Statistica e Laboratorio

5. Calcolo delle probabilità: variabili casuali

Paolo Vidoni

Dipartimento di Scienze Economiche e Statistiche
Università di Udine
via Tomadini 30/a - Udine
paolo.vidoni@uniud.it

https://elearning.uniud.it/

Sommario

Sommario

Variabili casuali

Per descrivere fenomeni o esperimenti aleatori si considera la nozione di variabile casuale, che fornisce un modello matematico utile anche per le applicazioni statistiche.

Lo spazio fondamentale Ω potrebbe non essere un insieme numerico oppure potrebbe non rappresentare in modo chiaro gli aspetti dell'esperimento a cui si è interessati. In alcuni casi Ω potrebbe essere astratto e molto complesso da specificare.

Le variabili casuali permettono di svincolarsi dallo spazio Ω e di operare in insiemi numerici dove le probabilità si calcolano mediante somme o integrali.

Esempio. L'esperimento del lancio di una moneta non da luogo ad un risultato numerico, poiché $\Omega = \{T, C\}$. Se si considera il numero di esiti "Testa" in $n = 1$ lanci, si ottiene una descrizione numerica del fenomeno in esame.

In una procedura di controllo della qualità si può non essere interessati all'esito completo dell'esperimento ma soltanto al numero di oggetti che soddisfano ad opportuni standard di qualità, tra quelli selezionati.

Nel lancio di due dadi, si può prestare attenzione non tanto alla coppia di valori che appaiono sulle facce superiori dei singoli dadi, ma alla somma di tali valori numerici. $\diamondsuit$

Dato un esperimento (fenomeno) aleatorio descritto da uno spazio fondamentale Ω e una probabilità P, si definisce **variabile casuale** (**aleatoria**) X una *applicazione* da Ω in $\mathbf{R}$ *misurabile*, cioè tale che sia possibile "probabilizzare" gli eventi ad essa riferiti.

Quindi una variabile casuale è una funzione che, a seconda del risultato dell'esperimento in esame, assume valori numerici a cui possibile attribuire una certa probabilità di realizzazione coerente con P.

Esempio. *Moneta.* Si consideri l'esperimento che consiste nel lanciare tre volte una moneta regolare e si supponga di essere interessati al numero totale degli esiti testa.

Quindi $\Omega = \{CCC, CCT, CTC, TCC, CTT, TCT, TTC, TTT\}$ e la variabile casuale $X : \Omega \to R$ associa ad ogni evento elementare di Ω il numero di esiti T. Ad esempio, $X(TTC) = 2$.

X assume valori in $\{0, 1, 2, 3\}$ e tali valori corrispondono a veri e propri eventi elementari, indicati con la scrittura simbolica $X = i$, $i = 0, 1, 2, 3$.

È immediato concludere che $X = i$ ha probabilità $1/8$, se $i = 0, 3$, e $3/8$, se $i = 1, 2$. Si noti che la somma delle probabilità riferite agli esiti di X è pari a 1. $\diamondsuit$

Sommario

Funzione di ripartizione

La misura di probabilità riferita agli eventi $X \in B$, $B \subseteq \mathbf{R}$, associati alla variabile casuale X, soddisfa gli assiomi di Kolmogorov ed è detta **distribuzione** (**legge**) di **probabilità** di X.

In genere, non si fa menzione dello spazio di partenza e si identifica una variabile casuale X con la sua distribuzione di probabilità.

Due variabili casuali X e Y sono dette **identicamente distribuite**, in simboli $X \sim Y$, se $P(X \in B) = P(Y \in B)$, per ogni $B \subseteq \mathbf{R}$.

Per specificare la distribuzione di probabilità di una variabile casuale X si considera la nozione di **funzione di ripartizione**, intesa come un'applicazione $F_X : \mathbf{R} \to [0, 1]$, tale che

$$F_X(x) = P(X \leq x), \quad x \in \mathbf{R}.$$

La conoscenza di F_X permette di calcolare, eventualmente con procedimenti di limite, tutte le probabilità $P(X \in B)$, $B \subseteq \mathbf{R}$.

In particolare, per ogni $a, b \in \mathbf{R}$, $a < b$,

$$P(a < X \leq b) = F_X(b) - F_X(a), \quad P(X > a) = 1 - F_X(a),$$

$$P(X = b) = F_X(b) - \lim_{x \to b^-} F_X(x).$$

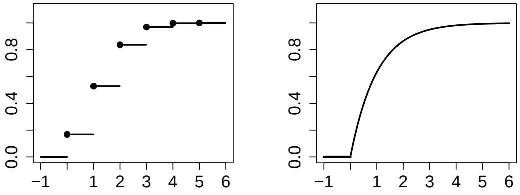
La funzione di ripartizione verifica le tre seguenti **proprietà caratterizzanti**:

- F_X è monotona non decrescente;
- F_X è continua da destra;
- F_X è tale che $\lim_{x \to -\infty} F_X(x) = 0$ e $\lim_{x \to +\infty} F_X(x) = 1$.

Perciò, F_X non è necessariamente continua anche da sinistra e quindi continua in ogni punto.

Si può dimostrare che F_X è continua nei punti in cui $P(X = x) = 0$ e discontinua nei punti in cui $P(X = x) > 0$, che sono al più un'infinità numerabile.

Vengono riportati due esempi di funzioni di ripartizione.



L'insieme di tutti i possibili valori della variabile casuale X corrisponde usualmente alla nozione di supporto.

Il **supporto** di X, indicato con S_X, è l'insieme dei punti $x \in \mathbf{R}$ i cui intorni sono eventi di probabilità strettamente positiva, cioè

$$S_X = \{x \in \mathbf{R} : \forall \varepsilon > 0, P(x - \varepsilon < X < x + \varepsilon) > 0\}.$$

Esempio. *Moneta* (continua). Si considera il lancio della moneta ripetuto per tre volte. In questo caso, $S_X = \{0, 1, 2, 3\}$ e $P(X = 0) = P(X = 3) = 1/8$, $P(X = 1) = P(X = 2) = 3/8$. $\Diamond$

Sommario

Funzione di probabilità

Tra le varie tipologie di variabili casuali si considerano quelle discrete, che possono assumere un numero finito o al più numerabile di valori, e quelle continue, che assumono valori in un insieme continuo.

Più precisamente, una **variabile casuale** X è **discreta** se esiste un insieme di numeri reali $\{x_i\}_{i \in I}$, **finito o al più numerabile**, tale che $P(X = x_i) = p_i > 0$ e $\sum_{i \in I} p_i = 1$; usualmente, $S_X = \{x_i, i \in I\}$.

La corrispondenza tra i possibili valori di X e le rispettive probabilità individua la **funzione di probabilità (massa)**

$$f_X(x) = \begin{cases} P(X = x_i) = p_i & \text{se } x = x_i, \quad \forall i \in I, \\ 0 & \text{altrimenti.} \end{cases}$$

Dalla conoscenza di f_X si risale facilmente alla funzione di ripartizione F_X e viceversa, quindi f_X caratterizza la variabile casuale X.

