

1 Problema delle Concordanze

Sia n un numero intero e si abbia a disposizione un mazzetto di n cartoncini. Essi vengono numerati da 1 a n sul fronte inizialmente bianco; sul dorso i cartoncini presentano la stessa immagine. L'esperimento \mathcal{E} consiste nel (i) mischiare i cartoncini, (ii) appoggiare su un tavolo il mazzetto mostrandone il dorso, (iii) mostrare il fronte del cartoncino più in alto "chiamando" il numero "uno" e riporre accanto al mazzetto il cartoncino mostrandone il fronte, (iv) procedere allo stesso modo fino all'ultimo cartoncino. Per $i = 1, 2, \dots, n$, si designa con C_i (concordanza) l'evento che si verifica quando si presenta il cartoncino con numero i alla i -ma chiamata.

2 Probabilità di osservare almeno una concordanza nelle n chiamate

L'evento "almeno una concordanza nelle n chiamate" risulta essere l'unione di tutte le concordanze: $\bigcup_{i=1}^n C_i$. Ovviamente, nulla vieta che nella stessa effettuazione dell'esperimento si possono verificare due o più concordanze; pertanto, è necessario applicare la formula di inclusione-esclusione:

$$\mathbb{P}\left(\bigcup_{i=1}^n C_i\right) = \sum_{i=1}^n \mathbb{P}(C_i) - \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{P}(C_i \cap C_j) + \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n \mathbb{P}(C_i \cap C_j \cap C_k) - \dots + (-1)^{n-1} \mathbb{P}(C_1 \cap C_2 \cap \dots \cap C_n) \quad (1)$$

Si osservi che ci sono $n!$ possibili mischiate; risulta:

$$i = 1, 2, \dots, n, \quad \mathbb{P}(C_i) = \frac{(n-1)!}{n!}.$$

Infatti, bloccando solo il cartoncino numerato con i all' i -mo posto dall'alto si verifica C_i ma tutti gli altri possono permutare in tutti i modi possibili. Con lo stesso ragionamento, si ha:

$$i, j = 1, 2, \dots, n : i < j, \quad \mathbb{P}(C_i \cap C_j) = \frac{(n-2)!}{n!}.$$

e

$$i, j, k = 1, 2, \dots, n : i < j < k, \quad \mathbb{P}(C_i \cap C_j \cap C_k) = \frac{(n-3)!}{n!}.$$

Allo stesso modo per i successivi addendi nella (1). In particolare, dal momento che $(n-n)! = 1$, si ha:

$$\mathbb{P}(C_1 \cap C_2 \cap \dots \cap C_n) = \frac{1}{n!}.$$

La formula (1) diventa:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n C_i\right) &= \frac{(n-1)!}{n!} \sum_{i=1}^n 1 - \frac{(n-2)!}{n!} \sum_{i=1}^n \sum_{j=i+1}^n 1 + \frac{(n-3)!}{n!} \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n 1 - \dots + (-1)^{n-1} \frac{1}{n!} \\ &= \frac{(n-1)!}{n!} \binom{n}{1} - \frac{(n-2)!}{n!} \binom{n}{2} + \frac{(n-3)!}{n!} \binom{n}{3} - \dots + (-1)^{n-1} \frac{1}{n!} \\ &= \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1} \frac{1}{n!}. \end{aligned} \quad (2)$$

Il passaggio intermedio nella (2) segue da questo ragionamento (valido per il secondo addendo ma facilmente generalizzabile agli altri): il numero degli addendi 1 è uguale al numero delle combinazioni semplici di lunghezza 2 ottenute dall'insieme $S_n = \{1, 2, \dots, n\}$ in quanto il valore dell'indice j è strettamente maggiore del valore dell'indice i .

In conclusione, si ha:

$$\mathbb{P}\left(\bigcup_{i=1}^n C_i\right) = \sum_{i=1}^n (-1)^{i-1} \frac{1}{i!}. \quad (3)$$

3 Probabilità di osservare zero concordanze nelle n chiamate

Si indichi con $E_{0,n}$ l'evento che si presenta quando in una esecuzione dell'esperimento \mathcal{E} non si osserva alcuna concordanza (ovvero, 0 concordanze); ovviamente, risulta che:

$$E_{0,n} = \left(\bigcup_{i=1}^n C_i\right)^c \implies \mathbb{P}(E_{0,n}) = 1 - \mathbb{P}\left(\bigcup_{i=1}^n C_i\right). \quad (4)$$

In definitiva, dalla (3) e dalla (4), si ottiene:

$$\mathbb{P}(E_{0,n}) = \sum_{i=0}^n (-1)^i \frac{1}{i!}. \quad (5)$$

Prima di terminare il paragrafo si osservi che, indicato con $N_{0,n}$ il numero delle mischiate che non fanno verificare alcuna concordanza, dalla (5) si ricava:

$$\mathbb{P}(E_{0,n}) = \frac{N_{0,n}}{n!} \implies N_{0,n} = n! \cdot \mathbb{P}(E_{0,n}). \quad (6)$$

4 Probabilità di osservare SOLO una concordanza nelle n chiamate

Si indichi con $F_{1,1,n}$ l'evento che si presenta quando in una esecuzione dell'esperimento \mathcal{E} si osserva esattamente una concordanza e questa avviene alla prima chiamata. Ovviamente, indicato con $M_{1,1,n}$ il numero delle mischiate che fanno verificare solo la concordanza nella prima chiamata e zero concordanze nelle successive, risulta che $M_{1,1,n} = N_{0,n-1}$: infatti, dopo la concordanza alla prima chiamata, dalla seconda alla n -chiamata non si verificano altre concordanze. Allora, in virtù della (6), si ha:

$$\mathbb{P}(F_{1,1,n}) = \frac{M_{1,1,n}}{n!} = \frac{N_{0,n-1}}{n!} = \frac{(n-1) \cdot \mathbb{P}(E_{0,n-1})}{n!} = \frac{1}{n} \mathbb{P}(E_{0,n-1}). \quad (7)$$

Si osservi che il risultato ottenuto nella (7) non dipende dal fatto che è stato richiesto di osservare la concordanza alla prima chiamata in quanto lo stesso ragionamento è valido per la richiesta di osservare un'unica concordanza nella generica chiamata. In altri termini, indicato con $F_{1,i,n}$ l'evento che si presenta quando in una esecuzione dell'esperimento \mathcal{E} si osserva esattamente una concordanza e questa avviene alla i -ma chiamata, allora si ha:

$$i = 1, 2, \dots, n, \quad \mathbb{P}(F_{1,i,n}) = \mathbb{P}(F_{1,1,n}) = \frac{1}{n} \mathbb{P}(E_{0,n-1}). \quad (8)$$

Dopo di ciò, se $E_{1,n}$ è l'evento che si presenta quando in una esecuzione dell'esperimento \mathcal{E} si osserva una sola concordanza, risulta

$$\mathbb{P}(E_{1,n}) = \sum_{i=1}^n \mathbb{P}(F_{1,i,n}) = \frac{1}{n} \mathbb{P}(E_{0,n-1}) \sum_{i=1}^n 1 = \mathbb{P}(E_{0,n-1}) = \sum_{i=0}^{n-1} (-1)^i \frac{1}{i!}. \quad (9)$$

5 Probabilità di osservare esattamente due concordanze nelle n chiamate

Si procede sulla falsariga del paragrafo precedente. Si lascia allo studente.

Fissato n in \mathbb{N} , si ricordi che

- a) un campione casuale semplice (c.c.s.) $\underline{X} = (X_1, X_2, \dots, X_n)$ è una n -pla di numeri aleatori congiuntamente distribuiti, indipendenti e somiglianti; le componenti di \underline{X} sono chiamate *osservazioni*;
- b) una variabile aleatoria X avente la stessa legge (distribuzione) delle osservazioni è detta essere *genitrice* del c.c.s. \underline{X} ;
- c) in altri termini le osservazioni sono repliche indipendenti della genitrice;
- d) nella legge di X , in generale, compare un *parametro* incognito θ di cui è noto l'insieme Θ dei suoi valori; Θ è la *regione parametrica*; nulla vieta che θ abbia dimensione maggiore di 1;
- e) una *statistica* T è una funzione misurabile e calcolabile di un campione casuale semplice: $T = g(\underline{X})$; nulla vieta che T abbia dimensione maggiore di 1;
- f) uno stimatore è una statistica che viene utilizzata ai fini della stima di una funzione $\psi(\theta)$ del parametro incognito θ .

Quando si vuole mettere in evidenza la dipendenza dal parametro incognito della media della genitrice si può usare la notazione $\mathbb{E}_X(\theta)$: in altri termini, la media deve essere vista non come un numero ma come una funzione definita nella regione parametrica. Allo stesso modo per la varianza, $\mathbb{D}_X^2(\theta)$ e per gli altri momenti teorici. Ovviamente, la stessa notazione e la stessa interpretazione si può usare per la media e la varianza di una qualsiasi statistica.

Si ricordi, infine, che uno stimatore T è corretto ai fini della stima di $\psi(\theta)$ se e solo se:

$$\theta \in \Theta, \quad \mathbb{E}_T(\theta) = \psi(\theta).$$

Nel caso in cui lo stimatore T non è corretto, la funzione

$$\theta \in \Theta, \quad d_T(\theta) := \mathbb{E}_T(\theta) - \psi(\theta).$$

è la sua *distorsione*.

Esempio 1. Sia $X \sim B(1, p)$ con $p \in (0, 1)$ e $n = 3$. La media campionaria \bar{X} è uno stimatore corretto per p . Infatti, si ha:

$$\begin{aligned} p \in (0, 1), \quad \mathbb{E}_{\bar{X}}(p) &= \frac{1}{3} \mathbb{E}_{X_1+X_2+X_3}(p) = \frac{1}{3} \left[\mathbb{E}_{X_1}(p) + \mathbb{E}_{X_2}(p) + \mathbb{E}_{X_3}(p) \right] \\ &= \frac{1}{3}(p + p + p) = \frac{1}{3} \cdot 3p = p. \end{aligned}$$

Invece, lo stimatore

$$V = \frac{X_1 + 2X_2 + X_3}{5}.$$

non è corretto; infatti, risulta:

$$\begin{aligned} p \in (0, 1), \quad \mathbb{E}_V(p) &= \frac{1}{5} \mathbb{E}_{X_1+2X_2+X_3}(p) = \frac{1}{5} \left[\mathbb{E}_{X_1}(p) + \mathbb{E}_{2X_2}(p) + \mathbb{E}_{X_3}(p) \right] \\ &= \frac{1}{5}(p + 2p + p) = \frac{1}{5} \cdot 4p = \frac{4}{5}p \neq p. \end{aligned}$$

Quindi, la distorsione dello stimatore V vale:

$$p \in (0, 1), \quad d_V(p) = \frac{4}{5}p - p = -\frac{1}{5}p. \quad (1)$$

◇

Ai fini del confronto di due stimatori T e S per la stessa funzione $\psi(\theta)$ si considera il *rischio quadratico medio* che per lo stimatore T è la seguente funzione:

$$\theta \in \Theta, \quad R_T(\theta) = \mathbb{E} \left\{ [T - \psi(\theta)]^2 \right\}.$$

Posto $\mu_T(\theta) := \mathbb{E}_T(\theta)$ e ricordando che è nulla la media di una variabile aleatoria scartata rispetto alla sua media, risulta :

$$\begin{aligned} \theta \in \Theta, \quad R_T(\theta) &= \mathbb{E} \left\{ [T - \psi(\theta)]^2 \right\} = \mathbb{E} \left\{ \left\{ [T - \mu_T(\theta)] + [\mu_T(\theta) - \psi(\theta)] \right\}^2 \right\} \\ &= \mathbb{E} \left\{ [T - \mu_T(\theta)]^2 \right\} + 2[\mu_T(\theta) - \psi(\theta)] \mathbb{E}[T - \mu_T(\theta)] + [\mu_T(\theta) - \psi(\theta)]^2 \\ &= \mathbb{D}_T^2(\theta) + d_T^2(\theta). \end{aligned}$$

In altri termini, il rischio quadratico medio è la somma della varianza dello stimatore con il quadrato della sua distorsione. Ne discende che per gli stimatori corretti il rischio coincide con la varianza.

Esempio 2. In riferimento all'Esempio 1, si ha:

$$\theta \in \Theta, \quad R_{\bar{X}}(p) = \mathbb{D}_{\bar{X}}(p) = \frac{1}{9} \left[\mathbb{D}_{X_1}(p) + \mathbb{D}_{X_2}(p) + \mathbb{D}_{X_3}(p) \right] = \frac{1}{9} \cdot 3p(1-p) = \frac{p(1-p)}{3}.$$

Inoltre, procedendo allo stesso modo, si ha:

$$\theta \in \Theta, \quad \mathbb{D}_V(p) = \frac{1}{25} \left[\mathbb{D}_{X_1}(p) + \mathbb{D}_{2X_2}(p) + \mathbb{D}_{X_3}(p) \right] = \frac{1}{25} \cdot 6p(1-p) = \frac{6p(1-p)}{25}.$$

Tenendo conto di quest'ultima e della (1), in definitiva si ha:

$$\theta \in \Theta, \quad R_V(p) = \mathbb{D}_V(p) + d_V^2(p) = \frac{6p(1-p)}{25} + \frac{1}{25}p^2.$$

◇

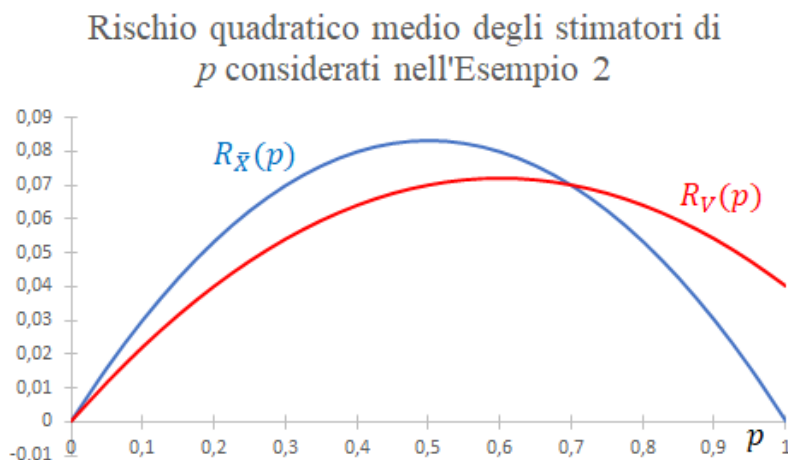
Definizione 1. Siano T e S due stimatori per $\psi(\theta)$. Si dice che S è preferibile a T e si scrive $S \leq T$ se e solo se, in tutta la regione parametrica, il rischio quadratico medio di S è non maggiore del rischio quadratico medio di T . In simboli:

$$S \leq T \iff \theta \in \Theta, \quad R_S(\theta) \leq R_T(\theta).$$

Definizione 2. Siano T e S due stimatori per $\psi(\theta)$. Si dice che S è strettamente preferibile a T e si scrive $S < T$ se e solo se $S \leq T$ ed esiste un $\theta_0 \in \Theta$ in corrispondenza del quale il rischio di S è strettamente minore del rischio di T . In simboli:

$$S < T \iff S \leq T \quad \text{e} \quad \exists \theta_0 \in \Theta : R_S(\theta_0) < R_T(\theta_0).$$

Esempio 3. In relazione all'Esempio 2 si presenta il grafico del rischio quadratico medio in funzione di p degli stimatori \bar{X} e V .



Nell'intervallo da 0 a 0,7 il rischio quadratico medio dello stimatore V è non maggiore di quello dello stimatore \bar{X} ma nell'intervallo da 0,7 a 1 è il rischio dello stimatore \bar{X} a essere non maggiore di quello di V ; se ne evince che i due stimatori considerati non sono confrontabili rispetto al rischio quadratico medio. \diamond

Si consiglia allo studente di verificare che, con riferimento all'Esempio 1 e ai fini della stima di θ , lo stimatore \bar{X} è strettamente preferibile allo stimatore corretto X_1 .

Definizione 3. Si dice che uno stimatore T è ammissibile ai fini della stima di $\psi(\theta)$ se non esiste un altro stimatore strettamente preferibile a T .

Si ricordi che una funzione X definita in uno spazio campione Ω e a valori in \mathbb{R} è un numero aleatorio (una variabile aleatoria) se la controimmagine tramite essa di un qualsiasi boreliano di \mathbb{R} è un elemento della famiglia \mathcal{F} degli eventi; in simboli:

$$B \in \mathcal{B}, \quad \{X \in B\} \in \mathcal{F}.$$

Inoltre, assegnata la misura di probabilità \mathbb{P} sulla famiglia degli eventi, ad ogni boreliano $B \in \mathcal{B}$, si assegna la probabilità $\mathbb{P}_X(B) := \mathbb{P}(X \in B)$, ottenendo in tal modo una misura di probabilità sulla famiglia \mathcal{B} . In particolare, se $B =] - \infty, x]$, si definisce la funzione di distribuzione di X ponendo:

$$x \in \mathbb{R}, \quad F_X(x) := \mathbb{P}_X(] - \infty, x]) = \mathbb{P}(X \leq x). \quad (1)$$

1 Proprietà della funzione di distribuzione

1) Una qualsiasi funzione di distribuzione F_X è non decrescente.

DIM

Sia $x_1 < x_2$; per la (1) si ha:

$$\begin{aligned} F_X(x_2) &= \mathbb{P}(X \leq x_2) = \mathbb{P}(\{X \leq x_1\} \cup \{x_1 < X \leq x_2\}) = \mathbb{P}(X \leq x_1) + \mathbb{P}(x_1 < X \leq x_2) \\ &= F_X(x_1) + \mathbb{P}(x_1 < X \leq x_2) \\ &\geq F_X(x_1). \end{aligned}$$

L'ultimo passaggio è dovuto al fatto che la funzione di distribuzione assume valori non negativi. \square

2) Una qualsiasi funzione di distribuzione F_X è continua a destra, ovvero che

$$x \in \mathbb{R}, \quad \lim_{\varepsilon \rightarrow 0^+} F_X(x + \varepsilon) = F_X(x). \quad (2)$$

DIM

Risulta,

$$x \in \mathbb{R}, n \in \mathbb{N}, \quad \left\{X \leq x + \frac{1}{n}\right\} = \{X \leq x\} \cup \left\{x < X \leq x + \frac{1}{n}\right\},$$

da cui discende che

$$x \in \mathbb{R}, n \in \mathbb{N}, \quad F_X\left(x + \frac{1}{n}\right) = F_X(x) + \mathbb{P}\left(x < X \leq x + \frac{1}{n}\right),$$

e, tenendo che la successione di termine generale $\{x < X \leq x + 1/n\}$ decresce all'evento impossibile, si ha:

$$\begin{aligned} x \in \mathbb{R}, \quad \lim_{n \rightarrow \infty} F_X\left(x + \frac{1}{n}\right) &= F_X(x) + \lim_{n \rightarrow \infty} \mathbb{P}\left(x < X \leq x + \frac{1}{n}\right) \\ &= F_X(x) + \mathbb{P}\left(\lim_{n \rightarrow \infty} \left\{x < X \leq x + \frac{1}{n}\right\}\right) \\ &= F_X(x) + \mathbb{P}(\emptyset) = F_X(x). \end{aligned} \quad (3)$$

In definitiva, la (2) segue dalla (3), dalla non decrescenza di F_X e dal teorema ponte \square .

3) Per una qualsiasi funzione di distribuzione F_X risulta:

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{e} \quad \lim_{x \rightarrow +\infty} F_X(x) = 1. \quad (4)$$

DIM

La successione di termine generale $\{X \leq -n\}$ decresce a \emptyset e questo comporta che:

$$\lim_{n \rightarrow \infty} F_X(-n) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq -n) = \mathbb{P}\left(\lim_{n \rightarrow \infty} \{X \leq -n\}\right) = \mathbb{P}(\emptyset) = 0.$$

La prima delle (4) segue da quest'ultima, dalla non decrescenza di F_X e dal teorema ponte.

Analogamente, la successione di termine generale $\{X \leq n\}$ cresce a Ω e questo comporta che:

$$\lim_{n \rightarrow \infty} F_X(n) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq n) = \mathbb{P}\left(\lim_{n \rightarrow \infty} \{X \leq n\}\right) = \mathbb{P}(\Omega) = 1.$$

La seconda delle (4) segue da quest'ultima, dalla non decrescenza di F_X e dal teorema ponte. \square .

Nelle precedenti dimostrazioni, lo scambio del segno di limite con quello di probabilità è lecito in quanto le successioni di eventi coinvolte sono tutte successioni monotone.

Si ricordi che, in generale, la legge di probabilità di un numero aleatorio X è caratterizzata da uno o più parametri. Quindi ogni momento teorico di X , se esistente, è funzione dei parametri. Ad esempio, nel caso $X \sim B(n, p)$ risulta $\mu'_1 = \mathbb{E}(X) = np$ e $\mu_2 = \mathbb{D}^2(X) = np(1-p)$.¹ Nel caso in cui il numero aleatorio X sia dotato di funzione generatrice dei momenti allora si può affermare che esistono finiti sia il momento che il momento centrale di qualsiasi ordine di X .

1 Il metodo dei momenti

Siano n, k e $p \geq k$ numeri interi positivi. Si supponga che la genitrice X di un campione casuale semplice $\underline{X} = (X_1, X_2, \dots, X_n)$ abbia la legge di probabilità caratterizzata da k parametri incogniti $\theta_1, \theta_2, \dots, \theta_k$. Allora, per quanto detto in premessa, scelti opportunamente p indici interi positivi, J_1, J_2, \dots, J_p risulta:

$$r = 1, 2, \dots, p, \quad \mu'_{j_r} = f_{j_r}(\theta_1, \theta_2, \dots, \theta_k). \quad (1)$$

La (1) rappresenta un sistema di p equazioni nei k parametri incogniti. L'obiettivo è quello di risolvere tale sistema per ottenere:

$$s = 1, 2, \dots, k, \quad \theta_s = g_s(\mu'_{j_1}, \mu'_{j_2}, \dots, \mu'_{j_p}). \quad (2)$$

Se si riesce ad ottenere la (2) allora lo stimatore fornito dal metodo dei momenti consiste nella sostituzione dei p momenti teorici considerati con i corrispondenti momenti campionari. Pertanto, si ha:

$$s = 1, 2, \dots, k, \quad \hat{\Theta}_{s,MM} := g_s(\bar{X}^{(j_1)}, \bar{X}^{(j_2)}, \dots, \bar{X}^{(j_p)}). \quad (3)$$

Sussistono i seguenti risultati.

Proposizione 1. *Per le genitrici dotate di funzioni generatrici dei momenti gli stimatori campionari sono stimatori consistenti dei rispettivi momenti teorici.*

Proposizione 2. *Siano consistenti gli stimatori campionari considerati nella (3). Se, per $s = 1, 2, \dots, k$, la funzione g_s nella (3) è continua allora $\hat{\Theta}_{s,MM}$ è uno stimatore consistente per θ_s .*

2 Alcuni esempi

Esempio 1. Sia $b > 0$ e $X \sim U(0, b)$ con b parametro incognito. Scegliendo $p = 1$ e $j_1 = 1$, si ha:

$$\mu'_1 = \frac{b}{2} \quad \Longleftrightarrow \quad b = 2\mu'_1.$$

Ne discende che

$$\hat{B}_{MM} := 2\bar{X}^{(1)} = 2\bar{X}.$$

Ovviamente, lo stimatore ottenuto è consistente per b .

¹ Il momento teorico di ordine k è indicato con il simbolo μ'_k e rappresenta la media di ordine k di X : $\mu'_k = \mathbb{E}(X^k)$. Quindi il primo momento teorico, se esistente e finito, coincide con la media μ di X . Invece, supposta esistente e finita la media μ di X il momento teorico centrale μ_k rappresenta la media di ordine k dello scarto di X dalla sua media: $\mu_k = \mathbb{E}[(X - \mu)^k]$.

Esempio 2. Sia $b > 0$ e $X \sim U(-b, b)$ con b parametro incognito. Scegliendo $p = 1$ e $j_1 = 1$, si ha:

$$\mu'_1 = 0.$$

Quindi la scelta $j_1 = 1$ non conduce ad una funzione di p . Scegliendo $j_1 = 2$ e tenendo conto che $b > 0$, si ha:

$$\mu'_2 = \frac{b^2}{3} \iff b = \sqrt{3\mu'_2}.$$

Ne discende che

$$\hat{B}_{MM} := \sqrt{3\bar{X}^{(2)}}.$$

Lo stimatore ottenuto è consistente per b in quanto la radice quadrata è una funzione continua.

Esempio 3. Siano $a < b$ due numeri reali tali che $a + b \neq 0$ e $a^2 + ab + b^2 \neq 0$. Sia $X \sim U(a, b)$ con a, b parametri incogniti. Risulta:

$$\begin{cases} \mu'_1 = \frac{a+b}{2}, \\ \mu'_2 = \frac{a^2+ab+b^2}{12}. \end{cases} \iff \begin{cases} a = \mu'_1 - \sqrt{3[\mu'_2 - (\mu'_1)^2]}, \\ b = \mu'_1 + \sqrt{3[\mu'_2 - (\mu'_1)^2]}. \end{cases}$$

Ne discende che

$$\begin{cases} \hat{A}_{MM} := \bar{X} - \sqrt{3[\bar{X}^{(2)} - (\bar{X})^2]}, \\ \hat{B}_{MM} := \bar{X} + \sqrt{3[\bar{X}^{(2)} - (\bar{X})^2]}. \end{cases}$$

Lo stimatore ottenuto è consistente per a (per b) in quanto la radice quadrata è una funzione continua e tali sono le operazioni di somma e prodotto di due funzioni continue.

1 Corrispondenze e Applicazioni

Siano S e T due insiemi. Un qualsiasi sottoinsieme del prodotto cartesiano $S \times T$ è una *corrispondenza* tra S e T . Quindi, gli elementi di una corrispondenza sono coppie ordinate in quanto il primo elemento della coppia deve appartenere a S e il secondo a T . Ad esempio, se $S = \{a, b, c, d, e\}$ e $T = \{1, 2\}$, alcune corrispondenze tra S e T sono elencate di seguito:

- $G_1 = \{(a, 1), (b, 1), (c, 2), (d, 2)\};$
- $G_2 = \{(b, 1), (b, 2)\};$
- $G_3 = \{(a, 1), (b, 1), (c, 1), (d, 1), (e, 1)\};$
- $G_4 = \{(a, 1), (b, 1), (c, 1), (d, 1), (e, 2)\}.$

Si osservi che in G_1 l'elemento $e \in S$ non compare come primo elemento di una coppia. In G_2 appare (due volte) un solo un elemento di S . Invece, in G_3 e in G_4 ci sono tutti gli elementi di S (una sola volta).

Nel caso in cui in una corrispondenza ogni elemento S appare una ed una sola volta la corrispondenza è chiamata *applicazione*. Per le applicazioni, l'insieme S è il *dominio*, l'insieme T è il *codominio* e il secondo elemento di ciascuna sua coppia si dice essere l'*immagine* del primo. Quindi, negli esempi elencati G_3 è un'applicazione (l'applicazione costante) e 1 è l'immagine di tutti gli elementi di S . Anche G_4 è un'applicazione e 2 è l'immagine di $e \in S$.

Di solito, per un'applicazione si usa la notazione $f : S \rightarrow T$ e se $s \in S$ allora $f(s) \in T$ è la sua immagine.

Le applicazioni possono essere *iniettive*: non ci sono due o più coppie con il secondo elemento uguale. In altri termini, un'applicazione è iniettiva quando elementi diversi del dominio, $s_1 \neq s_2$, hanno immagini diverse nel codominio: $f(s_1) \neq f(s_2)$. L'applicazione costante non è iniettiva.

Le applicazioni possono essere *suriettive*: ogni elemento di T è presente nella corrispondenza. In altri termini, per ogni elemento di $t \in T$ esiste almeno un elemento $s \in S$ per il quale $f(s) = t$. Quindi un'applicazione costante può essere suriettiva solo nel caso di un codominio costituito da un singoletto. Negli esempi elencati sopra, G_3 non è applicazione suriettiva mentre lo è G_4 .

Un'applicazione che è sia iniettiva che suriettiva è detta *biettiva*.

Infine, se tra due insiemi S e T si può definire un'applicazione biettiva allora essi hanno la stessa cardinalità (oppure, si dice che S e T sono *equipotenti*). Un insieme che è equipotente ad una sua parte propria si dice *infinito*.

2 Controimmagine di un'applicazione

Siano S e T due insiemi e f un'applicazione di S in T . La controimmagine tramite f di un qualsiasi sottoinsieme B di T , indicata con $f^{-1}(B)$ è l'insieme A costituito dagli elementi di S la cui immagine è in B :

$$B \subseteq T, \quad A = f^{-1}(B) = \{s \in S : f(s) \in B\} \subseteq S.$$

Per la controimmagine sussistono i seguenti risultati.

1. La controimmagine del vuoto è il vuoto.
2. La controimmagine del codominio è il dominio.
3. Le controimmagini di insiemi disgiunti in T sono insiemi disgiunti in S .

