un numero aleatorio X , discreto o continuo, sia $Y = e^{itX} = \cos(tX) + i\sin(tX)$, dove i è l'unità immaginaria e t è un fissato valore reale, e indichiamo con $\phi_X(t)$ la previsione di Y , che risulta essere una funzione di t . La funzione $\phi_X(t)$ si chiama *funzione caratteristica* di X .

Nel caso discreto, posto $P(X = x_h) = p_h$, si ha

$$\phi_X(t) = \sum_h p_h e^{itx_h},$$

mentre nel caso continuo, indicando con $f(x)$ la densità di X , si ha

$$\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx.$$

Alcune proprietà:

1. $\phi_X(0) = 1, (\sum_h p_h = 1, \int_{-\infty}^{+\infty} f(x)dx = 1);$
2. $|\phi_X(t)| \leq \phi_X(0) = 1, \forall t;$
3. Se $Y = aX + b$, si ha

$$\phi_Y(t) = \mathbb{P}(e^{itY}) = \mathbb{P}(e^{it(aX+b)}) = e^{ibt} \phi_X(at);$$

4. In particolare, se $Y = -X$, si ha:

$$\phi_Y(t) = \phi_{-X}(t) = \mathbb{P}(e^{-itX}) = \phi_X(-t) = \bar{\phi}_X(t).$$

Nota: Se $\phi_X(t)$ è una funzione reale, si ha $\bar{\phi}_X(t) = \phi_X(t)$. Allora $\phi_{-X}(t) = \phi_X(-t) = \phi_X(t)$ e quindi $\phi_X(t)$ è una funzione reale pari.

Esempi.

a) Dato un evento E di probabilità p , sia $X = |E|$.
Si ha

$$\phi_X(t) = \phi_{|E|}(t) = pe^{it \cdot 1} + qe^{it \cdot 0} = pe^{it} + q.$$

b) Dati n eventi E_1, \dots, E_n , indipendenti ed equiprobabili di probabilità p , consideriamo il n.a. $X = |E_1| + \dots + |E_n|$. Si ha $X \sim B(n, p)$; inoltre

$$\phi_X(t) = \sum_{h=0}^n p_h e^{ith} = \dots = (pe^{it} + q)^n.$$

c) Sia dato un numero aleatorio X con distribuzione di Poisson di parametro λ . Si ha

$$\phi_X(t) = \sum_{h=0}^{+\infty} p_h e^{ith} = \dots = e^{\lambda(e^{it}-1)}.$$

d) Sia dato un numero aleatorio X con distribuzione geometrica di parametro p . Si ha

$$\phi_X(t) = \sum_{h=1}^{+\infty} p_h e^{ith} = \sum_{h=1}^{+\infty} pq^{h-1} e^{ith} = \dots = \frac{pe^{it}}{1 - qe^{it}}.$$

e) Se X ha una distribuzione normale standard si

ha

$$\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \dots = e^{-\frac{t^2}{2}}.$$

Come si vede, essendo reale, la funzione caratteristica di una distribuzione normale standard è anche pari.

f) Se X ha una distribuzione normale di parametri m, σ , il n.a. $Y = \frac{X-m}{\sigma}$ ha una distribuzione normale standard e si ha $\phi_Y(t) = e^{-\frac{t^2}{2}}$. Allora, osservando che $X = \sigma Y + m$, applicando la proprietà 3), con $a = \sigma, b = m$, si ottiene

$$\phi_X(t) = e^{imt - \frac{\sigma^2 t^2}{2}}.$$

g) Se X ha una distribuzione esponenziale di parametro λ , si ha

$$\phi_X(t) = \int_0^{+\infty} e^{itx} \lambda e^{-\lambda x} dx = \dots = \frac{\lambda}{\lambda - it}.$$

Calcolo dei momenti.

Per ogni fissato intero $k = 1, 2, \dots$, la previsione di X^k , che indichiamo con $m^{(k)}$, si chiama momento di ordine k di X . Ricordiamo che, dato un numero aleatorio continuo X , con densità $f(x)$, si ha

$$\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx .$$

Derivando rispetto alla variabile t , si ha

$$\phi'_X(t) = \int_{-\infty}^{+\infty} ixe^{itx} f(x) dx ,$$

$$\phi''_X(t) = \int_{-\infty}^{+\infty} (ix)^2 e^{itx} f(x) dx ,$$

.....

$$\phi^{(k)}_X(t) = \int_{-\infty}^{+\infty} (ix)^k e^{itx} f(x) dx .$$

.....