Infatti, per ogni $x \in \mathbf{R}$,

$$F_X(x) = P(X \leq x) = \sum_{i:\, x_i \leq x} p_i.$$

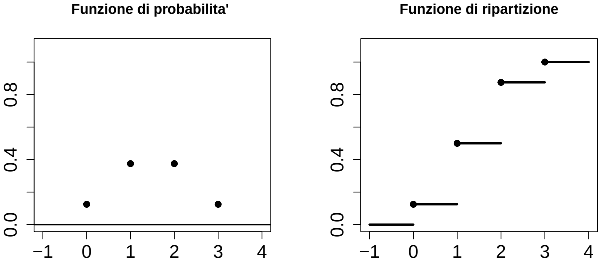
Il grafico di $F_X(x)$ è una *funzione a gradini*, continua da destra, con salti in corrispondenza degli elementi del supporto $x_i \in S_X$ e ampiezza del salto data da

$$p_i = f_X(x_i) = F_X(x_i) - F_X(x_{i-1}).$$

La conoscenza di f_X permette spesso una notevole semplificazione nel calcolo di probabilità di eventi relativi a X, dal momento che, per ogni $B \subseteq \mathbf{R}$,

$$P(X \in B) = \sum_{i\,:\, x_i \in B} f_X(x_i).$$

Esempio. *Moneta* (continua). Si considera la variabile casuale X, che conta il numero di esiti testa in tre lanci di una moneta regolare, e si rappresentano le associate funzioni di probabilità e di ripartizione



Inoltre, si calcolano facilmente probabilità riferite a X, come ad esempio

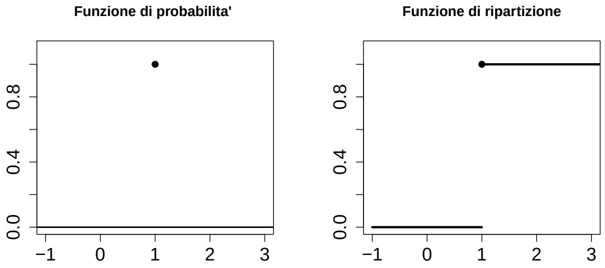
$$P(X \geq 1) = \sum_{i\,:\,x_i \geq 1} P(X = x_i) = 1 - P(X = 0) = 7/8,$$
$$P(X < 2) = \sum_{i\,:\,x_i < 2} P(X = x_i) = F_X(1) = 1/2.$$

$\diamondsuit$

Esempio. *Variabile casuale degenere.* Una **variabile casuale** X è **degenere** nel punto $c \in \mathbf{R}$, in simboli $X \sim D(c)$, se $P(X = c) = 1$. In questo caso $S_X = \{c\}$ e le funzioni di probabilità e di ripartizione corrispondono a

$$f_X(x) = \begin{cases} 1 & \text{se } x = c \\ 0 & \text{se } x \neq c, \end{cases} \qquad F_X(x) = \begin{cases} 0 & \text{se } x < c \\ 1 & \text{se } x \geq c, \end{cases}$$

con grafico, per il caso $c = 1$,

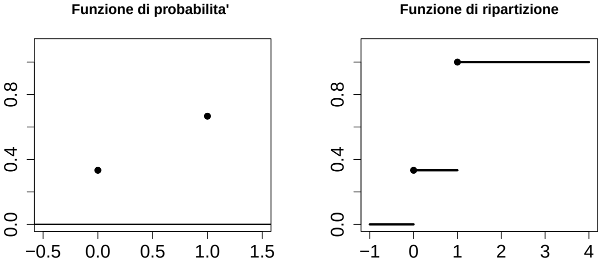


Una variabile casuale degenere descrive un esperimento non aleatorio. ◊

Esempio. *Variabile casuale Bernoulliana.* Una **variabile casuale** X è **Bernoulliana**, in simboli $X \sim Ber(p)$, con $p \in (0,1)$, se $S_X = \{0,1\}$ e $P(X=1)=p$, $P(X=0)=1-p$. Si ha che

$$f_X(x) = \begin{cases} 1-p & \text{se } x=0 \\ p & \text{se } x=1 \\ 0 & \text{altrimenti,} \end{cases} \qquad F_X(x) = \begin{cases} 0 & \text{se } x<0 \\ 1-p & \text{se } 0 \le x < 1 \\ 1 & \text{se } x \ge 1, \end{cases}$$

con grafico, per il caso $p = 2/3$, rispettivamente

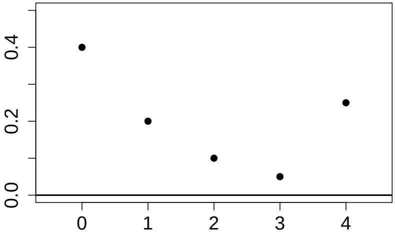


Esperimento aleatorio dicotomico, cioè con due possibili esiti, ad esempio, successo e insuccesso, quantificati in 1 e 0. $\Diamond$

Esempio. *Interruzioni.* Sia X una variabile casuale discreta che descrive il numero di interruzioni registrate in una linea di produzione di una certa azienda in una settimana. La sua funzione di probabilità è

$$f_X(x) = \begin{cases} 0.4 & \text{se } x = 0 \\ 0.2 & \text{se } x = 1 \\ 0.1 & \text{se } x = 2 \\ 0.05 & \text{se } x = 3 \\ 0.25 & \text{se } x = 4 \\ 0 & \text{altrimenti.} \end{cases}$$

con grafico



◇

Sommario

Funzione di densità

Una **variabile casuale** X è **continua** se la sua **funzione di ripartizione** F_X è **continua** ed è tale che esiste una funzione f_X, definita su $\mathbf{R}$, tale che

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad \forall x \in \mathbf{R}.$$

La f_X è chiamata **funzione di densità probabilità** ed è tale che

- $f_X(x) \geq 0$, per ogni $x \in \mathbf{R}$;
- $\int_{-\infty}^{+\infty} f_X(x)dx = 1$;
- $f_X(x) = \frac{d}{dx}F_X(x)$, per ogni $x \in \mathbf{R}$ in cui $f_X(x)$ continua.

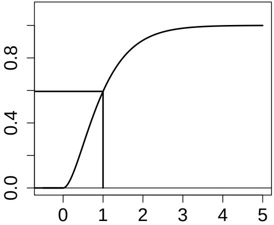
Quindi dalla conoscenza di f_X si ottiene F_X e viceversa; f_X caratterizza la variabile casuale X.

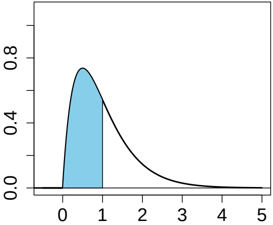
Il supporto S_X è un insieme continuo, ad esempio $\mathbf{R}$ o un intervallo o una semiretta di $\mathbf{R}$.