4. La controimmagine di una unione è l'unione delle controimmagini.
5. La controimmagine di una intersezione è l'intersezione delle controimmagini.
6. La controimmagine di una unione fatta da sottoinsiemi di T a due a due disgiunti è unione di sottoinsiemi di S a due a due disgiunti.
7. La controimmagine della differenza di due sottoinsiemi di T è la differenza delle rispettive controimmagini.
8. La controimmagine del complementare di un sottoinsieme di T è il complementare della controimmagine.

Nel seguito per indicare la controimmagine di un sottoinsieme B si userà la notazione abbreviata:

$$B \subseteq T, \quad \{f \in B\} := f^{-1}(B).$$

3 Applicazioni misurabili e numeri aleatori

Se \mathcal{F} è una sigma-algebra sull'insieme S , la coppia (S, \mathcal{F}) si dice essere uno *spazio misurabile* e gli elementi di \mathcal{F} sono chiamati *insiemi misurabili* o solo *misurabili*.

Siano (S_1, \mathcal{F}_1) e (S_2, \mathcal{F}_2) due spazi misurabili. Un'applicazione f di S_1 in S_2 si dice essere $\mathcal{F}_1/\mathcal{F}_2$ -misurabile se la controimmagine di qualsiasi elemento di \mathcal{F}_2 appartiene a \mathcal{F}_1 :

$$B \in \mathcal{F}_2, \quad f^{-1}(B) \in \mathcal{F}_1.$$

All'insieme \mathbb{R} è associata, di norma, la sigma-algebra generata dagli intervalli aperti e limitati che si designa con il simbolo \mathcal{B} . Gli elementi di \mathcal{B} sono chiamati *insiemi di Borel* oppure *boreliani*.

Si consideri ora uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e lo spazio misurabile $(\mathbb{R}, \mathcal{B})$. Un *numero aleatorio* è un'applicazione X di Ω in \mathbb{R} con la condizione che essa sia \mathcal{F}/\mathcal{A} -misurabile (oppure, dato che \mathcal{A} è fissata, che essa sia \mathcal{F} -misurabile). In altri termini, la controimmagine tramite X di un boreliano deve essere un evento:

$$B \in \mathcal{B}, \quad X^{-1}(B) \in \mathcal{F}.$$

Con il requisito della \mathcal{F} -misurabilità di X si può definire una misura di probabilità sulla sigma-algebra \mathcal{B} :

$$B \in \mathcal{B}, \quad \mathbb{P}_X(B) := \mathbb{P}(X \in B).$$

Infatti,

- $B \in \mathcal{B}, \quad \mathbb{P}_X(B) = \mathbb{P}(X \in B) \geq 0$;
- per il secondo punto del precedente elenco si ha che $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$;
- per il terzo e il quarto punto del precedente elenco si ha che

$$\begin{aligned} (B_n)_{n \in \mathbb{N}} \subseteq \mathcal{B} : B_i \cap B_j = \emptyset, \quad i \neq j, \quad \mathbb{P}_X \left(\bigcup_{n \in \mathbb{N}} B_n \right) &= \mathbb{P} \left(X \in \bigcup_{n \in \mathbb{N}} B_n \right) = \mathbb{P} \left(\bigcup_{n \in \mathbb{N}} \{X \in B_n\} \right) \\ &= \sum_{n \in \mathbb{N}} \mathbb{P}(X \in B_n) = \sum_{n \in \mathbb{N}} \mathbb{P}_X(B_n). \end{aligned}$$

La misura di probabilità \mathbb{P}_X è la *distribuzione* oppure la *legge* del numero aleatorio X . Essa fornisce tutte le informazioni probabilistiche relative al numero aleatorio X ed esse si basano sulla misura di probabilità \mathbb{P} della descrizione matematica dell'esperimento aleatorio sottostante.

1 Problema di analisi combinatoria numero 12 lettera (a)

Si vuole dimostrare in maniera combinatoria la seguente identità:

$$n \in \mathbb{N}, \quad \sum_{i=1}^n k \binom{n}{k} = n \cdot 2^{n-1}. \quad (1)$$

A tale scopo, fissato $n \in \mathbb{N}$, si consideri l'insieme dei soci di un'associazione e si supponga che esso abbia cardinalità n . L'identità (1) è vera in quanto:

- 1) $k = 1, 2, \dots, n$, $k \binom{n}{k}$ rappresenta il numero dei possibili Consigli Direttivi aventi k consiglieri e differenti anche per il consigliere Presidente;
- 2) $\sum_{i=1}^n k \binom{n}{k}$ rappresenta il numero N dei possibili Consigli Direttivi con qualsiasi numero di consiglieri e diversi anche per il consigliere Presidente;
- 3) N si può ottenere anche con il seguente ragionamento: si sceglie uno qualsiasi dei soci a fungere da Presidente (si può fare in n modi diversi) e si forma il Consiglio Direttivo con un qualsiasi sottoinsieme dei soci restanti (si può fare in 2^{n-1} modi diversi): $N = n \cdot 2^{n-1}$.¹

2 Problema di analisi combinatoria numero 12 lettera (b)

Si vuole dimostrare in maniera combinatoria la seguente identità:

$$n \in \mathbb{N}, \quad \sum_{i=1}^n k^2 \binom{n}{k} = n(n+1) \cdot 2^{n-2}. \quad (2)$$

A tale scopo, fissato $n \in \mathbb{N}$, si consideri l'insieme dei soci di un'associazione e si supponga che esso abbia cardinalità n . L'identità (2) è vera in quanto:

- 4) $k = 1, 2, \dots, n$, $k \cdot k \binom{n}{k}$ rappresenta il numero dei possibili Consigli Direttivi aventi k consiglieri e differenti anche per il consigliere Presidente e il consigliere Segretario (non è escluso il caso del consigliere Presidente e Segretario);
- 5) $\sum_{i=1}^n k \cdot k \binom{n}{k}$ rappresenta il numero M dei possibili Consigli Direttivi con qualsiasi numero di consiglieri e diversi anche per il consigliere Presidente e il consigliere Segretario (non è escluso il caso del consigliere Presidente e Segretario);
- 6) M si può ottenere anche come somma di M_1 (specificato nel punto 7) e M_2 (specificato nel punto 8);
- 7) M_1 si ottiene con il seguente ragionamento: si sceglie uno qualsiasi dei soci a fungere da Presidente e Segretario (si può fare in n modi diversi) e si forma il Consiglio Direttivo con un qualsiasi sottoinsieme dei soci restanti (si può fare in 2^{n-1} modi diversi): $M_1 = n \cdot 2^{n-1}$;
- 8) M_2 si ottiene con il seguente ragionamento: si scelgono due soci a fungere da Presidente e da Segretario (si può fare in $n(n-1)$ modi diversi) e si forma il Consiglio Direttivo con un qualsiasi sottoinsieme dei soci restanti (si può fare in 2^{n-2} modi diversi): $M_2 = n(n-1) \cdot 2^{n-2}$;

¹Si ricordi che se S è un insieme finito allora $|\mathcal{P}(S)| = 2^{|S|}$.

9) risulta che

$$\begin{aligned}M_1 + M_2 &= n \cdot 2^{n-1} + n(n-1) \cdot 2^{n-2} = 2n \cdot 2^{n-2} + n(n-1) \cdot 2^{n-2} \\&= n \cdot 2^{n-2} [2 + (n-1)] = n(n+1) \cdot 2^{n-2}.\end{aligned}$$

3 Problema di analisi combinatoria numero 12 lettera (c)

Si vuole dimostrare in maniera combinatoria la seguente identità:

$$n \in \mathbb{N}, \quad \sum_{i=1}^n k^3 \binom{n}{k} = n^2(n+3) \cdot 2^{n-3}. \quad (3)$$

A tale scopo, fissato $n \in \mathbb{N}$, si consideri l'insieme dei soci di un'associazione e si supponga che esso abbia cardinalità n . L'identità (3) è vera in quanto:

- 10) $k = 1, 2, \dots, n$, $k \cdot k \cdot k \binom{n}{k}$ rappresenta il numero dei possibili Consigli Direttivi aventi k consiglieri e differenti anche per il consigliere Presidente, il consigliere Segretario e il consigliere Tesoriere (non è escluso il caso del consigliere che svolge due o tre funzioni);
- 11) $\sum_{i=1}^n k \cdot k \binom{n}{k}$ rappresenta il numero P dei possibili Consigli Direttivi con qualsiasi numero di consiglieri e diversi anche per il consigliere Presidente, il consigliere Segretario e il consigliere Tesoriere (non è escluso il caso del consigliere che svolge due o tre funzioni);
- 12) P si può ottenere anche come somma di P_1 (specificato nel punto 13), P_2 (specificato nel punto 14) e P_3 (specificato nel punto 15);
- 13) lasciato allo studente ...
- 14) lasciato allo studente ...
- 15) lasciato allo studente ...
- 16) calcoli a cura dello studente ...

1 Problema delle Concordanze

Sia n un numero intero e si abbia a disposizione un mazzetto di n cartoncini. Essi vengono numerati da 1 a n sul fronte inizialmente bianco; sul dorso i cartoncini presentano la stessa immagine. L'esperimento \mathcal{E} consiste nel (i) mischiare i cartoncini, (ii) appoggiare su un tavolo il mazzetto mostrandone il dorso, (iii) mostrare il fronte del cartoncino più in alto "chiamando" il numero "uno" e riporre accanto al mazzetto il cartoncino mostrandone il fronte, (iv) procedere allo stesso modo fino all'ultimo cartoncino. Per $i = 1, 2, \dots, n$, si designa con C_i (concordanza) l'evento che si verifica quando si presenta il cartoncino con numero i alla i -ma chiamata.

2 Probabilità di osservare almeno una concordanza nelle n chiamate

L'evento "almeno una concordanza nelle n chiamate" risulta essere l'unione di tutte le concordanze: $\bigcup_{i=1}^n C_i$. Ovviamente, nulla vieta che nella stessa effettuazione dell'esperimento si possono verificare due o più concordanze; pertanto, è necessario applicare la formula di inclusione-esclusione:

$$\mathbb{P}\left(\bigcup_{i=1}^n C_i\right) = \sum_{i=1}^n \mathbb{P}(C_i) - \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{P}(C_i \cap C_j) + \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n \mathbb{P}(C_i \cap C_j \cap C_k) - \dots + (-1)^{n-1} \mathbb{P}(C_1 \cap C_2 \cap \dots \cap C_n) \quad (1)$$

Si osservi che ci sono $n!$ possibili mischiate; risulta:

$$i = 1, 2, \dots, n, \quad \mathbb{P}(C_i) = \frac{(n-1)!}{n!}.$$

Infatti, bloccando solo il cartoncino numerato con i all' i -mo posto dall'alto si verifica C_i ma tutti gli altri possono permutare in tutti i modi possibili. Con lo stesso ragionamento, si ha:

$$i, j = 1, 2, \dots, n : i < j, \quad \mathbb{P}(C_i \cap C_j) = \frac{(n-2)!}{n!}.$$

e

$$i, j, k = 1, 2, \dots, n : i < j < k, \quad \mathbb{P}(C_i \cap C_j \cap C_k) = \frac{(n-3)!}{n!}.$$

Allo stesso modo per i successivi addendi nella (1). In particolare, dal momento che $(n-n)! = 1$, si ha:

$$\mathbb{P}(C_1 \cap C_2 \cap \dots \cap C_n) = \frac{1}{n!}.$$

La formula (1) diventa:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n C_i\right) &= \frac{(n-1)!}{n!} \sum_{i=1}^n 1 - \frac{(n-2)!}{n!} \sum_{i=1}^n \sum_{j=i+1}^n 1 + \frac{(n-3)!}{n!} \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n 1 - \dots + (-1)^{n-1} \frac{1}{n!} \\ &= \frac{(n-1)!}{n!} \binom{n}{1} - \frac{(n-2)!}{n!} \binom{n}{2} + \frac{(n-3)!}{n!} \binom{n}{3} - \dots + (-1)^{n-1} \frac{1}{n!} \\ &= \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1} \frac{1}{n!}. \end{aligned} \quad (2)$$

Il passaggio intermedio nella (2) segue da questo ragionamento (valido per il secondo addendo ma facilmente generalizzabile agli altri): il numero degli addendi 1 è uguale al numero delle combinazioni semplici di lunghezza 2 ottenute dall'insieme $S_n = \{1, 2, \dots, n\}$ in quanto il valore dell'indice j è strettamente maggiore del valore dell'indice i .

In conclusione, si ha:

$$\mathbb{P}\left(\bigcup_{i=1}^n C_i\right) = \sum_{i=1}^n (-1)^{i-1} \frac{1}{i!}. \quad (3)$$

3 Probabilità di osservare zero concordanze nelle n chiamate

Si indichi con $E_{0,n}$ l'evento che si presenta quando in una esecuzione dell'esperimento \mathcal{E} non si osserva alcuna concordanza (ovvero, 0 concordanze); ovviamente, risulta che:

$$E_{0,n} = \left(\bigcup_{i=1}^n C_i\right)^c \implies \mathbb{P}(E_{0,n}) = 1 - \mathbb{P}\left(\bigcup_{i=1}^n C_i\right). \quad (4)$$

In definitiva, dalla (3) e dalla (4), si ottiene:

$$\mathbb{P}(E_{0,n}) = \sum_{i=0}^n (-1)^i \frac{1}{i!}. \quad (5)$$

Prima di terminare il paragrafo si osservi che, indicato con $N_{0,n}$ il numero delle mischiate che non fanno verificare alcuna concordanza, dalla (5) si ricava:

$$\mathbb{P}(E_{0,n}) = \frac{N_{0,n}}{n!} \implies N_{0,n} = n! \cdot \mathbb{P}(E_{0,n}). \quad (6)$$

4 Probabilità di osservare SOLO una concordanza nelle n chiamate

Si indichi con $F_{1,1,n}$ l'evento che si presenta quando in una esecuzione dell'esperimento \mathcal{E} si osserva esattamente una concordanza e questa avviene alla prima chiamata. Ovviamente, indicato con $M_{1,1,n}$ il numero delle mischiate che fanno verificare solo la concordanza nella prima chiamata e zero concordanze nelle successive, risulta che $M_{1,1,n} = N_{0,n-1}$: infatti, dopo la concordanza alla prima chiamata, dalla seconda alla n -chiamata non si verificano altre concordanze. Allora, in virtù della (6), si ha:

$$\mathbb{P}(F_{1,1,n}) = \frac{M_{1,1,n}}{n!} = \frac{N_{0,n-1}}{n!} = \frac{(n-1) \cdot \mathbb{P}(E_{0,n-1})}{n!} = \frac{1}{n} \mathbb{P}(E_{0,n-1}). \quad (7)$$

Si osservi che il risultato ottenuto nella (7) non dipende dal fatto che è stato richiesto di osservare la concordanza alla prima chiamata in quanto lo stesso ragionamento è valido per la richiesta di osservare un'unica concordanza nella generica chiamata. In altri termini, indicato con $F_{1,i,n}$ l'evento che si presenta quando in una esecuzione dell'esperimento \mathcal{E} si osserva esattamente una concordanza e questa avviene alla i -ma chiamata, allora si ha:

$$i = 1, 2, \dots, n, \quad \mathbb{P}(F_{1,i,n}) = \mathbb{P}(F_{1,1,n}) = \frac{1}{n} \mathbb{P}(E_{0,n-1}). \quad (8)$$

Dopo di ciò, se $E_{1,n}$ è l'evento che si presenta quando in una esecuzione dell'esperimento \mathcal{E} si osserva una sola concordanza, risulta

$$\mathbb{P}(E_{1,n}) = \sum_{i=1}^n \mathbb{P}(F_{1,i,n}) = \frac{1}{n} \mathbb{P}(E_{0,n-1}) \sum_{i=1}^n 1 = \mathbb{P}(E_{0,n-1}) = \sum_{i=0}^{n-1} (-1)^i \frac{1}{i!}. \quad (9)$$

5 Probabilità di osservare esattamente due concordanze nelle n chiamate

Si procede sulla falsariga del paragrafo precedente. Si lascia allo studente.

Con riferimento a [1]: **Cap. 3** – *Elementi di probabilità*

- Par. 3.1 – Introduzione
- Par. 3.2 – Spazio degli esiti ed eventi
- Par. 3.3 – I diagrammi di Venn e l'algebra degli eventi
- Par. 3.4 – Assiomi della probabilità
- Par. 3.5 – Spazi di esiti equiprobabili
- Par. 3.5.1 – Il coefficiente binomiale
- Par. 3.6 – Probabilità condizionata
- Par. 3.7 – Fattorizzazione di un evento e formula di Bayes
- Par. 3.8 – Eventi indipendenti

Con riferimento a [2]:

Zara, Calcolo_Combinatorio, Concorde, Insiemi_e_Classi, Definizioni_di_Probabilità, Impostazione_Assiomatica, Indipendenza_e_Condizionamento.

Con riferimento a [1]: **Cap. 4** – *Variabili aleatorie e valore atteso*

- Par. 4.1 – Variabili aleatorie
- Par. 4.2 – Variabili aleatorie discrete e continue
- Par. 4.3 Coppie e vettori di variabili aleatorie
- Par. 4.3.1 – Distribuzione congiunta per variabili aleatorie discrete
- Par. 4.3.2 – Distribuzione congiunta per variabili aleatorie continue
- Par. 4.3.3 – Variabili aleatorie indipendenti
- Par. 4.3.4 – Generalizzazione a più di due variabili aleatorie
- Par. 4.3.5 * – Distribuzioni condizionali
- Par. 4.4 – Valore atteso
- Par. 4.5 – Proprietà del valore atteso
- Par. 4.5.1 – Valore atteso della somma di variabili aleatorie
- Par. 4.6 – Varianza
- Par. 4.7 – La covarianza e la varianza della somma di variabili aleatorie
- Par. 4.8 – La funzione generatrice dei momenti
- Par. 4.9 – La legge debole dei grandi numeri

Con riferimento a [2]:

Misurabilità, Probabilità_Intervalli_e_FD, Proprietà_FD, Variabili_Aleatorie.

Con riferimento a [1]: **Cap. 5 – Modelli di variabili aleatorie**

Par. 5.1 – Variabili aleatorie di Bernoulli e binomiali

Par. 5.2 – Variabili aleatorie di Poisson

Par. 5.4 – Variabili aleatorie uniformi

Par. 5.5 – Variabili aleatorie normali o gaussiane

Par. 5.6 – Variabili aleatorie esponenziali

Par. 5.8.1 – Le distribuzioni chi-quadro (solo il caso di un grado di libertà)

Problema 20 – Cap. 5 – Variabili aleatorie geometriche

Con riferimento a [1]: **Cap. 6 – La distribuzione delle statistiche campionarie**

Par. 6.1 – Introduzione

Par. 6.2 – La media campionaria

Par. 6.3 – Il teorema del limite centrale

Par. 6.3.1 – Distribuzione approssimata della media campionaria

Par. 6.3.2 – Quando un campione è abbastanza numeroso?

Par. 6.4 – La varianza campionaria

Par. 6.5 – Le distribuzioni delle statistiche di popolazioni normali

Par. 6.5.1 – La distribuzione della media campionaria

Par. 6.5.2 – La distribuzione congiunta di \bar{X} e S^2 (fino al Teorema 6.5.1)

Con riferimento a [1]: **Cap. 7 – Stima parametrica**

Par. 7.1 – Introduzione

Par. 7.2 – Stimatori di massima verosimiglianza

Con riferimento a [2]:

Confronto_Stimatori, Metodo_dei_Momenti.

Con riferimento a [3]

Definizioni e classificazione dei caratteri, distribuzioni di frequenza, moda, mediana e quartili, rappresentazioni grafiche; medie analitiche, centri; indici di dispersione, diagramma scatola con baffi.

TESTI CONSIGLIATI

[1] Sheldon Ross – Probabilità e statistica per l'ingegneria e le scienze – APOGEO, 2003

[2] Aniello Buonocore – Appunti dalle lezioni:

Zara, Concorde, Calcolo_Combinatorio, Confronto_Stimatori, Definizioni_di_Probabilità, Impostazione_Assiomatica, Indipendenza_e_Condizionamento, Insiemi_e_Classi, Metodo_dei_Momenti, Misurabilità, Probabilità_Intervalli_e_FD, Proprietà_FD, Variabili_Aleatorie

[3] Aniello Buonocore – Dispensa di Statistica Descrittiva

Si ricordi che una funzione X definita in uno spazio campione Ω e a valori in \mathbb{R} è un numero aleatorio (una variabile aleatoria) se la controimmagine tramite essa di un qualsiasi boreliano di \mathbb{R} è un elemento della famiglia \mathcal{F} degli eventi; in simboli:

$$B \in \mathcal{B}, \quad \{X \in B\} \in \mathcal{F}.$$

Inoltre, assegnata la misura di probabilità \mathbb{P} sulla famiglia degli eventi, ad ogni boreliano $B \in \mathcal{B}$, si assegna la probabilità $\mathbb{P}_X(B) := \mathbb{P}(X \in B)$, ottenendo in tal modo una misura di probabilità sulla famiglia \mathcal{B} . In particolare, se $B =] - \infty, x]$, si definisce la funzione di distribuzione di X ponendo:

$$x \in \mathbb{R}, \quad F_X(x) := \mathbb{P}_X(] - \infty, x]) = \mathbb{P}(X \leq x). \quad (1)$$

1 Proprietà della funzione di distribuzione

1) Una qualsiasi funzione di distribuzione F_X è non decrescente.

DIM

Sia $x_1 < x_2$; per la (1) si ha:

$$\begin{aligned} F_X(x_2) &= \mathbb{P}(X \leq x_2) = \mathbb{P}(\{X \leq x_1\} \cup \{x_1 < X \leq x_2\}) = \mathbb{P}(X \leq x_1) + \mathbb{P}(x_1 < X \leq x_2) \\ &= F_X(x_1) + \mathbb{P}(x_1 < X \leq x_2) \\ &\geq F_X(x_1). \end{aligned}$$

L'ultimo passaggio è dovuto al fatto che la funzione di distribuzione assume valori non negativi. \square

2) Una qualsiasi funzione di distribuzione F_X è continua a destra, ovvero che

$$x \in \mathbb{R}, \quad \lim_{\varepsilon \rightarrow 0^+} F_X(x + \varepsilon) = F_X(x). \quad (2)$$

DIM

Risulta,

$$x \in \mathbb{R}, n \in \mathbb{N}, \quad \left\{X \leq x + \frac{1}{n}\right\} = \{X \leq x\} \cup \left\{x < X \leq x + \frac{1}{n}\right\},$$

da cui discende che

$$x \in \mathbb{R}, n \in \mathbb{N}, \quad F_X\left(x + \frac{1}{n}\right) = F_X(x) + \mathbb{P}\left(x < X \leq x + \frac{1}{n}\right),$$

e, tenendo che la successione di termine generale $\{x < X \leq x + 1/n\}$ decresce all'evento impossibile, si ha:

$$\begin{aligned} x \in \mathbb{R}, \quad \lim_{n \rightarrow \infty} F_X\left(x + \frac{1}{n}\right) &= F_X(x) + \lim_{n \rightarrow \infty} \mathbb{P}\left(x < X \leq x + \frac{1}{n}\right) \\ &= F_X(x) + \mathbb{P}\left(\lim_{n \rightarrow \infty} \left\{x < X \leq x + \frac{1}{n}\right\}\right) \\ &= F_X(x) + \mathbb{P}(\emptyset) = F_X(x). \end{aligned} \quad (3)$$

In definitiva, la (2) segue dalla (3), dalla non decrescenza di F_X e dal teorema ponte \square .

3) Per una qualsiasi funzione di distribuzione F_X risulta:

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{e} \quad \lim_{x \rightarrow +\infty} F_X(x) = 1. \quad (4)$$

DIM

La successione di termine generale $\{X \leq -n\}$ decresce a \emptyset e questo comporta che:

$$\lim_{n \rightarrow \infty} F_X(-n) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq -n) = \mathbb{P}\left(\lim_{n \rightarrow \infty} \{X \leq -n\}\right) = \mathbb{P}(\emptyset) = 0.$$

La prima delle (4) segue da quest'ultima, dalla non decrescenza di F_X e dal teorema ponte.

Analogamente, la successione di termine generale $\{X \leq n\}$ cresce a Ω e questo comporta che:

$$\lim_{n \rightarrow \infty} F_X(n) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq n) = \mathbb{P}\left(\lim_{n \rightarrow \infty} \{X \leq n\}\right) = \mathbb{P}(\Omega) = 1.$$

La seconda delle (4) segue da quest'ultima, dalla non decrescenza di F_X e dal teorema ponte. \square .

Nelle precedenti dimostrazioni, lo scambio del segno di limite con quello di probabilità è lecito in quanto le successioni di eventi coinvolte sono tutte successioni monotone.

Spazi di probabilità

1. Introduzione

Fenomeni aleatori (o **casuali**) – fenomeni dei quali non è possibile avere sotto controllo tutti i fattori che ne determinano lo svolgimento, e dei quali, dunque, è impossibile prevedere il risultato.

(Ω, A, P) – **Spazio di probabilità** usato per descrivere un esperimento aleatorio

- Ω – **Spazio campione** (contiene tutti i possibili risultati dell'esperimento)
 - ω – generico evento detto anche **punto campione**
- A – Famiglia degli eventi “interessanti” per l'esperimento ($A \subseteq \mathcal{P}(\Omega)$)
- P – funzione che fornisce la probabilità che un evento si verifichi ($P: A \rightarrow \mathbb{R}^+$)

Per modellizzare dal punto di vista matematico un fenomeno aleatorio bisogna:

1. Costruire lo spazio campione Ω
2. Stabilire una famiglia A di sottoinsiemi di Ω che rappresentano la famiglia di eventi “interessanti” per l'esperimento
3. Definire una funzione P che fornisca le valutazioni di probabilità per i singoli eventi

(1.1) **Assiomi di Kolmogorov** – Una famiglia A di sottoinsiemi di Ω è detta una **σ -algebra degli eventi** se verifica le seguenti proprietà:

- (i) $\Omega \in A$
- (ii) $A \in A \Rightarrow A^c \in A$
- (iii) Se $\{A_n\}_{n \in \mathbb{N}}$ è una successione di elementi di A , allora $\cup_n A_n \in A$

Osservazioni sugli assiomi di Kolmogorov

- (a) $A \in A$ si può esprimere a parole dicendo semplicemente che “ A è un evento”
- (b) Ω è detto anche **evento certo**
- (c) Se A è un evento, A^c (che è un evento per l'assioma (ii)) si chiama **evento contrario** di A
- (d) Dagli assiomi (i) e (ii) segue che $\emptyset = \Omega^c \in A$. Il sottoinsieme \emptyset è detto **evento impossibile**
- (e) L'assioma (iii) richiede in particolare che, se una famiglia finita di sottoinsiemi di Ω è “interessante” per l'esperimento che stiamo considerando, anche la loro unione lo sia (ciò si esprime dicendo che A è **stabile rispetto all'unione finita**). Più precisamente, nell'assioma (iii) si vuole che sia un evento anche l'unione di ogni successione numerabile di eventi (A è **stabile rispetto all'unione numerabile**)
- (f) Dagli assiomi (ii) e (iii) e dalle leggi di De Morgan discende facilmente la seguente proprietà: **stabilità di A rispetto all'intersezione numerabile**

(iv) Se $\{A_n\}_n$ è una successione numerabile di elementi di A , allora $\bigcap_n A_n \in A$ cioè:

$$\{A_n\}_n \Rightarrow \bigcap_{n=1}^{\infty} A_n \in A$$

Grazie alle formule di De Morgan $\bigcap_{n=1}^{\infty} A_n$ può essere scritta come $(\bigcup_{n=1}^{\infty} A_n^c)^c$

Analizzando quest'ultima espressione notiamo che:

- ogni $A_n^c \in A$ (per l'assioma (ii)¹)
- l'unione degli $A_n^c \in A$, in quanto è l'unione di elementi appartenenti ad A (assioma (iii)¹)
- il complemento dell'unione degli $A_n^c \in A$ (per l'assioma (ii)¹)

(1.2) Siano Ω un insieme, A una σ -algebra di sottoinsiemi di Ω . Un'applicazione $P: A \rightarrow R^+$ si chiama **probabilità** (su A) se verifica le due condizioni seguenti:

- (i) $P(\Omega) = 1$
- (ii) Se $\{A_n\}_{n \in N}$ è una successione di elementi di A due a due disgiunti (cioè, tali che $\forall i \neq j$ si abbia $A_i \cap A_j = \emptyset$) risulta $P(\bigcup_n A_n) = \sum_n P(A_n)$ (proprietà di **finita additività**; nel caso di una successione numerabile, si parla di **σ -additività**)

Nota: eventi disgiunti vengono anche detti **eventi incompatibili**.