Allora, se esistono i vari momenti di X , si ha

$$\phi'_X(0) = i \int_{-\infty}^{+\infty} xf(x) dx = im^{(1)} ,$$

$$\phi_X''(0) = i^2 \int_{-\infty}^{+\infty} x^2 f(x) dx = i^2 m^{(2)},$$

.....

$$\phi_X^{(k)}(0) = i^k \int_{-\infty}^{+\infty} x^k f(x) dx = i^k m^{(k)}.$$

.....

Pertanto, si ha $m^{(k)} = \frac{\phi_X^{(k)}(0)}{i^k}$. Un ragionamento analogo si può fare se X è un n.a. discreto. In molti casi, dovendo calcolare $m^{(k)}$, conviene sfruttare tale formula anzichè applicare la definizione

$$m^{(k)} = \int_{-\infty}^{+\infty} x^k f(x) dx,$$

nel caso continuo, oppure

$$m^{(k)} = \sum_n p_n x_n^k,$$

nel caso discreto.

La proprietà più importante delle funzioni caratteristiche è la seguente:

dati n numeri aleatori X_1, \dots, X_n stocasticamente indipendenti e posto $Y = X_1 + \dots + X_n$, si ha

$$\phi_Y(t) = \dots = \phi_{X_1}(t) \cdot \dots \cdot \phi_{X_n}(t).$$

Ad esempio, dati n eventi E_1, \dots, E_n , indipendenti ed equiprobabili di probabilità p , e posto

$$X_1 = |E_1|, \dots, X_n = |E_n|,$$

si ha

$$\phi_{X_1}(t) = \dots = \phi_{X_n}(t) = pe^{it} + q.$$

Quindi

$$\phi_{X_1 + \dots + X_n}(t) = \phi_{X_1}(t) \cdot \dots \cdot \phi_{X_n}(t) = (pe^{it} + q)^n.$$

Ritroviamo in questo modo la funzione caratteristica del numero aleatorio $|E_1| + \dots + |E_n|$, che ha distribuzione binomiale di parametri n, p .

Altri due aspetti teorici importanti relativi alle funzioni caratteristiche sono:

1. La corrispondenza tra funzioni caratteristiche e distribuzioni di probabilità è biunivoca; quindi la funzione caratteristica $\phi_X(t)$ determina univocamente la distribuzione di probabilità di X .

Ad esempio, ad una distribuzione normale di parametri m, σ corrisponde la funzione caratteristica $e^{imt - \frac{\sigma^2 t^2}{2}}$ e quindi, se $X \sim N(x)$, si ha $\phi_X(t) = e^{-\frac{t^2}{2}}$. Allora, se $Y = 2X + 3$, si ha

$$\phi_Y(t) = \dots = e^{3it - 2t^2},$$

e quindi $Y \sim N_{3,2}$.

Altro esempio: se $X \sim N_{m_1, \sigma_1}$ e $Y \sim N_{m_2, \sigma_2}$, con X, Y stocasticamente indipendenti, si ha

$$\phi_X(t) = e^{im_1 t - \frac{\sigma_1^2 t^2}{2}}, \quad \phi_Y(t) = e^{im_2 t - \frac{\sigma_2^2 t^2}{2}}.$$

Inoltre, per il n.a. $Z = aX + bY$ si ha

$$\phi_Z(t) = \dots = e^{im_3 t - \frac{\sigma_3^2 t^2}{2}},$$

con

$$m_3 = am_1 + bm_2, \quad \sigma_3 = \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}.$$

Pertanto $Z \sim N_{m_3, \sigma_3}$. Si noti che, volendo evitare l'uso della funzione caratteristica, il calcolo della distribuzione di Z richiederebbe un ragionamento probabilistico molto più complicato.

(\rightarrow integrale di convoluzione)

2. (*Convergenza in legge*) Data una successione di n.a. X_1, \dots, X_n, \dots , siano F_1, \dots, F_n, \dots e $\phi_1, \dots, \phi_n, \dots$ le corrispondenti successioni di funzioni di ripartizione e di funzioni caratteristiche. Allora, data una funzione di ripartizione F e la corrispondente funzione caratteristica ϕ , la successione F_1, \dots, F_n, \dots converge ad F se e solo se la successione $\phi_1, \dots, \phi_n, \dots$ converge a ϕ .