Invece che assegnare probabilità a valori puntuali (si ricordi che, essendo F_X continua, $P(X = x) = 0$, per ogni $x \in \mathbf{R}$), si assegna probabilità agli intervalli, semirette, ecc. di $\mathbf{R}$.

Gli eventi $(X < a)$ e $(x \leq a)$, $a \in \mathbf{R}$, hanno la stessa probabilità.

Il valore della funzione di ripartizione in $x = 1$, $F_X(1)$, (grafico di sinistra) corrisponde all'area sottesa dalla funzione di densità con riferimento a $(-\infty, 1]$ (grafico di destra).





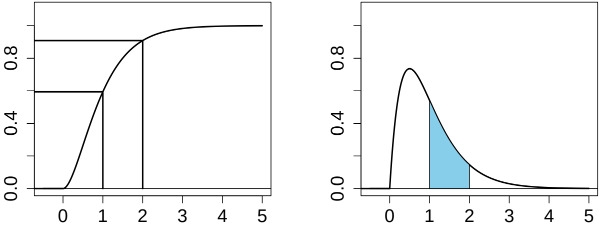
Inoltre, come conseguenza dei risultati di probabilità elementare,

$$P(X > a) = 1 - P(X \leq a) = 1 - F_X(a), \text{ per ogni } a \in \mathbf{R}.$$

Vale il seguente risultato: per ogni $a, b \in \mathbf{R}$, $a < b$,

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx,$$

che corrisponde all'area sottesa dalla funzione di densità con riferimento all'intervallo $[a, b]$. Graficamente, se $[a, b] = [1, 2]$,

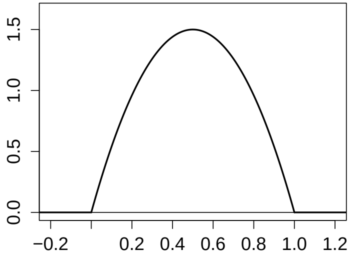


In generale, la probabilità associata all'evento $X \in B$ corrisponde a

$$P(X \in B) = \int_B f_X(x)dx.$$

Si noti che f_X non definisce la probabilità associata all'evento $X = x$, che risulta essere nulla, ma è direttamente proporzionale alla probabilità che X assuma valori in un intorno di x.

Esempio. *Internet.* Una compagnia telefonica ha riscontrato che la durata, in un'ora, dei collegamenti internet dei propri utenti è descritta da una variabile casuale continua X con funzione di densità $f_X(x) = 6x(1-x)$, se $x \in [0, 1]$, e nulla altrove



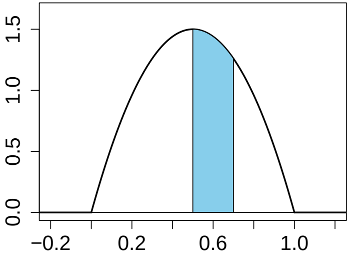
Si verifica facilmente che $\int_{-\infty}^{+\infty} f_X(x)dx = 1$ e che f_X è non negativa.

La funzione di ripartizione è tale che per $x \in [0,1]$, $F_X(x) = 3x^2 - 2x^3$, mentre, se $x < 0$, $F_X(x) = 0$ e, se $x > 1$, $F_X(x) = 1$.

La probabilità che X assuma valori in $[0.5, 0.7]$ è

$$P(0.5 \leq X \leq 0.7) = F_X(0.7) - F_X(0.5) = \int_{0.5}^{0.7} 6x(1-x)dx = 0.284$$

e corrisponde all'area evidenziata nel grafico sottostante



◊

Esempio. *Variabile casuale esponenziale.* Una **variabile casuale** X è **esponenziale**, in simboli $X \sim Esp(\lambda)$, con $\lambda > 0$, se $S_X = [0, +\infty)$ e

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \in S_X \\ 0 & \text{altrimenti.} \end{cases}$$

La funzione di ripartizione è

$$F_X(x) = \int_{-\infty}^{x} \lambda e^{-\lambda t} dt = \int_{0}^{x} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x},$$

se $x \in S_X$, mentre $F_X(x) = 0$, se $x \notin S_X$.

Si calcolano le probabilità

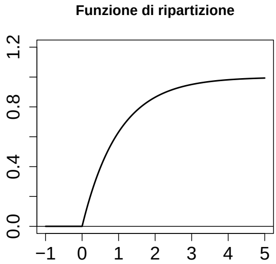
$$P(X > 1) = 1 - F_X(1) = e^{-\lambda},$$

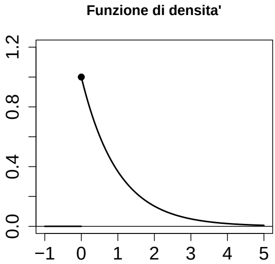
$$P(1 \leq X \leq 3) = F_X(3) - F_X(1) = e^{-\lambda} - e^{-3\lambda},$$

che, se $\lambda = 1$, corrispondono rispettivamente a e^{-1} e $e^{-1} - e^{-3}$.

La variabile casuale esponenziale viene utilizzata soprattutto per rappresentare durate e tempi di vita o di funzionamento, nel caso in cui si ipotizza assenza di memoria o di usura.

Si presentano, rispettivamente, i grafici della funzione di ripartizione e della funzione di densità nel caso $\lambda = 1$.





◊

Esempio. *Variabile casuale uniforme.* Una **variabile casuale continua** X è **uniforme** in $[0, 1]$, in simboli $X \sim U(0, 1)$, se $S_X = [0, 1]$ e

$$f_X(x) = \begin{cases} 1 & \text{se } x \in S_X \\ 0 & \text{altrimenti.} \end{cases}$$

La funzione di ripartizione è

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1, \end{cases}$$

Si noti che, se gli intervalli $[a, b]$ e $[c, d]$ del supporto, con $a < b$ e $c < d$, hanno uguale ampiezza h, allora

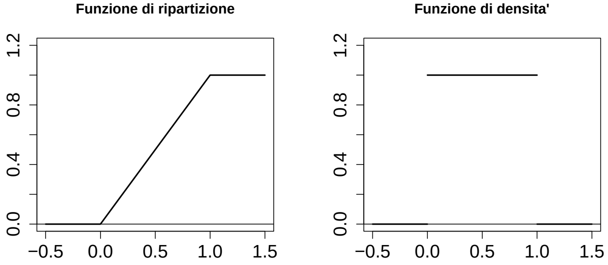
$$P(a \leq X \leq b) = P(c \leq X \leq d) = h \cdot 1 = h.$$

Dunque, tutti gli intervalli del supporto di uguale lunghezza hanno la stessa probabilità di contenere un valore di X.

La variabile casuale uniforme continua viene utilizzata per esperimenti aleatori che possono essere rappresentati come un'estrazione casuale di un numero da un certo intervallo di $\mathbf{R}$.