(1.3) **Teorema** – $P(\emptyset) = 0$

Dimostrazione – Scriviamo ora l'insieme \emptyset come unione di un numero infinito di insiemi \emptyset , cioè $\emptyset = \emptyset \cup \emptyset \cup \dots \cup \emptyset \dots$

Applicando l'assioma (ii)² si ha

$$P(\emptyset) = P\left(\bigcup_{n=1}^{\infty} \emptyset_n\right) = \sum_{n=1}^{\infty} P(\emptyset_n)$$

dove, per $n = 1, 2, \dots$, $P(\emptyset_1) = P(\emptyset_2) = \dots$

Ponendo $P(\emptyset_n) = \varepsilon$, si ha

$$(*) \quad P(\emptyset) = \sum_{n=1}^{\infty} \varepsilon$$

Per $\varepsilon > 0$, si ha che la serie (*) diverge (ciò non può verificarsi, in quanto andrebbe contro l'assioma (i)²). Dunque l'unico valore che ε può assumere è 0 e si ha che

$$P(\emptyset) = \sum_{n=1}^{\infty} 0 = 0$$

¹ – Assiomi di Kolmogorov per le σ -algebre degli eventi (Assiomi (1.1))

² – Assiomi di Kolmogorov per la probabilità (Assiomi (1.2))

(1.4) **Teorema** – La σ -additività comporta anche la finita additività

Dimostrazione – Siano

$$B_1 = A_1$$

$$B_2 = A_2$$

...

$$B_n = A_n$$

$$B_{n+1} = B_{n+2} = \dots = \emptyset$$

sequenze di eventi, dove $B_i \cap B_j = \emptyset, \forall i \neq j$.

Per l'assioma (ii)¹ abbiamo

$$(*) \quad P\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} P(B_k)$$

Scomponendo la serie (*) nella somma di due serie, in cui la prima va da 1 a n e la seconda va da $n+1$ a ∞ , si ha

$$(**) \quad P\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^n P(B_k) + \sum_{k=n+1}^{\infty} P(B_k)$$

La serie $\sum_{k=n+1}^{\infty} P(B_k)$ è uguale a 0 in quanto ogni B_k , con $k = n+1, n+2, \dots$, è pari a \emptyset e la somma di

infiniti $P(\emptyset)$ è uguale a 0 (vedi (1.3))

Da qui, sostituendo nella (*) si ha

$$(***) \quad P\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^n P(B_k)$$

Dalla (***) scomponiamo l'unione in

$$(***) \quad P\left(\bigcup_{k=1}^{\infty} B_k\right) = P\left(\bigcup_{k=1}^n B_k \cup \bigcup_{k=n+1}^{\infty} B_k\right)$$

La $\bigcup_{k=n+1}^{\infty} B_k$ è uguale a \emptyset in quanto ogni B_k , con $k = n+1, n+2, \dots$, è pari a \emptyset

Dunque $P\left(\bigcup_{k=1}^{\infty} B_k\right) = P\left(\bigcup_{k=1}^n B_k\right) = \sum_{k=1}^n P(B_k)$. Essendo $B_i = A_i$, per $i = 1, 2, \dots, n$, si ha

$$P\left(\bigcup_{k=1}^n B_k\right) = P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k) = \sum_{k=1}^n P(B_k)$$

¹ – Assioma di Kolmogorov per la probabilità (Assiomi (1.2))

Nota: Nel caso di insiemi Ω a cardinalità finita o numerabile, tutti i sottoinsiemi di Ω sono considerati eventi, cioè si pone sempre $\mathcal{A} = \mathcal{P}(\Omega)$.

2. Alcune conseguenze degli assiomi

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità. Siano A e B due eventi. Possiamo scrivere dunque

$$B = B \cup \emptyset = B \cup (A \cap A^c) = (A \cap B) \cup (A^c \cap B)$$

Gli eventi $(A \cap B)$ e $(A^c \cap B)$ sono evidentemente disgiunti, pertanto è possibile applicare la proprietà additiva di P , e abbiamo

$$(2.1) \quad P(B) = P(A \cap B) + P(A^c \cap B)$$

Questa formula è di per sé molto importante: capita spesso di non saper calcolare direttamente la probabilità di un evento (in questo caso B), ma di saper “spezzare” B in due parti disgiunte tramite un evento ausiliario (in questo caso A) la cui probabilità è facilmente calcolabile.

La (2.1) ha alcune conseguenze importanti:

- (i) Per $B = \Omega$ e per ogni $A \in \mathcal{A}$, ricordando l’assioma (i) di P , la (2.1) diventa

$$1 = P(\Omega) = P(A) + P(A^c)$$

ovvero

$$(2.2) \quad P(A^c) = 1 - P(A)$$

La (2.2) permette di calcolare la probabilità dell’evento contrario di A se è nota la probabilità di A .

- (ii) Per $A \subseteq B$, la (2.1) diventa

$$(2.3) \quad P(B) = P(A) + P(A^c \cap B)$$

poiché $P(A^c \cap B) \geq 0$, si ricava per $A \subseteq B$

$$(2.4) \quad P(A) \leq P(B)$$

Questa proprietà si chiama proprietà di *isotonia* della probabilità.

Una sua semplice conseguenza è il fatto che ogni evento ha probabilità inferiore o uguale a 1: infatti, se A è un evento, dalla ovvia relazione $A \subseteq \Omega$ segue

$$P(A) \leq P(\Omega) = 1$$

In altre parole, ogni funzione di probabilità assume valori nell’intervallo $[0, 1]$.

- (iii) Se $A \subseteq B$, si usa scrivere

$$A^c \cap B = B \setminus A$$

e la (2.3) diventa

$$P(B \setminus A) = P(B) - P(A)$$

- (iv) Siano ora A e B due eventi qualsiasi. Possiamo scrivere

$$A \cup B = A \cup (A^c \cap B)$$

essendo i due eventi A e $(A^c \cap B)$ disgiunti, dalla proprietà di additività della probabilità e dalla (2.1) si deduce

$$(2.5) \quad P(A \cup B) = P(A) + P(A^c \cap B) = P(A) + P(B) - P(A \cap B)$$

Questa formula permette di calcolare la probabilità dell'unione di due eventi A e B, a patto di conoscere già le probabilità degli eventi A, B e della loro intersezione.

La (2.5) è un caso particolare di una formula (nota come la **formula di inclusione-esclusione**) che si utilizza per il calcolo della probabilità dell'unione di una famiglia finita di eventi. Per completezza diamo anche la formula per il caso di tre eventi A, B e C

$$P(A \cup B \cup C) = P((A \cup B) \cup C)$$

per la (2.5) si ha

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

applicando la (2.5) a $P(A \cup B)$

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

applicando la proprietà distributiva dell'intersezione rispetto all'unione a $(A \cup B) \cap C$

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

applicando la (2.5) a $P((A \cap C) \cup (B \cap C))$

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

ovvero

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

- (v) Sia $\{A_n\}_n$ una successione di eventi. Poiché

$$\cup_n A_n = (\cap_n A_n^c)^c$$

dalla (2.2) si ricava

$$(2.6) \quad P(\cup_n A_n) = P((\cap_n A_n^c)^c) = 1 - P(\cap_n A_n^c)$$

Questa formula è spesso utile perchè tipicamente la probabilità di un'intersezione di eventi risulta più facile da calcolare della probabilità di un'unione.

Tutte le proprietà viste finora sono conseguenza della sola proprietà di additività finita della probabilità. Queste ultime due, invece, utilizzano la σ -additività.

- (vi) (***proprietà di passaggio al limite sulle successioni crescenti***) Sia $\{A_n\}_n$ una successione crescente di eventi (cioè, $\forall n \in N$, risulta $A_n \subseteq A_{n+1}$). Si ha allora

$$P(\cup_n A_n) = P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$$

- (vii) (***proprietà di passaggio al limite sulle successioni decrescenti***) Sia $\{A_n\}_n$ una successione decrescente di eventi (cioè, $\forall n \in N$, risulta $A_n \supseteq A_{n+1}$). Si ha allora

$$P(\cap_n A_n) = P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$$

3. Spazi di probabilità uniformi

Supponiamo di essere in presenza di un esperimento aleatorio che si svolge in **condizioni di equiprobabilità** (cioè, come nel caso dei lanci della moneta equilibrata, in modo che non ci sia motivo di pensare che un particolare risultato sia più o meno frequente degli altri), e sia Ω lo spazio campione associato ad esso. La condizione di equiprobabilità impone evidentemente che ci sia

$$P(\omega) = \text{costante} \equiv p, \quad \forall \omega \in \Omega$$

Determiniamo allora la costante p . Si ha l'eguaglianza ovvia

$$\Omega = \cup_{\omega \in \Omega} \{\omega\}$$

per gli assiomi (i)¹ e (ii)¹ si deduce

$$1 = P(\Omega) = \sum_{\omega \in \Omega} P(\omega) = \sum_{\omega \in \Omega} p = p (\text{card } \Omega)$$

da questa relazione segue allora

$$p = \frac{1}{\text{card } \Omega}$$

Questo determina la probabilità di ciascun punto campione. Per trovare ora la probabilità di un qualunque evento A si ragiona in modo sostanzialmente identico: dall'eguaglianza

$$A = \cup_{\omega \in A} \{\omega\}$$

si deduce

$$(3.1) \quad P(A) = \sum_{\omega \in A} P(\omega) = \sum_{\omega \in A} \frac{1}{\text{card } \Omega} = \frac{\text{card } A}{\text{card } \Omega}$$

Questa formula definisce allora la funzione probabilità su tutta la σ -algebra degli eventi.

Sia (Ω, A, P) uno spazio di probabilità, tale che $\text{card } \Omega < +\infty$. (Ω, A, P) si dice uno **spazio di probabilità uniforme** se

$$P(\omega) = \frac{1}{\text{card } \Omega}$$

per ogni $\omega \in \Omega$.

La formula (3.1) dice semplicemente che, nel caso di uno spazio uniforme, la probabilità di un evento si calcola come il rapporto tra il “numero dei casi favorevoli” e il “numero dei casi possibili”. E' bene ricordare, comunque, che solo in alcuni casi lo spazio di probabilità da costruire sarà uniforme.

¹ – Assioma di Kolmogorov per la probabilità (Assiomi (1.2))

4. Probabilità condizionale e indipendenza

(4.1) Siano A, B due eventi, con $P(B) > 0$. Si chiama **probabilità condizionale di A , dato B** (oppure “*sapendo che si è verificato B* ”) la quantità

$$(4.2) \quad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Si sottolinea il fatto che la scrittura $P(A|B)$ non rappresenta la probabilità dell'evento “ $A|B$ ”, ma solo un modo per indicare il rapporto che compare al secondo membro della (4.2).

(4.3) **Teorema** – Sia B un evento, con $P(B) > 0$. L'applicazione $Q: A \rightarrow R^+$ definita da $Q(A) = P(A|B)$ è una probabilità.

Dimostrazione – Dalla definizione di probabilità (1.2), l'applicazione Q è una probabilità (su A) se verifica le seguenti due condizioni:

- (i) $Q(\Omega) = 1$
- (ii) Se $\{A_n\}_n$ è una successione di elementi di A due a due disgiunti (cioè, tali che $\forall i \neq j$ si abbia $A_i \cap A_j = \emptyset$) risulta $Q(\cup_n A_n) = \sum_n Q(A_n)$
- (i) per come è stata definita l'applicazione

$$Q(\Omega) = P(\Omega|B)$$

per la (4.2) si ha

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

essendo $\Omega \cap B = B$

$$\frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

- (ii) per come è stata definita l'applicazione

$$Q(\cup_n A_n) = P(\cup_n A_n|B)$$

cioè

$$P(\cup_n A_n|B) = P((A_1 \cup A_2 \cup \dots \cup A_n \cup \dots)|B)$$

applicando la (4.2) otteniamo

$$P((A_1 \cup A_2 \cup \dots \cup A_n \cup \dots)|B) = \frac{P((A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) \cap B)}{P(B)}$$

per la proprietà distributiva di \cap rispetto a \cup , abbiamo

$$\frac{P((A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B) \cup \dots)}{P(B)}$$

gli eventi $A_i \cap B$, per $i = 1, 2, \dots$, sono evidentemente disgiunti; dunque è possibile applicare l'assioma (ii)¹ e si ha

$$\frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B) \cup \dots)}{P(B)} = \frac{P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B) + \dots}{P(B)}$$

le varie $P(A_i \cap B)$ possono essere raggruppate tutte in una serie, ovvero

$$(*) \quad \frac{P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B) + \dots}{P(B)} = \sum_n \frac{P(A_n \cap B)}{P(B)}$$

per la (4.2) si ha

$$(**) \quad \frac{P(A_n \cap B)}{P(B)} = P(A_n | B)$$

dove, per l'ipotesi del teorema,

$$(***) \quad P(A_n | B) = Q(A_n)$$

applicando, infine, la (**) e (***) alla (*) si ottiene

$$\sum_n \frac{P(A_n \cap B)}{P(B)} = \sum_n P(A_n | B) = \sum_n Q(A_n)$$

Può capitare che l'informazione supplementare "si è verificato B" non cambi la nostra valutazione della probabilità del verificarsi di A.

Siano A e B due eventi, con $P(B) > 0$. A e B si dicono *indipendenti (tra loro)* se

$$(4.4) \quad P(A | B) = P(A)$$

Usando la (4.2), la (4.4) si può scrivere nella forma

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

o anche, moltiplicando per $P(B)$,

$$(4.5) \quad P(A \cap B) = P(A)P(B)$$

La (4.5) è chiaramente simmetrica in A e B (del resto lo è anche l'espressione A e B sono indipendenti (tra loro) usata nella definizione, mentre non è affatto evidente che lo sia anche la (4.4)). Inoltre, mentre la (4.4) perde di significato se B è un evento con probabilità nulla, la (4.5) ha ancora senso. Essa risulta più comoda come definizione di indipendenza; dunque sostituiremo la definizione provvisoria con la seguente:

Due eventi A e B si dicono *indipendenti (tra loro)* se vale la (4.5)

Sarà necessario estendere la definizione di indipendenza ad una famiglia qualsiasi di eventi. Per iniziare, consideriamo il caso di tre eventi A, B e C. Per analogia con il caso di due eventi, potremmo pensare che sia corretto dire che essi sono indipendenti se

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

Purtroppo questa condizione non garantisce che gli eventi siano anche “due a due” indipendenti, come ci dice invece l’intuizione. La definizione ragionevole è allora la seguente

Tre eventi A, B e C si dicono **indipendenti** se valgono tutte le seguenti condizioni

- (i) $P(A \cap B \cap C) = P(A)P(B)P(C)$
- (ii) $P(A \cap B) = P(A)P(B)$
- (iii) $P(A \cap C) = P(A)P(C)$
- (iv) $P(B \cap C) = P(B)P(C)$

La definizione precedente si generalizza in modo ovvio ad una famiglia di qualsiasi eventi: essi si diranno indipendenti se per ogni sottofamiglia finita la probabilità dell’intersezione è uguale al prodotto delle singole probabilità; in termini precisi

(4.6) Gli eventi $\{A_i\}_{i \in I}$ si dicono **indipendenti (tra loro)** se per ogni $k \in N$ e per ogni successione finita i_1, i_2, \dots, i_k di indici di I , si ha

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$$

Nota: L’ipotesi di indipendenza consente di costruire un particolare tipo di spazio di probabilità, utilizzabile nella pratica in una varietà di situazioni solo apparentemente differenti tra loro.

(4.7) Siano A e B due eventi. Le affermazioni seguenti sono equivalenti:

- (a) A e B sono indipendenti
- (b) A e B^c sono indipendenti
- (c) A^c e B sono indipendenti
- (d) A^c e B^c sono indipendenti

Dimostrazione – Supponiamo che i due eventi siano indipendenti (affermazione (a)), ovvero che

$$(*) \quad P(A \cap B) = P(A)P(B)$$

Dimostriamo che le rimanenti affermazioni sono equivalenti.

(b) Scomponiamo $P(A)$, come in (2.1), tramite l’evento ausiliario B ed abbiamo

$$(b.1) \quad P(A) = P(A \cap B) + P(A \cap B^c)$$

sostituendo la (*) in (b.1) si ha

$$P(A \cap B^c) = P(A) - P(A)P(B)$$

mettendo in evidenza $P(A)$ al secondo membro, otteniamo

$$(b.2) \quad P(A \cap B^c) = P(A)(1 - P(B))$$

per la (2.2), la (b.2) diventa

$$P(A \cap B^c) = P(A)P(B^c)$$

(c) Scomponiamo $P(B)$, come in (2.1), tramite l'evento ausiliario A ed abbiamo

$$(c.1) \quad P(B) = P(A \cap B) + P(A^c \cap B)$$

sostituendo la (*) in (c.1) si ha

$$P(A^c \cap B) = P(B) - P(A)P(B)$$

mettendo in evidenza $P(B)$ al secondo membro, otteniamo

$$(c.2) \quad P(A^c \cap B) = P(B)(1 - P(A))$$

per la (2.2), la (c.2) diventa

$$P(A^c \cap B) = P(A^c)P(B)$$

(d) Dalle formule di De Morgan otteniamo

$$P(A^c \cap B^c) = P((A \cup B)^c)$$

applicando la (2.2)

$$P(A^c \cap B^c) = 1 - P(A \cup B)$$

dalla (2.5) e (*) abbiamo

$$P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A)P(B)$$

applicando la (2.2) a $(1 - P(A))$

$$P(A^c \cap B^c) = P(A^c) - P(B) + P(A)P(B)$$

mettendo in evidenza $-P(B)$ si ha

$$P(A^c \cap B^c) = P(A^c) - P(B)(1 - P(A))$$

applicando la (2.2) a $(1 - P(A))$

$$P(A^c \cap B^c) = P(A^c) - P(B)P(A^c)$$

mettendo in evidenza $P(A^c)$

$$P(A^c \cap B^c) = P(A^c)(1 - P(B))$$

applicando la (2.2) a $(1 - P(B))$

$$P(A^c \cap B^c) = P(A^c)P(B^c)$$

5. Schema delle prove indipendenti

Supponiamo di dover eseguire n volte (in condizioni di indipendenza) un certo esperimento che produce due soli possibili risultati, che chiameremo convenzionalmente **successo** e **insuccesso** e che indicheremo rispettivamente con i simboli 1 e 0. Supponiamo inoltre di sapere che il successo (risp. l'insuccesso) si presenta con probabilità p (risp. $1 - p$) nella generica ripetizione dell'esperimento (con $0 < p < 1$). Lo spazio campione è il seguente

$$\Omega = \{0, 1\}^n$$

Sia dunque $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ (con $\omega_i \in \{0, 1\}$) un elemento di Ω ; osserviamo che le quantità

$$\sum_{i=1}^n \omega_i, \quad n - \sum_{i=1}^n \omega_i$$

Non sono altro che il numero di simboli uguali a 1 e il numero i simboli uguali a 0 che formano la sequenza ω . Pertanto avremo

$$P(\omega) = p^{(\sum_{i=1}^n \omega_i)} (1 - p)^{(n - \sum_{i=1}^n \omega_i)}$$

6. La formula di Bayes

(6.1) Siano (Ω, A, P) uno spazio di probabilità fissato, e A_1, \dots, A_n n eventi in A . Si dice che A_1, \dots, A_n è una **partizione** di (Ω, A, P) se valgono entrambe le seguenti proprietà:

- (i) $\bigcup_{i=1}^n A_i = \Omega$
- (ii) $A_i \cap A_j = \emptyset \quad \forall (i, j) \text{ con } i \neq j$

Gli eventi di una partizione vanno pensati come una successione di n eventualità che possono presentarsi durante lo svolgimento dell'esperimento; esse esauriscono tutte le possibilità (sicuramente una di esse si verifica) e si escludono a vicenda (due di esse non possono verificarsi contemporaneamente).

(6.2) **Teorema (Formula di Bayes)** – Sia (Ω, A, P) uno spazio di probabilità. Sia A_1, \dots, A_n una partizione di Ω , tale che $P(A_i) > 0$ per ogni $i = 1, 2, \dots, n$. Sia infine B un evento con $P(B) > 0$. Allora per ogni $k = 1, 2, \dots, n$ risulta

$$P(A_k | B) = \frac{P(B | A_k)P(A_k)}{\sum_{i=1}^n P(B | A_i)P(A_i)}$$

Dimostrazione – Per la (4.2) abbiamo che

$$(*) \quad P(A_k | B) = \frac{P(A_k \cap B)}{P(B)}$$

Ora soffermiamoci sul numeratore e sul denominatore della (*)

- (numeratore) per la proprietà simmetrica dell'intersezione

$$P(A_k \cap B) = P(B \cap A_k)$$

applicando la (4.2) a $P(B \cap A_k)$ si ottiene

$$(**) \quad P(B \cap A_k) = P(B | A_k)P(A_k)$$

- (denominatore) possiamo scrivere

$$B = B \cap \Omega$$

per il punto (i) della definizione (6.1) otteniamo che

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^n A_i\right)$$

applicando la proprietà distributiva di \cap rispetto a \cup , abbiamo

$$B \cap \left(\bigcup_{i=1}^n A_i\right) = \left(\bigcup_{i=1}^n (B \cap A_i)\right)$$

poiché gli eventi $(B \cap A_i)$ sono tra loro a due a due disgiunti, per l'assioma di additività della probabilità si ottiene

$$(***) \quad P(B) = \sum_{i=1}^n P(B \cap A_i)$$

come abbiamo fatto in precedenza per $P(B \cap A_k)$ (vedi (**)) si può scrivere

$$P(B \cap A_i) = P(B | A_i)P(A_i)$$

e quindi la (***) diventa

$$P(B) = \sum_{i=1}^n P(B | A_i)P(A_i)$$

applicando la (**) e la (****) alla (*) si ricava allora

$$P(A_k | B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(B | A_k)P(A_k)}{\sum_{i=1}^n P(B | A_i)P(A_i)}$$

ovvero la formula di Bayes che cercavamo, in cui

1. $P(A_k | B)$ è detta ***probabilità a posteriori***
2. $P(B | A_k)$ è detta ***probabilità di verosimiglianza***
3. $P(A_k)$ è detta ***probabilità a priori***

DEFINIZIONI E CLASSIFICAZIONE DEI CARATTERI

Con il termine statistica descrittiva si intende un insieme di tecniche e strumenti finalizzati ad assolvere uno dei principali compiti assegnati alla Statistica:

descrivere, rappresentare e sintetizzare in maniera opportuna un insieme di dati relativamente ad una o più caratteristiche di una popolazione di interesse.

Per *popolazione* (o collettivo statistico) si intende la totalità dei casi, ovvero dei *membri* (o unità statistiche) sui quali è possibile *rilevare* uno o più *caratteri* che rivestono particolare importanza per il fenomeno che si sta studiando.

Esempio 1

Durante il semestre viene proposto agli studenti dei corsi di Laurea offerti in Ateneo il questionario sulla valutazione della struttura didattica nella sua complessità.

Qui la *popolazione* è costituita dagli studenti con o senza obbligo di frequenza ciascuno dei quali è un *membro*. Per lo studio in questione ci si serve di un *questionario* avente un certo numero di domande raggruppate per sezioni. In questo contesto ciascuna domanda del questionario corrisponde ad un *carattere*. Per le maggior parte delle domande lo studente può rispondere scegliendo una tra 4 possibili risposte tra loro alternative:

Decisamente no, Più no che sì, Più sì che no, Decisamente sì.

Per completezza si riportano alcune di queste domande.

Le aule dove si svolgono le lezioni sono adeguate? Le modalità con le quali si è svolto l'insegnamento (lezioni, diapositive, audiovisivi, ecc.) sono soddisfacenti? Il carico di studio dell'insegnamento è proporzionale ai crediti assegnati? Il docente è reperibile per chiarimenti e spiegazioni?

◇

Esempio 2

In un allevamento di bufale da latte si vuole mettere in relazione la produzione giornaliera di latte con la grandezza delle mammelle e con lo stato di salute delle stesse.

Qui la *popolazione* è costituita da tutte le bufale dell'allevamento ciascuna delle quali è un *membro*. Per lo studio in questione i *caratteri* da *rilevare* su ciascun membro sono: la produzione di latte, una misura lineare delle mammelle, lo stato di salute delle mammelle. Per la rilevazione sono necessari due *strumenti di misura* e la diagnosi di un veterinario determinata su una prestabilita *scala* di valori.

◇

Le diverse espressioni con le quali si manifesta un carattere si chiamano *modalità*.

Nell'Esempio 1 le modalità, per ciascun carattere, sono rappresentate dalle 4 risposte alternative tra loro. Ciascuno dei caratteri, ovvero una domanda proposta nel questionario, è di tipo *qualitativo ordinale*. Infatti, si osservi che esse sono delle etichette e che in relazione al gradimento espresso dallo studente risulta:

Decisamente no < Più no che sì < Più sì che no < Decisamente sì.

Invece nell'Esempio 2, i caratteri “produzione giornaliera di latte” e “grandezza delle mammelle” sono *quantitativi continui* in quanto per la loro determinazione sono necessari uno strumento di misurazione di una capacità (volume) e uno strumento per la misurazione di una lunghezza e pertanto i dati ottenuti sono numeri decimali appartenenti ad un conveniente intervallo. Nello stesso esempio, il carattere “stato di salute delle mammelle” è invece di tipo qualitativo ordinale e le modalità sono i diversi valori della scala prescelta (ad esempio, mammelle sane, infiammazione lieve, infiammazione moderata, infiammazione grave). Il fatto stesso che si parla di scala comporta la presenza di un ordinamento tra le etichette:

sana < infiammazione lieve < infiammazione moderata < infiammazione grave.

Per un esempio di carattere *qualitativo nominale* si pensi al gruppo sanguigno che, prescindendo dal fattore Rh, si manifesta con le modalità: A, B, AB e 0.

Per un esempio di carattere *quantitativo discreto* si pensi al numero delle persone presenti nello stato di famiglia dei residenti nel comune di Napoli alla data del più recente censimento ISTAT.

DISTRIBUZIONI DI FREQUENZA

Si può senz'altro pensare ad una popolazione come ad un insieme. La cardinalità della popolazione rilevata è detta *taglia*; essa, di solito, si designa con la lettera N . Un carattere viene designato con una lettera latina maiuscola mentre i valori rilevati vengono rappresentati con la stessa lettera ma in minuscolo. Se, allora, Y rappresenta il carattere sotto studio non continuo, la sequenza

$$\underline{y} = (y_1, y_2, \dots, y_N)$$

rappresenta l'intera rilevazione dei dati. Denotiamo ora con x_1, x_2, \dots, x_k (con $k \leq N$) le modalità del carattere Y . Per $j = (1, 2, \dots, k)$, il numero n_j che rappresenta quante volte è presente la modalità x_j in \underline{y} è detto *frequenza assoluta* (o semplicemente *frequenza*) della modalità x_j . In aggiunta,

$$j = (1, 2, \dots, k), \quad f_j = \frac{n_j}{N}$$

è detto *frequenza relativa* della modalità x_j .

La frequenza relativa è più informativa della frequenza assoluta in quanto tiene conto anche della taglia. È del tutto ovvio che la somma delle k frequenze assolute vale N mentre la somma delle k frequenze relative vale 1.

Infine,

$$j = (1, 2, \dots, k), \quad F_j = f_1 + f_2 + \dots + f_j = F_{j-1} + f_j$$

è detta *frequenza assoluta cumulata* di x_j . È del tutto ovvio che $F_k = 1$.

La rappresentazione tabellare di quanto appena esposto è detta *distribuzione di frequenza* della rilevazione dati $\underline{y} = (y_1, y_2, \dots, y_N)$:

modalità del carattere	frequenza assoluta	frequenza relativa	frequenza relativa cumulata
x_1	n_1	$f_1 = n_1 / N$	$F_1 = f_1$
x_2	n_2	$f_1 = n_1 / N$	$F_2 = F_1 + f_2$
x_{k-1}	n_{k-1}	$f_{k-1} = n_{k-1} / N$	$F_{k-1} = F_{k-2} + f_{k-1}$
x_k	n_k	$f_k = n_k / N$	1
	<hr/>	<hr/>	
	N	1	

Per un carattere quantitativo continuo è necessario dapprima procedere alla suddivisione in *classi di modalità* dell'intervallo nel quale si manifesta il carattere stesso. Qui bisogna fare attenzione a rendere le classi contigue ma senza ingenerare dubbi di collocazione di un dato nelle classi stesse. Allo scopo, se i dati (y_1, y_2, \dots, y_N) sono riportati con s cifre decimali, è sufficiente tenere conto dell'operazione di arrotondamento e rappresentare gli estremi delle classi con $s + 1$ cifre decimali. Ad esempio, supponiamo che una rilevazione di un peso è effettuata con bilancia digitale precisa all'ettogrammo. Sia 3,2 kg il peso minore rilevato: 3,2 rappresenta tutte le misurazioni comprese nell'intervallo $[3,15; 3,25[$. Allo stesso modo sia 4,4 kg il peso maggiore rilevato: 4,4 rappresenta tutte le misurazioni comprese nell'intervallo $[4,35; 4,45[$. Quindi se è vero che l'intervallo nel quale si osservano i dati è $[3,2; 4,4]$ è a maggior ragione vero che senza l'operazione di arrotondamento esso sarebbe stato $[3,15; 4,45[$. Allora, è quest'ultimo intervallo che deve essere suddiviso nel numero desiderato di classi e queste devono avere come estremi dei numeri aventi due cifre decimali. Dopo di ciò, per un carattere quantitativo, la prima colonna contiene le classi di modalità così individuate.

MODA E QUARTILI

Definizione

Si consideri un carattere (di qualsiasi tipo). La modalità corrispondente alla frequenza (assoluta o relativa) più grande viene detta *moda* (M_0) della rilevazione dati.

