
Calcolo delle Probabilità e Statistica

Avon Alberto, 30 gennaio 2011

Sommario

| | |
|---|----|
| Esperimenti casuali e spazio campionario | 3 |
| Eventi e classe degli eventi | 3 |
| Misura di probabilità, spazio probabilizzato, spazio probabilizzabile e assiomi di Kolmogorov | 3 |
| Teoremi elementari del Calcolo delle Probabilità | 4 |
| Probabilità condizionale e formula della probabilità composta | 5 |
| Formula della probabilità totale | 5 |
| Eventi indipendenti | 5 |
| Teorema di Bayes | 6 |
| Variabile casuale univariata: legge di probabilità e supporto, variabili casuali identicamente distribuite. | 6 |
| Funzione di ripartizione di una variabile casuale | 7 |
| Variabili casuali con legge di probabilità di tipo discreto e funzione di massa di probabilità | 7 |
| Variabili casuali univariate con legge di probabilità di tipo continuo e funzione di densità di probabilità | 8 |
| I principali indici di posizione (valore atteso, moda, mediana, quantili) e la disuguaglianza di Markov. | 9 |
| Gli indici di variabilità (varianza, scarto quadratico medio) e la disuguaglianza di Cebyshev | 11 |
| Momenti e funzione generatrice dei momenti di una variabile casuale | 12 |
| Trasformazioni di variabili casuali | 13 |
| Distribuzione uniforme discreta e distribuzione degenere | 14 |
| Distribuzione binomiale e distribuzione di Bernoulli (o binomiale elementare) | 15 |
| Distribuzione ipergeometrica | 16 |
| Distribuzione geometrica | 16 |
| Distribuzione di Poisson | 17 |
| Distribuzione uniforme continua | 18 |
| Distribuzione esponenziale | 19 |
| La funzione tasso di guasto e le leggi di Weibull | 19 |
| La distribuzione normale | 19 |
| Variabili casuali bivariate e multivariate con componenti indipendenti | 20 |
| Legge di probabilità di massimo, minimo e somma di variabili casuali indipendenti | 21 |
| La distribuzione della media campionaria di normali indipendenti e identicamente distribuite | 21 |
| Distribuzioni marginali di una variabile casuale con legge discreta | 22 |
| Le leggi condizionali | 22 |
| Proprietà additive | 23 |

Esperimenti casuali e spazio campionario

Un **esperimento casuale** è un fenomeno o esperimento in riferimento al quale le conoscenze inducono a ritenere possibile una pluralità di esiti. Non è quindi possibile stabilire con certezza quale risultato si osserverà prima di eseguire l'esperimento.

Lo **spazio campionario** (o fondamentale) Ω di un esperimento casuale è l'insieme formato da tutti i suoi possibili risultati chiamati **eventi elementari** e supposti disgiunti in senso insiemistico.

Due esempi di esperimento casuale sono il **lancio di una moneta** e il **lancio di un dado**.

Eventi e classe degli eventi

Si chiama **evento** ogni sottoinsieme di Ω , di conseguenza è un evento ogni elemento di $\mathcal{P}(\Omega)$ (insieme delle parti di Ω). Si può definire l'evento anche come collezione di eventi elementari che si realizza se e solo se si realizza uno degli eventi elementari che lo definiscono.

Nell'esperimento del lancio di un dado, un esempio di evento è "*esce un numero dispari*", formato dagli eventi elementari "*esce il numero 1*", "*esce il numero 3*" ed "*esce il numero 5*".

Con **evento certo** s'intende l'evento definito da tutti gli eventi elementari dell'intero spazio fondamentale e si indica anch'esso con Ω ; con **evento impossibile** s'intende una collezione di eventi elementari che non appartengono a Ω e si indica con il simbolo \emptyset .

Sempre nel lancio del dado l'evento certo è "*esce un numero compreso tra 1 e 6*", mentre un possibile evento impossibile è "*esce il numero 7*" poiché non appartiene a Ω .

Essendovi una stretta analogia tra eventi dello spazio fondamentale e sottoinsiemi di un dato insieme ambiente, si possono definire delle **operazioni logiche** su eventi. Se $A \subseteq \Omega$ allora \bar{A} viene chiamato **evento complementare** di A in Ω ; se $A, B \subseteq \Omega$ si definisce **evento unione**, e viene indicato con $A \cup B$, quell'evento che si realizza se e solo se si realizza uno fra gli eventi elementari di A oppure di B ; se $A, B \subseteq \Omega$ si chiama, invece, **evento intersezione** quell'evento che si realizza se e solo se si realizza un evento elementare che definisce sia A che B e si indica con $A \cap B$. Se si verifica che $A \cap B = \emptyset$, allora gli eventi A e B si dicono **incompatibili**. Si può infine definire l'**evento differenza** $A \setminus B$ che si verifica se si realizza un evento elementare di A , ma non di B .