È un modello che descrive l'equiprobabilità nel continuo.

Si presentano, rispettivamente, i grafici della funzione di ripartizione e della funzione di densità.



Sommario

Indici sintetici per variabili casuali

La distribuzione di probabilità di una variabile casuale X viene descritta in modo completo dalla associata funzione di ripartizione o dalla corrispondente funzione (di densità) di probabilità.

Nonostante ciò, spesso si interessati a conoscere soltanto alcuni aspetti parziali della distribuzione di probabilità di X, quali

- la **posizione**, cioè il *centro* della distribuzione di probabilità;
- la **variabilità**, cioè la *dispersione* della distribuzione di probabilità attorno ad un centro;
- la forma della distribuzione di probabilità, considerando la **simmetria** e la **curtosi**.

Si riprendono sostanzialmente gli stessi concetti presentati in Statistica descrittiva, modificando il contesto di applicazione e gli elementi interpretativi.

Indici di posizione: valore atteso

Data una variabile casuale discreta o continua X, con supporto S_X e funzione (di densità) di probabilità f_X, si chiama **valore atteso (medio)** o **media** di X, in simboli $E(X)$, la media dei suoi possibili valori ponderati con le relative probabilità (la relativa funzione di densità di probabilità), ovvero

$$E(X) = \sum_{x \in S_X} x f_X(x) = \sum_{x \in S_X} x P(X = x), \quad \text{se } X \text{ è } \textbf{discreta},$$

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx, \quad \text{se } X \text{ è } \textbf{continua},$$

purché la serie o l'integrale siano convergenti.

È l'indice di posizione più noto. Usualmente si pone $E(X) = \mu$ e si intende tacitamente che tale valore atteso esista finito.

Esempio. *Voti*. La seguente tabella di frequenza sintetizza i voti ottenuti da 30 alunni in un compito in classe.

voto	4	5	6	7	8
no. alunni	2	3	10	11	4

Si può calcolare la media aritmetica (statistica descrittiva) che corrisponde a 6.4.

Si supponga di avere un'urna con 30 palline, ciascuna contenente il voto di un alunno, e si estragga a caso una pallina.

La variabile casuale X, che indica il voto ottenuto con l'estrazione, ha distribuzione di probabilità

x	4	5	6	7	8
$P(X = x)$	$2/30$	$3/30$	$10/30$	$11/30$	$4/30$

e valore atteso $E(X) = 6.4$ (calcolo delle probabilità). Il valore è lo stesso, ma l'interpretazione è evidentemente diversa. $\Diamond$

Esempio. *Gioco*. Si lancia una moneta che da testa con probabilità $p \in (0, 1)$; se esce testa Tizio paga a Caio un euro, se esce croce è Caio a dover dare a Tizio la stessa somma.

Indicata con X la variabile casuale che descrive il guadagno di Tizio, si ha che $E(X) = (-1)p + 1(1 - p) = 1 - 2p$.

Quindi, $E(X)$ è positivo, nullo o negativo se, rispettivamente, $p < 1/2$, $p = 1/2$ (moneta regolare) o $p > 1/2$. $\diamondsuit$

Esempio. *Variabile casuale esponenziale* (continua). Si consideri la variabile casuale $X \sim Esp(\lambda)$. Poiché la funzione di densità è nulla fuori dal supporto $S_X = [0, +\infty)$,

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{+\infty} t e^{-t} dt = \frac{1}{\lambda},$$

avendo operato il cambio di variabile $t = \lambda x$ e poi integrato per parti. $\diamondsuit$

Esempio. *Internet* (continua). Si considera la variabile casuale X che misura la durata, in un'ora, dei collegamenti internet degli utenti di una certa compagnia telefonica. La funzione di densità di X è pari a $f_X(x) = 6x(1-x)$, se $x \in [0,1]$, e nulla altrove.

Poiché la funzione di densità è nulla fuori dal supporto $S_X = [0,1]$,

$$E(X) = \int_0^1 x6x(1-x)dx = \int_0^1 6x^2 - 6x^3 dx = \frac{1}{2}.$$

Si noti che, in questo caso, $f_X(x)$ è simmetrica rispetto a $x = 1/2$.

$\diamondsuit$

Esempio. *Variabile casuale uniforme* (continua). Si consideri la variabile casuale $X \sim U(0,1)$. Poiché la funzione di densità è nulla fuori dal supporto $S_X = [0,1]$,

$$E(X) = \int_0^1 x \cdot 1\, dx = \frac{1}{2}.$$

$\diamondsuit$

Sia X una variabile casuale e $Y = g(X)$ una variabile casuale ottenuta come trasformata della X, tramite l'applicazione $g(\cdot)$.

Nota la distribuzione di probabilità di X, si può calcolare il valore atteso di Y, ovvero $E(Y) = E(g(X))$, senza conoscere la legge di Y; infatti,

$$E(Y) = \sum_{x \in S_X} g(x) f_X(x), \quad \text{se } X \text{ e } Y \text{ sono } \textbf{discrete},$$

$$E(Y) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx, \quad \text{se } X \text{ e } Y \text{ sono } \textbf{continue}.$$

Sulla nozione di valore atteso si possono fare considerazioni analoghe a quelle fatte in statistica descrittiva con riferimento alla media aritmetica.

Valgono inoltre le seguenti proprietà, per le quali si omettono le dimostrazioni essendo sostanzialmente analoghe, per lo meno con riferimento al caso discreto, a quelle viste per la media aritmetica:

1) **Proprietà di Cauchy**: $\inf\{x \in S_X\} \leq E(X) \leq \sup\{x \in S_X\}$.
2) **Proprietà di baricentro**: $E(X - E(X)) = 0$.
3) **Proprietà di linearità**: $E(aX + b) = aE(X) + b$, per ogni $a, b \in \mathbf{R}$.

Inoltre, si può dimostrare che vale la seguente estensione della proprietà di linearità: date due variabili casuali X e Y, per ogni $a, b \in \mathbf{R}$,

$$E(aX + bY) = aE(X) + bE(Y);$$

tale proprietà si può estendere anche al caso di combinazioni lineari di più di due variabili casuali.

Oltre al valore atteso esistono altri indici di posizione. Tra questi verranno ricordati la mediana e la moda.

Indici di posizione: mediana

La **mediana** della distribuzione di probabilità di X, o più semplicemente la mediana di X, indicata con $x_{0.5}$, è quel valore $x_{0.5} \in \mathbf{R}$ tale che

$$P(X \leq x_{0.5}) \geq 1/2 \quad \text{e} \quad P(X \geq x_{0.5}) \geq 1/2.$$

Quindi, $x_{0.5}$ ripartisce la massa unitaria di probabilità, di modo che gli eventi $X \leq x_{0.5}$ e $X \geq x_{0.5}$ abbiano probabilità pari a $1/2$, o anche maggiore di $1/2$ se $P(X = x_{0.5}) > 0$.