◇

Definizione

Si consideri un carattere (non qualitativo nominale). La modalità corrispondente alla più piccola frequenza relativa cumulata maggiore o uguale a 0,5 viene detta *mediana* (M_1) oppure *secondo quartile* (Q_2) della rilevazione dati. La mediana suddivide la rilevazione dati ordinata

$$(y_{(1)}, y_{(2)}, \dots, y_{(N)}) \quad \text{con} \quad y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(N)}$$

in due parti: i dati minori della mediana sono nello stesso numero dei dati maggiori della mediana.

◇

Definizione

Si consideri un carattere (non qualitativo nominale). La modalità corrispondente alla più piccola frequenza relativa cumulata maggiore o uguale a 0,25 viene detta *primo quartile* (Q_1) della rilevazione dati. Il primo quartile corrisponde anche alla mediana dei dati minori della mediana.

◇

Definizione

Si consideri un carattere (non qualitativo nominale). La modalità corrispondente alla più piccola frequenza relativa cumulata maggiore o uguale a 0,75 viene detta *terzo quartile* (Q_3) della rilevazione dati. Il terzo quartile corrisponde anche alla mediana dei dati maggiori della mediana.

◇

Definizione

Si consideri un carattere (non qualitativo nominale). Il dato più piccolo $y_{(1)}$ è il *quartile di ordine zero* (Q_0) della rilevazione dati.

◇

Definizione

Si consideri un carattere (non qualitativo nominale). Il dato più grande $y_{(N)}$ è il *quarto quartile* (Q_4) della rilevazione dati.

◇

RAPPRESENTAZIONI GRAFICHE

Le distribuzioni di frequenza possono essere rappresentate in forma grafica con scelta eseguita opportunamente rispetto al tipo di carattere. Per i caratteri qualitativi e quantitativi discreti, oltre al *diagramma circolare*, la rappresentazione grafica più usata è quella del *diagramma a barre verticali*: ogni barra verticale è centrata attorno ad una modalità ed ha altezza pari alla frequenza assoluta (oppure relativa).

Per i soli caratteri quantitativi continui è opportuno utilizzare la rappresentazione grafica detta *istogramma*.

La differenza qualitativa tra un istogramma ed un diagramma a barre verticali è che nell'istogramma le barre verticali devono essere contigue (in effetti sono dei rettangoli e, quindi, dotati di base e altezza). Ma la differenza sostanziale è che nell'istogramma la frequenza della classe di modalità rappresenta l'area del rettangolo (e non l'altezza come nel diagramma a barre).

Pertanto, per qualsiasi $j \in \{1, 2, \dots, k\}$, se b_j rappresenta l'ampiezza della j -ma classe di modalità, l'altezza h_j del relativo rettangolo è ottenuto mediante la formula inversa per l'area e quindi:

$$j = (1, 2, \dots, k), \quad h_j = \frac{n_j}{b_j}.$$

Alcuni esempi di rappresentazioni grafiche con ulteriori dettagli sono stati forniti nella dispensa *Statistica_Descrittiva_1*.

INDICI DI POSIZIONE

Si vuole ora considerare il problema di *sintetizzare* la rilevazione dati

$$\underline{y} = (y_1, y_2, \dots, y_N)$$

con un unico valore di *tendenza centrale*, ossia un valore che fornisca un'indicazione di massima sulla localizzazione di \underline{y} . Ciò è utile non solo per una più immediata comprensione dei risultati dell'indagine ma anche per istituire un confronto del fenomeno studiato con altri fenomeni dello stesso tipo.

Per i caratteri quantitativi è possibile fare ricorso ai due indici media e mediana. Nel caso di taglia N dispari, un modo pratico per ottenere la mediana senza costruire la distribuzione di frequenza è quello di ordinare i dati dal più piccolo al più grande e poi, ricorsivamente, depennare il minimo e il massimo fino a quando resta un unico elemento che è la mediana. Se invece N è pari, alla fine di tutti i depennamenti restano due elementi. In tal caso, bisogna separare il caso (i) i due elementi restanti sono uguali (il loro valore comune coincide con la mediana) dal caso (ii) i due elementi restanti sono diversi. Nel caso (ii) con un carattere quantitativo la mediana è la semisomma dei

due elementi restanti. Nel caso (ii) con un carattere qualitativo ordinale la mediana è indeterminata.

Per i caratteri quantitativi ci sono almeno altri due approcci teorici in grado di far ottenere indici di tendenza centrale:

- le *medie analitiche*;
- i *centri*.

MEDIE ANALITICHE

Sia $\underline{y} = (y_1, y_2, \dots, y_N)$ una rilevazione dati su un carattere quantitativo. Per ottenere una *media analitica* bisogna dapprima specificare un criterio C (o *funzione di circostanza*) rispetto al quale si vuole ottenere la valutazione di tendenza centrale. Dopo di ciò bisogna determinare un numero reale y per il quale, indicata con $\underline{y}^* = (y, y, \dots, y)$ una rilevazione dati (fittizia) di taglia N aventi tutti gli elementi uguali a y , la valutazione della funzione di circostanza C su \underline{y} deve coincidere con la valutazione di C su \underline{y}^* . In simboli:

$$C(y_1, y_2, \dots, y_N) = C(y, y, \dots, y).$$

Esempio: media aritmetica

Si scelga come funzione di circostanza C la somma dei dati:

$$C(y_1, y_2, \dots, y_N) = y_1 + y_2 + \dots + y_N.$$

Dopo di ciò,

$$\begin{aligned} C(y_1, y_2, \dots, y_N) &= C(y, y, \dots, y) \\ \Leftrightarrow y_1 + y_2 + \dots + y_N &= \underbrace{y + y + \dots + y}_{N\text{-volte}} \\ \Leftrightarrow y_1 + y_2 + \dots + y_N &= N \cdot y. \end{aligned}$$

In definitiva, la soluzione dell'equazione di circostanza è la *media aritmetica* dei dati:

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_N}{N} = \frac{1}{N} \sum_{i=1}^N y_i.$$

◇

Esempio: media geometrica

Si scelga come funzione di circostanza C il prodotto dei dati (che devono essere rilevati da un carattere positivo):

$$C(y_1, y_2, \dots, y_N) = y_1 \cdot y_2 \cdots y_N.$$

Dopo di ciò,

$$C(y_1, y_2, \dots, y_N) = C(y, y, \dots, y)$$

$$\Leftrightarrow y_1 \cdot y_2 \cdot \dots \cdot y_N = \underbrace{y \cdot y \cdot \dots \cdot y}_{N\text{-volte}} \Leftrightarrow y_1 \cdot y_2 \cdot \dots \cdot y_N = y^N.$$

In definitiva, la soluzione dell'equazione di circostanza è la *media geometrica* dei dati:

$$M_g = \sqrt[N]{y_1 \cdot y_2 \cdot \dots \cdot y_N} = \sqrt[N]{\prod_{i=1}^N y_i}.$$

◇

Esempio: media armonica

Si scelga come funzione di circostanza C la somma dei reciproci dei dati (che devono essere da un carattere non nullo):

$$C(y_1, y_2, \dots, y_N) = \frac{1}{y_1} + \frac{1}{y_2} + \dots + \frac{1}{y_N}.$$

Dopo di ciò,

$$C(y_1, y_2, \dots, y_N) = C(y, y, \dots, y)$$

$$\Leftrightarrow \frac{1}{y_1} + \frac{1}{y_2} + \dots + \frac{1}{y_N} = \underbrace{\frac{1}{y} + \frac{1}{y} + \dots + \frac{1}{y}}_{N\text{-volte}}$$

$$\Leftrightarrow \left(\frac{1}{y_1} + \frac{1}{y_2} + \dots + \frac{1}{y_N} \right) = \frac{N}{y}.$$

In definitiva, la soluzione dell'equazione di circostanza è la *media armonica* dei dati:

$$M_a = \frac{N}{\frac{1}{y_1} + \frac{1}{y_2} + \dots + \frac{1}{y_N}} = \left(\frac{\frac{1}{y_1} + \frac{1}{y_2} + \dots + \frac{1}{y_N}}{N} \right)^{-1}$$

$$= \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{y_i} \right)^{-1}.$$

Si può allora dire che la *media armonica* M_a di dati non nulli è uguale al reciproco della *media aritmetica* dei loro reciproci.

◇

CENTRI

Sia $\underline{y} = (y_1, y_2, \dots, y_N)$ una rilevazione dati su un carattere quantitativo Y e sia

$$\forall x \in \mathbb{R}, \quad d(x, \underline{y}) \geq 0$$

una funzione che si ritiene adatta a rappresentare la *distanza* di un generico valore reale x da tutti gli elementi della rilevazione dati \underline{y} . Si definisce *centro* di una rilevazione dati \underline{y} , e lo si indica con $\xi(\underline{y})$, il punto di minimo assoluto della funzione $d(x, \underline{y})$; in simboli:

$$\xi(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} d(x, \underline{y}).$$

In particolare, sono molto spesso considerate le seguenti funzioni di distanza di *tipo potenze*:

$$\forall x \in \mathbb{R}, \quad d_0(x, \underline{y}) = \frac{1}{N} \sum_{i=1}^N |x - y_i|^0$$

e

$$\forall r \in \mathbb{N}, \quad \forall x \in \mathbb{R}, \quad d_r(x, \underline{y}) = \sqrt[r]{\frac{1}{N} \sum_{i=1}^N |x - y_i|^r}.$$

La funzione $d_r(x, \underline{y})$ è detta anche *distanza di ordine r* e il suo punto di minimo assoluto è detto *centro di ordine r* . Quindi il centro di ordine r di una rilevazione dati \underline{y} è il numero reale $\xi_r(\underline{y})$ che rende minima la distanza di ordine r . In simboli:

$$\xi_r(\underline{y}) := \operatorname{argmin}_{x \in \mathbb{R}} d_r(x, \underline{y}).$$

Teorema 1

Il centro di ordine 0 della rilevazione dati \underline{y} , ovvero $\xi_0(\underline{y})$, coincide con la moda della rilevazione dati.

Dimostrazione

Per definizione, la funzione

$$d_0(x, \underline{y}) = \frac{1}{N} \sum_{i=1}^N |x - y_i|^0$$

rappresenta la distanza di ordine 0 di x dall'intera rilevazione dati \underline{y} , mentre $|x - y_i|^0$ rappresenta la distanza tra x e il generico dato y_i .

Se ne ricava che quando x coincide con y_i la distanza è nulla ovvero l'addendo i -mo non porta contributo alla distanza complessiva. Pertanto, si ha:

$$d_0(x, \underline{y}) = \begin{cases} 1, & \text{se } x \notin \{y_1, y_2, \dots, y_n\}, \\ 1 - \frac{n_x}{N} < 1, & \text{se } x \in \{y_1, y_2, \dots, y_n\}. \end{cases}$$

Nella precedente formula n_x rappresenta il numero delle volte che si presenta il dato x . Allora, il minimo si trova tra gli elementi di \underline{y} e precisamente è quel dato al quale compete la frequenza maggiore che per definizione è la moda della rilevazione dati:

$$\xi_0(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} \frac{1}{N} \sum_{i=1}^N |x - y_i|^0 = M_0.$$

Teorema 2 (senza dimostrazione)

Il centro di ordine 1 della rilevazione dati \underline{y} , ovvero $\xi_1(\underline{y})$, coincide con la mediana:

$$\xi_1(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} d_1(x, \underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} \frac{1}{N} \sum_{i=1}^N |x - y_i| = M_1 \equiv Q_2.$$

In più, il valore minimo assunto dalla distanza di ordine 1 vale:

$$d_1(Q_2, \underline{y}) = \frac{1}{N} \sum_{i=1}^N |Q_2 - y_i|.$$

◇

Teorema 3

Il centro di ordine 2 della rilevazione dati \underline{y} , $\xi_2(\underline{y})$, coincide con la media aritmetica \bar{y} . In più, il valore minimo assunto dalla distanza di ordine 2:

$$d_2(\bar{y}, \underline{y}) = \frac{1}{N} \sum_{i=1}^N (\bar{y} - y_i)^2.$$

Dimostrazione

Per definizione,

$$d_2(x, \underline{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N |x - y_i|^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x - y_i)^2}.$$

D'altra parte, dal momento che la funzione radice quadrata è strettamente crescente nel suo dominio $[0, +\infty[$, il minimo della distanza di ordine due viene raggiunto in corrispondenza del minimo del radicando, ovvero della funzione:

$$x \in \mathbb{R}, \quad f(x) = \frac{1}{N} \sum_{i=1}^N (x - y_i)^2.$$

Pertanto,

$$\xi_2(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} d_2(x, \underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} f(x).$$

Ricerca dei punti stazionari di $f(x)$:

$$\begin{aligned} x \in \mathbb{R}, \quad f'(x) = \frac{2}{N} \sum_{i=1}^N (x - y_i) &\Rightarrow f'(x) = 0 \Leftrightarrow \sum_{i=1}^N (x - y_i) = 0 \\ &\Leftrightarrow x = \frac{1}{N} \sum_{i=1}^N y_i \equiv \bar{y}. \end{aligned}$$

Ricerca dei punti di minimo e massimo relativo di $f(x)$:

$$f''(x) = \frac{2}{N} \sum_{i=1}^N 1 = 2 \Rightarrow f''(\bar{y}) = 2 > 0,$$

la qual cosa implica che \bar{y} è un punto di minimo relativo per $f(x)$.

Ricerca del minimo assoluto di $f(x)$:

$$\bar{y} = \operatorname{argmin}_{x \in \mathbb{R}} f(x)$$

in quanto $\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow -\infty} f(x) = +\infty$, la derivata prima di $f(x)$ è definita in \mathbb{R} , e $f(x)$ ammette un unico punto di minimo relativo.

In definitiva,

$$\xi_2(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} d_2(x, \underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} f(x) = \bar{y}.$$

Definizione

Il *centro di ordine infinito* di una rilevazione dati \underline{y} si indica con il simbolo $\xi_{\infty}(\underline{y})$ ed è definito dalla posizione:

$$\xi_{\infty}(\underline{y}) := \lim_{r \rightarrow +\infty} \xi_r(\underline{y}).$$

◇

Definizione

Il *valore centrale* di una rilevazione dati \underline{y} è la semisomma tra il dato minimo e il dato massimo:

$$\frac{y_{(1)} + y_{(N)}}{2} = \frac{Q_{(0)} + Q_{(4)}}{2}$$

◇

Teorema (senza dimostrazione)

Il *centro di ordine infinito della rilevazione dati \underline{y}* coincide con il *valore centrale*:

$$\xi_{\infty}(\underline{y}) = \frac{y_{(1)} + y_{(N)}}{2}.$$

◇

INDICI DI DISPERSIONE

Un *indice di dispersione* (o indicatore di dispersione o indice di variabilità o indice di variazione) serve per descrivere sinteticamente la misura con la quale una rilevazione dati di un carattere quantitativo è distante da una sua tendenza centrale. La dispersione esprime la bontà o la inadeguatezza di un indice di tendenza centrale quale descrittore di una distribuzione di frequenza.

Per i caratteri qualitativi si usano gli *indici di diversità* dei quali quello maggiormente usato è l'*indice di ricchezza* che opera un semplice conteggio del numero delle modalità presenti nella rilevazione dati.

Sia $\underline{y} = (y_1, y_2, \dots, y_N)$ una rilevazione dati su un carattere quantitativo. Sono indici di dispersione i seguenti.

- a) La differenza tra il dato più grande da quello più piccolo (*campo o intervallo di variazione*):

$$\Gamma = y_{(N)} - y_{(1)} = Q_4 - Q_0.$$

Il campo di variazione è associato al valore centrale $\frac{y_{(1)} + y_{(N)}}{2}$.

- b) La differenza tra il terzo e il primo quartile (*differenza interquartilica*):

$$\gamma = Q_3 - Q_1.$$

La differenza interquartilica è associata alla mediana Q_2 .

- c) La media del valore assoluto delle differenze dei dati dalla loro mediana Q_2 (*scarto mediano assoluto*):

$$S_{Q_2} = \frac{1}{N} \sum_{i=1}^N |y_i - Q_2|.$$

Lo scarto mediano assoluto è associato alla mediana Q_2 : si veda il Teorema 2 in questa stessa dispensa.

- d) La media del valore assoluto delle differenze dei dati dalla loro media aritmetica \bar{y} (*scarto medio assoluto*):

$$S_{\bar{y}} = \frac{1}{N} \sum_{i=1}^N |y_i - \bar{y}|.$$

Lo scarto medio assoluto è associato alla media aritmetica \bar{y} .

- e) La media del quadrato delle differenze dei dati dalla loro media aritmetica M_2 (*varianza*):

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2.$$

La varianza è associata alla media aritmetica \bar{y} : si veda il Teorema 3 in questa stessa dispensa

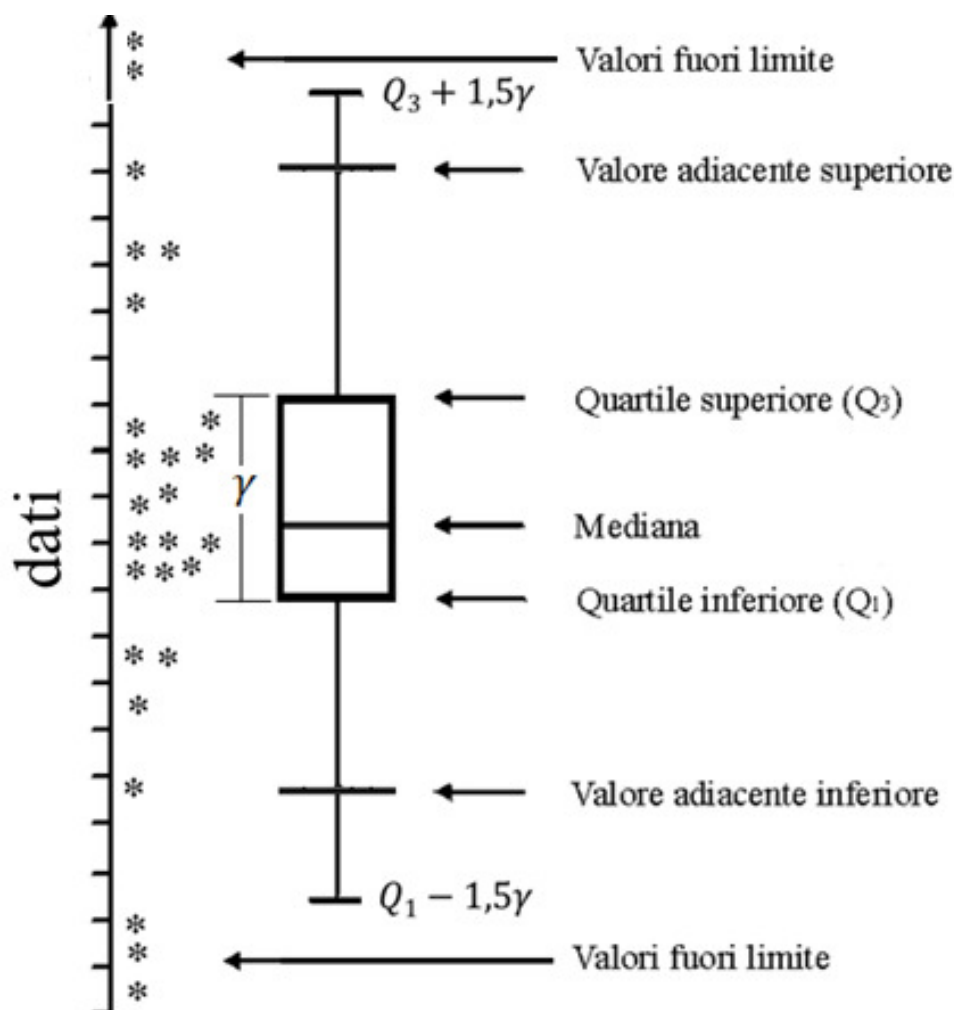
- f) La radice quadrata della varianza (*scarto tipo o deviazione standard*):

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}.$$

La deviazione standard è associata alla media aritmetica \bar{y} .

DIAGRAMMA SCATOLA CON BAFFI

Un metodo grafico per rappresentare una distribuzione di frequenze che mette in risalto anche la dispersione intorno alla mediana è il *grafico della scatola con baffi* (*Box Plot*):



La linea interna alla scatola rappresenta la *Mediana* della distribuzione. Le linee estreme della scatola rappresentano il primo ed il terzo quartile.

La distanza interquartilica γ , è una misura della **dispersione** della distribuzione. Il 50% dei dati si trovano comprese tra questi due valori. Se l'intervallo interquartilico è piccolo, tale metà delle osservazioni si trova fortemente concentrata intorno alla mediana; all'aumentare della distanza interquartilica aumenta la dispersione del 50% dei dati centrali intorno alla mediana.

Le distanze tra ciascun quartile e la mediana forniscono informazioni relativamente alla **forma** della distribuzione. Se una distanza è diversa dall'altra allora la distribuzione è asimmetrica.

Le linee che si allungano dai bordi della scatola (*baffi*) individuano gli intervalli in cui sono posizionati i valori rispettivamente minori di Q_1 e maggiori di Q_3 ; i punti estremi dei "baffi" evidenziano i *valori adiacenti*. Il *valore adiacente inferiore* (VAI) è il valore più piccolo tra i dati che risulta maggiore o uguale a $Q_1 - 1,5\gamma$. Il *valore adiacente superiore* (VAS), invece, è il valore più grande tra i dati che risulta minore o uguale a $Q_3 + 1,5\gamma$.

I valori esterni ai valori adiacenti (chiamati in genere *valori fuori limiti oppure valori anomali*), vengono segnalati individualmente nel box-plot per meglio evidenziarne la presenza e la posizione. Questi valori infatti costituiscono una “anomalia” rispetto alla maggior parte dei valori osservati e pertanto è necessario identificarli per poterne analizzare le caratteristiche e le eventuali cause che li hanno determinati. Essi forniscono informazioni ulteriori sulla dispersione e sulla forma della distribuzione.

Quando il valore adiacente superiore coincide con il dato più grande e il valore adiacente inferiore coincide con il dato più piccolo, allora non comparirà alcun valore anomalo.

Variabili aleatorie

1. Definizioni e prime proprietà

Molto spesso, l'esperimento che viene effettuato produce una quantità numerica X che deve essere studiata; anzi, si può dire che in moltissime situazioni è proprio la quantità X l'oggetto interessante per chi esegue l'esperimento, (più che l'esperimento in sé per sé).

Esempio – Un programma informatico produce sequenze aleatorie di lunghezza 5, composte dai simboli 0 e 1. Questo esperimento si modella con uno schema $n = 5$ prove indipendenti (p = probabilità di ottenere un simbolo “1”).

E' noto che il tempo impiegato per produrre un simbolo “1” (risp. “0”) è 2 (risp. 3) μsec . Il tecnico che utilizza il programma è interessato a sapere quanto tempo è necessario per produrre una sequenza; in altre parole, le domande a cui vuole saper rispondere sono del tipo “qual è la probabilità che il programma impieghi più di $x \mu\text{sec}$ ”, oppure “qual è la probabilità che il programma impieghi un tempo compreso fra $x \mu\text{sec}$ e $y \mu\text{sec}$ ”, e così via, (piuttosto che domande come “qual è la probabilità che il programma produca (per esempio) la sequenza (1, 0, 0, 1, 1)”).

Cioè in questa situazione siamo interessati a ottenere informazioni “di tipo probabilistico” sulla quantità X = tempo (in μsec) impiegato dal programma per produrre una sequenza.

Cerchiamo ora di formalizzare domande del tipo “qual è la probabilità che il programma impieghi un tempo compreso fra $x \mu\text{sec}$ e $y \mu\text{sec}$ ”. Questa domanda equivale a cercare la probabilità del sottoinsieme costituito dai possibili risultati ω dell'esperimento per i quali risulti $x < X(\omega) \leq y$, ovvero la probabilità di

$$\{\omega \in \Omega : x < X(\omega) \leq y\}$$

Pertanto una richiesta ragionevole sarà che i sottoinsiemi di Ω del tipo precedente siano eventi.

E' facile far vedere che, se si suppone che siano eventi i sottoinsiemi

$$\{\omega \in \Omega : x < X(\omega) \leq y\}$$

allora sono eventi tutti i sottoinsiemi del tipo

$$\{\omega \in \Omega : X(\omega) \in I\}$$

dove I è un qualsiasi intervallo (limitato o no, chiuso o no) di \mathbb{R} .

Intervalli di questo tipo sono ad esempio $(-\infty, b)$, $(-\infty, b]$, (a, b) , $(a, b]$, $[a, b)$, $[a, b]$, $(a, +\infty)$ e $[a, +\infty)$ con $a, b \in \mathbb{R}$.

Osserviamo che è del tipo considerato anche ogni sottoinsieme quale

$$\{\omega \in \Omega : X(\omega) = x\}$$

(basta prendere $I = \{x\}$, intervallo chiuso ridotto al solo punto x)

Più in generale si può far vedere che $\{X \in I\}$ è un evento per una famiglia più vasta di sottoinsiemi I di \mathbb{R} (cioè non solo per gli intervalli). Tali insiemi sono detti **misurabili**. E' opportuno sapere che non tutti i sottoinsiemi di \mathbb{R} sono misurabili.

D'ora in avanti l'evento $\{\omega \in \Omega : X(\omega) \in I\}$ sarà indicato con la notazione abbreviata $\{X \in I\}$. Allo stesso modo scriveremo $\{x < X \leq y\}$ al posto di $\{\omega \in \Omega : x < X(\omega) \leq y\}$, $\{X \leq y\}$ invece di $\{\omega \in \Omega : X(\omega) \leq y\}$ e così via.

Per evitare possibili equivoci, insistiamo sul fatto che quando si scrive, per esempio, $\{1 \leq X \leq 2\}$, non si intende l'intervallo $[1, 2]$ ma un evento.

(1.2) Sia (Ω, A, P) uno spazio di probabilità. Una **variabile aleatoria** (o anche **random variable**) è una funzione $X : \Omega \rightarrow \mathbb{R}$ tale che, per ogni $t \in \mathbb{R}$ risulti

$$\{X \leq t\} \in A$$

La precedente definizione è giustificata dal fatto che, affinché tutti i sottoinsiemi del tipo $\{X \in I\}$, con I insieme misurabile, siano eventi, è necessario e sufficiente supporre che lo siano anche i sottoinsiemi $\{X \leq y\}$.

Per brevità scriveremo semplicemente $P(X \in I)$ anziché (come sarebbe corretto) $P(\{X \in I\})$.

(1.3) Sia X una variabile aleatoria su (Ω, A, P) . Si chiama **legge** di X l'applicazione

$$I \rightarrow P(X \in I)$$

(definita sulla classe degli insiemi misurabili di \mathbb{R} e a valori in $[0, 1]$).

La legge di una variabile aleatoria X va pensata come una “fotografia” delle varie probabilità assegnate a tutti gli eventi del tipo interessante (cioè $\{X \in I\}$), e se ne intuisce dunque l'importanza. D'altra parte è evidente che nella pratica sarà impossibile “elencare” tutte le probabilità $P(X \in I)$, per ciascun I . Si capisce dunque la necessità di procurarsi alcuni strumenti non troppo complicati, che ci consentano di “risparmiare” calcoli.

2. Funzione di ripartizione di una variabile aleatoria: definizione e proprietà

(2.1) Sullo spazio di probabilità (Ω, A, P) sia X una variabile aleatoria assegnata. Si chiama **funzione di ripartizione** (o **di distribuzione**) (f.d.r.) di X la funzione $F: \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$F(t) = P(X \leq t)$$

Quando saranno in gioco più variabili aleatorie X, Y, \dots , le rispettive funzioni di ripartizione verranno indicate con F_X, F_Y, \dots .