Una **classe di eventi** \mathcal{F} è una "famiglia" di eventi associata a Ω che contiene gli eventi di Ω più altri in modo tale che sia chiusa rispetto alle operazioni insiemistiche. Inoltre una classe di eventi \mathcal{F} viene chiamata **σ -algebra** se:

- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \rightarrow \bar{A} \in \mathcal{F}$
- $A_i \in \mathcal{F} \forall i \in I \subseteq \mathbb{N} \rightarrow \bigcup_{i \in I} A_i \in \mathcal{F}$

Conseguenze immediate della σ -algebra sono:

- $\emptyset = \bar{\Omega} \in \mathcal{F}$
- Se $A_i \in \mathcal{F} \forall i \in I \subseteq \mathbb{N} \rightarrow \bigcup_{i \in I} A_i \in \mathcal{F}$ allora $\bigcap_{i \in I} A_i = \overline{\bigcup_{i \in I} \bar{A}_i} \in \mathcal{F}$ (De Morgan)
- Se $A, B \in \mathcal{F}$ allora $A \setminus B = A \cap \bar{B} \in \mathcal{F}$

Se Ω ha cardinalità finita o numerabile, allora si considera come σ -algebra $\mathcal{F} = \mathcal{P}(\Omega)$. Se si considera il lancio di una moneta, la classe degli eventi di $\Omega = \{T, C\}$ è $\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{T\}, \{C\}, \Omega\}$.

Misura di probabilità, spazio probabilizzato, spazio probabilizzabile e assiomi di Kolmogorov

Una **misura di probabilità** P su una σ -algebra \mathcal{F} di sottoinsiemi di Ω è una funzione $P: \mathcal{F} \rightarrow [0, 1]$ tale che

- $P(A) \geq 0, \forall A \in \mathcal{F}$ assioma di *non negatività*;
- $P(\Omega) = 1$ assioma di *normalizzazione*;

- se $A_i \in \mathcal{F}$, $\forall i \in I \subseteq \mathbb{N}$ tali che $A_i \cap A_j = \emptyset$, $\forall i, j \in I$, allora

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i)$$

assioma di σ -additività.

L'assioma di non negatività, quello di normalizzazione e quello di additività sono detti **assiomi di Kolmogorov**.

Si chiama **spazio probabilizzato** la terna (Ω, \mathcal{F}, P) dove Ω è lo spazio campionario, \mathcal{F} è una σ -algebra di parti di Ω e P è una misura di probabilità per \mathcal{F} . La coppia (Ω, \mathcal{F}) si chiama, invece, **spazio probabilizzabile**.

Teoremi elementari del Calcolo delle Probabilità

1. $P(\emptyset) = 0$

Dimostrazione:

Poiché $\bar{\Omega} = \emptyset$ e $\bar{\Omega} \cap \Omega = \emptyset$, gli eventi Ω e \emptyset sono incompatibili, pertanto per il terzo assioma si ha che

$$1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + P(\emptyset)$$

da cui $P(\emptyset) = 0$.

2. $P(\bar{A}) = 1 - P(A)$

Dimostrazione:

Poiché $A \cap \bar{A} = \emptyset$ e $A \cup \bar{A} = \Omega$ cioè $\{A, \bar{A}\}$ è una partizione di Ω , si ha che

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

da cui $P(\bar{A}) = 1 - P(A)$.

3. $0 \leq P(A) \leq 1$

Dimostrazione:

$P(A) \geq 0$ per definizione (assioma di non negatività), mentre per il teorema precedente (2) si ha che

$$1 = P(A) + P(\bar{A})$$

Se, per assurdo, fosse $P(A) > 1$ allora $P(\bar{A}) < 0$ in contrasto col primo assioma.

4. $A \subseteq B \rightarrow P(A) \leq P(B)$

Dimostrazione:

$$A \subseteq B \rightarrow B = A \cup (B \setminus A)$$

e

$$A \cap (B \setminus A) = \emptyset$$

per il terzo assioma si ha

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

e dato che sicuramente $P(B \setminus A) \geq 0$ per il primo assioma, allora posso dedurre che $P(B) \geq P(A)$.