Tale risultato teorico permette di dimostrare il seguente teorema:

Teorema limite centrale.

Data una successione di numeri aleatori X_1, \dots, X_n, \dots , indipendenti ed ugualmente di-

stribuiti, con $\mathbb{P}(X_i) = m$, $\text{Var}(X_i) = \sigma^2$, si consideri la successione delle medie aritmetiche

$$Y_1 = X_1, Y_2 = \frac{X_1 + X_2}{2}, Y_n = \frac{X_1 + \dots + X_n}{n}, \dots,$$

e quella delle medie aritmetiche ridotte Z_1, \dots, Z_n . Ovviamente $\mathbb{P}(Y_n) = m$, $\text{Var}(Y_n) = \frac{\sigma^2}{n}$ e quindi $Z_n = \frac{Y_n - m}{\sigma/\sqrt{n}}$. Indicando con F_i la funzione di ripartizione di Z_i , la successione F_1, \dots, F_n, \dots converge alla funzione di ripartizione (di una distribuzione normale standard) Φ , ovvero si ha

$$\begin{aligned} \lim_{n \rightarrow +\infty} F_n(z) &= \lim_{n \rightarrow +\infty} P(Z_n \leq z) = \\ &= \Phi(z) = \int_{-\infty}^z N(t) dt, \quad \forall z \in \mathbb{R}. \end{aligned}$$

Il risultato precedente si ottiene dimostrando che la successione $\phi_1, \dots, \phi_n, \dots$ (di funzioni caratteristiche dei numeri aleatori Z_1, \dots, Z_n, \dots) converge alla funzione caratteristica (della distribuzione normale standard) $\phi(t) = e^{-\frac{t^2}{2}}$.

Somme di numeri aleatori.

Dato un vettore aleatorio continuo (X, Y) , con densità $f(x, y)$, sia

$$Z = X + Y, \quad G(z) = P(Z \leq z), \quad z \in \mathbb{R}.$$

Si ha

$$\begin{aligned} G(z) &= P(X + Y \leq z) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} f(x, y) dy = \\ &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^z f(x, t - x) dt = \\ &= \int_{-\infty}^z \left(\int_{-\infty}^{+\infty} f(x, t - x) dx \right) dt = \int_{-\infty}^z g(t) dt, \end{aligned}$$

dove g è la densità di Z data da

$$g(z) = G'(z) = \int_{-\infty}^{+\infty} f(x, z - x) dx.$$

Caso notevole: se X e Y sono indipendenti, ovvero $f(x, y) = f_1(x)f_2(y)$, si ottiene

$$g(z) = \int_{-\infty}^{+\infty} f_1(x)f_2(z-x)dx,$$

che rappresenta l'*integrale di convoluzione* di f_1, f_2 e si indica con il simbolo: $g = f_1 * f_2$ (si noti che: $f_1 * f_2 = f_2 * f_1$).

Esempi:

(i) $f_1 = N_{m_1, \sigma_1} = N_1$, $f_2 = N_{m_2, \sigma_2} = N_2$; si ha

$$g(z) = \int_{-\infty}^{+\infty} N_1(x)N_2(z-x)dx = \dots = N_3(z),$$

con

$$N_3 = N_{m_3, \sigma_3}, \quad m_3 = m_1 + m_2, \quad \sigma_3^2 = \sigma_1^2 + \sigma_2^2.$$

Pertanto, dalla convoluzione di due distribuzioni normali si ottiene una distribuzione ancora normale.

(ii) $f_1 = G_{c_1, \lambda}$, $f_2 = G_{c_2, \lambda_2}$; si ha

$$g = G_{c_1, \lambda} * G_{c_2, \lambda} = G_{c_1 + c_2, \lambda}.$$

In particolare:

$$\text{Exp}(\lambda) * \text{Exp}(\lambda) = G_{1, \lambda} * G_{1, \lambda} = G_{2, \lambda}.$$

(iii) $X \sim U([0, a])$, $Y \sim U([0, a])$; si ha

$$g = U([0, a]) * U([0, a]) = T,$$

dove

$$T(z) = \begin{cases} \frac{z}{a^2}, & 0 \leq z \leq a, \\ \frac{2a-z}{a^2}, & a < z \leq 2a, \\ 0, & \text{altrove.} \end{cases}$$