(2.2) Teorema – Proprietà caratteristiche di una funzione di ripartizione

a) Siano X una variabile aleatoria, F la sua funzione di ripartizione. Valgono le seguenti proprietà:

- (i) $0 \leq F(t) \leq 1, \quad \forall t \in \mathbb{R}$ (ovvio, in quanto $F(t)$ rappresenta la probabilità di un evento, più precisamente dell'evento $\{X \leq t\}$)
- (ii) F è non decrescente, cioè, per ogni coppia x, y di numeri reali, con $x \leq y$ risulta

$$F(x) \leq F(y)$$

Dimostrazione – $F(x)$ rappresenta la probabilità dell'evento $\{X \leq x\}$, mentre $F(y)$ rappresenta la probabilità dell'evento $\{X \leq y\}$. Essendo $x \leq y$, si deduce che $\{X \leq x\} \subseteq \{X \leq y\}$ e quindi, per la proprietà di isotonia (la (2.4)¹), si ha che $F(x) \leq F(y)$

(iii) valgono le relazioni

$$\text{a) } \lim_{n \rightarrow -\infty} F(n) = 0 \qquad \text{b) } \lim_{n \rightarrow +\infty} F(n) = 1$$

(Dimostrazione di b) – Osserviamo che $F(n)$ è la probabilità dell'evento $A_n = \{X \leq n\}$. La successione di eventi $(A_n)_{n \in \mathbb{N}}$ è crescente ed inoltre $\bigcup_{n \in \mathbb{N}} A_n = \Omega$. Dunque, per la proprietà di passaggio al limite delle successioni crescenti di eventi (vedi punto (vi) – paragrafo 2)¹ si ha

$$\lim_{n \rightarrow +\infty} F(n) = \lim_{n \rightarrow +\infty} P(A_n) = P(\bigcup_{n \in \mathbb{N}} A_n) = P(\Omega) = 1$$

(Dimostrazione di a) – Osserviamo che $F(n)$ è la probabilità dell'evento $A_n = \{X \leq -n\}$. La successione di eventi $(A_n)_{n \in \mathbb{N}}$ è decrescente ed inoltre $\bigcap_{n \in \mathbb{N}} A_n = \emptyset$. Dunque, per la proprietà di passaggio al limite delle successioni decrescenti di eventi (vedi punto (vii) – paragrafo 2)¹ si ha

$$\lim_{n \rightarrow -\infty} F(n) = \lim_{n \rightarrow -\infty} P(A_n) = P(\bigcap_{n \in \mathbb{N}} A_n) = P(\emptyset) = 0$$

¹ – Riferimento al capitolo 1 (Spazi di probabilità)

(iv) F è continua a destra, cioè per ogni $x \in \mathbb{R}$ si ha

$$F(x^+) = \lim_{n \rightarrow +\infty} F(x + \frac{1}{n}) = \lim_{t \rightarrow x^+} F(t) = F(x)$$

Per la definizione di funzione di ripartizione, possiamo scrivere

$$F(x^+) = \lim_{n \rightarrow +\infty} F(x + \frac{1}{n}) = \lim_{n \rightarrow +\infty} P(X \leq x + \frac{1}{n})$$

scomponiamo ora l'evento $\{X \leq x + \frac{1}{n}\}$ in un'unione di due eventi disgiunti

$$\lim_{n \rightarrow +\infty} P(X \leq x + \frac{1}{n}) = \lim_{n \rightarrow +\infty} P\left((X \leq x) \cup (x < X \leq x + \frac{1}{n})\right)$$

applicando la proprietà di finita additività della probabilità, otteniamo

$$\lim_{n \rightarrow +\infty} P\left((X \leq x) \cup (x < X \leq x + \frac{1}{n})\right) = \lim_{n \rightarrow +\infty} \left[P(X \leq x) + P(x < X \leq x + \frac{1}{n}) \right]$$

per le proprietà dei limiti si ricava che

$$\lim_{n \rightarrow +\infty} \left[P(X \leq x) + P(x < X \leq x + \frac{1}{n}) \right] = \lim_{n \rightarrow +\infty} P(X \leq x) + \lim_{n \rightarrow +\infty} P(x < X \leq x + \frac{1}{n})$$

ove

$$\lim_{n \rightarrow +\infty} P(X \leq x) = \lim_{n \rightarrow +\infty} F(x) = F(x)$$

e

$\lim_{n \rightarrow +\infty} P(x < X \leq x + \frac{1}{n}) = 0$ in quanto la probabilità che dell'evento $\{X \in (x, x + \frac{1}{n}]\}$ tende a 0 al crescere di n

Per quanto evidente, si sottolinea il fatto che con la scrittura $F(x^+)$ non si intende il valore di F nel punto (inesistente) x^+ , ma il limite da destra verso x .

b) sia $F: \mathbb{R} \rightarrow \mathbb{R}$ una funzione, che gode di tutte le proprietà (i), (ii), (iii), e (iv) elencate sopra. Allora esiste una variabile aleatoria X definita su un opportuno spazio di probabilità, che ammette F come funzione di ripartizione.

Il teorema precedente è utile quando si debba decidere se un'assegnata funzione F è una funzione di ripartizione: lo sarà se gode di tutte le proprietà (i)–(iv), non lo sarà se anche una sola di esse non è verificata.

(2.3) **Teorema** – La funzione di ripartizione individua (univocamente) la legge della variabile aleatoria. In altre parole se due variabili aleatorie hanno la stessa funzione di ripartizione, allora hanno la stessa legge (le due variabili si dicono *somiglianti* o anche *equamente distribuite*).

Vediamo in concreto il calcolo della legge di X a partire dalla sua funzione di ripartizione F in alcuni casi particolari (cioè il calcolo della quantità $P(X \in I)$ per qualche tipo particolare di insieme misurabile I).

(i) $I = (a, b]$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

(ii) $I = (-\infty, b)$

$$P(X < b) = \lim_{n \rightarrow +\infty} P(X \leq b - \frac{1}{n}) = \lim_{t \rightarrow b^-} F(t) = F(b^-)$$

con il simbolo $F(b^-)$ si è indicato il limite verso b da sinistra, (e non il valore nel punto inesistente b^-), e il calcolo precedente dice che tale limite esiste e vale $P(X < b)$.

(iii) $I = \{x\}$

$$P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(x^-)$$

(iv) $I = [a, b)$

$$P(a \leq X < b) = P(X < b) - P(X < a) = F(b^-) - F(a^-)$$

(v) $I = (a, b)$

$$P(a < X < b) = P(X < b) - P(X \leq a) = F(b^-) - F(a)$$

(vi) $I = [a, b]$

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a^-)$$

(vii) $I = (a, +\infty)$

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a)$$

(viii) $I = [a, +\infty)$

$$P(X \geq a) = 1 - P(X < a) = 1 - F(a^-)$$

La formula al punto (iii) è interessante. Ricordando che ogni funzione di ripartizione è continua a destra, essa può essere scritta anche nella forma equivalente

$$(2.4) \quad P(X = x) = F(x^+) - F(x^-)$$

ed interpretata dicendo che la probabilità che X assuma il valore x è esattamente uguale al “salto” che la funzione di ripartizione di X presenta nel punto di ascissa x .

I punti in cui F è continua sono tutti e soli i valori assunti dalla variabile X con probabilità nulla: questo segue immediatamente dalla relazione (2.4), osservando che F è continua in x se e soltanto se risulta $F(x^+) - F(x^-) = 0$.

Ciò giustifica la seguente definizione:

(2.5) La variabile aleatoria X è detta **continua** se la sua funzione di ripartizione F è continua, cioè se X assume con probabilità nulla ogni valore reale x (in simboli, $P(X = x) = 0 \quad \forall x \in \mathbb{R}$)

Esempio – La funzione di ripartizione definita da

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 & x \geq 1 \end{cases}$$

non è continua. In particolare, la variabile aleatoria X assume il solo valore 1 con probabilità uguale a 1 (il “salto della funzione di ripartizione”).

3. Variabili aleatorie discrete

(3.1) Sia X una variabile aleatoria definita sullo spazio (Ω, A, P) . X si dice **discreta** se il suo dominio S , detto anche **spettro**, è un sottoinsieme finito o numerabile di \mathbb{R} .

(3.2) Sia X una variabile aleatoria discreta. L'applicazione $p : \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$p(x) = P(X = x)$$

si chiama **densità (discreta)** della variabile aleatoria X .

Quando saranno in gioco più variabili aleatorie X, Y, \dots useremo i simboli p_X, p_Y, \dots per indicare le rispettive densità.

(3.3) Sia p la densità di un'assegnata variabile aleatoria discreta X . p gode delle seguenti proprietà:

- (i) $p(x) \geq 0 \quad \forall x \in \mathbb{R}$
- (ii) $p(x) = 0$, tranne quando $x \in S$
- (iii) $\sum_{x \in \mathbb{R}} p(x) = 1$

4. Principali densità discrete

(4.1) **La densità binomiale** – Siano $n \in \mathbb{N}$ e $p \in (0, 1)$ due numeri fissati. Si chiama **densità binomiale** di parametri n e p (in simboli, $B(n, p)$) la funzione così definita

$$p(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k = 0, 1, 2, \dots, n \\ 0 & \text{altrimenti} \end{cases}$$

Vediamo adesso una tipica situazione in cui compare la densità binomiale.

Sia (Ω, A, P) lo schema delle prove indipendenti di parametri (noti) n e p (vedi 5.¹).

Ricordiamo che ciò significa che

$$\Omega = \{0, 1\}^n$$

ed inoltre, se $\omega = (\omega_1, \dots, \omega_n)$ è un evento elementare (con $\omega_i \in \{0, 1\}$ per ogni $i = 1, \dots, n$), si pone

$$P(\omega) = p^{(\sum_{i=1}^n \omega_i)} (1-p)^{(n - \sum_{i=1}^n \omega_i)}$$

Poniamo ora, per ogni $i = 1, \dots, n$

$$X_i(\omega) = \omega_i$$

$$X(\omega) = \sum_{i=1}^n X_i(\omega) = \sum_{i=1}^n \omega_i$$

Sia dunque k un intero fissato, con $k \in \{0, 1, 2, \dots, n\}$. L'evento $\{X = k\}$ è costituito dalle sequenze ω che contengono il simbolo “1” esattamente k volte, cioè tali che

$$\sum_{i=1}^n \omega_i = k$$

Pertanto, per calcolare la probabilità di $\{X = k\}$, basterà sommare le probabilità di tutte le sequenze del tipo sopra detto, cioè

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

E' importante il caso particolare di densità $B(1, p)$; la densità ottenuta si chiama anche **densità Bernoulliana**. Tale è la densità di una variabile aleatoria che assume i soli valori 1 e 0 con probabilità rispettive pari a p e $1-p$.

¹ – Riferimento al capitolo 1 (Spazi di probabilità)

(4.2) **La densità geometrica** – Sia p un numero assegnato, e supponiamo che sia $0 < p < 1$. Si chiama **densità geometrica** la funzione

$$p(k) = \begin{cases} (1-p)^{k-1} p & k = 1, 2, \dots \\ 0 & \text{altrove} \end{cases}$$

Una situazione in cui si presenta la densità geometrica è la seguente.

Supponiamo di eseguire una successione infinita di lanci di una stessa moneta, per la quale è p (assegnata) la probabilità di ottenere “1” in un singolo lancio. Ovviamente, al k -esimo lancio, avremo $k-1$ “0” (la cui probabilità è $1-p$) seguiti dal singolo simbolo “1” con $k = 1, 2, \dots$.

Da questo ricaviamo che

$$p(k) = (1-p)^{k-1} p \quad k = 1, 2, \dots$$

Nell’esperimento degli infiniti lanci di una moneta, potrebbe verificarsi l’evento B: “il simbolo “1” non esce mai”. Ma con quale probabilità avviene ciò?

Iniziamo a porre

A_k : “esce “1” al k -esimo lancio”

Si può scrivere ora

$$B^c = \bigcup_{k=1}^{\infty} A_k$$

poiché gli eventi A_k sono due a due disgiunti, per la proprietà di σ -additività di P si ha

$$P(B) = 1 - P\left(\bigcup_{k=1}^{\infty} A_k\right) = 1 - \sum_{k=1}^{\infty} P(A_k) = 1 - \sum_{k=1}^{\infty} (1-p)^{k-1} p$$

ricordando la somma della serie geometrica, abbiamo

$$1 - \sum_{k=1}^{\infty} (1-p)^{k-1} p = 1 - p \frac{1}{1-(1-p)} = 0$$

Cioè: per quanto piccola (purché strettamente positiva) sia la probabilità p di ottenere “1” in un singolo lancio, l’eventualità che “1” non esca mai è “quasi impossibile” (cioè di probabilità nulla).

(4.3) **La densità di Poisson** – Sia $\lambda > 0$ un numero reale assegnato.

La funzione

$$p(k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & k = 0, 1, 2, \dots \\ 0 & \text{altrimenti} \end{cases}$$

è una densità. Essa va sotto il nome di **densità di Poisson** di parametro λ .

Per ottenere in modo naturale questa densità, consideriamo la seguente situazione:

per ogni intero n , sia X_n una variabile aleatoria avente densità $B(n, p = \lambda/n)$, cioè

$$P(X_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}, \quad k = 0, 1, 2, \dots, n$$

Ci interessa trovare il limite di questa probabilità al tendere di n all'infinito (cioè, ci interessa trovare a quali numeri si avvicinano i valori della densità binomiale quando il numero n di prove ripetute aumenta e la probabilità λ/n di successo in una singola prova diminuisce).

Fissato ora k , sia n abbastanza grande in modo che sia $n \geq k$. Si ha

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} = \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{n!}{(n-k)! n^k} \end{aligned}$$

ricordando, dalla (6)¹, che $\frac{n!}{(n-k)!} = (n)_k = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$, abbiamo

$$P(X_n = k) = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{n!}{(n-k)! n^k} = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \underbrace{\frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n}}_{k \text{ fattori}}$$

Osservando che

- (i) $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$
- (ii) $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1$
- (iii) $\lim_{n \rightarrow \infty} \frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n} = 1$

se ne deduce che

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

cioè la densità di Poisson di parametro λ , calcolata in k .

¹ – Riferimento al capitolo 2 (Richiami di calcolo combinatorio)

(4.4) **La densità ipergeometrica** – Assegnati tre numeri interi strettamente positivi a , b e n , con $n \leq a+b$, poniamo $k_1 = \max(0, n-b)$ e $k_2 = \min(a, n)$ e osserviamo che è $k_1 \leq k_2$.

La funzione definita nel modo seguente

$$p(k) = \begin{cases} \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}} & k = k_1, k_1+1, \dots, k_2 \\ 0 & \text{altrove} \end{cases}$$

è una densità discreta. Essa prende il nome di **densità ipergeometrica** di parametri a , b e n .

Vediamo una situazione in cui questa densità si presenta in modo naturale.

Supponiamo di avere un'urna contenente a palline rosse e b palline verdi. Si eseguono n estrazioni senza rimpiazzo e poniamo X = numero di palline rosse (tra le n estratte).

Costruiamo uno spazio di probabilità (Ω, A, P) con cui modellizzare l'esperimento: ogni possibile risultato è un gruppo di n palline prese dalle $a+b$ palline presenti nell'urna; quindi l'insieme Ω degli eventi può essere identificato con l'insieme delle combinazioni di $a+b$ oggetti a n a n (vedi (10)¹). Inoltre, poiché non c'è motivo di pensare che un particolare risultato sia più o meno probabile di ogni altro, assegniamo a ciascun $\omega \in \Omega$ la stessa probabilità, pari a

$$P(\omega) = \frac{1}{\text{card}\Omega} = \frac{1}{\binom{a+b}{n}}$$

Premesso ciò, è immediato verificare che sullo spazio così costruito X è una variabile aleatoria, e ciò che resta da fare è calcolarne la densità, cioè le quantità $P(X = k)$ per ogni intero k .

E' chiaro che $P(X = k)$ sarà diversa da zero solo per gli interi k compresi tra k_1 e k_2 (infatti: il numero di palline rosse estratte non può superare né il numero di palline rosse a disposizione ($= a$), né il numero totale di estrazioni effettuate ($= n$). Inoltre, se $n \geq b$, $n-b$ è il minimo numero di palline rosse che possono essere estratte).

L'evento $\{X = k\}$ corrisponde al sottoinsieme di Ω costituito dai gruppi di n palline delle quali k sono rosse e $n-k$ verdi, e dunque ha cardinalità pari a

$$\binom{a}{k} \binom{b}{n-k}$$

Dato che stiamo lavorando su uno spazio uniforme si conclude che

$$P(X = k) = \frac{\text{card}\{X = k\}}{\text{card}\Omega} = \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}}$$

¹ – Riferimento al capitolo 2 (Richiami di calcolo combinatorio)

5. Vettori aleatori discreti

Può capitare che l'esperimento sotto osservazione produce più variabili interessanti, e che di esse si debba studiare il comportamento "congiunto": ad esempio, in uno schema di 5 prove indipendenti, consideriamo le due quantità X = numero di successi nelle 5 prove e Y = numero di successi nelle prime due prove; in tal caso una domanda interessante a cui rispondere potrebbe essere: quanto vale la probabilità che su un totale di 3 successi, due di essi si siano presentati nelle prime due prove? In simboli, questo significa calcolare la probabilità congiunta

$$P(\{X = 3\} \cap \{Y = 2\})$$

Supporremo fissato una volta per tutte lo spazio di probabilità (Ω, A, P) .

(5.1) Siano X_1, X_2, \dots, X_m m variabili aleatorie discrete (tutte definite su (Ω, A, P)). Si chiama **vettore aleatorie discreto** m -dimensionale (su (Ω, A, P)) la m -upla di variabili aleatorie

$$X = (X_1, X_2, \dots, X_m)$$

si tratta della funzione $X : \Omega \rightarrow \mathbb{R}^m$ definita da

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_m(\omega))$$

L'attributo *aleatorio* dato al vettore X è giustificato dal fatto seguente: per ogni $i = 1, 2, \dots, m$ sia $x_i \in \mathbb{R}$. Come sappiamo, il sottoinsieme di Ω $\{X_i = x_i\}$ è un evento, dunque, per la proprietà di stabilità della σ -algebra A rispetto all'intersezione finita, è un evento anche il sottoinsieme

$$\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \dots \cap \{X_m = x_m\}$$

che viene indicato più brevemente con la scrittura

$$\{X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\}$$

Ha quindi senso considerare la probabilità degli eventi del tipo precedente e si ha la seguente definizione

(5.2) Indichiamo con $x = (x_1, x_2, \dots, x_m)$ il generico punto di \mathbb{R}^m . Si chiama **densità** del vettore X (o **densità congiunta** delle variabili X_1, X_2, \dots, X_m) la funzione $p : \mathbb{R}^m \rightarrow \mathbb{R}$ definita da

$$p(x) = p(x_1, x_2, \dots, x_m) = P(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m)$$

Riprendiamo ora l'esempio citato all'inizio del paragrafo.

Siano $S_X = \{0, 1, 2, 3, 4, 5\}$ e $S_Y = \{0, 1, 2\}$ i rispettivi spettri delle due variabili aleatorie.

La seguente matrice, in cui ogni $p(i, j) = P(X = i, Y = j)$, è detta **matrice delle probabilità congiunte**.

X/Y	0	1	2
0	$p(0,0)$	$p(0,1)$	$p(0,2)$
1	$p(1,0)$	$p(1,1)$	$p(1,2)$
2	$p(2,0)$	$p(2,1)$	$p(2,2)$
3	$p(3,0)$	$p(3,1)$	$p(3,2)$
4	$p(4,0)$	$p(4,1)$	$p(4,2)$
5	$p(5,0)$	$p(5,1)$	$p(5,2)$

Le quantità del tipo

$$p(., j) = \sum_{i=0}^5 p(i, j) = P(Y = j), \quad p(i, .) = \sum_{j=0}^2 p(i, j) = P(X = i)$$

vengono chiamate anche **densità marginali**.

Da ciò ricaviamo che è possibile ottenere le densità marginali quando si conosce la densità congiunta (mentre, in generale, le densità marginali non individuano in modo univoco la congiunta). Inoltre,

$$\sum_{i=0}^5 \sum_{j=0}^2 p(i, j) = \sum_{i=0}^5 p(i, .) = \sum_{i=0}^5 P(X = i) = 1$$
$$\sum_{j=0}^2 \sum_{i=0}^5 p(i, j) = \sum_{j=0}^2 p(., j) = \sum_{j=0}^2 P(Y = j) = 1$$

cioè, la somma delle componenti della matrice è pari a 1.

6. Variabili aleatorie indipendenti

Supporremo al solito fissato lo spazio di probabilità (Ω, A, P) . Può accadere che due variabili aleatorie X e Y siano tali che il sapere X assume un valore in un certo insieme A non cambi la valutazione di probabilità che Y assuma un valore in un dato insieme B : in altre parole, tali che gli eventi $\{X \in A\}$ e $\{Y \in B\}$ siano indipendenti, e ciò per ogni coppia di sottoinsiemi A e B di R .

Due variabili aleatorie X e Y si dicono tra loro **indipendenti** se, per ogni coppia di sottoinsiemi A e B di R , risulta

$$(6.1) \quad P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Più in generale, le variabili aleatorie X_1, X_2, \dots, X_m si dicono tra loro **indipendenti** se, per ogni famiglia A_1, A_2, \dots, A_m di sottoinsiemi di R , risulta

$$(6.2) \quad P(X_1 \in A_1, X_2 \in A_2, \dots, X_m \in A_m) = \prod_{i=1}^m P(X_i \in A_i)$$

Siano X_1, X_2, \dots, X_m m variabili aleatorie discrete, aventi densità congiunta $p(x_1, x_2, \dots, x_m)$ e densità marginali $p_{X_1}(x_1), p_{X_2}(x_2), \dots, p_{X_m}(x_m)$. Allora X_1, X_2, \dots, X_m sono indipendenti tra loro se e solo se, $\forall (x_1, x_2, \dots, x_m) \in R^m$ vale la relazione

$$(6.3) \quad p_{X_1}(x_1), p_{X_2}(x_2), \dots, p_{X_m}(x_m) = p_{X_1}(x_1) \times p_{X_2}(x_2) \times \dots \times p_{X_m}(x_m)$$

Un'ultima definizione importante è la seguente:

Siano X e Y due variabili aleatorie discrete, aventi densità congiunta $p(x, y)$. Si chiama **densità condizionale** di Y , dato $X = x$, la funzione definita come segue

$$y \mapsto p_{Y|X}(y|x) = \begin{cases} \frac{P(Y=y, X=x)}{P(X=x)} = \frac{p(x, y)}{p_X(x)} & \text{se } p_X(x) \neq 0 \\ 0 & \text{altrimenti} \end{cases}$$

7. Riepilogo delle principali densità discrete

Nome	Densità	Legge	Situazione in cui si applica (breve descrizione)
Binomiale	$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$	$X \sim B(n, p)$	Esperimento di n prove ripetute (indipendenti); p rappresenta la probabilità di successo in una singola prova
Bernoulliana	$p(k) = p^k (1-p)^{1-k} \quad k = 0, 1$	$X \sim B(1, p)$	Caso particolare della binomiale in cui il numero di prove è pari a 1
Geometrica	$p(k) = (1-p)^{k-1} p \quad k = 1, 2, \dots$	$X \sim G(p)$	Esperimento di prove (indipendenti) ripetute fino ad ottenere il successo; p rappresenta la probabilità di successo in una singola prova
Poisson	$p(k) = \frac{a^k}{k!} e^{-a} \quad k = 0, 1, \dots, n$	$X \sim \text{Poisson}(a)$	Caso particolare della binomiale in cui: n , il numero di prove, viene fatto tendere a $+\infty$; p , la probabilità di successo di una singola prova, a 0; a , il parametro della densità, è pari a np
Ipergeometrica	$p(k) = \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}}$ $k = \max\{0, n-b\}, \dots, \min\{a, n\}$	$X \sim IG(a, b, n)$	Esperimento di n prove ripetute (non indipendenti);

1 Esempio 4.1.1 del libro di testo consigliato

L'esempio è esemplificativo della descrizione formale di un *esperimento aleatorio* e dei *numeri non deterministici*.

Si consideri l'esperimento \mathcal{E}_2 che consiste nel lanciare due dadi *onesti* (ossia, non sbilanciati) di colore differente in maniera tale che si può parlare di primo e secondo dado. Lo *spazio campione*, ovvero l'insieme dei possibili risultati di \mathcal{E}_2 , è:

$$\Omega_2 := \{(i, j) : i, j = 1, 2, \dots, 6\}.$$

Ne risulta che $|\Omega_2| = 6^2$, ovvero che in Ω_2 ci sono 36 coppie ordinate ognuna delle quali è un *punto campione*. D'altra parte, l'ipotesi dell'onestà dei due dadi porta a ritenere che

$$i, j = 1, 2, \dots, 6, \quad \mathbb{P}_2(\{(i, j)\}) = \frac{1}{6^2} = \frac{1}{36},$$

ovvero che in questo specifico contesto è possibile utilizzare la definizione di probabilità introdotta da Laplace (rapporto tra gli esiti favorevoli e gli esiti possibili). Allora, considerando *evento* ogni sottoinsieme di Ω_2 , si ha:

$$A \in \mathcal{P}(\Omega_2), \quad \mathbb{P}_2(A) = \frac{|A|}{36}. \quad (1)$$

In definitiva, la terna $(\Omega_2, \mathcal{P}(\Omega_2), \mathbb{P}_2)$ con \mathbb{P}_2 definita nella (1) descrive completamente e in maniera formale l'esperimento \mathcal{E}_2 .

Raramente, però, si è interessati al risultato (esito, punto campione) dell'esperimento ma, piuttosto, il valore rilevante è una sua funzione. In questo contesto, ad esempio, ha senso considerare la somma dei punteggi dei due dadi:

$$Z_2 : (i, j) \in \Omega_2 \rightarrow (i + j) \in \mathbb{R}.$$

La funzione Z_2 assume con probabilità positiva i valori: 2, 3, ..., 11, 12. Per ognuno di tali valori è possibile determinare agevolmente la sua *controimmagine* tramite la funzione Z_2 :

$$\begin{aligned} \{Z_2 = 2\} &= \{(1, 1)\}, & \{Z_2 = 12\} &= \{(6, 6)\}, \\ \{Z_2 = 3\} &= \{(1, 2), (2, 1)\}, & \{Z_2 = 11\} &= \{(5, 6), (6, 5)\}, \\ \{Z_2 = 4\} &= \{(1, 3), (2, 2), (3, 1)\}, & \{Z_2 = 10\} &= \{(4, 6), (5, 5), (6, 4)\}, \\ \{Z_2 = 5\} &= \{(1, 4), (2, 3), (3, 2), (4, 1)\}, & \{Z_2 = 9\} &= \{(3, 6), (4, 5), (5, 4), (6, 3)\}, \\ \{Z_2 = 6\} &= \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}, & \{Z_2 = 8\} &= \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}, \\ \{Z_2 = 7\} &= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}. \end{aligned} \quad (2)$$

Si osservi che la classe costituita dai sottoinsiemi individuati nella (2) costituisce una *partizione* di Ω_2 .

Dalle (1) e (2) se ne deduce che

$$\begin{aligned} \mathbb{P}_2(Z_2 = 2) &= \mathbb{P}_2(Z_2 = 12) = \frac{1}{36}, & \mathbb{P}_2(Z_2 = 3) &= \mathbb{P}_2(Z_2 = 11) = \frac{2}{36}, \\ \mathbb{P}_2(Z_2 = 4) &= \mathbb{P}_2(Z_2 = 10) = \frac{3}{36}, & \mathbb{P}_2(Z_2 = 5) &= \mathbb{P}_2(Z_2 = 9) = \frac{4}{36}, \\ \mathbb{P}_2(Z_2 = 6) &= \mathbb{P}_2(Z_2 = 8) = \frac{5}{36}, & \mathbb{P}_2(Z_2 = 7) &= \frac{6}{36}, \end{aligned} \quad (3)$$

e, inoltre, risulta

$$\sum_{s=2}^{12} \mathbb{P}_2(Z_2 = s) = \frac{2(1 + 2 + 3 + 4 + 5) + 6}{36} = 1.$$

2 Estensione al caso di tre dadi

Si consideri l'esperimento \mathcal{E}_3 che consiste nel lanciare tre dadi onesti di colore differente in maniera tale che si possa parlare di primo, secondo e terzo dado. Lo spazio campione, ovvero l'insieme dei possibili risultati di \mathcal{E}_3 , è:

$$\Omega_3 := \{(i, j, k) : i, j, k = 1, 2, \dots, 6\}.$$

Ne risulta che $|\Omega_3| = 6^3$, ovvero che in Ω_3 ci sono 216 coppie ordinate ognuna delle quali è un punto campione. D'altra parte, l'ipotesi dell'onestà dei tre dadi porta a ritenere che

$$i, j, k = 1, 2, \dots, 6, \quad \mathbb{P}_3(\{(i, j, k)\}) = \frac{1}{6^3} = \frac{1}{216},$$

ovvero che è possibile utilizzare la definizione di probabilità introdotta da Laplace. Allora, considerando evento ogni sottoinsieme di Ω_3 , si ha:

$$A \in \mathcal{P}(\Omega_3), \quad \mathbb{P}_3(A) = \frac{|A|}{216}. \quad (4)$$

In definitiva, la terna $(\Omega_3, \mathcal{P}(\Omega_3), \mathbb{P}_3)$ con \mathbb{P}_3 definita nella (4) descrive completamente e in maniera formale l'esperimento \mathcal{E}_3 .

In questo contesto, ad esempio, ha senso considerare la somma dei punteggi dei tre dadi:

$$Z_3 : (i, j, k) \in \Omega_3 \rightarrow (i + j + k) \in \mathbb{R}.$$

La funzione Z_3 assume con probabilità positiva i seguenti valori: 3, 4, ..., 17, 18. Per ognuno di tali valori è possibile determinare agevolmente la sua controimmagine tramite la funzione Z_3 . A tale scopo, per sintetizzare le formule, è utile individuare le classi di terne compatibili con la somma in questione e che sono diverse per almeno un punteggio con le rispettive dimensioni. Ad esempio, per $\{Z_3 = 9\}$ ci sono le seguenti 6 classi terne:

$$[(1, 2, 6); \textcolor{red}{6}], [(1, 3, 5); \textcolor{red}{6}], [(1, 4, 4); \textcolor{red}{3}], [(2, 2, 5); \textcolor{red}{3}], [(2, 3, 4); \textcolor{red}{6}], [(3, 3, 3); \textcolor{red}{1}].^1$$

Dopo di ciò, si ha:

$$\mathbb{P}_3(Z_3 = 9) = \frac{\textcolor{red}{6} + \textcolor{red}{6} + \textcolor{red}{3} + \textcolor{red}{3} + \textcolor{red}{6} + \textcolor{red}{1}}{6^3} = \frac{25}{216}, \quad (5)$$

Allo stesso modo, in $\{Z_3 = 10\}$ ci sono le seguenti 6 classi terne:

$$[(1, 3, 6); \textcolor{red}{6}], [(1, 4, 5); \textcolor{red}{6}], [(2, 2, 6); \textcolor{red}{3}], [(2, 3, 5); \textcolor{red}{6}], [(2, 4, 4); \textcolor{red}{3}], [(3, 3, 4); \textcolor{red}{3}],$$

e, pertanto, risulta:

$$\mathbb{P}_3(Z_3 = 10) = \frac{\textcolor{red}{6} + \textcolor{red}{6} + \textcolor{red}{3} + \textcolor{red}{6} + \textcolor{red}{3} + \textcolor{red}{3}}{6^3} = \frac{27}{216}. \quad (6)$$

Le formule (5) e (6) forniscono la soluzione al problema posto dai gentiluomini fiorentini a Galileo Galilei: in totale, ci sono 25 terne di punteggi che fanno realizzare il 9 e 27 terne di punteggi che fanno realizzare il 10.