Può non essere unica e, in alcuni casi, può corrispondere anche ad un intervallo di valori reali.

Se X è una **continua**, la mediana $x_{0.5}$ è tale che

$$F_X(x_{0.5}) = 1/2.$$

È il valore dove la funzione di ripartizione vale $1/2$ e che ripartisce a metà l'area unitaria sottesa dalla funzione di densità.

Se X è una **discreta**, la mediana $x_{0.5}$ è quel valore (anche più di uno) dove la funzione di ripartizione raggiunge o supera per la prima volta $1/2$.

Esempio. *Moneta* (continua). Si consideri la variabile casuale X che conta gli esiti testa in tre lanci di una moneta regolare.

Le condizioni $P(X \leq x_{0.5}) \geq 1/2$ e $P(X \geq x_{0.5}) \geq 1/2$ risultano verificate per $x_{0.5} = 1$, $x_{0.5} = 2$ e per ogni valore reale in $(1, 2)$.

La variabile casuale X presenta come mediana tutti i valori dell'intervallo $[1, 2]$. La mediana convenzionale è 1.5. $\diamondsuit$

Esempio. Sia X una variabile casuale tale che $S_X = \{-2, 0, 1, 2\}$, $P(X = -2) = P(X = 2) = 1/4$, $P(X = 0) = 1/6$ e $P(X = 1) = 1/3$. Si ha allora che

$$P(X \leq 1) = P(X = -2) + P(X = 0) + P(X = 1) > 1/2,$$

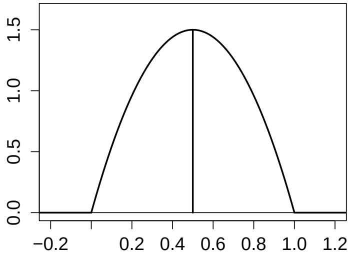
$$P(X \geq 1) = P(X = 1) + P(X = 2) > 1/2.$$

Soltanto il valore $x_{0.5} = 1$ soddisfa le due condizioni della definizione ed è quindi la mediana di X. $\diamondsuit$

Esempio. *Variabile casuale esponenziale* (continua). Si considera la variabile casuale $X \sim Esp(\lambda)$, che ha funzione di ripartizione $F_X(x) = 1 - e^{-\lambda x}$, per $x \geq 0$, e nulla altrove, con $\lambda > 0$.

La mediana di X si ottiene risolvendo l'equazione $1 - e^{-\lambda x_{0.5}} = 1/2$. In particolare, si ha che $x_{0.5} = \lambda^{-1} \log 2$. $\lozenge$

Esempio. *Internet* (continua). Si considera la variabile casuale X che misura la durata, in un'ora, dei collegamenti internet. La funzione di densità di X è rappresentata nella figura sottostante.



Poiché è simmetrica rispetto a $x = 1/2$, si ha che $x_{0.5} = 1/2$. $\lozenge$

Indici di posizione: moda

La **moda** della distribuzione di probabilità di X, o più semplicemente la moda di X, indicata con x_{mo}, è quel valore $x_{mo} \in \mathbf{R}$ per cui è massima la funzione (di densità) di probabilità.

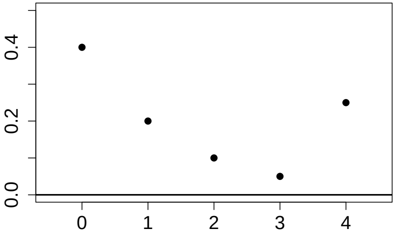
La moda non è necessariamente unica e può anche non esistere. Se esiste, appartiene al supporto S_X e individua i valori più probabili, se X discreta, o i cui intorni sono gli eventi più probabili, se X continua.

Nel caso in cui $f_X(x)$ ha un unico massimo, la **distribuzione di probabilità** di X è detta **unimodale**; se ci sono due o più punti di massimo, si parla di **distribuzioni bimodali o multimodali**.

Esempio. *Internet* (continua). Si considera la variabile casuale X che misura la durata, in un'ora, dei collegamenti internet. Dalla analisi del grafico della funzione di densità si conclude che $x_{mo} = 1/2$. $\Diamond$

Esempio. *Interruzioni* (continua). Sia X una variabile casuale discreta che descrive il numero di interruzioni registrate in una certa linea di produzione.

Dalla analisi della funzione di massa si conclude che $x_{mo} = 0$.



$\Diamond$

Esempio. *Variabile casuale esponenziale* (continua). Si considera la variabile casuale $X \sim Esp(\lambda)$. Dalla analisi del grafico della funzione di densità si conclude che $x_{mo} = 0$. $\Diamond$

Esempio. *Variabile casuale uniforme* (continua). Si consideri la variabile casuale $X \sim U(0, 1)$. Dalla analisi del grafico della funzione di densità si conclude che la moda x_{mo} corrisponde ad ogni punto dell'intervallo $S_X = [0, 1]$. $\Diamond$

Quantili

Sia $\alpha \in (0, 1)$, si chiama **quantile di livello** α della distribuzione di probabilità di X, o più semplicemente quantile di livello α di X, indicato con x_α, quel valore $x_\alpha \in \mathbf{R}$ tale che

$$P(X \leq x_\alpha) \geq \alpha \quad \text{e} \quad P(X \geq x_\alpha) \geq 1 - \alpha.$$

Quindi, a meno di effetti legati alla discretezza, x_α ripartisce la massa unitaria di probabilità lasciando una porzione pari ad α alla propria sinistra e pari a $1 - \alpha$ alla propria destra.

Può non essere unico e, in alcuni casi, può corrispondere anche ad un intervallo di valori reali.

Se X è una **continua**, x_α è tale che

$$F_X(x_\alpha) = \alpha.$$

È il valore dove la funzione di ripartizione vale α e che ripartisce in due porzioni pari ad α e $1 - \alpha$ l'area unitaria sottesa dalla funzione di densità.

Se X è una **discreta**, il quantile x_α è quel valore (anche più di uno) dove la funzione di ripartizione raggiunge o supera per la prima volta α.

Quindi la mediana corrisponde al quantile di livello $\alpha = 1/2$. Se α è espresso in termini decimali o percentuali e si parla allora di **decili** o di **percentili**. Se $\alpha = 1/4, 1/2, 3/4$, si hanno i **quartili**.