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

Dimostrazione:

Poiché $A \cup B = (A \cap B) \cup [B \setminus (A \cap B)] \cup [A \setminus (A \cap B)]$ si ha che:

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

In generale questo risultato si può estendere a più eventi (**formula di Poincarè**):

$$P\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \dots + (-1)^{k+1} P(A_1 \cap \dots \cap A_k) \quad k \in \mathbb{N}^+.$$

6. $P\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k P(A_i)$

disuguaglianza di Boole

Dimostrazione:

Per Poincarè si ha che

$$P\left(\bigcup_{i=1}^k A_i\right) \leq P(A_1) + P\left(\bigcup_{i=2}^k A_i\right) \leq P(A_1) + P(A_2) + P\left(\bigcup_{i=3}^k A_i\right) \leq \dots \leq \sum_{i=1}^k P(A_i).$$

$$7. \quad P(B) = \sum_{i \in I} P(B \cap A_i) \quad \{A_i, i \in I \subseteq \mathbb{N}\} \text{ partizione di } \Omega$$

Dimostrazione:

Per il terzo assioma, poiché $(B \cap A_i) \cap (B \cap A_j) = \emptyset$, si ha

$$P(B) = P(B \cap \Omega) = P\left(B \cap \left(\bigcup_{i \in I} A_i\right)\right) = P\left(\bigcup_{i \in I} B \cap A_i\right) = \sum_{i \in I} P(B \cap A_i)$$

Probabilità condizionale e formula della probabilità composta

Dati due eventi A e B , con $P(A) > 0$, l'evento $B|A$ si traduce con “ B si verifica condizionatamente al fatto che A si è verificato” e la sua probabilità, chiamata **probabilità condizionale** di B dato A , è definita ponendo

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

Infatti se A si è verificato, l'unica parte di B che può ancora realizzarsi è quella comune ad A e tutto ciò che è incompatibile con A non ha più alcuna probabilità di realizzarsi. Il parametro $P(A)$ a denominatore permette di ristabilire le proporzioni, assicurando la normalizzazione. In altri termini lo spazio fondamentale passa da Ω ad A .

Diretta conseguenza della formula della probabilità condizionale è la **formula della probabilità composta o formula di moltiplicazione**

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

estendibile al caso di tre eventi

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

con $P(A \cap B) > 0$.

Formula della probabilità totale

Sia $\{A_i, i = 1 \dots n\}$ partizione di Ω costituita da un numero finito di eventi di \mathcal{F} non trascurabili, si definisce **formula della probabilità totale** la seguente

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

Dimostrazione:

dal momento che gli eventi $A_i, i = 1 \dots n$ sono disgiunti e tali che $\bigcup_{i=1}^n A_i = \Omega$ (cioè sono una partizione dello spazio campionario) si ha per il “teorema” numero (7)

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

per la *formula della probabilità composta*.

Eventi indipendenti

Due eventi A, B sono detti (**stocasticamente**) **indipendenti** se

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

Se, ciò non succede, allora gli eventi si dicono **dipendenti**. Se due eventi A, B sono indipendenti, allora anche (\bar{A}, B) , (A, \bar{B}) e (\bar{A}, \bar{B}) sono indipendenti.

In generale se ho gli eventi A_i con $i \in I \subseteq \mathbb{N}^+$, essi si dicono (**mutuamente**) **indipendenti** se

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

Come esempio si consideri un circuito costituito da 6 componenti elettronici indipendenti tra loro. La probabilità di rottura in un certo intervallo di tempo è 0,5 per il primo componente, 0,2 per il secondo e 0,1 per i restanti quattro. Si consideri inoltre l'evento A_i "il componente i si rompe nell'intervallo di tempo prefissato" e l'evento B "il circuito s'interrompe nell'intervallo di tempo prefissato". Gli eventi A_i sono indipendenti.

Se i componenti sono collegati in serie, il circuito s'interrompe se *almeno uno* dei componenti si guasta, quindi la probabilità che il circuito si blocchi è

$$P(B) = 1 - P(\bar{B}) = 1 - P(\bar{A}_1 \cap \dots \cap \bar{A}_6) = 1 - \prod_{i=1}^6 P(\bar{A}_i) \approx 0,738$$

Se invece i componenti sono collegati in parallelo, il circuito s'interrompe se *tutti* i componenti si guastano, quindi $P(B)$ in questo caso è

$$P(B) = P(A_1 \cap \dots \cap A_6) = \prod_{i=1}^6 P(A_i) \approx 0,00001$$

Comunque per garantire la mutua indipendenza non è necessario che gli eventi siano a due a due indipendenti.

Teorema di Bayes

Sia (Ω, \mathcal{F}, P) uno spazio probabilizzato e $\{C_i, i \in I\}$ una partizione di Ω tale che $P(C_i) > 0 \quad \forall i \in I$. Per ogni evento non trascurabile $A \in \mathcal{F}$, si ha che $\forall i$

$$P(C_i|A) = \frac{P(C_i)P(A|C_i)}{P(A)} = \frac{P(C_i)P(A|C_i)}{\sum_{j \in I} P(C_j)P(A|C_j)}$$

La quantità $P(A|C_i)$ è chiamata **verosimiglianza** di C_i .