Si lascia allo studente la verifica che

$$\sum_{s=3}^{18} \mathbb{P}_3(Z_3 = s) = 1.$$

¹La scrittura $[(1, 2, 6); \textcolor{red}{6}]$ indica le **sei** terne che si ottengono permutando i punteggi 1, 2 e 6 in tutti i modi possibili; analogamente, la scrittura $[(1, 4, 4); \textcolor{red}{3}]$ indica le **tre** terne che si ottengono permutando i punteggi 1, 4 e 4 in tutti i modi possibili.

Osservazione 1. Il numero aleatorio Z_2 può essere considerato anche nello spazio di probabilità $(\Omega_3, \mathcal{P}(\Omega_3), \mathbb{P}_3)$:

$$Z_2 : (i, j, k) \in \Omega_3 \rightarrow (i + j + 0 \cdot k) \in \mathbb{R}.$$

Ovviamente, questo comporta un aggravio nelle formule in quanto, ad esempio,

$$\{Z_2 = 2\} = \{(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 1, 4), (1, 1, 5), (1, 1, 6)\} \implies |Z_2 = 2| = 6.$$

Questo fatto non altera i risultati individuati nel paragrafo 1; infatti, relativamente all'evento $\{Z_2 = 2\}$, tenendo presente la prima delle formule (3) si ha:

$$\mathbb{P}_3(Z_2 = 2) = \frac{6}{216} = \frac{1}{36} = \mathbb{P}_2(Z_2 = 2).$$

La statistica inferenziale

1. Introduzione ai problemi statistici

Ogni rilevamento statistico produce un campione di dati (relativamente piccolo).

La statistica descrittiva si occupa di organizzare e riassumere in modo significativo questi dati.

La statistica inferenziale, invece, utilizzando metodi e nozioni del calcolo della probabilità, cerca di fare previsioni sul futuro, o di ottenere risultati estendibili all'intera popolazione (a partire solo dal piccolo campione osservato).

Nella pratica, tipicamente si deve effettuare un esperimento che produce una variabile aleatoria X di cui non si conosce la legge, e si vogliono ricavare informazioni su di essa.

2. Il concetto di stimatore

Uno dei problemi che si presentano più frequentemente allo statistico che voglia ottenere informazioni su una data variabile aleatoria X , è quello di deciderne la legge: ad esempio, può succedere che egli sappia che tale legge appartiene ad una data famiglia, dipendente da un parametro θ non noto. Nel caso generale il parametro da studiare è indicato con θ . In casi specifici il simbolo usato potrà essere diverso.

La cosa più naturale, che uno sperimentatore può fare, è quella di “procurarsi delle osservazioni” del fenomeno di cui X è l'espressione, effettuando qualche tipo di esperimento, e decidere (in statistica, “fare inferenza”) in base ai risultati ottenuti.

Tipicamente come risultato del suo esperimento egli otterrà dei numeri x_1, x_2, \dots, x_n . Essi vanno pensati come valori assunti (nel corso dell'esperimento) da certe variabili aleatorie X_1, X_2, \dots, X_n aventi una legge congiunta dipendente da θ . Le chiameremo **osservazioni** di X . In corrispondenza i numeri x_1, x_2, \dots, x_n (che sono i valori assunti dalle osservazioni dopo che l'esperimento è stato effettuato) si chiameranno **valori osservati**.

Un caso molto frequente è quello in cui X_1, X_2, \dots, X_n sono tra loro indipendenti ed hanno tutte la stessa legge di X : è la formalizzazione matematica del caso in cui lo sperimentatore decide di ripetere n volte, in condizioni di indipendenza, proprio l'esperimento che produce X . Si dice allora che X_1, X_2, \dots, X_n costituiscono un **campione di taglia n estratto** dalla legge di X .

Comunque sia, lo sperimentatore userà i numeri trovati per calcolare, a partire da essi, una stima del parametro incognito; in altri termini sceglierà un'opportuna funzione t di n variabili reali e stimerà il parametro θ con il numero $t(x_1, x_2, \dots, x_n)$. Ovviamente la funzione t andrà scelta non dipendente dal parametro incognito (dato che essa va usata appunto per stimarlo).

(2.1) Sia $t: \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione non dipendente da θ .

- (i) Si chiama **stimatore** di θ la variabile aleatoria $T = t(X_1, X_2, \dots, X_n)$, dove X_1, X_2, \dots, X_n sono n osservazioni di X
- (ii) Si chiama **stima** di θ il numero $t = t(x_1, x_2, \dots, x_n)$, dove x_1, x_2, \dots, x_n sono gli n valori osservati (corrispondenti alle osservazioni del punto (i))

(2.2) Siano X_1, X_2, \dots, X_n n osservazioni aventi tutte la stessa legge, dipendente da un parametro θ . Sia $T = t(X_1, X_2, \dots, X_n)$ uno stimatore di una funzione $\psi(\theta)$ del parametro. La situazione “ideale” sarebbe che valesse l’uguaglianza

$$(2.2.1) \quad T(\omega) = \psi(\theta) \quad \forall \omega \in \Omega$$

cioè che lo stimatore fornisca sempre e esattamente la quantità da stimare. Ciò non è ovviamente possibile; più ragionevole è chiedersi se l’uguaglianza (2.2.1) possa valere almeno in media; in effetti una buona proprietà di uno stimatore è la seguente

(2.3) Lo stimatore $T = t(X_1, X_2, \dots, X_n)$ di $\psi(\theta)$ si dice **corretto** (o **non distorto**) se vale la relazione

$$\mathbf{E}^\theta[T] = \psi(\theta) \quad \forall \theta \in \Theta$$

Nota: In tutti gli esempi che seguono, indicheremo con μ e σ^2 rispettivamente la media e la varianza della comune legge delle X_i . Nelle scritture del tipo \mathbf{E}^θ o \mathbf{Var}^θ o simili sottintenderemo il parametro θ , (cioè scriveremo semplicemente \mathbf{E} o \mathbf{Var}).

(2.3.1) La media campionaria \bar{X} è uno stimatore corretto di μ . Infatti

$$\mathbf{E}[\bar{X}] = \mathbf{E}\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{1}{n} \mathbf{E}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu$$

(2.3.2) Se μ è nota, lo stimatore

$$T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

è uno stimatore corretto di σ^2 . Infatti

$$\begin{aligned} \mathbf{E}\left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{n}\right] &= \frac{1}{n} \mathbf{E}\left[\sum_{i=1}^n (X_i - \mu)^2\right] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[(X_i - \mu)^2] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{Var} X_i = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \frac{1}{n} n\sigma^2 = \sigma^2 \end{aligned}$$

(2.3.3) Supponiamo ora in più che le osservazioni X_1, X_2, \dots, X_n siano tra loro indipendenti. Vogliamo trovare uno stimatore corretto di σ^2 nel caso, molto frequente, che μ non sia nota. L'idea è quella di sostituire μ con il suo stimatore \bar{X} nella formula della (2.2.3), cioè di usare lo stimatore

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

Calcoliamo dunque $E[Z]$, cominciando dal numeratore della frazione.

$$\begin{aligned} (*) \quad E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] &= E\left[\sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + \sum_{i=1}^n \bar{X}^2\right] = E\left[\sum_{i=1}^n X_i^2 - 2n\bar{X} \frac{\sum_{i=1}^n X_i}{n} + n\bar{X}^2\right] \\ &= E\left[\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right] = E\left[\sum_{i=1}^n X_i^2 - n\bar{X}^2\right] = \sum_{i=1}^n E[X_i^2] - nE[\bar{X}^2] \end{aligned}$$

D'altra parte

$$(**) \quad E[X_i^2] = \text{Var}X_i + E[X]^2 = \sigma^2 + \mu^2$$

$$\begin{aligned} (***) \quad E[\bar{X}^2] &= \text{Var}\bar{X} + E[\bar{X}]^2 = \text{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) + \mu^2 = \frac{1}{n^2} \sum_{i=1}^n \text{Var}X_i + \mu^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 + \mu^2 \\ &= \frac{1}{n^2} n\sigma^2 + \mu^2 = \frac{1}{n} \sigma^2 + \mu^2 \end{aligned}$$

Usando le relazioni (**) e (***) nella (*) si ottiene

$$E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = \sum_{i=1}^n (\sigma^2 + \mu^2) - n\left(\frac{1}{n} \sigma^2 + \mu^2\right) = (n-1) \sigma^2$$

Dunque

$$E[Z] = \frac{n-1}{n} \sigma^2$$

che è uno stimatore distorto. Tuttavia per $n \rightarrow \infty$, $E[Z]$ converge a σ^2 (in questo caso diremo che lo stimatore è *asintoticamente corretto* e la quantità $\delta = \sigma^2 - E[Z] = \frac{\sigma^2}{n}$ è detta *distorsione*).

Il calcolo appena fatto per $E[Z]$ suggerisce che lo stimatore corretto di σ^2 è

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

(2.4) Sostituendo il vero valore di $\psi(\theta)$ con il suo stimatore $T = t(X_1, X_2, \dots, X_n)$ si commette un errore, che è opportuno misurare. Si chiama **rischio quadratico** dello stimatore T la funzione $\theta \mapsto R_T(\theta)$ definita su Θ da

$$R_T(\theta) = \mathbf{E}^\theta[(T - \psi(\theta))^2]$$

(2.4.1) Se lo stimatore è corretto (cioè quando $\mathbf{E}^\theta[T] = \psi(\theta)$), si ha evidentemente

$$R_T(\theta) = \mathbf{E}^\theta[(T - \mathbf{E}^\theta[T])^2] = \mathbf{Var}^\theta T$$

(2.4.2) In presenza di due stimatore S e T della quantità $\psi(\theta)$, preferiremo ovviamente lo stimatore con rischio più piccolo, per ogni valore di θ . Si dice che S è **preferibile** a T se $R_S(\theta) \leq R_T(\theta) \forall \theta \in \Theta$; se in più $\exists \theta_0 \in \Theta$ tale che $R_S(\theta_0) < R_T(\theta_0)$, allora si dice che S è **strettamente preferibile** a T . Uno stimatore è detto **ammissibile** se non esistono stimatori ad esso strettamente preferibili.

(2.5) Un'altra proprietà asintotica che può essere importante per uno stimatore $T_n = t_n(X_1, X_2, \dots, X_n)$ è la “consistenza” (o anche “coerenza”).

(2.5.1) Una successione (T_n) di stimatori di $\psi(\theta)$ si dice **fortemente consistente** se T_n converge quasi certamente verso $\psi(\theta)$.

(2.5.2) Una successione (T_n) di stimatori di $\psi(\theta)$ si dice **debolmente consistente** se T_n converge verso $\psi(\theta)$ in probabilità (cioè, $\lim_{n \rightarrow \infty} P(|T_n - \psi(\theta)| < \varepsilon) = 1, \forall \varepsilon > 0$).

(2.6) Sia (T_n) una successione di stimatori fortemente consistente del parametro θ , e supponiamo che $\theta \mapsto \psi(\theta)$ sia una funzione continua. Allora $U_n \mapsto \psi(T_n)$ è una successione di stimatori fortemente consistente di $\psi(\theta)$.

(2.7) Sia X una variabile aleatoria la cui legge dipende da un certo numero di parametri $\theta_1, \theta_2, \dots, \theta_r$ (il parametro θ è in generale un vettore $\theta = (\theta_1, \theta_2, \dots, \theta_r)$). Supponiamo che $\theta_1, \theta_2, \dots, \theta_r$ siano non noti, e come al solito il nostro scopo è darne una stima dipendente dalle osservazioni (X_1, X_2, \dots, X_n) . Sia k un numero intero ≥ 1 fissato. Si chiama **momento teorico di ordine k** di X il numero

$$m_k = E[X^k]$$

(a patto che esso sia finito, e cioè che $E[|X|^k] < \infty$).

(2.7.1) Dato che la legge di X dipende da $\theta_1, \theta_2, \dots, \theta_r$, lo stesso accadrà per il momento teorico m_k ; in altre parole esisterà una funzione f_k di r variabili tale che

$$m_k = f_k(\theta_1, \theta_2, \dots, \theta_r)$$

Supponiamo che la variabile aleatoria X ammetta i primi q momenti (q è un numero intero ≥ 1). Ciò significa che possiamo scrivere la relazione precedente $\forall k = 1, \dots, q$, ottenendo così il sistema

$$\begin{cases} m_1 = f_1(\theta_1, \theta_2, \dots, \theta_r) \\ m_2 = f_2(\theta_1, \theta_2, \dots, \theta_r) \\ \vdots \\ m_q = f_q(\theta_1, \theta_2, \dots, \theta_r) \end{cases}$$

Si tratta evidentemente di un sistema di q equazioni nelle r incognite $\theta_1, \theta_2, \dots, \theta_r$ che si può cercare di risolvere. Se questo è possibile, otterremo r espressioni del tipo seguente

$$(2.7.2) \quad \begin{cases} \theta_1 = g_1(m_1, m_2, \dots, m_q) \\ \theta_2 = g_2(m_1, m_2, \dots, m_q) \\ \vdots \\ \theta_r = g_r(m_1, m_2, \dots, m_q) \end{cases}$$

(2.7.3) Supponiamo che le osservazioni (X_1, X_2, \dots, X_n) siano tra loro indipendenti. Si definisce **momento empirico** di X la quantità aleatoria

$$\hat{m}_k = \frac{\sum_{i=1}^n X_i^k}{n}$$

Per la legge dei grandi numeri, si ha

$$(2.7.4) \quad \hat{m}_k \rightarrow_{n \rightarrow \infty} m_k$$

Questa relazione spiega i nomi di momento teorico e momento empirico dati alle due quantità sopra definite; ci dice anche che i momenti empirici sono stimatori consistenti di quelli teorici.

La relazione (2.7.4) suggerisce il procedimento seguente: dato che per n grande $\hat{m}_k \cong m_k$, nel sistema (2.7.2) sostituiamo \hat{m}_k al posto di m_k per ogni $k = 1, \dots, q$, ottenendo le relazioni

$$\begin{cases} \theta_1 \cong g_1(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q) \\ \theta_2 \cong g_2(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q) \\ \vdots \\ \theta_r \cong g_r(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q) \end{cases}$$

ovvero avremo espresso (approssimativamente) ciascuno dei parametri in termini delle variabili aleatorie $\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q$ che sono note, perché dipendenti solo dalle osservazioni.

Dunque, a sua volta, ciascuna delle funzioni $g_i(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q)$, per ogni $i = 1, \dots, r$, dipende solo dalle osservazioni, ed è dunque uno stimatore $\theta_i = \theta_i(X_1, X_2, \dots, X_n)$ di θ_i .

(2.7.5) Il vettore di variabili aleatorie $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)$ si chiama **stimatore dei momenti** del parametro $\theta = (\theta_1, \theta_2, \dots, \theta_q)$.

(2.7.6) Chiameremo **stime dei momenti** i numeri che si ottengono mettendo i valori osservati, cioè x_1, x_2, \dots, x_n al posto delle osservazioni X_1, X_2, \dots, X_n nelle espressioni, sopra definite, $\theta_i(X_1, X_2, \dots, X_n)$.

Gli stimatori dei momenti sono spesso distorti, ma consistenti. Tuttavia essi non sono dei buoni stimatori. Stimatori migliori si ottengono con il **metodo della massima verosimiglianza**.

(2.8) Sia X una variabile aleatoria la cui legge dipende da un parametro θ . Disponiamo di n osservazioni di X (non necessariamente indipendenti) che indichiamo con (X_1, X_2, \dots, X_n) . Per semplicità, in questo momento supporremo che la legge congiunta delle osservazioni sia discreta, e indicheremo con P^θ la loro densità congiunta. Dopo aver effettuato l'esperimento, il campione di osservazioni avrà prodotto un campione di n valori osservati. La probabilità (in funzione di $\theta \in \Theta$) che il risultato sia quello effettivamente osservato è

$$\theta \mapsto P^\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n), \quad \theta \in \Theta$$

Essa è evidentemente una funzione di θ dipendente dai valori osservati x_1, x_2, \dots, x_n . Nel caso generale essa si chiama **funzione di verosimiglianza** e il simbolo usato sarà

$$\theta \mapsto L(\theta | x_1, x_2, \dots, x_n), \quad \theta \in \Theta$$

Supponiamo di essere riusciti a calcolare (in qualche modo) il massimo di questa funzione (al variare di $\theta \in \Theta$); il corrispondente punto di massimo dipenderà anch'esso dai parametri x_1, x_2, \dots, x_n ; indichiamo con $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$. Questo numero si chiama **stima di massima verosimiglianza** di θ . La variabile aleatoria $\hat{\theta}(X_1, X_2, \dots, X_n)$ si chiama **stimatore di massima verosimiglianza**.

(2.8.1) Diamo ora qualche dettaglio di alcuni metodi per il calcolo della funzione di verosimiglianza. Una situazione semplice è quella in cui il vettore delle osservazioni ha densità (congiunta) discreta. Un caso particolare di questa situazione si ha quando le osservazioni costituiscono un campione, cioè sono tra loro indipendenti. Supponiamo cioè che la variabile aleatoria X abbia densità discreta $p_\theta(x) = P^\theta(X = x)$. Allora

$$P^\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = p_\theta(x_1) \cdot p_\theta(x_2) \cdot \dots \cdot p_\theta(x_n) = L(\theta | x_1, x_2, \dots, x_n)$$

Per analogia, se X è assolutamente continua, con densità $f_\theta(x)$ (e naturalmente le osservazioni sono ancora tra loro indipendenti), la funzione di verosimiglianza è definita dalla formula

$$L(\theta | x_1, x_2, \dots, x_n) = f_\theta(x_1) \cdot f_\theta(x_2) \cdot \dots \cdot f_\theta(x_n)$$

Per concludere, diamo un rapido elenco di alcuni metodi usati per trovare il punto di massimo di L .

- 1) Supponiamo che l'insieme Θ sia costituito da un numero finito di elementi $\theta_1, \theta_2, \dots, \theta_M$. Se M non è troppo grande, un modo del tutto elementare consiste nel calcolare $L(\theta_i) \forall i = 1, \dots, M$ e scegliere il valore θ_i per cui $L(\theta_i)$ risulta massimo. Se M è grande, bisognerà ricorrere a qualche trucco, da vedere caso per caso.
- 2) Se Θ è un intervallo della retta (eventualmente non limitato), si possono applicare i metodi studiati in Analisi I per trovare i punti di minimo e di massimo di una funzione: in particolare, la stima di massima verosimiglianza, se interna all'intervallo Θ , è soluzione dell'equazione

$$\frac{d}{d\theta} L(\theta | x_1, x_2, \dots, x_n) = 0$$

Si ricordi che per gli (eventuali) estremi dell'intervallo, non si può applicare il criterio dei punti stazionari.

- 3) Nel caso che la stima di massima verosimiglianza sia stata ottenuta con il solo criterio dei punti stazionari, necessario ricordare che il fatto che un punto sia stazionario non garantisce automaticamente che esso sia di massimo. Di regola, bisognerebbe proseguire l'indagine studiando la derivata seconda di L .
- 4) In molti casi, prima di effettuare la derivazione di L , conviene passare al logaritmo, cioè derivare non la funzione $\theta \mapsto L(\theta | x_1, x_2, \dots, x_n)$, ma $\theta \mapsto \log L(\theta | x_1, x_2, \dots, x_n)$. Il passaggio al logaritmo non cambia i punti stazionari, né la loro natura. In questi casi, dunque, la stima di massima verosimiglianza viene trovata come soluzione dell'equazione

$$\frac{d}{d\theta} \log L(\theta | x_1, x_2, \dots, x_n) = 0$$

che è nota come *equazione di verosimiglianza*.

(2.8.2) Esempio – Calcolare la stima e lo stimatore di massima verosimiglianza del parametro p della legge bernoulliana $B(1, p)$, basato sul campione (X_1, X_2, \dots, X_n) . Si suppone che $p \in (0, 1)$. Sia X una variabile aleatoria di legge $B(1, p)$; la sua densità è

$$P^p(X = x) = p^x (1 - p)^{1-x}, \quad x = 0, 1$$

Dunque, la funzione di verosimiglianza è

$$p \mapsto L(p \mid x_1, x_2, \dots, x_n) = p^{x_1} (1 - p)^{1-x_1} \cdot p^{x_2} (1 - p)^{1-x_2} \cdot \dots \cdot p^{x_n} (1 - p)^{1-x_n}$$

per $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$. Per semplicità poniamo

$$\alpha = x_1 + x_2 + \dots + x_n$$

L'equazione di verosimiglianza è

$$0 = \frac{d}{dp} \log L(p \mid x_1, x_2, \dots, x_n) = \frac{d}{dp} \log(p^\alpha (1 - p)^{n-\alpha}) = \frac{d}{dp} (\alpha \log p + (n - \alpha) \log(1 - p)) = \frac{\alpha}{p} - \frac{n - \alpha}{1 - p}$$

la cui unica soluzione è

$$\hat{p} = \frac{\alpha}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} = \bar{x}$$

(media del campione di valori osservati). Lo stimatore di massima verosimiglianza è dunque

$$\hat{p} = \frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}$$

ovvero la media campionaria delle osservazioni.

3. Intervalli di fiducia

Ora ogni funzione delle osservazioni, non dipendente dal parametro incognito θ , sarà detta **statistica**. Dunque il concetto di statistica non è diverso da quello di **stimatore**, che abbiamo dato in precedenza. Quello che cambia è solo il nome: ora non si tratta di dare un valore approssimato al parametro, ma di utilizzare la nostra funzione delle osservazioni per confronti più specifici.

(3.1) Sia $\alpha \in (0, 1)$ un numero fissato. Date due statistiche $T_1 = t_1(X_1, X_2, \dots, X_n)$ e $T_2 = t_2(X_1, X_2, \dots, X_n)$, si dice che $I_X = [T_1, T_2]$ è un **intervallo di fiducia per $\psi(\theta)$ di livello $1 - \alpha$** se, $\forall \theta \in \Theta$ si ha

$$P^\theta(\psi(\theta) \in I_X) \geq 1 - \alpha$$

(3.1.1) Tipicamente il valore di α è piccolo ($\alpha = 0.05; \alpha = 0.01 \dots$).

(3.1.2) Il significato della definizione (3.1) è il seguente: in base alle osservazioni che abbiamo, possiamo dire che $T_1 \leq \psi(\theta) \leq T_2$ con probabilità almeno $1 - \alpha$, e questo $\forall \theta$.

(3.1.3) Costruiamo un intervallo di fiducia di livello 0.95 per il parametro λ dell'esponenziale, basato su una sola osservazione X . Significa che si devono trovare due funzioni $t_1(X)$ e $t_2(X)$ tali che

$$P^\lambda(t_1(X) \leq \lambda \leq t_2(X)) \geq 0.95$$

Partiamo da questa semplice osservazione: se $X \sim \text{Esp}(\lambda)$ (con media pari a $1/\lambda$), allora la variabile $Y = \lambda X \sim \text{Esp}(1)$. Infatti, per $t > 0$ si ha

$$P(Y \leq t) = P(\lambda X \leq t) = P(X \leq t/\lambda) = F_X(t/\lambda) = 1 - e^{-t}$$

Di conseguenza, $\forall a, b > 0$ con $a < b$ si ottiene

$$P^\lambda(a \leq Y \leq b) = (1 - e^{-b}) - (1 - e^{-a}) = e^{-a} - e^{-b}$$

il che equivale a

$$P^\lambda(a \leq \lambda X \leq b) = P^\lambda\left(\frac{a}{X} \leq \lambda \leq \frac{b}{X}\right) = e^{-a} - e^{-b}$$

Allora poniamo

$$t_1(X) = \frac{a}{X}, \quad t_2(X) = \frac{b}{X}$$

dove le costanti a e b sono scelte in modo che $e^{-a} - e^{-b} = 0.95$.

Il metodo qui seguito è quello della **quantità pivotale**.

Nota: D'ora in avanti, gli intervalli che costruiremo saranno tutti di livello $1 - \alpha$ fissato. Inoltre, ai punti (3.2), (3.3), (3.4) e (3.5) (X_1, X_2, \dots, X_n) sarà un campione di taglia n estratto dalla legge $N(\mu, \sigma^2)$.

(3.2) **Intervalli di fiducia per la media della normale con varianza nota** – Si parte dall'osservazione che

$$Y = \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N(0, 1)$$

(a) **Intervallo bilaterale** – Sia

$$(*) \quad P^\mu(a \leq Y \leq b) = \Phi(b) - \Phi(a)$$

Pertanto ora basterà trovare a e b in modo che $\Phi(b) - \Phi(a) = 1 - \alpha$.

Siano ora β e γ due numeri reali $\in [0, 1]$ tali che $b = \phi_\beta$ e $a = \phi_\gamma$ (si pensi a ϕ come la funzione inversa di Φ); allora si ha

$$\Phi(b) = \beta, \quad \Phi(a) = \gamma$$

Pertanto basterà scegliere β e γ in modo tale che $\beta - \gamma = 1 - \alpha$. Una scelta possibile è $\beta = 1 - \alpha/2$, $\gamma = \alpha/2$. Dalla (*) si ricava

$$\begin{aligned} \Phi(b) - \Phi(a) &= P^\mu(a \leq Y \leq b) = P^\mu\left(a \leq \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq b\right) = P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} b \leq \mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}} a\right) \\ &= P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_\beta \leq \mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}} \phi_\gamma\right) = P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \leq \mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{\alpha/2}\right) = \\ &P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2}\right) \end{aligned}$$

L'intervallo è dunque

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \right)$$

(b) **Intervallo unilaterale destro** – Il termine significa che si vuole trovare una limitazione per μ solo dal basso. Questa volta partiamo dalla relazione

$$(**) \quad P^\mu(Y \leq b) = \Phi(b)$$

Sia ora β come al punto (a). Pertanto basterà trovare b in modo che $\Phi(b) = 1 - \alpha = \beta$. Ovvero, semplicemente, $b = \phi_{1-\alpha}$. Dalla (**) si ricava che

$$\begin{aligned} \Phi(b) &= P^\mu(Y \leq b) = P^\mu\left(\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq b\right) = P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} b \leq \mu\right) = P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_\beta \leq \mu\right) \\ &= P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha} \leq \mu\right) \end{aligned}$$

L'intervallo è dunque

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha}, +\infty \right)$$

(c) **Intervallo unilaterale sinistro** – Analogamente al caso dell'intervallo unilaterale destro troviamo che l'intervallo è

$$\left(-\infty, \bar{X} - \frac{\sigma}{\sqrt{n}}\phi_{\alpha}\right) = \left(-\infty, \bar{X} + \frac{\sigma}{\sqrt{n}}\phi_{1-\alpha}\right)$$

ricordando che $-\phi_{\alpha} = \phi_{1-\alpha}$.