Inferenza statistica

In molte applicazioni statistiche si studiano popolazioni in cui una o più caratteristiche numeriche sono incognite. Tali caratteristiche costituiscono quindi un numero (o un vettore) aleatorio, che possiamo indicare con Θ , a cui viene assegnata (sulla base dell'informazione iniziale) una distribuzione iniziale, in particolare una densità iniziale $\beta(\theta)$ nel caso continuo. Per ridurre l'incertezza sul parametro Θ si procede all'osservazione di un vettore (a priori aleatorio) $\mathbf{X} = (X_1, \dots, X_n)$ estratto dalla popolazione in oggetto. Fare inferenza su Θ significa applicare il procedimento bayesiano, che consiste nel determinare la distribuzione finale di Θ condizionata al vettore osservato $\mathbf{x} = (x_1, \dots, x_n)$, ovvero $\beta(\theta|\mathbf{x})$. Se, riferendoci al caso continuo, indichiamo con $c(\mathbf{x}, \theta)$ la densità congiunta del vettore aleatorio (\mathbf{X}, Θ) e con $\alpha(\mathbf{x}|\theta) = \alpha(x_1, \dots, x_n|\theta)$ la densità di \mathbf{X} condizionata a un fissato valore θ di Θ , si ha

$$c(\mathbf{x}, \theta) = \beta(\theta)\alpha(\mathbf{x}|\theta) = \alpha(\mathbf{x})\beta(\theta|\mathbf{x}).$$

Quindi (*teorema di Bayes per vettori aleatori*)

$$\begin{aligned}\beta(\theta|\mathbf{x}) &= \frac{c(\mathbf{x},\theta)}{\alpha(\mathbf{x})} = \frac{\beta(\theta)\alpha(\mathbf{x}|\theta)}{\int_{\Theta} c(\mathbf{x},\theta)d\theta} = \\ &= \frac{\beta(\theta)\alpha(x_1,\dots,x_n|\theta)}{\int_{\Theta} \beta(\theta)\alpha(x_1,\dots,x_n|\theta)d\theta} = k(\mathbf{x})\beta(\theta)\alpha(\mathbf{x}|\theta),\end{aligned}$$

con

$$k(\mathbf{x}) = \frac{1}{\int_{\Theta} \beta(\theta)\alpha(\mathbf{x}|\theta)d\theta}.$$

Una situazione tipica nelle applicazioni statistiche è quella in cui, per il vettore delle osservazioni (o misure) $\mathbf{X} = (X_1, \dots, X_n)$, valgono le seguenti proprietà:

(i) indicando con $f_i(x_i)$ la densità di X_i , si ha $f_1(\cdot|\theta) = \dots = f_n(\cdot|\theta) = f(\cdot|\theta)$, ovvero per ogni fissato θ , i numeri aleatori X_1, \dots, X_n sono ugualmente distribuiti condizionatamente a θ ;

(ii) per ogni fissato θ , i numeri aleatori X_1, \dots, X_n sono stocasticamente indipendenti condizionatamente a θ .

Da (i) e (ii) segue

$$\begin{aligned}\alpha(x_1, \dots, x_n | \theta) &= f_1(x_1 | \theta) \cdots f_n(x_n | \theta) = \\ &= f(x_1 | \theta) \cdots f(x_n | \theta),\end{aligned}$$

e quindi

$$\beta(\theta | x_1, \dots, x_n) = k(x_1, \dots, x_n) \beta(\theta) f(x_1 | \theta) \cdots f(x_n | \theta),$$

con

$$k(x_1, \dots, x_n) = \frac{1}{\int_{\Theta} \beta(\theta) f(x_1 | \theta) \cdots f(x_n | \theta) d\theta}.$$

Quando valgono (i) e (ii) il vettore aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ si dice un campione casuale.

La distribuzione finale di Θ (condizionata al vettore delle osservazioni \mathbf{x}) può essere utilizzata per determinare un insieme A , ad esempio un intervallo $[\theta_1, \theta_2]$ (possibilmente di lunghezza minima), tale che per un opportuno valore α risulti $P(\Theta \in A | \mathbf{x}) = \alpha$. In particolare, se vale $P(\theta_1 \leq \Theta \leq \theta_2 | \mathbf{x}) = \alpha$ l'intervallo

$[\theta_1, \theta_2]$ si dice un *intervallo di confidenza* al $100\alpha\%$.