Esempio. Sia X una variabile casuale tale che $S_X = \{-2, 0, 1, 2\}$, $P(X = -2) = P(X = 2) = 1/4$, $P(X = 0) = 1/6$ e $P(X = 1) = 1/3$. Si cerca il quantile di livello $\alpha = 0.4$. Si ha che

$$P(X \leq 0) = P(X = -2) + P(X = 0) > 0.4,$$

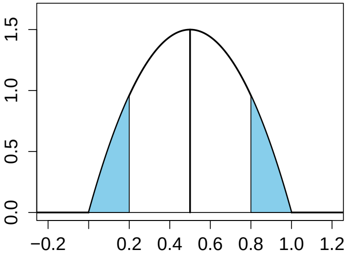
$$P(X \geq 0) = P(X = 0) + P(X = 1) + P(X = 2) > 0.6.$$

Poiché soltanto il valore 0 soddisfa le due condizioni della definizione, si conclude che $x_{0.4} = 0$. $\diamondsuit$

Esempio. *Variabile casuale esponenziale* (continua). Si considera la variabile casuale $X \sim Esp(\lambda)$, che ha funzione di ripartizione $F_X(x) = 1 - e^{-\lambda x}$, per $x \geq 0$, e nulla altrove, con $\lambda > 0$.

Il quantile x_α si ottiene risolvendo l'equazione $1 - e^{-\lambda x_\alpha} = \alpha$. In particolare, si ha che $c = -\lambda^{-1} \log(1 - \alpha)$. $\Diamond$

Esempio. *Internet* (continua). Si considera la variabile casuale X che misura la durata, in un'ora, dei collegamenti internet degli utenti di una certa compagnia telefonica. La funzione di densità di X è rappresentata nella figura sottostante.



Poiché è simmetrica rispetto a $x = 1/2$, si può concludere che, per ogni $\alpha \in (0, 0.5)$, l'area della coda alla sinistra di x_α coincide con l'area della coda alla destra di $x_{1-\alpha}$. $\Diamond$

Indici di variabilità: varianza

Data una variabile casuale discreta o continua X, con supporto S_X e funzione (di densità) di probabilità f_X, si chiama **varianza** di X, in simboli $V(X)$, la quantità

$$V(X) = E((X - E(X))^2),$$

se esiste finita, ovvero

$$V(X) = \sum_{x \in S_X} (x - E(X))^2 f_X(x), \quad \text{se } X \text{ è } \textbf{discreta},$$

$$V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x) dx, \quad \text{se } X \text{ è } \textbf{continua},$$

purché la serie o l'integrale siano convergenti.

È l'indice di variabilità più noto. Usualmente si pone $V(X) = \sigma^2$ e si intende tacitamente che il valore atteso della definizione esista finito.

La varianza è il valore atteso della variabile casuale scarto $X - E(X)$ elevata al quadrato e misura la dispersione distribuzione di probabilità attorno alla media.

Lo **scarto quadratico medio** di X, indicato con σ, è la radice quadrata aritmetica (l'unica positiva) della varianza, $\sigma = \sqrt{V(X)}$.

Valgono inoltre le seguenti proprietà, per le quali si omettono le dimostrazioni essendo sostanzialmente analoghe a quelle viste per la varianza in statistica descrittiva:

1) **Proprietà di non negatività**: $V(X) \geq 0$, con $V(X) = 0$ se e solo se X è degenere.
2) **Formula per il calcolo**: $V(X) = E(X^2) - (E(X))^2$.
3) **Proprietà di invarianza per traslazioni**: $V(X + b) = V(X)$, $b \in \mathbf{R}$.
4) **Proprietà di omogeneità di secondo grado**: $V(aX) = a^2 V(X)$, $a \in \mathbf{R}$.

Dalle proprietà 3) e 4) discende che $V(aX+b) = a^2V(X)$, con $a, b \in \mathbf{R}$.

Inoltre, data una variabile casuale X, con media $\mu = E(X)$ e varianza $\sigma^2 = V(X)$, la variabile casuale trasformata

$$Y = \frac{X - \mu}{\sigma}$$

è tale che $E(Y) = 0$ e $V(Y) = 1$ ed è detta **variabile casuale standardizzata**.

Viceversa, a partire da una variabile casuale Y, con $E(Y) = 0$ e $V(Y) = 1$, utilizzando la trasformata

$$X = \sigma Y + \mu$$

si ottiene una variabile casuale X con prefissati valor medio μ e varianza σ^2.

Ulteriori indici di variabilità

Se X è **positiva**, più precisamente se $P(X > 0) = 1$, si può definire la quantità σ/μ, chiamata **coefficiente di variazione**.

Nel caso di variabili casuali non necessariamente positive, il coefficiente di variazione si definisce come $\sigma/|\mu|$.

Poiché non dipende dalla unità di misura con cui viene studiato il fenomeno, può risultare utile per confrontare la dispersione di due o più variabili casuali.

Lo **scarto medio assoluto dalla mediana**, definito come $E(|X - x_{0.5}|)$, se esiste finito, esprime la distanza attesa tra i valori di X e la mediana $x_{0.5}$.

Lo **scarto interquartilico** $SI = x_{3/4} - x_{1/4}$, corrisponde alla differenza tra il terzo e il primo quartile.

Il **campo di variazione (range)** $R = \sup\{x \in S_X\} - \inf\{x \in S_X\}$, corrisponde sostanzialmente alla differenza tra il valore più grande e più piccolo del supporto.

Esempio. *Moneta* (continua). Si considera la variabile casuale X che conta il numero di esiti testa in tre lanci di una moneta regolare. In questo caso, $S_X = \{0, 1, 2, 3\}$ e $P(X = 0) = P(X = 3) = 1/8$, $P(X = 1) = P(X = 2) = 3/8$ ed è facile verificare che

$$E(X) = 0 + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = \frac{3}{2},$$

$$E(X^2) = 0 + 1\frac{3}{8} + 4\frac{3}{8} + 9\frac{1}{8} = 3.$$

Con la regola per il calcolo, si ha $V(X) = 3 - (3/2)^2 = 3/4$. $\Diamond$

Esempio. *Variabile casuale uniforme* (continua). Si consideri la variabile casuale $X \sim U(0, 1)$. Poiché $E(X) = 1/2$ e

$$E(X^2) = \int_0^1 x^2 \cdot 1\, dx = \frac{1}{3},$$

si conclude che $V(X) = 1/3 - (1/2)^2 = 1/12$. $\Diamond$

Esempio. *Variabile casuale esponenziale* (continua). Si consideri la variabile casuale $X \sim Esp(\lambda)$. Poiché $E(X) = 1/\lambda$ e, integrando per parti,

$$E(X^2) = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = \frac{2}{\lambda} \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \frac{2}{\lambda^2},$$

si conclude che $V(X) = 2/\lambda^2 - (1/\lambda)^2 = 1/\lambda^2$. $\Diamond$

Esempio. *Internet* (continua). Si considera la variabile casuale X che misura la durata, in un'ora, dei collegamenti internet degli utenti di una certa compagnia telefonica. Poiché $E(X) = 1/2$ e

$$E(X^2) = \int_0^1 x^2 6x(1-x) dx = \int_0^1 6x^3 - 6x^2 dx = \frac{3}{10},$$

si conclude che $V(X) = 3/10 - (1/2)^2 = 1/20$. $\Diamond$

Esempio. *Costante di normalizzazione.* Si consideri la variabile casuale continua X con funzione di densità di probabilità

$$f_X(x) = \begin{cases} k\,x^2 & \text{se } x \in [0,2] \\ 0 & \text{altrimenti,} \end{cases}$$