(3.3) Intervalli di fiducia per la media della normale con varianza non nota – Nella pratica gli intervalli di (3.2) sono di scarsa utilità, perché nelle formule che li definiscono interviene la varianza σ^2 , che in genere non si conosce. In questo caso si può sostituire σ^2 con

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

(che ne è uno stimatore corretto) e applicare di nuovo il metodo della quantità pivotale partendo dalla variabile aleatoria

$$Z = \sqrt{n} \frac{\bar{X} - \mu}{S} \sim t(n-1)$$

Osservando che l'unica proprietà della legge normale standard che abbiamo usato è stata la simmetria e ricordando che anche la legge t di *Student* è simmetrica, risulta chiaro che tutto ciò che abbiamo detto in (3.2) si può ripetere, semplicemente sostituendo σ con S e i quantili della normale standard con quelli della t di *Student*. Pertanto avremo

(a) **Intervallo bilaterale**

$$\left(\bar{X} - \frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1), \bar{X} + \frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1)\right)$$

(b) **Intervallo unilaterale destro**

$$\left(\bar{X} - \frac{S}{\sqrt{n}}t_{1-\alpha}(n-1), +\infty\right)$$

(c) **Intervallo unilaterale sinistro**

$$\left(-\infty, \bar{X} + \frac{S}{\sqrt{n}}t_{1-\alpha}(n-1)\right)$$

(3.4) **Intervalli di fiducia per la varianza della normale con media nota** – Qui si parte osservando che la variabile aleatoria

$$Y = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$$

ha legge $\chi^2(n)$. Posto

$$U^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

(che è uno stimatore corretto della varianza) si può scrivere

$$Y = \frac{nU^2}{\sigma^2}$$

(a) **Intervallo bilaterale** – Indichiamo con F_n la funzione di ripartizione della $\chi^2(n)$. Allora si ha

$$F_n(b) - F_n(a) = P^{\sigma^2}(a \leq Y \leq b) = P^{\sigma^2}\left(a \leq \frac{nU^2}{\sigma^2} \leq b\right) = P^{\sigma^2}\left(\frac{nU^2}{b} \leq \sigma^2 \leq \frac{nU^2}{a}\right)$$

se, al solito come visto in (3.2), poniamo $b = \chi^2_{\beta}(n)$, $a = \chi^2_{\gamma}(n)$, avremo

$$1 - \alpha = F_n(b) - F_n(a) = \beta - \gamma$$

Una scelta possibile (come in 3.2) è $\beta = 1 - \alpha/2$, $\gamma = \alpha/2$ e si ottiene l'intervallo

$$\left(\frac{nU^2}{\chi^2_{1-\alpha/2}(n)}, \frac{nU^2}{\chi^2_{\alpha/2}(n)} \right)$$

(b) **Intervallo unilaterale destro** – Esattamente come per la (3.2) abbiamo che l'intervallo è

$$\left(\frac{nU^2}{\chi^2_{1-\alpha}(n)}, +\infty \right)$$

(c) **Intervallo unilaterale sinistro** – Analogamente abbiamo che l'intervallo è

$$\left(-\infty, \frac{nU^2}{\chi^2_{\alpha}(n)} \right)$$

(3.5) **Intervalli di fiducia per la varianza della normale con media non nota** – Poiché la media non è nota, useremo, al posto di μ , il suo stimatore (corretto) \bar{X} . Prendiamo in considerazione ora la quantità

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

(che, come è noto, è uno stimatore corretto della varianza). Quindi, applichiamo il metodo della quantità pivotale a partire dalla variabile aleatoria

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

che, dal teorema di Cochran, risulta avere legge $\chi^2(n-1)$ (quindi i ragionamenti fatti al punto (3.4) valgono anche qui). Dunque, per ricavare i tre nuovi intervalli basterà sostituire, nelle formule di (3.4), $n-1$ al posto di n (e, naturalmente, i quantili della $\chi^2(n-1)$ al posto di quelli della $\chi^2(n)$). Si ottiene cioè

(a) **Intervallo bilaterale**

$$\left(\frac{(n-1)U^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)U^2}{\chi^2_{\alpha/2}(n-1)} \right)$$

(b) **Intervallo unilaterale destro**

$$\left(\frac{(n-1)U^2}{\chi^2_{1-\alpha}(n-1)}, +\infty \right)$$

(c) **Intervallo unilaterale sinistro**

$$\left(-\infty, \frac{(n-1)U^2}{\chi^2_{\alpha}(n-1)} \right)$$

(3.6) **Intervalli di fiducia per il parametro della bernoulliana** – Sia (X_1, X_2, \dots, X_n) un campione dalla legge $B(1, p)$. Osservando che per n abbastanza grande, la variabile aleatoria

$$Y = \frac{X - p}{\sqrt{p(1-p)}} \sqrt{n}$$

(dove p è la media e $p(1-p)$ è la varianza della variabile aleatoria bernoulliana) è approssimativamente di legge $N(0,1)$ per il Teorema Limite Centrale, possiamo pensare di usufruire di quanto detto a proposito di campioni gaussiani. Per la stima bilaterale, avremo l'intervallo

$$\left(\bar{X} - \frac{\sqrt{p(1-p)}}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sqrt{p(1-p)}}{\sqrt{n}} \phi_{1-\alpha/2} \right)$$

Purtroppo, l'intervallo risultante dipende dal parametro p che è appunto l'incognita da stimare, e dunque non è un buon intervallo di fiducia. L'idea è allora quella di sostituire p con il suo stimatore \bar{X} . E si ha

(a) **Intervallo bilaterale**

$$\left(\bar{X} - \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha/2} \right)$$

(b) **Intervallo unilaterale destro**

$$\left(\bar{X} - \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha}, +\infty \right)$$

(c) **Intervallo unilaterale sinistro**

$$\left(-\infty, \bar{X} - \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha} \right)$$

Momenti di una variabile aleatoria

Consideriamo X v.a. e k intero positivo tali che X^k sia una v.a. con speranza matematica finita. Diciamo

- **momento di ordine k di X** la quantità:

$$\boxed{\mu'_k = E[X^k]};$$

- **momento centrato di ordine k di X** la quantità:

$$\boxed{\mu_k = E[(X - E[X])^k]}.$$

Indicata con $p(x_i)$ oppure $f(x)$ le densità di X (a seconda che sia discreta o continua), si ha:

$$\mu'_k = E[X^k] = \begin{cases} \sum_{i=0}^{\infty} x_i^k p(x_i) \\ \int_{-\infty}^{+\infty} x^k f(x) dx \end{cases} ;$$

$$\mu_k = E[(X - E[X])^k] = \begin{cases} \sum_{i=0}^{\infty} (x_i - E[X])^k p(x_i) \\ \int_{-\infty}^{+\infty} (x - E[X])^k f(x) dx \end{cases}$$

E' immediato verificare che

- $\mu'_1 = E[X]$,
- $\mu_1 = 0$ per qualsiasi v.a. X ,
- $\mu_2 = \sigma^2 = Var(X)$.

Funzione generatrice dei momenti

Definizione.

Consideriamo X v.a. tale che e^{tX} sia una v.a. con speranza matematica finita, per ogni scelta di $t \in \mathbf{R}$.

Diciamo

funzione generatrice dei momenti di X :

$$\begin{aligned}\phi : \mathbf{R} &\longmapsto \mathbf{R} \\ t &\longmapsto \boxed{\phi(t) = E[e^{tX}]}\end{aligned}$$

(attenzione la v.a. X è fissata dunque la funzione dipende dalla variabile numerica $t \in \mathbf{R}$)

Se indichiamo con $p(x_i)$ o $f(x)$ la densità di X (a seconda che sia discreta o continua), si ottiene

$$\phi(t) = E[e^{tX}] = \begin{cases} \boxed{\sum_{i=1}^{\infty} e^{tx_i} p(x_i)} \\ \boxed{\int_{-\infty}^{+\infty} e^{tx} f(x) dx} \end{cases}$$

Proprietà.

Detta $\phi(t)$ la funzione generatrice dei momenti della v.a. X , si ha

$$\phi'(0) = E[X]$$

$$\left[\text{infatti} \right. \\ \left. \phi'(t) = \frac{d}{dt}\phi(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}e^{tX}\right] = E[Xe^{tX}] \right]$$

Si può verificare che:

$$\phi''(0) = E[X^2]$$

e in generale:

$$\phi^{(n)}(0) = E[X^n], \text{ per ogni } n \in \mathbb{N}.$$

Dunque, data la v.a. X con funzione generatrice $\phi(t)$, si ha:

- $E[X] = \phi'(0),$
- $Var(X) = \phi''(0) - (\phi'(0))^2.$

Esempio: distribuzione di Poisson

Dat X v.a. di Poisson di parametro $\lambda > 0$:
 $X \sim P_0(\lambda)$.

la sua densità è $P(k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, \dots, \infty$.
Otteniamo la funzione generatrice:

$$\begin{aligned}\phi(t) &= E[e^{tX}] = e^{-\lambda} \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = \\ &= \boxed{e^{-\lambda} e^{\lambda e^t}}.\end{aligned}$$

Per derivazione si ha:

$$\phi'(t) = \lambda e^t e^{\lambda(e^t-1)},$$

$$\phi''(t) = (\lambda e^t)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}.$$

Concludiamo allora

$$E[X] = \phi'(0) = \boxed{\lambda}$$

$$Var(X) = \phi''(0) - (\phi'(0))^2 = \lambda^2 + \lambda - \lambda^2 = \boxed{\lambda}$$

Esempio distribuzione esponenziale.

Consideriamo la v.a. esponenziale di parametro $\lambda > 0$: $X \sim Esp(\lambda)$, con densità:

$$f(t) = \begin{cases} \lambda e^{-\lambda t}; & t \geq 0 \\ 0; & t < 0 \end{cases}$$

La sua **funzione generatrice** sarà data per $t < \lambda$ da:

$$\begin{aligned} \phi(t) : &= E[e^{tX}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \\ &= \lambda \int_0^\infty e^{-(\lambda-t)x} dx = \boxed{\frac{\lambda}{\lambda-t}} \end{aligned}$$

Poiché $\phi'(t) = \frac{\lambda}{(\lambda-t)^2}$ e $\phi''(t) = \frac{2\lambda}{(\lambda-t)^3}$, otteniamo:

$$E[X] = \phi'(0) = \boxed{\frac{1}{\lambda}};$$

$$E[X^2] = \phi''(0) = \frac{2}{\lambda^2};$$

$$Var(X) = E[X^2] - E[X]^2 = \boxed{\frac{1}{\lambda^2}}.$$

Richiami di calcolo combinatorio

(1) Dati due insiemi finiti A e B, con $\text{card } A = m$, $\text{card } B = n$, si ha

$$\text{card}(A \times B) = m \cdot n$$

(2) Siano A_1, \dots, A_n n insiemi finiti, con $\text{card } A_i = m_i$. Si deduce facilmente per induzione da (1) che

$$\text{card}(A_1 \times A_2 \times \dots \times A_n) = m_1 \cdot m_2 \cdot \dots \cdot m_n$$

(3) E' importante il caso particolare di (2) in cui sia

$$A_1 = A_2 = \dots = A_n = A$$

Posto $m = \text{card } A$, si trova

$$\text{card}(A^n) = (\text{card } A)^n = m^n$$

(4) Siano dati due insiemi X e A, con $\text{card } X = n$, $\text{card } A = m$. Allora il numero delle applicazioni di X in A è dato da m^n .

(5) Ogni elemento del prodotto cartesiano $A^n = A \times A \times \dots \times A$ (n volte può essere visto come una n -upla di elementi di A, dei quali, eventualmente, qualcuno sia ripetuto. Le n -uple di elementi di A si chiamano anche **sequenze con ripetizione** (di elementi di A) di lunghezza n . Il numero di tali sequenze è dunque, come abbiamo appena visto, $(\text{card } A)^n$.

(6) Dato un insieme A formato da m oggetti, siamo interessati anche al numero di **sequenze senza ripetizione** (di elementi di A) di lunghezza n ($0 < n \leq m$). Esse vengono chiamate anche **disposizioni** di m oggetti a n a n . E' facile vedere che questo numero è dato da

$$(m)_n = m \cdot (m-1) \cdot (m-2) \cdot \dots \cdot (m-n+1) = \frac{m!}{(m-n)!}$$

(7) Siano X e A due insiemi come in (4), e supponiamo $0 < n \leq m$. Il numero delle applicazioni iniettive di X in A è dato da $(m)_n$.

(8) E' importante il caso particolare di (6) in cui sia $n = m$; $(m)_m$ è il numero di modi diversi in cui si possono disporre gli m elementi di A, o, come si dice, il numero delle **permutazioni** degli m oggetti che formano l'insieme A. $(m)_m$ si indica più frequentemente con il simbolo $m!$, e vale

$$m! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot m$$

(9) Caso particolare di (8) si ha quando le permutazioni sono **circolari**. Il numero di quest'ultime è $m-1!$

(10) Sia A come in (6). Due sequenze senza ripetizione di elementi di A possono differire per gli oggetti che le compongono, oppure anche soltanto per l'ordine in cui questi oggetti vengono presi; in molte situazioni, tuttavia, basterà contare il numero delle sequenze che differiscono per la natura degli elementi (identificando cioè quelle formate dagli stessi oggetti, presi in ordine diverso); ci interessa dunque trovare il numero di gruppi costituiti da n elementi, che si possono formare a partire da un insieme di cardinalità m . Questi gruppi si chiamano **combinazioni** di m oggetti n a n . Il numero che cerchiamo vale

$$\frac{(m)_n}{n!} = \frac{m \cdot (m-1) \cdot (m-2) \cdot \dots \cdot (m-n+1)}{n!} = \frac{m!}{n!(m-n)!} = \binom{m}{n}$$

$\binom{m}{n}$ si chiama **coefficiente binomiale** m, n . Osserviamo che

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

(11) Caso particolare di (10) si ha quando le combinazioni sono con ripetizione. Il numero di quest'ultime è dato da

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

(12) E' necessario dare un senso anche al simbolo $0!$ (il caso $m = n = 0$) non rientra nella situazione considerata in (8). Possiamo ragionare così: se in (10) prendiamo $n = m$, il coefficiente binomiale risultante dovrà essere il numero di gruppi costituiti da m oggetti, formati a partire da un insieme di cardinalità m . E' chiaro che di questi gruppi ne potremo formare solo uno (l'insieme dato) e quindi dovremo avere

$$\binom{m}{m} = 1$$

d'altra parte, sviluppando formalmente la formula che definisce il coefficiente binomiale, si trova

$$\binom{m}{m} = \frac{m!}{m!0!} = \frac{1}{0!}$$

Dovremo pertanto porre $0! = 1$.

(13) Sia A come in (6), e siano n_1, n_2, \dots, n_r degli interi maggiori o uguali a 0, tali che $n_1 + n_2 + \dots + n_r = m$. In quanti modi si possono formare r gruppi (a partire dagli elementi di A) G_1, G_2, \dots, G_r , in modo che il gruppo G_i contenga n_i elementi?

Per rispondere alla domanda, possiamo pensare di procedere così: formiamo prima un gruppo di n_1 elementi, poi, con gli $m - n_1$ elementi restanti, un gruppo di n_2 , e così via. Il numero di gruppi possibili è quindi dato da

$$\binom{m}{n_1} \cdot \binom{m-n_1}{n_2} \cdot \dots \cdot \binom{m-n_1-\dots-n_{r-1}}{n_r}$$

Si vede facilmente che la quantità scritta sopra è uguale a

$$\frac{m!}{n_1! \cdot n_2! \cdot \dots \cdot n_r!}$$

e si indica col simbolo

$$\binom{m}{n_1, n_2, \dots, n_r},$$

detto **coefficiente multinomiale** m, n_1, n_2, \dots, n_r .

Per $r = 2$ si ha

$$\binom{m}{n_1, n_2} = \binom{m}{n_1} = \binom{m}{n_2}$$

(14) Se $1 \leq n \leq m-1$ si ha

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

(15) **Il binomio di Newton.** Siano a, b due numeri reali, n un intero non negativo. Si ha allora

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

(16) Un caso particolare della formula precedente: prendendo $a = b = 1$, si trova

$$\sum_{k=0}^n \binom{n}{k} = 2^n$$

Ricordando il significato combinatorio dei coefficienti binomiali $\binom{n}{k}$, la formula scritta sopra si può interpretare dicendo che il numero dei sottoinsiemi che si possono formare a partire da un insieme di n elementi è pari a 2^n .

Schema riassuntivo

Riferimento	Descrizione	Proprietà	Numero
(5)	Sequenze con ripetizioni	$O - R$	m^n
(6)	Sequenze senza ripetizioni (o anche disposizioni)	$O - \overline{R}$	$(m)_n$
(8)	Permutazioni (caso particolare di (6))	$O - \overline{R}$	$m!$
(9)	Permutazioni circolari (caso particolare di (8))	$O - \overline{R}$	$m-1!$
(10)	Combinazioni	$\overline{O} - \overline{R}$	$\binom{m}{n}$
(11)	Combinazioni con ripetizione	$\overline{O} - R$	$\binom{m+n-1}{n}$
(13)	Permutazioni con ripetizione	$\overline{O} - R$	$\binom{m}{n_1, n_2, \dots, n_r}$

O – Ordine (cioè due sequenze composte dagli stessi simboli, ma posti in un ordine diverso, non sono considerate la medesima)

R – Ripetizione (è possibile che uno o più elementi vengano ripetuti una o più volte)

Speranza matematica e varianza

1. Definizioni e proprietà

Sia X una assegnata variabile aleatoria discreta sullo spazio (Ω, A, P) , avente densità p_X . Si dice che X ha *speranza matematica finita* se

$$(1.1) \quad \sum_{x \in \mathbb{R}} |x| p_X(x) < +\infty$$

In tal caso si chiama *speranza matematica* di X la quantità

$$(1.2) \quad \mathbf{E}[X] = \sum_{x \in \mathbb{R}} x p_X(x)$$

La speranza matematica si chiama anche, semplicemente, *speranza*. Altri nomi usati sono *attesa*, *valore atteso*, *media*, *valore medio*.

(1.3) Sia $X = (X_1, X_2, \dots, X_m)$ un vettore aleatorie discreto m -dimensionale, avente densità congiunta $p(x) = p(x_1, x_2, \dots, x_m)$. Sia poi $\phi: \mathbb{R}^m \rightarrow \mathbb{R}$ una funzione, e consideriamo la variabile aleatoria $Z = \phi \circ X$. Allora Z ha speranza matematica finita se e solo se

$$(1.4) \quad \sum_{x \in \mathbb{R}} |\phi(x)| p_X(x) < +\infty$$

In tal caso si ha

$$(1.5) \quad \mathbf{E}[Z] = \sum_{x \in \mathbb{R}} \phi(x) p_X(x)$$

(1.6) Siano X e Y due variabili aleatorie discrete aventi entrambe speranza matematica finita. Allora

(i) per ogni $c \in \mathbb{R}$ la variabile aleatoria cX ha speranza matematica finita e si ha

$$\mathbf{E}[cX] = c\mathbf{E}[X]$$

Dimostrazione – Per la definizione di speranza abbiamo

$$\mathbf{E}[cX] = \sum_{x \in \mathbb{R}} c x p_X(x)$$

portando la c fuori dalla sommatoria

$$\mathbf{E}[cX] = c \sum_{x \in \mathbb{R}} x p_X(x) = c\mathbf{E}[X]$$

(ii) la variabile aleatoria $X + Y$ ha speranza matematica finita e si ha

$$\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$$

Dimostrazione – Per la definizione di speranza abbiamo

$$\mathbf{E}[X + Y] = \sum_{x,y} (x + y) p_{X,Y}(x, y) = \sum_x \sum_y (x + y) p_{X,Y}(x, y)$$

separando le due sommatorie si ha

$$\mathbf{E}[X + Y] = \sum_x \sum_y (x + y) p_{X,Y}(x, y) = \sum_x \sum_y x p_{X,Y}(x, y) + \sum_y \sum_x y p_{X,Y}(x, y)$$

portando la x fuori dalla sommatoria $\sum_y x p_{X,Y}(x, y)$ e y fuori da $\sum_x y p_{X,Y}(x, y)$

$$\mathbf{E}[X + Y] = \sum_x x \sum_y p_{X,Y}(x, y) + \sum_y y \sum_x p_{X,Y}(x, y)$$

notando che $\sum_y p_{X,Y}(x, y) = p_X(x)$ e $\sum_x p_{X,Y}(x, y) = p_Y(y)$

$$\mathbf{E}[X + Y] = \sum_x x p_X(x) + \sum_y y p_Y(y) = \mathbf{E}[X] + \mathbf{E}[Y]$$

Le proprietà (i) e (ii) precedenti si possono riassumere dicendo che l'operatore $X \mapsto \mathbf{E}[X]$ è lineare. In particolare, se X è una variabile aleatoria e c e d due costanti numeriche, si ha

$$(1.7) \quad \mathbf{E}[cX + d] = c\mathbf{E}[X] + d$$

(1.8) Siano X e Y due variabili aleatorie indipendenti, entrambe con speranza matematica finita. Allora la variabile aleatoria $Z = XY$ ha speranza matematica finita e

$$\mathbf{E}[XY] = \mathbf{E}[X] \mathbf{E}[Y]$$

Dimostrazione – Per la definizione di speranza abbiamo

$$\mathbf{E}[XY] = \sum_{x,y} x y p_X(x) p_Y(y) = \sum_x \sum_y x y p_X(x) p_Y(y) = \sum_x \sum_y x p_X(x) y p_Y(y)$$

portando $x p_X(x)$ fuori dalla sommatoria $\sum_y x p_X(x) y p_Y(y)$

$$\mathbf{E}[XY] = \sum_x \sum_y x p_X(x) y p_Y(y) = \sum_x x p_X(x) \sum_y y p_Y(y) = \mathbf{E}[X] \mathbf{E}[Y]$$

(1.9) Vediamo ora il calcolo della speranza in alcuni casi importanti.

- (i) **Densità Bernoulliana** – Sia X una variabile aleatoria avente densità $B(1, p)$. Cioè

$$X = \begin{cases} 1 & \text{con prob. } p \\ 0 & \text{con prob. } (1-p) \end{cases}$$

Applicando allora la definizione di speranza, si ha

$$\mathbf{E}[X] = 1 \times p + 0 \times (1-p) = p$$

- (ii) **Densità Binomiale** – Sia X una variabile aleatoria avente densità binomiale $B(n, p)$. Ciò significa che X assume i valori interi $0, 1, \dots, n$, e che, per $k = 0, 1, \dots, n$ si ha

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Il valore della somma precedente si può calcolare, e si trova che è np .

Un'altra via, più semplice, per calcolare la speranza della densità binomiale è la seguente:

Sia Y una variabile aleatoria avente densità $B(n, p)$, cioè che conta il numero di successi in n prove indipendenti (tutte aventi legge $B(1, p)$). Dunque possiamo scrivere

$$Y = \sum_{i=1}^n X_i$$

dove le X_i hanno legge $B(1, p)$: poiché la speranza di una variabile dipende esclusivamente dalla sua legge, possiamo utilizzare la variabile aleatoria Y per il calcolo che ci interessa. Per la proprietà di linearità dell'operatore speranza e per il punto precedente si ha allora

$$\mathbf{E}[Y] = \sum_{i=1}^n \mathbf{E}[X_i] = \sum_{i=1}^n p = np$$

- (iii) **Densità di Poisson** – Sia X una variabile aleatoria con densità di Poisson: X assume i valori interi e

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

Si ha dunque

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \lambda \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \sum_{h=0}^{\infty} \lambda \frac{\lambda^h}{h!} e^{-\lambda} = \lambda \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} e^{-\lambda} = \lambda$$

Si noti che l'ultima somma scritta sopra non è altro che la somma dei valori della densità di Poisson (e quindi vale 1).

La speranza di una variabile aleatoria X è una delle cosiddette **misure di centralità** della legge di X : ciò significa che, in un certo senso, la speranza ci dà una valutazione del valore centrale di X ; tuttavia fino a questo momento non abbiamo nessuna indicazione su quanto sia grande la “tendenza” dei valori di X a disperdersi attorno a tale indice di centralità.

Una prima idea potrebbe essere la seguente: lo scarto tra X e $\mathbf{E}[X]$ è dato da $|X - \mathbf{E}[X]|$; il valore medio di tale scarto è dunque

$$\mathbf{E}[|X - \mathbf{E}[X]|]$$

Questo numero potrebbe essere un ragionevole indice di dispersione; il problema è che il suo uso non è agevole perché i calcoli con esso risultano in genere troppo complicati. Si può allora rimediare così: per misurare lo scarto tra X e $\mathbf{E}[X]$ usiamo la quantità $|X - \mathbf{E}[X]|^2 = (X - \mathbf{E}[X])^2$.

La media di questa nuova variabile è

$$\mathbf{E}[(X - \mathbf{E}[X])^2]$$

Sia X una variabile aleatoria (discreta), avente media finita. Supponiamo che sia anche

$$\mathbf{E}[(X - \mathbf{E}[X])^2] < \infty$$

In tal caso la quantità precedente si chiama **varianza** di X e si indica con **VarX**. Si pone cioè

$$\mathbf{VarX} = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

Inoltre, la quantità

$$\sqrt{\mathbf{VarX}}$$

è detta **deviazione standard** di X .

Se X ha densità p_X , dalla (1.3) segue la formula

$$(1.10) \quad \mathbf{VarX} = \sum_{x \in \mathbf{R}} [(x - \mathbf{E}[X])^2] p_X(x)$$

Vediamo adesso le principali proprietà della varianza.

Dalla formula della definizione si ricava

$$(1.11) \quad \begin{aligned} \mathbf{VarX} &= \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2 - 2\mathbf{E}[X]X + \mathbf{E}[X]^2] \\ &= \mathbf{E}[X^2] - 2\mathbf{E}[X]^2 + \mathbf{E}[X]^2 = \mathbf{E}[X^2] - \mathbf{E}[X]^2 \end{aligned}$$

questa è una seconda forma con cui calcolare la varianza.

(1.12) Sia X una variabile aleatoria avente varianza finita e sia c una costante. Allora

(i) $\mathbf{Var}(cX) = c^2 \mathbf{Var}X$

Dimostrazione

$$\begin{aligned}\mathbf{Var}(cX) &= \mathbf{E}[c^2 X^2] - (\mathbf{E}[cX])^2 = c^2 \mathbf{E}[X^2] - (c\mathbf{E}[X])^2 = c^2 \mathbf{E}[X^2] - c^2 \mathbf{E}[X]^2 \\ &= c^2 (\mathbf{E}[X^2] - \mathbf{E}[X]^2) = c^2 \mathbf{Var}(X)\end{aligned}$$

(ii) $\mathbf{Var}(X + c) = \mathbf{Var}X$

Dimostrazione

$$\mathbf{Var}(X + c) = \mathbf{E}[(X + c - \mathbf{E}[X + c])^2] = \mathbf{E}[(X + c - \mathbf{E}[X] - c)^2] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{Var}(X)$$

(iii) $\mathbf{Var}(X) = 0 \Leftrightarrow X = c$

Dimostrazione

- $X = c \Rightarrow \mathbf{Var}(X) = 0$ – Ovvio in quanto, essendo X una costante, l'unico valore che assume X è c e pertanto la “dispersione” attorno alla media è nulla.
- $\mathbf{Var}(X) = 0 \Rightarrow X = c$ – Supponiamo che sia $\mathbf{Var}(X) = 0$. Allora per la (1.10) abbiamo

$$0 = \mathbf{Var}X = \sum_{x \in \mathbb{R}} [(x - \mathbf{E}[X])^2] p_X(x)$$

Nella somma precedente tutti gli addendi sono non negativi. Pertanto la loro somma può essere nulla se e solo se ciascun addendo è nullo. Ciò può accadere se $p_X(x) = 0$ oppure se $x - \mathbf{E}[X] = 0$, cioè se il valore x coincide con il valore medio. Dunque l'unico valore assunto da X con probabilità non nulla è $\mathbf{E}[X]$, e X è dunque costante.

Calcoliamo ora la varianza della somma di due variabili X e Y . Risulta

$$\begin{aligned}\mathbf{Var}(X + Y) &= \mathbf{E}[(X + Y - \mathbf{E}[X + Y])^2] = \mathbf{E}[(X + Y - \mathbf{E}[X] - \mathbf{E}[Y])^2] \\ &= \mathbf{E}[(X - \mathbf{E}[X] + Y - \mathbf{E}[Y])^2] = \mathbf{E}[(X - \mathbf{E}[X])^2 + (Y - \mathbf{E}[Y])^2 + 2(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}[(X - \mathbf{E}[X])^2] + \mathbf{E}[(Y - \mathbf{E}[Y])^2] + 2\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{Var}(X) + \mathbf{Var}(Y) + 2\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]\end{aligned}$$

Si chiama **covarianza** di X e Y la quantità

$$\mathbf{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

Sia $p_{X,Y}$ la densità congiunta della coppia (X, Y) . Dalla (1.3) segue la formula

$$\mathbf{Cov}(X, Y) = \sum_{x,y} (x - \mathbf{E}[X])(y - \mathbf{E}[Y]) p_{X,Y}(x, y)$$

Inoltre risulta

$$\begin{aligned}\mathbf{Cov}(X, Y) &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}[XY - X\mathbf{E}[Y] - Y\mathbf{E}[X] + \mathbf{E}[X]\mathbf{E}[Y]] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] - \mathbf{E}[X]\mathbf{E}[Y] + \mathbf{E}[X]\mathbf{E}[Y] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]\end{aligned}$$

Se ne deduce che

$$\mathbf{Cov}(X, Y) = 0 \Leftrightarrow \mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$$

(1.13) Due variabili aleatorie X e Y si dicono *non correlate* se $\mathbf{Cov}(X, Y) = 0$, oppure, equivalentemente, se $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.

(1.14) Dalla (1.13) si deduce che due variabili aleatorie indipendenti con varianza finita sono non correlate. Si ricordi che non vale il viceversa, cioè: esistono coppie di variabili aleatorie non correlate ma non indipendenti.

(1.15) Se X e Y sono non correlate, allora

$$\mathbf{Var}(X + Y) = \mathbf{Var}X + \mathbf{Var}Y$$

(1.16) Sia X una variabile aleatoria. Dalla definizione di covarianza abbiamo

$$\mathbf{Cov}(X, X) = \mathbf{E}[(X - \mathbf{E}[X])(X - \mathbf{E}[X])] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{Var}X$$

(1.17) Calcoliamo ora la varianza delle densità fondamentali che conosciamo.