Applicazioni:

(a) *campionamento da una popolazione normale con parametro m incognito.*

Indichiamo con Θ il parametro incognito m di una distribuzione normale e supponiamo che sia $\beta(\theta) = N_{m_0, \sigma_0}(\theta)$. Supponiamo inoltre che, per $i = 1, \dots, n$, sia

$$f_i(x_i|\theta) = f(x_i|\theta) = N_{\theta, \sigma}(x_i).$$

Ad esempio, Θ potrebbe rappresentare la lunghezza incognita di una sbarra, a cui viene assegnata come densità iniziale (sulla base dell'informazione a priori) una distribuzione normale di parametri m_0, σ_0 . Tale sbarra viene misurata n volte in modo da ottenere un vettore $\mathbf{x} = (x_1, \dots, x_n)$ che costituisce il valore osservato di un campione casuale $\mathbf{X} = (X_1, \dots, X_n)$, per le cui componenti si assume (in base a certe ipotesi sullo strumento di misura) che sia $X_i|\theta \sim N_{\theta, \sigma}$.

Si ha quindi

$$\begin{aligned}\beta(\theta|x_1, \dots, x_n) &= k(x_1, \dots, x_n)\beta(\theta)f(x_1|\theta) \cdots f(x_n|\theta) = \\ &= k(x_1, \dots, x_n)N_{m_0, \sigma_0}(\theta)N_{\theta, \sigma}(x_1) \cdots N_{\theta, \sigma}(x_n) = \\ &= \cdots = N_{m_n, \sigma_n}(\theta),\end{aligned}$$

dove, indicando con \bar{x} la media aritmetica dei valori x_1, \dots, x_n , è

$$\begin{aligned}\frac{1}{\sigma_n^2} &= \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}, \\ m_n &= \frac{\frac{1}{\sigma_0^2} \cdot m_0 + \frac{n}{\sigma^2} \cdot \bar{x}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} = \lambda m_0 + (1 - \lambda)\bar{x},\end{aligned}$$

dove

$$\lambda = \frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \quad 1 - \lambda = \frac{\frac{n}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}.$$

Come si vede, m_n è una media ponderata dei valori m_0, \bar{x} con rispettivi pesi $\lambda, 1 - \lambda$. Inoltre, per n grande si ha $\sigma_n \ll \sigma_0$, ovvero la distribuzione finale di Θ è molto più concentrata di quella iniziale (infatti σ_n

tende a 0 per $n \rightarrow +\infty$).

In particolare, per $n = 1$, si ha

$$\begin{aligned}\beta(\theta|x_1) &= k(x_1)\beta(\theta)f(x_1|\theta) = \\ &= k(x_1)N_{m_0,\sigma_0}(\theta)N_{\theta,\sigma}(x_1) = \\ &= \dots = N_{m_1,\sigma_1}(\theta),\end{aligned}$$

con

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}, \quad m_1 = \frac{\frac{1}{\sigma_0^2} \cdot m_0 + \frac{1}{\sigma^2} \cdot x_1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}.$$

(b) *Campionamento da una popolazione binomiale con parametro p incognito.*

Si tratta del caso in cui una proporzione incognita p di individui di una data popolazione possiede una certa caratteristica. Indicando con Θ il parametro incognito p , si ha intanto $\Theta \in [0, 1]$. L'obiettivo è quello di fare inferenza su Θ estraendo n individui dalla popolazione ed osservando quanti di essi possiedono la data

caratteristica. Possiamo considerare un campione casuale $\mathbf{X} = (X_1, \dots, X_n)$, dove $(X_i = 1)$ significa che l' i -mo individuo osservato possiede la data caratteristica, mentre $(X_i = 0)$ significa che l' i -mo individuo osservato non possiede tale caratteristica. Pertanto

$$P(X_i = 1 | \theta) = f_i(1 | \theta) = \theta,$$

$$P(X_i = 0 | \theta) = f_i(0 | \theta) = 1 - \theta,$$

e quindi, osservando che $f_i(x_i | \theta) = \theta^{x_i}(1 - \theta)^{1-x_i}$, si ha

$$\begin{aligned} \alpha(x_1, \dots, x_n | \theta) &= f_1(x_1 | \theta) \cdots f_n(x_n | \theta) = \\ &= f(x_1 | \theta) \cdots f(x_n | \theta) = \dots = \theta^{\sum_i x_i} (1 - \theta)^{n - \sum_i x_i}. \end{aligned}$$