Variabile casuale univariata: legge di probabilità e supporto, variabili casuali identicamente distribuite.

Intuitivamente, una variabile casuale è il risultato numerico di un esperimento quando questo non è prevedibile con certezza. Per esempio l'esperimento del lancio del dado può essere modellizzato con una variabile casuale che può assumere i valori 1, 2, 3, 4, 5, 6. La definizione formale invece è diversa.

Si considerino lo spazio probabilizzato (Ω, \mathcal{F}, P) e lo spazio probabilizzabile $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ dove $\mathcal{B}_{\mathbb{R}}$ è la σ -algebra di Borel associata ad \mathbb{R} .

Una **variabile casuale (univariata)** definita su Ω con valori in \mathbb{R} , è una *funzione*

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

tale che per ogni elemento $B \in \mathcal{B}_{\mathbb{R}}$ si ha che

$$X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in B\} \in \mathcal{F}$$

X deve quindi essere una funzione misurabile rispetto alle σ -algebre \mathcal{F} e $\mathcal{B}_{\mathbb{R}}$. La collezione delle semirette $\{]-\infty, x], x \in \mathbb{R}\}$ è un sistema di generatori per $\mathcal{B}_{\mathbb{R}}$ quindi la definizione può venire espressa considerando solo gli eventi $B =]-\infty, x]$ con $x \in \mathbb{R}$.

Una variabile casuale X definita su Ω con valori in \mathbb{R} è tale da indurre sullo spazio probabilizzabile una **misura di probabilità** P_X definita mediante la seguente regola:

$$P_X(B) = P(X \in B) = P(X^{-1}(B)) \quad \forall B \in \mathcal{B}_{\mathbb{R}}$$

$P(X \in B)$ è la misura di probabilità che X assuma valori in B riferita allo spazio probabilizzato di partenza. In particolare se $B = \{x\}$, $x \in \mathbb{R}$ (si ricordi che B è un sottoinsieme di \mathbb{R} in quanto elemento della σ -algebra di Borel, classe degli eventi per \mathbb{R})

$$P_X(\{x\}) = P(X = x) = P(X^{-1}(\{x\})) = P(\{\omega \in \Omega : X(\omega) = x\})$$

La funzione P_X prende il nome di **distribuzione di probabilità** di X . Se due variabili casuali X e Y anche definite su spazi diversi assumono valori $P_X = P_Y$ sono dette **identicamente distribuite**, in simboli $X \sim Y$.

Quindi una variabile casuale X pone in relazione uno spazio probabilizzato (di partenza) (Ω, \mathcal{F}, P) con un altro (di arrivo) $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P_X)$ e attua un trasferimento della probabilità da Ω alla retta reale. Solitamente non si considera lo spazio di partenza, pertanto una VC viene caratterizzata (e studiata) in base alla distribuzione di probabilità che induce.

Se al posto di considerare \mathbb{R} avessimo considerato \mathbb{R}^d avremmo avuto una variabile **casuale d-variata** (o **multivariata**).

Infine, si definisce **supporto** di una variabile casuale X , indicato con S_X , il più piccolo sottoinsieme chiuso di $\mathcal{B}_{\mathbb{R}}$ tale che

$$P_X(S_X) = P(X \in S_X) = 1$$

Altro non è che l'insieme di tutti i possibili valori che può assumere la variabile casuale X .

Funzione di ripartizione di una variabile casuale

Data una variabile casuale X , la funzione $F_X : \mathbb{R} \rightarrow [0, 1]$ definita come

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

è chiamata **funzione di ripartizione** della variabile casuale X ed ha le seguenti proprietà:

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

Una variabile casuale d-variata invece ha una funzione di ripartizione $F_X : \mathbb{R}^d \rightarrow [0, 1]$ così definita

$$F_X(x) = P((-\infty, x_1], (-\infty, x_2] \dots (-\infty, x_d])$$

Variabili casuali con legge di probabilità di tipo discreto e funzione di massa di probabilità

Una variabile casuale X con supporto S_X e funzione di ripartizione $F_X(x)$ è detta **discreta** se

1. Il supporto S_X è una successione finita o numerabile di punti $x \in \mathbb{R}^d$ tali che
 - a. $p_X(x) > 0$;
 - b. $\sum_{x \in S_X} p_X(x) = 1$.
2. $F_X(x) = P(X \leq x) = \sum_{t \in S_X: t \leq x} p_X(t)$

dove la funzione p_X viene chiamata **funzione di massa di probabilità** della variabile casuale discreta X ed è definita da

$$p_X(x) = P(X = x)$$

Si tenga presente che la funzione di ripartizione $F_X(x)$ esprime la probabilità che $P(X \leq x)$ che X assuma valori minori o uguali ad x e il suo grafico, nel caso di variabili casuali discrete, è una funzione a gradini continua da destra con salti in corrispondenza degli elementi del supporto. L'ampiezza del salto è data da $P(X = x_i) = p_i$ dove $x_i \in S_X$.