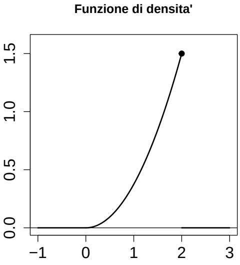
con k una opportuna costante reale. Si vuole determinare k di modo che $f_X(x) \geq 0$, per ogni $x \in \mathbf{R}$, e che

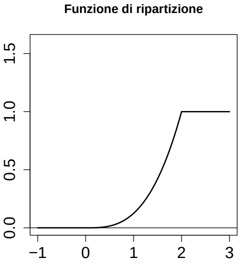
$$\int_{-\infty}^{+\infty} f_X(x)dx = \int_0^2 k\,x^2 dx = 1;$$

pertanto che $k = 3/8$. La funzione di ripartizione è tale che, per $x \in [0,2]$,

$$F_X(x) = \frac{3}{8}\int_0^x t^2 dt = \frac{x^3}{8},$$

mentre è nulla per $x < 0$ e vale 1 per $x > 2$. Si riportano i grafici.





Come si vede dal grafico di sinistra, $x_{mo} = 2$, mentre valore atteso è

$$E(X) = \frac{3}{8}\int_0^2 x^3 dx = \frac{3}{2}.$$

La mediana è il valore $x_{0.5} \in \mathbf{R}$ tale che $F_X(x_{0.5}) = x_{0.5}^3/8 = 1/2$, quindi $x_{0.5} = \sqrt[3]{4}$.

Infine, utilizzando la formula per il calcolo, si ottiene la varianza

$$V(X) = E(X^2) - (E(X))^2 = \frac{3}{8}\int_0^2 x^4 dx - \frac{9}{4} = \frac{3}{20}.$$

$\lozenge$

Indici di simmetria e di curtosi

In modo analogo a quanto considerato in statistica decrittiva, è possibile studiare la simmetria e la curtosi della funzione di densità (probabilità) di una variabile casuale X. L'indice **indice di simmetria** più utilizzato è

$$\gamma = \frac{E[(X - E(X))^3]}{\sigma^3},$$

dove $\sigma = \sqrt{V(X)}$ è lo scarto quadratico medio di X.

Se la funzione di densità (probabilità) è **simmetrica**, $\gamma = 0$; se c'è **asimmetria negativa**, $\gamma < 0$; se c'è **asimmetria positiva**, $\gamma > 0$.

L'indice **indice di curtosi** più utilizzato è

$$\beta = \frac{E[(X - E(X))^4]}{\sigma^4}.$$

Se la funzione di densità (probabilità) è **normocurtica**, $\beta = 3$; se è **leptocurtica** (*code pesanti*), $\beta > 3$; se è **platicurtica** (*code leggere*), $\beta < 3$.

Sommario

Variabili casuali bivariate

Nelle applicazioni, è assai frequente dover prendere in considerazione più di una variabile casuale contemporaneamente.

Dal punto di vista concettuale la trattazione è molto simile al caso univariato, tuttavia dal punto di vista matematico vi sono delle difficoltà aggiuntive, basti pensare che sia la funzione di ripartizione che quella di densità (probabilità) sono funzioni di più variabili.

Una **variabile casuale multivariata (vettore aleatorio)** $(X_1, \dots, X_n)$ è un vettore i cui elementi sono variabili casuali.

Esempio. *Campione bernoulliano.* Si svolgono n esperimenti bernoulliani indipendenti con la stessa probabilità di successo p. L'esperimento nel suo complesso è descritto da una variabile casuale multivariata $(X_1, \dots, X_n)$, dove $X_i \sim Ber(p)$, $i = 1, \dots, n$. Questa è la situazione che si presenta tipicamente nella statistica inferenziale. $\diamondsuit$

Si limita la trattazione alle **variabili casuali bivariate**, che tuttavia risulta sufficiente per introdurre tutti i concetti che sono importanti per il caso generale.

Una **variabile casuale bivariata** (X, Y) risulta specificata dalla sua **funzione di ripartizione congiunta**

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y), \quad (x, y) \in \mathbf{R}^2.$$

Inoltre, il **supporto congiunto** $S_{X,Y}$ è dato dall'insieme dei punti $(x, y) \in \mathbf{R}^2$ nei cui intorni si possono osservare valori per (X, Y) con probabilità strettamente positiva.

Dalla conoscenza della funzione di ripartizione congiunta si ottiene la **funzione di ripartizione marginale** delle due componenti X e Y; ad esempio, per la componente marginale X

$$F_X(x) = \lim_{y \to +\infty} F_{X,Y}(x, y), \quad x \in \mathbf{R}$$

e analogamente per $F_Y(y)$.

Si considera il **caso discreto**; quanto presentato si può estendere al caso continuo con opportune attenzioni.

Una **variabile casuale bivariata** (X, Y) è **discreta** se esiste un insieme di coppie di numeri reali $\{(x_i, y_j)\}_{(i,j)\in I\times J}$, *finito o al più numerabile*, tale che $P(X = x_i, Y = y_j) = p_{ij} > 0$ e $\sum_{i,j} p_{ij} = 1$; usualmente, $S_{X,Y} = \{(x_i, y_j), (i, j) \in I \times J\}$.

Analogamente al caso univariato, risulta definita la **funzione di probabilità (massa) congiunta**

$$f_{X,Y}(x, y) = \begin{cases} p_{ij} & \text{se } (x, y) = (x_i, y_j), \quad \forall (i, j) \in I \times J, \\ 0 & \text{altrimenti.} \end{cases}$$

Dalla conoscenza di $f_{X,Y}$ si risale facilmente alla funzione di ripartizione congiunta $F_{X,Y}$ e viceversa.

Quindi anche $f_{X,Y}$ caratterizza la variabile casuale bivariata (X, Y) e permette il calcolo delle probabilità di eventi ad essa associati.

$f_{X,Y}$ può essere rappresentata mediante una *tabella a doppia entrata*, analoga alle tabelle di contingenza, che fornisce le probabilità p_{ij} riferite alle coppie (x_i, y_j), $(i, j) \in I \times J$.

Una variabile casuale bivariata (X, Y) è un vettore che ha, come **componenti marginali**, le variabili casuali univariate X e Y.

In precedenza si è visto come determinare la **funzione di ripartizione marginale** di X e di Y a partire dalla funzione di ripartizione congiunta. Il **supporto marginale** S_X corrisponde, intuitivamente, a tutti i possibili valori della componente X; analogamente per S_Y.