Se, come distribuzione iniziale di Θ si sceglie una distribuzione beta di parametri r_0, s_0 , si può verificare che la distribuzione finale è ancora di tipo beta e, posto

$\sum_i x_i = h$, risulta

$$\begin{aligned} \beta(\theta|x_1, \dots, x_n) &= k(x_1, \dots, x_n) B_{r_0, s_0}(\theta) \theta^h (1 - \theta)^{n-h} = \\ &= \dots = B_{r_n, s_n}(\theta), \quad (r_n = r_0 + h, \quad s_n = s_0 + n - h). \end{aligned}$$

(c) *Campionamento da una popolazione esponenziale con parametro λ incognito.*

Supponiamo che la durata aleatoria fino al guasto di un certo tipo di dispositivi abbia una distribuzione esponenziale con parametro incognito, che indichiamo con Θ (anzichè λ).

Sia (X_1, \dots, X_n) il vettore aleatorio costituito dall'osservazione delle durate fino al guasto di n di tali dispositivi. Si ha

$$X_i | \theta \sim f = \text{Exp}(\theta) = G_{1, \theta}, \quad i = 1, \dots, n;$$

$$f(x | \theta) = \theta e^{-\theta x}, \quad x \geq 0; \quad f(x | \theta) = 0, \quad x < 0;$$

$$\begin{aligned}\alpha(x_1, \dots, x_n | \theta) &= f(x_1 | \theta) \cdots f(x_n | \theta) = \\ &= \theta e^{-\theta x_1} \cdots \theta e^{-\theta x_n} = \theta^n e^{-\theta \sum_i x_i};\end{aligned}$$

$$\beta(\theta | x_1, \dots, x_n) = k(x_1, \dots, x_n) \beta(\theta) \theta^n e^{-\theta \sum_i x_i}.$$

Se per Θ si assume una distribuzione iniziale di tipo Gamma, ad esempio $\beta(\theta) = G_{c_0, \lambda_0}(\theta)$, avendo osservato un vettore $\mathbf{x} = (x_1, \dots, x_n)$ si ha

$$\begin{aligned}\beta(\theta | \mathbf{x}) &= k(\mathbf{x}) \frac{\lambda_0^{c_0}}{\Gamma(c_0)} \theta^{c_0-1} e^{-\lambda_0 \theta} \theta^n e^{-\theta \sum_i x_i} = \\ &= k(\mathbf{x}) \frac{\lambda_0^{c_0}}{\Gamma(c_0)} \theta^{c_0+n-1} e^{-\theta(\lambda_0 + \sum_i x_i)} = G_{c_n, \lambda_n}(\theta),\end{aligned}$$

con $c_n = c_0 + n$, $\lambda_n = \lambda_0 + \sum_i x_i$.

Pertanto la distribuzione finale di Θ è ancora di tipo Gamma e si ha

$$\mathbb{P}(\Theta | x_1, \dots, x_n) = \frac{c_n}{\lambda_n} = \frac{c_0 + n}{\lambda_0 + \sum_i x_i};$$

$$Var(\Theta | x_1, \dots, x_n) = \frac{c_n}{\lambda_n^2} = \frac{c_0 + n}{(\lambda_0 + \sum_i x_i)^2}.$$

(d) *Campionamento da una popolazione di Poisson con parametro λ incognito.*

Supponiamo che il numero aleatorio di "arrivi" in un certo fenomeno abbia una distribuzione di Poisson con parametro incognito Θ .

Gli "arrivi" potrebbero, ad esempio, essere le telefonate che giungono ad un centralino in un fissato intervallo di tempo durante il giorno.