È possibile determinare la funzione di ripartizione anche a partire dalla funzione di massa di probabilità

$$p_X(x_i) = \sum_{t \in S_X: t \leq x_i} p_X - \sum_{t \in S_X: t < x_i} p_X = F_X(x_i) - F_X(x_{i-1})$$

Ciò implica che si può definire univocamente una variabile casuale discreta specificandone o la funzione di massa di probabilità $p_X(x)$ o la funzione di ripartizione $F_X(x)$ ed il supporto S_X .

La conoscenza della fmp permette spesso una semplificazione nei calcoli di probabilità di eventi relativi a X dato che $\forall B \in \mathcal{B}_{\mathbb{R}}$

$$P(X \in B) = P_X(B) = \sum_{x \in B \cap S_X} p_X(x)$$

Ad esempio se si considera l'esperimento che consiste nel lanciare tre volte una moneta equilibrata e si studi la variabile X che riporta il numero di teste realizzato. Il supporto di X è $S_X = \{0, 1, 2, 3\}$ e le probabilità sono:

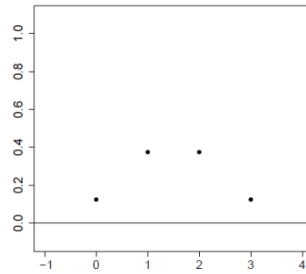
$$P_X(\{0\}) = P(X = 0) = P(\{CCC\}) = \frac{1}{8}$$

$$P_X(\{1\}) = P(X = 1) = P(\{TCC, CTC, CCT\}) = \frac{3}{8}$$

$$P_X(\{2\}) = P(X = 2) = P(\{CTT, TCT, TTC\}) = \frac{3}{8}$$

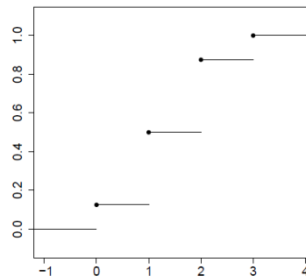
$$P_X(\{3\}) = P(X = 3) = P(\{TTT\}) = \frac{1}{8}$$

La $p_X(x)$ è



Come si può vedere dal grafico, la funzione di massa di probabilità assume i valori $\frac{1}{8}$ e $\frac{3}{8}$ solo in corrispondenza dei punti 0, 3, 1, 2. Esprime i valori $P_X(x_i)$ per ogni $x_i \in S_X$.

La $F_X(x)$ invece è la seguente:



Qui invece notiamo che i valori sono dati dal salto di due gradini. L'ampiezza dell'intervallo in corrispondenza di x_i e x_{i+1} è la $P_X(x_i)$.

Variabili casuali univariate con legge di probabilità di tipo continuo e funzione di densità di probabilità

Una variabile casuale X ha legge **continua** se esiste una funzione reale $p_X(x)$ definita su \mathbb{R} con $\int_{-\infty}^{+\infty} p_X(x) dx = 1$ tale che

$$F_X(x) = \int_{-\infty}^x p_X(t) dt$$

Questa funzione è detta **funzione di densità di probabilità** della variabile X . Dato che $F_X(x)$ è continua, allora

$$p_X(x) = \frac{d}{dx} F_X(x) = \lim_{\epsilon \rightarrow 0^+} \frac{F_X(x + \epsilon) - F_X(x - \epsilon)}{2\epsilon} = \lim_{\epsilon \rightarrow 0^+} \frac{P(x - \epsilon \leq X \leq x + \epsilon)}{2\epsilon}$$

A punti del supporto a cui corrisponde un'elevata funzione di densità indicano un'alta probabilità che X assuma valori in un loro intorno.

Si considerino $x_1, x_2 \in \mathbb{R}$ tali che $x_1 < x_2$,

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} p_X(x) dx$$

Dunque la probabilità che X assuma valori compresi nell'intervallo $(x_1, x_2]$ corrisponde all'area sottostante la curva $p_X(x)$ compresa tra gli estremi x_1 e x_2 . Essendo X continua tale probabilità coincide

con quella degli intervalli $[x_1, x_2]$, $[x_1, x_2)$, $(x_1, x_2]$. In generale la probabilità associata all'evento $X \in B$ corrisponde a

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

I principali indici di posizione (valore atteso, moda, mediana, quantili) e la disuguaglianza di Markov.