Data una **variabile casuale bivariata discreta** (X, Y), con funzione di probabilità congiunta $f_{X,Y}$, si può ricavare facilmente la **funzione di probabilità marginale** di X, poiché per ogni $x_i \in S_X$

$$P(X = x_i) = \sum_{j \in J} P(X = x_i, Y = y_j) = \sum_{j \in J} p_{ij} = p_{i+}.$$

Analogamente, per la componente Y

$$P(Y = y_j) = \sum_{i \in I} p_{ij} = p_{+j}, \quad y_j \in S_Y.$$

Se si considera la rappresentazione di $f_{X,Y}$ mediante una tabella a doppia entrata, la funzione di probabilità marginale di X (di Y) si ottiene calcolando i totali di riga (di colonna).

A partire dalla distribuzione marginale delle due componenti si possono calcolare valore atteso e varianza, che vengono chiamati **valore atteso marginale** e **varianza marginale** di X e di Y.

Viene ora rappresentata in generale una tabella a doppia entrata con le probabilità congiunte e marginali associate a (X, Y).

	y_1	y_2	$\dots$	y_k	
x_1	p_{11}	p_{12}	$\dots$	p_{1k}	p_{1+}
x_2	p_{21}	p_{22}	$\dots$	p_{2k}	p_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\dots$	p_{mk}	p_{m+}
	p_{+1}	p_{+2}	$\dots$	p_{+k}	1

Condizionamento e indipendenza

Le componenti marginali X e Y sono indipendenti se ogni evento associato a X è indipendente da ogni evento associato a Y.

Formalmente, X e Y sono (**stocasticamente**) **indipendenti** se

$$F_{X,Y}(x,y) = F_X(x)F_Y(y), \quad \text{per ogni } (x,y) \in \mathbf{R}^2.$$

Se, invece, esiste almeno un punto (x,y) per cui questo non vale, X e Y vengono dette **dipendenti**. Se X e Y sono indipendenti, il supporto congiunto è il prodotto cartesiano dei supporti marginali, cioè $S_{X,Y} = S_X \times S_Y$.

Se (X,Y) è **discreta**, la definizione di indipendenza è equivalente a chiedere che, per ogni $(x_i, y_j) \in S_{X,Y}$,

$$f_{X,Y}(x_i, y_j) = f_X(x_i)f_Y(y_j).$$

Utilizzando la notazione introdotta in precedenza, ciò corrisponde a chiedere che, per ogni $(x_i, y_j) \in S_{X,Y}$, $p_{ij} = p_{i+}p_{+j}$.

Data una variabile casuale bivariata (X, Y), può essere interessante determinare la distribuzione di probabilità di una componente condizionatamente ai valori assunti dall'altra.

Se (X, Y) è **discreta**, in accordo con la definizione di probabilità condizionata, si ottiene la funzione di probabilità della **variabile casuale** X **condizionata a (dato)** $Y = y_j$, dove $P(Y = y_j) > 0$, in simboli $X \mid Y = y_j$.

In particolare, per ogni $x_i \in S_{X|Y=y_j}$, si ha

$$f_{X|Y=y_j}(x_i) = P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{+j}},$$

mentre la funzione è nulla altrove. $S_{X|Y=y_j}$ è il **supporto** della variabile casuale condizionata, definito come l'insieme dei valori che X può assumere se $Y = y_j$.

Si ottengono definizioni analoghe per $Y \mid X = x_i$.

Se X e Y sono *indipendenti*, tutte le distribuzioni condizionate di $X \mid Y = y_j$, al variare di y_j sono uguali e coincidono con la distribuzione marginale di X; analogamente per $Y \mid X = x_i$.

A partire dalla distribuzione di probabilità della variabile casuale condizionata $X \mid Y = y_j$, è possibile determinare, con le formule usuali, il **valore atteso condizionato**

$$E(X|Y = y_j) = \sum_{x_i} x_i f_{X|Y=y_j}(x_i)$$

e la **varianza condizionata**

$$V(X|Y = y_j) = \sum_{x_i} (x_i - E(X|Y = y_j))^2 f_{X|Y=y_j}(x_i).$$

Se X e Y sono *indipendenti*, valore atteso e varianza condizionati sono costanti e coincidono con $E(X)$ e $V(X)$.

Si ottengono definizioni analoghe per $Y \mid X = x_i$.

Covarianza e correlazione

Analogamente a quanto visto in statistica descrittiva, una misura della dipendenza lineare fra due variabili casuali X e Y, con media $E(X)$ e $E(Y)$, è data dalla **covarianza**

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))].$$

Nel caso di **variabili casuali discrete**

$$Cov(X,Y) = \sum_{x_i} \sum_{y_j} (x_i - E(X))(y_j - E(Y)) f_{X,Y}(x_i, y_j).$$

In alternativa, si può calcolare utilizzando la *formula per il calcolo*

$$Cov(X,Y) = E(XY) - E(X)E(Y),$$

dove, nel caso *discreto*, $E(XY) = \sum_{x_i} \sum_{y_j} x_i y_j f_{X,Y}(x_i, y_j)$.

Spesso si indica con σ_{XY}, che ne richiama il legame con la varianza che corrisponde a $V(X) = \sigma_X^2 = \sigma_{XX} = Cov(X,X)$.

Una misura normalizzata della dipendenza lineare è il **coefficiente di correlazione lineare** definito da

$$\rho_{XY} = Cor(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Dalla diseguaglianza di Cauchy-Schwarz si ha che $-1 \leq \rho_{XY} \leq 1$.

Se $\rho_{XY} = 0$, c'è assenza di legame lineare tra X e Y, che sono dette **incorrelate** (ma non necessariamente indipendenti).

Se $\rho_{XY} > 0$ c'è **relazione lineare crescente** fra X e Y; nel caso in cui $\rho_{XY} = 1$ la relazione è esattamente lineare crescente con probabilità 1.

Se $\rho_{XY} < 0$ c'è **relazione lineare decrescente** fra X e Y; nel caso in cui $\rho_{XY} = -1$ la relazione è esattamente lineare decrescente con probabilità 1.

L'assenza di legame lineare non assicura l'indipendenza tra le variabili. Tuttavia esistono alcune situazioni particolari dove si verifica che, eccezionalmente, se $Cov(X, Y) = 0$ si può concludere che X e Y sono anche indipendenti.

In particolare questo accade con:

- (X, Y) con componenti $X \sim Ber(p)$, $Y \sim Ber(p)$;
- (X, Y) variabile casuale gaussiana bivariata con componenti $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$.

Infine, si evidenzia che la covarianza è coinvolta nell'espressione della varianza di una combinazione lineare di X e Y. Infatti, per ogni $a, b \in \mathbf{R}$,

$$V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2ab\, Cov(X, Y).$$

Casi particolari sono

$$\begin{aligned} V(X+Y) &= V(X) + V(Y) + 2Cov(X,Y), \\ V(X-Y) &= V(X) + V(Y) - 2Cov(X,Y). \end{aligned}$$

Se X e Y sono *incorrelate*, o a maggior ragione *indipendenti*, le relazioni presentate continuano a valere con $Cov(X, Y) = 0$.