Sia (X_1, \dots, X_n) il vettore aleatorio costituito dall'osservazione del numero di telefonate che arrivano a tale centralino nel fissato intervallo di tempo in n giorni distinti. Si ha

$$X_i | \theta \sim f = \mathcal{P}(\theta), \quad i = 1, \dots, n;$$

$$f(x | \theta) = \frac{\theta^x}{x!} e^{-\theta}, \quad x = 0, 1, \dots, k, \dots$$

$$\alpha(x_1, \dots, x_n | \theta) = f(x_1 | \theta) \cdots f(x_n | \theta) =$$

$$= \frac{\theta^{x_1}}{x_1!} e^{-\theta} \cdots \frac{\theta^{x_n}}{x_n!} e^{-\theta} = \frac{\theta^{\sum_i x_i}}{x_1! \cdots x_n!} e^{-n\theta};$$

$$\beta(\theta | x_1, \dots, x_n) = \frac{k(x_1, \dots, x_n)}{x_1! \cdots x_n!} \beta(\theta) \theta^{\sum_i x_i} e^{-n\theta}.$$

Se per Θ si assume una distribuzione iniziale (di tipo Gamma) $\beta(\theta) = G_{c_0, \lambda_0}(\theta)$, condizionatamente ad un vettore osservato $\mathbf{x} = (x_1, \dots, x_n)$ si ha

$$\begin{aligned} \beta(\theta | \mathbf{x}) &= \frac{k(\mathbf{x})}{x_1! \cdots x_n!} \frac{\lambda_0^{c_0}}{\Gamma(c_0)} \theta^{c_0-1} e^{-\lambda_0 \theta} \theta^{\sum_i x_i} e^{-n\theta} = \\ &= \frac{k(\mathbf{x})}{x_1! \cdots x_n!} \frac{\lambda_0^{c_0}}{\Gamma(c_0)} \theta^{c_0 + \sum_i x_i - 1} e^{-\theta(\lambda_0 + n)} = G_{c_n, \lambda_n}(\theta), \end{aligned}$$

con $c_n = c_0 + \sum_i x_i$, $\lambda_n = \lambda_0 + n$.

Pertanto la distribuzione finale di Θ è ancora di tipo Gamma e si ha

$$\mathbb{P}(\Theta | x_1, \dots, x_n) = \frac{c_n}{\lambda_n} = \frac{c_0 + \sum_i x_i}{\lambda_0 + n};$$

$$\text{Var}(\Theta | x_1, \dots, x_n) = \frac{c_n}{\lambda_n^2} = \frac{c_0 + \sum_i x_i}{(\lambda_0 + n)^2}.$$

Distribuzione Chi-Quadro

Un numero aleatorio X , continuo e non negativo, si dice che ha una distribuzione Chi-Quadro, con n gradi di libertà, che indichiamo con il simbolo $X \sim \chi_n^2$, se la densità di X è

$$f(x) = \chi_n^2(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Come si può verificare, la distribuzione Chi-Quadro è una particolare distribuzione Gamma: $\chi_n^2 = G_{\frac{n}{2}, \frac{1}{2}}$.

Pertanto

$$\mathbb{P}(X) = \frac{\frac{n}{2}}{\frac{1}{2}} = n, \quad \text{Var}(X) = \frac{\frac{n}{2}}{\frac{1}{4}} = 2n.$$

Osservazioni:

(i) dato un numero aleatorio con distribuzione $N_{m,\sigma}$, si ha

$$Z = \frac{X - m}{\sigma} \sim N_{0,1}.$$

Allora, posto $Y = Z^2$, risulta $Y \geq 0$; inoltre, indicando con G la funzione di ripartizione di Y e con g la sua densità, per ogni $y > 0$ si ha

$$G(y) = P(Y \leq y) = P(-\sqrt{y} \leq Z \leq \sqrt{y}) = 2\Phi(\sqrt{y}) - 1,$$

e quindi

$$g(y) = G'(y) = 2N(\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, \quad y > 0.$$

Come si vede, risulta: $g(y) = \mathcal{X}_1^2(y) = G_{\frac{1}{2}, \frac{1}{2}}(y)$.

(ii) dati n numeri aleatori X_1, \dots, X_n , indipendenti e con distribuzione normale di parametri

$$m_i = m, \quad \sigma_i = \sigma, \quad i = 1, \dots, n,$$

i numeri aleatori

$$\left(\frac{X_1 - m}{\sigma}\right)^2, \dots, \left(\frac{X_n - m}{\sigma}\right)^2$$

hanno tutti distribuzione χ_1^2 .