Si chiama **valore atteso** (o **media**) $E(X)$ di una variabile casuale X con supporto S_X e funzione di probabilità $p_X(x)$, la media dei suoi possibili valori ponderati con le relative probabilità, ossia

Per X **discreta**

$$E(X) = \sum_{x \in S_X} x \cdot p_X(x) = \sum_{x \in S_X} x \cdot P(X = x)$$

Per X **assolutamente continua**

$$E(X) = \int_{-\infty}^{+\infty} x \cdot p_X(x) dx$$

purché la serie o l'integrale siano assolutamente convergenti.

In particolare, per X discreta e S_X con cardinalità numerabile si richiede che

$$\sum_{x \in S_X} |x| \cdot p_X(x) < +\infty$$

mentre se X è continua si richiede che

$$\int_{-\infty}^{+\infty} |x| \cdot p_X(x) dx < +\infty$$

altrimenti il valore atteso non esiste finito. Usualmente si pone $E(X) = \mu$ e si intende tacitamente che esista finito.

Ad esempio se si considera l'esperimento che consiste nel lanciare una moneta ripetutamente fino a quando non si osserva testa per la prima volta. Si è interessati alla variabile casuale X che conta il numero di lanci necessari per ottenere testa per la prima volta. Il supporto di X è $S_X = \mathbb{N}^+$ e la sua funzione di probabilità è

$$p_X(x) = \begin{cases} (1-p)^{x-1}p & \text{se } x \in \mathbb{N}^+ \\ 0 & \text{se } x \notin \mathbb{N}^+ \end{cases}$$

dove $p \in (0, 1)$ indica la probabilità di ottenere testa in ogni singolo lancio della moneta. $(1-p)^{x-1}$ indica i primi x lanci in cui ho ottenuto croce, mentre p indica l'ultimo lancio in cui ho ottenuto testa.

$$\begin{aligned} E(X) &= \sum_{x=1}^{+\infty} x \cdot p_X(x) = \sum_{x=1}^{+\infty} x(1-p)^{x-1}p = \sum_{x=1}^{+\infty} (1-p)^{x-1}p + \sum_{x=1}^{+\infty} (x-1)(1-p)^{x-1}p \\ &= \sum_{x=1}^{+\infty} p_X(x) + \sum_{x=1}^{+\infty} (x-1)(1-p)^{x-1}p = 1 + \sum_{x=1}^{+\infty} x(1-p)^x p \\ &= 1 + (1-p) \sum_{x=1}^{+\infty} x(1-p)^{x-1}p \\ &= 1 + (1-p) E(X) \end{aligned}$$

da cui

$$E(X) = \frac{1}{p}$$

Si considerino le variabili casuali X e $Y = g(X)$ ottenuta come **trasformata** della X tramite $g(\cdot)$. Si può verificare che $E(Y) = E(g(x))$ è dato da:

$$E(Y) = \sum_{y \in S_Y} y \cdot p_Y(y) = \sum_{x \in S_X} g(x) \cdot p_X(x) \quad \text{se } X \text{ e } Y \text{ discrete}$$

$$E(Y) = \int_{-\infty}^{+\infty} y \cdot p_Y(y) dy = \int_{-\infty}^{+\infty} x \cdot p_X(x) dx \quad \text{se } X \text{ e } Y \text{ continue}$$

Date le costanti reali $a, b \in \mathbb{R}$ e la variabile casuale X , con supporto S_X e funzione di probabilità $p_X(x)$ che ammette valore atteso $E(X)$ finito, si ha che

1. **$\inf\{x \in S_X\} \leq E(X) \leq \sup\{x \in S_X\}$** **proprietà di Cauchy**

Dimostrazione:

Si consideri, per semplicità, X discreta. $S_X = \{x_1, \dots, x_n\}$ con $x_1 < \dots < x_n$. Poiché

$$x_1 \leq x_i \leq x_n \quad i = 1 \dots n$$

e

$$x_1 p_X(x_i) \leq x_i p_X(x_i) \leq x_n p_X(x_i)$$

si ottiene che

$$\sum_{i=1}^n x_1 p_X(x_i) \leq \sum_{i=1}^n x_i p_X(x_i) \leq \sum_{i=1}^n x_n p_X(x_i)$$

da cui $x_1 \leq E(X) \leq x_n$.

2. **$Y = aX + b \rightarrow E(Y) = aE(X) + b$** **proprietà di linearità**

Dimostrazione:

Si consideri X continua. Poiché $Y = g(X) = aX + b$ si ha

$$E(Y) = \int_{-\infty}^{+\infty} g(x) p_X(x) dx = \int_{-\infty}^{+\infty} (aX + b) \cdot p_X(x) dx$$

da cui si ottiene la tesi per la proprietà di linearità dell'integrale definito.