- (i) **Densità Bernoulliana** – Sia X una variabile aleatoria avente legge $B(1, p)$. Allora, poiché $X^2 = X$, per il punto (1.9)(i) si ha

$$\mathbf{Var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \mathbf{E}[X] - \mathbf{E}[X]^2 = p - p^2 = p(1 - p)$$

- (ii) **Densità Binomiale** – Sia X una variabile avente legge $B(n, p)$. Allora come già sappiamo, X ha la stessa legge di

$$\sum_{i=1}^n X_i$$

dove le X_i sono tra loro indipendenti ed hanno tutte legge $B(1, p)$. Allora per la (1.15) e per il punto precedente si ha

$$\mathbf{Var}X = \mathbf{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbf{Var}X_i = \sum_{i=1}^n p(1 - p) = np(1 - p)$$

- (iii) **Densità di Poisson** – Sia X una variabile aleatoria avente densità di Poisson di parametro λ . Per il calcolo della varianza, adoperiamo anche in questo caso l'espressione

$$\mathbf{Var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

cominciando dal calcolo di $\mathbf{E}[X^2]$ risulta

$$\begin{aligned} \mathbf{E}[X^2] &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \mathbf{E}[X+1] = \lambda(\lambda+1) = \lambda^2 + \lambda \end{aligned}$$

Di conseguenza

$$\mathbf{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

2. Riepilogo delle speranze e varianze delle principali densità

Nome	Legge	Speranza	Varianza
Bernoulliana	$X \sim B(1, p)$	p	$p(1 - p)$
Binomiale	$X \sim B(n, p)$	np	$np(1 - p)$
Poisson	$X \sim \text{Poisson}(\lambda)$	λ	λ

Variabili aleatorie assolutamente continue

1. Definizioni e prime proprietà

Abbiamo visto che, se X è una variabile aleatoria discreta di densità p , allora, per ogni $A \subseteq \mathbb{R}$, si ha

$$(1.1) \quad P(X \in A) = \sum_{x \in A} p(x)$$

Una variabile aleatoria sullo spazio (Ω, A, P) si dice **assolutamente continua** se esiste una funzione $f: \mathbb{R} \rightarrow \mathbb{R}^+$ integrabile su \mathbb{R} e tale che, per ogni insieme misurabile A di \mathbb{R} , risulta

$$(1.2) \quad P(X \in A) = \int_A f(x) dx$$

In tal caso la funzione f si chiama **densità** di X .

Come al solito, quando saranno in gioco più variabili X, Y, \dots , indicheremo le rispettive densità con f_X, f_Y, \dots .

Si noti l'analogia tra le formule (1.1) e (1.2). La seconda si può formalmente ottenere dalla prima sostituendo il simbolo di sommatoria con quello di integrale.

Dalla formula (1.2) si deduce in particolare che la funzione di ripartizione F di X assuma la forma

$$F(t) = P(X \leq t) = P(X \in (-\infty, t]) = \int_{-\infty}^t f(x) dx$$

Inoltre si ha

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

E ancora

$$P(X \in \{a\}) = \int_{\{a\}} f(x) dx = 0$$

da cui segue che la variabile aleatoria X è continua.

Infine si ha (notare l'analogia con il caso discreto)

$$\int_{-\infty}^{+\infty} f(x) dx = P(X \in \mathbb{R}) = P(\Omega) = 1$$

Due variabili aleatorie si dicono tra loro indipendenti se per ogni coppia A, B di sottoinsiemi misurabili di \mathbb{R} , risulta

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

oppure se, per ogni coppia di numeri reali s, t , risulta

$$P(X \leq s, Y \leq t) = P(X \leq s)P(Y \leq t) = F_X(s)F_Y(t)$$

Sia X una variabile aleatoria assolutamente continua avente densità f .

(1.3) Si dice che X ha *speranza matematica finita* se

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

in tal caso si chiama *speranza matematica* di X il numero

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

(1.4) Se X ha speranza $\mathbf{E}[X]$ finita e se anche

$$\int_{-\infty}^{+\infty} (x - \mathbf{E}[X])^2 f(x) dx < \infty$$

si dice che X ha *varianza finita* e si pone

$$\mathbf{Var}X = \mathbf{E}[(X - \mathbf{E}[X])^2] = \int_{-\infty}^{+\infty} (x - \mathbf{E}[X])^2 f(x) dx$$

Per la speranza e la varianza delle variabili aleatorie assolutamente continue valgono tutte le proprietà che abbiamo visto nel caso discreto. Si noti l'analogia delle definizioni: il simbolo di sommatoria che usavamo nel caso discreto è stato sostituito dal simbolo di integrale.

2. Alcune densità di variabili aleatorie assolutamente continue

(2.1) **Le leggi uniformi** – Si sceglie un punto a caso nell'intervallo $[a, b]$ della retta \mathbb{R} . Indichiamo con X l'ascissa del punto scelto. La funzione di ripartizione di X è

$$F(t) = \begin{cases} 0 & \text{se } t < a \\ \frac{t-a}{b-a} & \text{se } a \leq t \leq b \\ 1 & \text{se } t > b \end{cases}$$

e la densità è

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$$

La legge qui descritta si chiama **legge uniforme** (sull'intervallo $[a, b]$ e si indica con il simbolo $U(a, b)$); ogni variabile aleatoria che ammetta questa legge si dice **uniforme** (su $[a, b]$).

(2.1.1) Calcolo della speranza matematica di X :

Per la definizione di speranza abbiamo

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x \frac{1}{b-a} dx$$

da cui si ha

$$\mathbf{E}[X] = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

Si noti che $\int_{-\infty}^{+\infty}$ diventa semplicemente \int_a^b in quanto non vengono considerati quei punti non appartenenti all'intervallo $[a, b]$, per i quali la f risulta pari a 0.

(2.1.2) Calcolo della varianza di X :

Per la (1.11)¹ abbiamo

$$\mathbf{Var}X = \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

cominciando dal calcolo di $\mathbf{E}[X^2]$

$$\mathbf{E}[X^2] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(a^2 + b^2 + ab)}{3(b-a)} = \frac{a^2 + b^2 + ab}{3}$$

Di conseguenza

$$\mathbf{Var}X = \frac{a^2 + b^2 + ab}{3} - \frac{a^2 + b^2 + 2ab}{4} = \frac{a^2 + b^2 - 2ab}{12} = \frac{(b-a)^2}{12}$$

Da ciò segue

$$\sqrt{\mathbf{Var}X} = \frac{b-a}{2\sqrt{3}}$$

(2.1.3) Consideriamo il caso in cui $a=0$ e $b=1$, cioè $X \sim U(0, 1)$.

Per quello detto finora, avremo

$$F(t) = \begin{cases} 0 & \text{se } t < 0 \\ t & \text{se } 0 \leq t \leq 1 \\ 1 & \text{se } t > 1 \end{cases}$$

$$f(x) = \begin{cases} 1 & \text{se } 0 \leq x \leq 1 \\ 0 & \text{altrimenti} \end{cases}$$

$$\mathbf{E}[X] = \frac{1}{2} \text{ e } \mathbf{Var}X = \frac{1}{12}$$

¹ – Riferimento al capitolo 4 (Speranza matematica e varianza)

(2.2) **La legge esponenziale** – Sia λ un numero reale positivo assegnato, X una variabile aleatoria. Si dice che X ha **legge esponenziale** (di parametro λ) (il simbolo è $\mathcal{E}(\lambda)$ oppure $Esp(\lambda)$) se ammette come densità la funzione

$$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \lambda e^{-\lambda x} & \text{se } x \geq 0 \end{cases}$$

La funzione di ripartizione è

$$F(t) = \begin{cases} 0 & \text{per } t \leq 0 \\ 1 - e^{-\lambda t} & \text{per } t > 0 \end{cases}$$

(2.2.1) Calcolo della speranza matematica di X :

Per la definizione di speranza abbiamo

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x \lambda e^{-\lambda x} dx$$

da cui si ha

$$\begin{aligned} \mathbf{E}[X] &= \int_0^{+\infty} x \lambda e^{-\lambda x} dx = x - e^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\ &= 0 + \frac{1}{\lambda} \cdot \int_0^{+\infty} \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \cdot 1 = \frac{1}{\lambda} \end{aligned}$$

Si noti che $\int_{-\infty}^{+\infty}$ diventa semplicemente $\int_0^{+\infty}$ in quanto non vengono considerati quei punti non appartenenti all'intervallo $[0, +\infty)$, per i quali la f risulta pari a 0.

(2.2.2) Calcolo della varianza di X :

Come al solito abbiamo $\mathbf{Var}X = \mathbf{E}[X^2] - \mathbf{E}[X]^2$; cominciando dal calcolo di $\mathbf{E}[X^2]$

$$\begin{aligned} \mathbf{E}[X^2] &= \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = x^2 - e^{-\lambda x} \Big|_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\lambda x} dx \\ &= 0 + \frac{1}{\lambda} \cdot 2 \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \cdot \frac{2}{\lambda} = \frac{2}{\lambda^2} \end{aligned}$$

Di conseguenza

$$\mathbf{Var}X = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Da ciò segue

$$\sqrt{\mathbf{Var}X} = \frac{1}{\lambda} = \mathbf{E}[X]$$

Ogni variabile X avente legge esponenziale verifica la seguente relazione

$$(2.2.3) \quad P(X > t + s \mid X > t) = P(X > s)$$

dove s e t sono due numeri positivi assegnati.

Dimostrazione – Per la definizione di probabilità condizionata abbiamo

$$P(X > t + s \mid X > t) = \frac{P(X > t + s, X > t)}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)}$$

dove, per la definizione di densità di una variabile aleatoria assolutamente continua, vale

$$\frac{P(X > t + s)}{P(X > t)} = \frac{\int_{t+s}^{+\infty} \lambda e^{-\lambda x} dx}{\int_t^{+\infty} \lambda e^{-\lambda x} dx}$$

da cui si ha

$$\frac{P(X > t + s)}{P(X > t)} = \frac{\int_{t+s}^{+\infty} \lambda e^{-\lambda x} dx}{\int_t^{+\infty} \lambda e^{-\lambda x} dx} = \frac{-e^{-\lambda x} \Big|_{t+s}^{+\infty}}{-e^{-\lambda x} \Big|_t^{+\infty}} = \frac{-e^{-\lambda(t+s)}}{-e^{-\lambda t}} = \frac{e^{-\lambda t} \cdot e^{-\lambda s}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s)$$

La relazione (2.2.3) viene detta **proprietà di assenza di memoria** o anche **proprietà di assenza di usura**. Per capire il significato di queste espressioni, immaginiamo che X rappresenti la durata di una lampadina, messa in funzione all'istante 0. Allora la (2.2.3) si interpreta così: sapendo che la lampadina non si è guastata fino all'istante t , la probabilità che essa duri per ulteriori s istanti è uguale alla probabilità che la lampadina messa in funzione all'istante 0, non si guasti fino all'istante s . In altri termini, la lampadina non è soggetta all'usura causata dai primi t istanti di funzionamento.

(2.3) **La legge di Weibull** – Siano $\alpha > 0$ e $\lambda > 0$ due numeri fissati, e consideriamo la funzione

$$f(x) = \begin{cases} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

Osserviamo anche che, per $\alpha = 1$, si ottiene la densità esponenziale di parametro λ di cui abbiamo già parlato ($\mathcal{E}(\lambda) = W(\lambda, 1)$ oppure $\text{Esp}(\lambda) = W(\lambda, 1)$). Nel caso generale, f viene chiamata **densità di Weibull** (di parametri α e λ , in simboli $W(\lambda, \alpha)$).

Sia X una variabile aleatoria avente f come densità. Ci interessa studiare la quantità $P(X > t+s \mid X > t)$. In questo caso si ha

$$P(X > t+s \mid X > t) = \frac{\int_t^{t+s} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} dx}{\int_t^{+\infty} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} dx} = \frac{-e^{-\lambda x^\alpha} \Big|_t^{t+s}}{-e^{-\lambda x^\alpha} \Big|_t^{+\infty}} = \frac{-e^{-\lambda (t+s)^\alpha}}{-e^{-\lambda t^\alpha}} = e^{-\lambda [(t+s)^\alpha - t^\alpha]}$$

Per ogni s fissato, la funzione $t \mapsto P(X > t+s \mid X > t)$ è crescente (risp. costante, decrescente) quando la funzione $t \mapsto (t+s)^\alpha - t^\alpha$ è decrescente (risp. costante, crescente). Ciò accade quando

$$\frac{d}{ds} [(t+s)^\alpha - t^\alpha] = \alpha [(t+s)^{\alpha-1} - t^{\alpha-1}] < 0$$

(risp. $= 0, > 0$) ovvero quando $\alpha < 1$ (risp. $= 1, > 1$). Ricapitolando

$$t \mapsto P(X > t+s \mid X > t) \begin{cases} \text{crescente per } \alpha < 1 \\ \text{costante per } \alpha = 1 \\ \text{decrescente per } \alpha > 1 \end{cases}$$

Supponiamo che X rappresenti la durata di una lampadina soggetta ad usura. E' naturale allora scegliere per X una densità per la quale la funzione $t \mapsto P(X > t+s \mid X > t)$ sia decrescente, e per tale motivo in questo caso viene usata una variabile X avente densità di Weibull di parametro $\alpha > 1$. Una situazione in cui sarebbe invece indicata una densità di Weibull con parametro $\alpha < 1$ è ad esempio quella in cui X rappresenta la durata di vita di un neonato (è noto che la probabilità di mortalità infantile è massima nei primi giorni di vita del bambino, per poi andare a diminuire man mano che i giorni passano).

3. Le principali densità della statistica

(3.1) **La densità normale** – Siano $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$ due parametri assegnati. Si chiama **densità normale** (o **gaussiana**) la funzione

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

La legge di cui f è una densità si indica con il simbolo $N(\mu, \sigma^2)$; ogni variabile aleatoria avente tale legge si chiama variabile aleatoria **normale** o **gaussiana**.

Nella famiglia delle leggi normali, sopra definite, particolare rilevanza ha la $N(0, 1)$ (cioè la legge di parametri $\mu = 0$ e $\sigma^2 = 1$). Scriviamone esplicitamente la densità:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Appunto per la sua importanza, questa legge viene detta **normale standard** (o **gaussiana standard**). Una prima proprietà della famiglia delle leggi normali è la sua stabilità rispetto alle trasformazioni affini; precisamente si ha:

(3.1.1) **Proposizione** – Sia X una variabile aleatoria avente legge gaussiana $N(m, s^2)$; siano inoltre a e b due numeri reali. Allora la variabile aleatoria $Y = aX + b$ ha legge $N(am + b, a^2 s^2)$.

Dimostrazione

(i) Per $a > 0$

$$\begin{aligned} P(Y \leq y) &= P(aX + b \leq y) = P(aX \leq y - b) = P\left(X \leq \frac{y - b}{a}\right) \\ P\left(X \leq \frac{y - b}{a}\right) &= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}s} e^{-\frac{(t-m)^2}{2s^2}} dt = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}as} e^{-\frac{(x-am-b)^2}{2a^2s^2}} dx \end{aligned}$$

dove l'ultima eguaglianza si ottiene effettuando il cambiamento di variabile

$$t = \frac{x - b}{a}$$

(ii) Per $a < 0$

$$\begin{aligned} P(Y \leq y) &= P(aX + b \leq y) = P(aX \leq y - b) = P\left(X \geq \frac{y - b}{a}\right) \\ P\left(X \geq \frac{y - b}{a}\right) &= \int_{\frac{y-b}{a}}^{+\infty} \frac{1}{\sqrt{2\pi}s} e^{-\frac{(t-m)^2}{2s^2}} dt = \int_y^{+\infty} \frac{1}{\sqrt{2\pi}as} e^{-\frac{(x-am-b)^2}{2a^2s^2}} dx = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}as} e^{-\frac{(x-am-b)^2}{2a^2s^2}} dx \end{aligned}$$

Dalla (3.1.1) si deduce il

(3.1.2) **Corollario** – Siano $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$ due numeri assegnati.

(i) Se $X \sim N(0, 1)$, allora $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$

(ii) Se $X \sim N(\mu, \sigma^2)$, allora $Y = \frac{X - \mu}{\sigma} \sim N(0, 1)$

Calcoliamo ora la speranza e la varianza di una variabile aleatoria X avente legge $N(\mu, \sigma^2)$, cominciamo dal caso particolare $\mu = 0$ e $\sigma^2 = 1$.

Sia dunque $X \sim N(0, 1)$.

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} = 0$$

Passiamo ora al calcolo della varianza (che, poiché $\mathbf{E}[X] = 0$, in questo caso coincide con $\mathbf{E}[X^2]$).

$$\begin{aligned} \mathbf{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{+\infty} x \left(x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right) dx = \int_{-\infty}^{+\infty} x \frac{d}{dx} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right) dx \\ &= -x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0 + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 \end{aligned}$$

A questo punto siamo in grado di calcolare $\mathbf{E}[X]$ e $\mathbf{Var}X$ anche per una variabile aleatoria X avente legge $N(\mu, \sigma^2)$. Per far questo usiamo il punto (ii) del corollario (3.1.2), che ci assicura che la variabile aleatoria definita da

$$Y = \frac{X - \mu}{\sigma}$$

ha legge $N(0, 1)$.

D'altra parte si può scrivere

$$X = \sigma Y + \mu$$

e quindi, per le note proprietà della speranza e della varianza, risulta

$$\mathbf{E}[X] = \mathbf{E}[\sigma Y + \mu] = \sigma \mathbf{E}[Y] + \mu = \mu$$

$$\mathbf{Var}X = \mathbf{Var}(\sigma Y + \mu) = \sigma^2 \mathbf{Var}(Y) = \sigma^2$$

(3.1.3) **Teorema** – Siano X_1, X_2, \dots, X_n n variabili aleatorie indipendenti, tali che, per ogni i , con $i = 1, 2, \dots, n$ si abbia

$$X_i \sim N(\mu_i, \sigma_i^2)$$

Allora la variabile aleatoria

$$Z = X_1 + X_2 + \dots + X_n$$

ha legge $N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.

(3.1.4) Siano X_1, X_2, \dots, X_n n variabili aleatorie indipendenti e tutte con la stessa legge (non necessariamente gaussiana). Poniamo

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

La variabile aleatoria \bar{X} viene detta **media campionaria** ed è di uso comune in statistica.

Calcoleremo ora la legge di \bar{X} nel caso particolare in cui le $X_i, i = 1, 2, \dots, n$ abbiano legge $N(\mu, \sigma^2)$. Posto $Z = X_1 + X_2 + \dots + X_n$, per il teorema precedente si ha

$$Z \sim N(n\mu, n\sigma^2)$$

dunque, per la (3.1.1) applicata a Z (con valori $a = \frac{1}{n}$ e $b = 0$) si ha

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Come conseguenza, per il punto (ii) di (3.1.2) si deduce che la variabile aleatoria centrata e ridotta

$$\begin{aligned} \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} &= \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} = \sqrt{n} \frac{\frac{X_1 + X_2 + \dots + X_n}{n} - \mu}{\sigma} = \frac{\sqrt{n}}{\sqrt{n}} \sqrt{n} \frac{\frac{X_1 + X_2 + \dots + X_n}{n} - \mu}{\sigma} \\ &= n \frac{\frac{X_1 + X_2 + \dots + X_n}{n} - \mu}{\sigma \sqrt{n}} = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma \sqrt{n}} \end{aligned}$$

ha legge normale standard.

D'ora in avanti la funzione di ripartizione della legge $N(0, 1)$ verrà indicata con il simbolo Φ ; abbiamo cioè

$$\Phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad t \in \mathbb{R}$$

Abbiamo visto che, se si hanno n variabili indipendenti X_1, X_2, \dots, X_n tutte con legge $N(\mu, \sigma^2)$ e a partire da esse si costruisce la variabile aleatoria centrata e ridotta associata alla loro media campionaria, cioè

$$S_n^* = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

la variabile S_n^* risulta avere legge normale standard.

(3.1.5) **Teorema limite centrale** (abbreviato con la sigla **CLT** nella letteratura di lingua anglosassone) – Sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie indipendenti e tutte con la stessa legge (si abbrevia con la sigla: i.i.d. = indipendenti identicamente distribuite), aventi media μ e varianza σ^2 finite. Per ogni intero n poniamo

$$S_n^* = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

Allora, per ogni $x \in \mathbb{R}$ si ha

$$\lim_{n \rightarrow \infty} P(S_n^* \leq x) = \Phi(x)$$

Si ha cioè il fatto che la legge di S_n^* tende ad avvicinarsi alla legge normale standard qualunque sia la legge di partenza delle variabili (X_i) (non necessariamente gaussiane).

In genere basta $n > 50$ per avere un'approssimazione soddisfacente.

(3.2) **La legge dei 3σ** – Sia X una variabile aleatoria con legge $N(\mu, \sigma^2)$. Cerchiamo di calcolare le seguenti probabilità:

- (i) $P(\mu - \sigma^2 \leq X \leq \mu + \sigma^2)$
- (ii) $P(\mu - 2\sigma^2 \leq X \leq \mu + 2\sigma^2)$
- (iii) $P(\mu - 3\sigma^2 \leq X \leq \mu + 3\sigma^2)$

Passando alla variabile aleatoria normale standard $Y = \frac{X - \mu}{\sigma}$, abbiamo

- (i) $P(\mu - \sigma^2 \leq X \leq \mu + \sigma^2) = P(-1 \leq Y \leq 1) = 0.6826$
- (ii) $P(\mu - 2\sigma^2 \leq X \leq \mu + 2\sigma^2) = P(-2 \leq Y \leq 2) = 0.9544$
- (iii) $P(\mu - 3\sigma^2 \leq X \leq \mu + 3\sigma^2) = P(-3 \leq Y \leq 3) = 0.9974$

Da questo si nota che è minore dello 0.3% la probabilità, che una variabile aleatoria con legge $N(\mu, \sigma^2)$, assuma un valore che cada al di fuori dell'intervallo semidimensionale 3σ avente centro in μ . Questa è la cosiddetta “**legge dei 3σ** ” secondo cui è altamente improbabile che una variabile aleatoria normale assuma valori che distano dalla sua media per più di tre deviazioni standard.

4. La funzione di ripartizione della legge normale standard

(4.1) Sia X una variabile aleatoria. Si dice che X è simmetrica se X e $-X$ hanno la stessa legge, o, in modo equivalente, la stessa funzione di ripartizione:

$$F_X(t) = P(X \leq t) = P(-X \leq t) = F_{-X}(t)$$

La precedente relazione si può scrivere anche nella forma

$$P(X \leq -t) = P(X \geq t) = 1 - P(X \leq t)$$

ovvero anche

$$(4.1.1) \quad P(X \leq t) + P(X \geq t) = F_X(t) + F_X(-t) = 1$$

Dalla relazione precedente si deduce in particolare che, se X è simmetrica allora $2F_X(0) = 1$, e cioè

$$(4.1.2) \quad F_X(0) = \frac{1}{2}$$

Se X ammette densità f , allora X è simmetrica se e solo se, $\forall t \in \mathbb{R}$, si ha

$$\int_{-\infty}^{-t} f(x) dx = \int_t^{+\infty} f(x) dx$$

(4.1.3) Sia X una variabile aleatoria assolutamente continua. Allora X è simmetrica se e solo se ammette una densità f pari, cioè tale che

$$f(x) = f(-x), \quad \forall x \in \mathbb{R}$$

Segue dalla proposizione (4.1.3) che ogni variabile aleatoria X avente legge normale standard è simmetrica. Le relazioni (4.1.1) e (4.1.2) applicate alla funzione di ripartizione Φ danno allora rispettivamente

$$\Phi(x) + \Phi(-x) = 1$$

$$\Phi(0) = \frac{1}{2}$$

Vediamo ora in che modo si usano le tavole della legge normale standard. Esistono vari tipi di tavole. Le più comuni riportano il valore di $\Phi(x)$ per $x \in (0, 3)$. Il valore x si legge fino al primo decimale nella colonna di sinistra; se x ha anche un secondo decimale, esso si legge nella riga in alto. Allora $\Phi(x)$ si trova, nella tavola, nella posizione in cui si incrociano la riga e la colonna prescelte. Ad esempio, se $x = 1.64$, il valore 1.6 si legge nella colonna di sinistra. Ad esso si aggiunge il numero 0.04, che troviamo nella riga in alto. Dunque $\Phi(x) = 0.9495$ (è il numero che si trova all'intersezione della riga marcata 1.6 con la colonna marcata 0.04).

Questo per quanto riguarda i numeri $x \in (0, 3)$. Se invece $x \in (-3, 0)$ si usa la relazione (4.1.1). Ad esempio, per $x = -1.64$ si trova

$$\Phi(-1.64) = 1 - \Phi(1.64) = 1 - 0.9495 = 0.0505$$

Infine, se $x \geq 3$, il valore di $\Phi(x)$ si approssima con 1 (e, di conseguenza, $\Phi(x) \approx 0$ per $x \leq -3$).

(4.2) Siano X_1, X_2, \dots, X_n n variabili aleatorie indipendenti e tutte con densità normale standard. Poniamo

$$Y = X_1^2 + X_2^2 + \dots + X_n^2$$

La densità di Y viene chiamata **densità del chi quadro a n gradi di libertà** e viene indicata con il simbolo $\chi^2(n)$.

(4.2.1) Calcolo della speranza di $\chi^2(n)$:

Per come abbiamo definito sopra la Y abbiamo

$$\mathbf{E}[Y] = \mathbf{E}[X_1^2 + X_2^2 + \dots + X_n^2]$$

per le note proprietà della speranza

$$\mathbf{E}[Y] = \mathbf{E}[X_1^2] + \mathbf{E}[X_2^2] + \dots + \mathbf{E}[X_n^2]$$

ricordando che la varianza di una variabile aleatoria standard coincide con la media del quadrato della variabile, otteniamo

$$\mathbf{E}[Y] = \mathbf{Var}X_1 + \mathbf{Var}X_2 + \dots + \mathbf{Var}X_n = n$$

(4.3) Siano $X \sim N(0, 1)$ e $Y \sim \chi^2(n)$ due variabili aleatorie indipendenti. Poniamo

$$Z = \sqrt{n} \frac{X}{\sqrt{Y}}$$

La densità di Z si chiama **densità t di Student a n gradi di libertà**: la indicheremo con $t(n)$.

Per i nostri scopi ci basterà sapere che si tratta di una densità pari e che per $n \rightarrow \infty$ il grafico della densità $t(n)$ tende a diventare quello della gaussiana standard.

Le tre densità $N(0, 1)$, $\chi^2(n)$ e $t(n)$ sono collegate da un importante teorema.

(4.4) **Teorema di Cochran** – Siano X_1, X_2, \dots, X_n n variabili aleatorie indipendenti, tutte con legge $N(\mu, \sigma^2)$. Indicheremo al solito con \bar{X} la loro media campionaria e porremo inoltre

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

Consideriamo le due variabili aleatorie

$$Z = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

e

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

Allora Z ha densità $N(0, 1)$, W ha densità $\chi^2(n-1)$ ed inoltre Z e W sono tra loro indipendenti.

(4.5) **Corollario** – Nelle ipotesi e con le notazioni del teorema precedente la variabile aleatoria

$$T = \sqrt{n-1} \frac{Z}{\sqrt{W}}$$

ha densità $t(n-1)$. La variabile T sopra definita si può scrivere anche in una forma che risulterà utile in statistica. Si ha

$$T = \sqrt{n-1} \frac{Z}{\sqrt{W}} = \sqrt{n-1} \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \frac{\sigma}{S \sqrt{n-1}} = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

dove con S si è indicata la radice quadrata aritmetica di S^2 .

5. La legge dei grandi numeri

Sullo spazio di probabilità (Ω, A, P) siano $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie, X una variabile aleatoria.

(5.1) Si dice che (X_n) converge *quasi certamente* verso X se

$$(i) \quad \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} \in A$$

$$(ii) \quad P(\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)) = 1$$

Vale il seguente teorema

(5.2) **Teorema (Legge forte dei grandi numeri)** – Sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie i.i.d., con media μ e varianza σ^2 finite. Allora la successione delle medie campionarie

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

converge quasi certamente a μ per $n \rightarrow \infty$.

6. Riepilogo delle principali densità

Nome	Funzione di ripartizione	Densità	Legge
Uniforme	$F(t) = \begin{cases} 0 & \text{se } t < a \\ \frac{t-a}{b-a} & \text{se } a \leq t \leq b \\ 1 & \text{se } t > b \end{cases}$	$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$	$X \sim U(a, b)$
Esponenziale	$F(t) = \begin{cases} 0 & \text{per } t \leq 0 \\ 1 - e^{-\lambda t} & \text{per } t > 0 \end{cases}$	$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \lambda e^{-\lambda x} & \text{se } x \geq 0 \end{cases}$	$X \sim \mathcal{E}(\lambda)$
Weibull	–	$f(x) = \begin{cases} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$	$X \sim W(\lambda, \alpha)$
Normale	–	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}$	$X \sim N(\mu, \sigma^2)$
Normale standard	$F(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad t \in \mathbb{R}$	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$X \sim N(0, 1)$

Nome	Speranza	Varianza
Uniforme	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Esponenziale	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Weibull	–	–
Normale	μ	σ^2
Normale standard	0	1