Allora, posto

$$Y = \left(\frac{X_1 - m}{\sigma}\right)^2 + \dots + \left(\frac{X_n - m}{\sigma}\right)^2,$$

ricordando la proprietà (iii) della distribuzione Gamma risulta

$$Y \sim G_{\frac{n}{2}, \frac{1}{2}} = \chi_n^2.$$

Inoltre, dati due numeri aleatori W e Y , con

$$W \sim \chi_k^2, \quad Y \sim \chi_n^2,$$

risulta

$$W + Y \sim \chi_{k+n}^2.$$

Statistiche campionarie

Dati n numeri aleatori X_1, \dots, X_n , indipendenti e ugualmente distribuiti, con

$$IP(X_i) = m, \quad Var(X_i) = \sigma^2, \quad i = 1, \dots, n,$$

consideriamo la *media campionaria* e la *varianza campionaria* così definite:

$$\begin{aligned} \bar{X} &= \frac{X_1 + \dots + X_n}{n}, \\ S^2 &= \frac{(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}. \end{aligned} \tag{23}$$

Allora, si ha

$$IP(\bar{X}) = IP\left(\frac{X_1 + \dots + X_n}{n}\right) = m,$$

$$Var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n},$$

$$IP(S^2) = \frac{n-1}{n} \sigma^2.$$

La previsione di S^2 si ottiene osservando che

$$\sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n X_i - n\bar{X} = \sum_{i=1}^n X_i - \sum_{i=1}^n X_i = 0;$$

inoltre

$$\begin{aligned} \sum_{i=1}^n (X_i - m)^2 &= \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - m)]^2 = \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + 2 \sum_{i=1}^n (X_i - \bar{X})(\bar{X} - m) + \\ &+ \sum_{i=1}^n (\bar{X} - m)^2 = \\ &= nS^2 + 2(\bar{X} - m) \sum_{i=1}^n (X_i - \bar{X}) + n(\bar{X} - m)^2 = \\ &= nS^2 + n(\bar{X} - m)^2; \end{aligned}$$

quindi

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2 - (\bar{X} - m)^2;$$

pertanto

$$\begin{aligned} IP(S^2) &= \frac{1}{n} \sum_{i=1}^n IP[(X_i - m)^2] - IP[(\bar{X} - m)^2] = \\ &= \frac{1}{n} \cdot n\sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2. \end{aligned}$$

Nota: dal risultato precedente segue anche

$$IP\left(\frac{n}{n-1}S^2\right) = IP\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right] = \sigma^2.$$

Se si vuole che la previsione della varianza campionaria coincida con σ^2 , nella (23) a denominatore occorre sostituire n con $n - 1$, definendo (come fanno alcuni autori) la varianza campionaria nel seguente modo:

$$S^{*2} = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}.$$

Distribuzione t di Student

Un numero aleatorio continuo X si dice che ha una distribuzione t di Student con n gradi di libertà, che indichiamo con il simbolo $X \sim t_n$, se ha la seguente densità:

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n} \Gamma(\frac{n}{2})} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}, \quad x \in \mathbb{R}.$$

Proprietà fondamentale:

dati due numeri aleatori indipendenti Z e W , con

$$Z \sim N_{0,1}, \quad W \sim \chi_n^2,$$

posto $X = \frac{Z}{\sqrt{W/n}}$, risulta: $X \sim t_n$.

Per $n > 2$, utilizzando la funzione Gamma, si ottiene

$$\mathbb{P}\left(\frac{1}{W}\right) = \frac{1}{n-2}; \quad \mathbb{P}(X) = \mathbb{P}(Z) \mathbb{P}\left(\frac{1}{\sqrt{W/n}}\right) = 0,$$

$$\text{Var}(X) = \mathbb{P}(X^2) = \mathbb{P}(Z^2) \mathbb{P}\left(\frac{1}{W/n}\right) = \frac{n}{n-2}.$$

Applicazione:

dato un campione X_1, \dots, X_n da una distribuzione normale $N_{m, \sigma}$, siano \bar{X}, S^{*2} la media campionaria e la varianza campionaria. Posto $t = \frac{\bar{X} - m}{S^*/\sqrt{n}}$, si ha: $t \sim t_{n-1}$.

Infatti

$$\frac{\bar{X} - m}{\sigma/\sqrt{n}} = Z \sim N_{0,1}, \quad \frac{(n-1)S^{*2}}{\sigma^2} = W \sim \chi_{n-1}^2;$$

allora Z e W sono indipendenti e si ha

$$t = \frac{\bar{X} - m}{S^*/\sqrt{n}} = \frac{Z}{\sqrt{W/(n-1)}} \sim t_{n-1}.$$