3. **$c \forall \in \mathbb{R} \quad E((X - c)^2) \geq E((X - E(X))^2)$** **proprietà dei minimi quadrati**

Dimostrazione:

$E((X - c)^2) = E\left(\left((X - E(X)) + (E(X) - c)\right)^2\right)$ sviluppando il quadrato e applicando la proprietà di linearità si ottiene

$$E((X - c)^2) = E\left((X - E(X))^2\right) + (E(X) - c)^2 \geq E\left((X - E(X))^2\right).$$

Dalla proprietà di linearità discende la proprietà di **baricentro** secondo cui $E(X - E(X)) = 0$ e che $\forall a \in \mathbb{R}, \quad E(a) = a$ e che quindi $E(aX) = aE(X)$.

Data una variabile casuale X si definisce **moda** di X , e si indica con x_{mo} , il valore reale per cui è massima la funzione di probabilità $p_X(x)$, cioè il valore x_{mo} tale che $p_X(x_{mo}) \geq p_X(x) \quad \forall x \in \mathbb{R}$.

La moda può non essere unica oppure può addirittura non esistere. Se esiste appartiene a S_X e individua i valori più probabili, se X è discreta, oppure i cui intorno sono gli eventi più probabili se X è continua. Nel caso in cui $p_X(x)$ ha un unico massimo, la distribuzione di probabilità di X è *unimodale*; se ce n'è più d'uno la distribuzione si chiama distribuzione *multimodale*.

Data una variabile casuale X , si chiama **mediana** di X il valore $x_{0.5} \in \mathbb{R}$ tale che

$$\begin{cases} P(X \leq x_{0.5}) \geq \frac{1}{2} \\ P(X \geq x_{0.5}) \geq \frac{1}{2} \end{cases}$$

ossia è quel valore reale che ripartisce in modo equilibrato la massa unitaria di probabilità riferita a X , nel senso che gli eventi

$$\begin{aligned} X &\in (-\infty, x_{0.5}] \\ X &\in [x_{0.5}, +\infty) \end{aligned}$$

hanno probabilità pari a $\frac{1}{2}$ o maggiore di $\frac{1}{2}$ se $P(X = x_{0.5}) > 0$. Anche la mediana può non essere unica e, in alcuni casi, può corrispondere ad un'infinità più che numerabile di valori reali.

Si può verificare che X ha un'unica mediana se e solo se

$$F_X(x) = \frac{1}{2} \quad \text{ha un'unica soluzione}$$

oppure

$$F_X(x) = \frac{1}{2} \quad \text{non ha soluzioni}$$

La definizione di mediana è generale quindi si applica sia a variabili discrete che continue. In particolare se X è una variabile casuale continua, si ha

$$P(X = x_{0.5}) = 0$$

$$P(X \leq x_{0.5}) = P(X \geq x_{0.5}) = \frac{1}{2}$$

La mediana soddisfa la seguente proprietà (dei **minimi scarti assoluti**)

Sia X variabile casuale, allora $\forall c \in \mathbb{R}$

$$E(|X - c|) \geq E(|X - x_{0.5}|)$$

Sia $\alpha \in (0, 1)$ e X variabile casuale. Si chiama **quantile α -esimo** di X il valore $x_\alpha \in \mathbb{R}$ tale che

$$\begin{cases} P(X \leq x_\alpha) \geq \alpha \\ P(X \geq x_\alpha) \geq 1 - \alpha \end{cases}$$

Anche i quantili non sono necessariamente unici. Analogamente alla mediana, nel caso di variabile casuale continua si ha che il quantile α -esimo x_α è tale che

$$P(X \leq x_\alpha) = 1 - P(X \geq x_\alpha) = 1 - \alpha$$

ed è equivalente a porre

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

Il quantile è interpretabile come quel valore x_α che ripartisce la massa unitaria di probabilità lasciando una porzione pari ad α alla propria sinistra e una porzione pari a $1 - \alpha$ alla propria destra. Se α è espresso in termini percentuali si parla di **percentili**. La mediana quindi costituisce il 50-esimo percentile o, il quantile con $\alpha = \frac{1}{2}$. I quantili x_α con $\alpha = \frac{1}{2}, \frac{1}{4}, \frac{3}{4}$ vengono chiamati **quartili**.

La seguente disuguaglianza viene chiamata disuguaglianza di **Markov**:

sia X variabile casuale non negativa, cioè tale che $P(X \geq 0) = 1$ con valor medio $\mu = E(X) > 0$ finito. Per ogni costante reale $c > 0$ vale che

$$P(X \geq c\mu) \leq \frac{1}{c}$$

Gli indici di variabilità (varianza, scarto quadratico medio) e la disuguaglianza di Cebyshev

Data una variabile casuale X discreta o continua con *valore atteso* $E(X)$ finito, si chiama **varianza** di X la quantità definita da

$$V(X) = E\left((X - E(X))^2\right)$$

se esiste finita.

Usualmente si pone $V(X) = \sigma^2$ e si intende che i valori attesi coinvolti nella definizione esistano finiti.

La varianza gode delle seguenti proprietà:

1. $V(X) = E(X^2) - (E(X))^2$ **formula per il calcolo**

Dimostrazione:

Per le proprietà del valore atteso si ha che:

$$\begin{aligned} V(X) &= E\left((X - E(X))^2\right) = E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E\left((E(X))^2\right) = E(X^2) - 2(E(X))^2 + (E(X))^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

2. $V(aX) = a^2V(X)$ **omogeneità di secondo grado**

Dimostrazione:

Poiché

$$\begin{aligned} V(aX) &= E\left((aX - E(aX))^2\right) = E\left((aX - aE(X))^2\right) = E\left(a^2(X - E(X))^2\right) \\ &= a^2E\left((X - E(X))^2\right) \end{aligned}$$

3. $V(X + a) = V(X)$ **invarianza per traslazione**

Dimostrazione:

$$V(X + a) = E((X + a - E(X + a))^2) = E((X + a - E(X) - a)^2) = E((X - E(X))^2) = V(X)$$

La radice non negativa σ della varianza è detta **scarto quadratico medio** (oppure **deviazione standard**) e assieme alla varianza è utile per misurare la dispersione dei valori di X attorno ad $E(X)$ ed è spesso preferita alla varianza in quanto è espressa nella stessa unità di misura di X e del suo valore atteso $E(X)$.

Dalle proprietà di omogeneità e invarianza segue che

$$V(aX + b) = a^2 V(X)$$

inoltre data X con $E(X) = \mu$ e $\sigma^2 = V(X)$ la variabile casuale *trasformata*

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

è tale che $E(Y) = 0$ e $V(Y) = 1$. Y in tal caso è detta **standardizzata**. Il ragionamento vale anche al contrario, quindi data Y tale che $E(Y) = 0$ e $V(Y) = 1$ si può ottenere una variabile casuale con μ e σ^2 prefissati utilizzando la trasformata $X = \sigma Y + \mu$.

Viene definita disuguaglianza di **Cebyshev** la seguente:

sia X una variabile con valor medio μ e varianza $\sigma^2 > 0$ finita per ogni costante reale $k > 0$ vale che

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

o equivalentemente ponendo $\epsilon = k\sigma$

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

e quindi la disuguaglianza assume il seguente significato:

la probabilità che X assuma valori esterni all'intervallo $[\mu - \epsilon, \mu + \epsilon]$ è superiormente limitata dalla quantità $\frac{\sigma^2}{\epsilon^2}$. Quindi tanto più è piccola è la varianza $V(X) = \sigma^2$ (cioè la dispersione di X attorno a $E(X)$), tanto più piccola sarà la *limitazione superiore della probabilità* di cadere al di fuori di un qualsiasi intervallo prefissato centrato sulla media.

Momenti e funzione generatrice dei momenti di una variabile casuale

Si chiama **momento** di ordine $r \in \mathbb{N}^+$ di una variabile casuale univariata X e si indica con μ_r il valore

$$\mu_r = E(X^r)$$

Il *momento primo* è quindi il valore atteso stesso poiché

$$\mu_1 = \mu = E(X^1) = E(X)$$

Il *momento secondo* di X è

$$\mu_2 = E(X^2)$$

Abbiamo che la varianza è

$$V(X) = E(X^2) - (E(X))^2 = \mu_2 - \mu_1^2$$

detta anche momento secondo centrato rispetto a μ .

La successione μ_r dei momenti di una variabile casuale è associabile ad una funzione reale di variabile reale.

Si chiama **funzione generatrice dei momenti** di X la funzione

$$M_X(t) = E(e^{tx}) = \begin{cases} \sum_{x \in S_X} e^{tx} p_X(x) & \text{se } X \text{ è discreta} \\ \int_{\mathbb{R}} e^{tx} p_X(x) & \text{se } X \text{ è assolutamente continua} \end{cases}$$

Qualunque sia la legge di probabilità di X vale che nell'origine

$$M_X(t = 0) = E(e^{0x}) = E(e^0) = 1$$

Se $\forall t \in (-\epsilon, \epsilon) \epsilon > 0$ vale che $M_X(t) < +\infty$ allora si dice che X ha funzione generatrice dei momenti **propria**. Se la funzione generatrice dei momenti di X ha questa proprietà, tutti i momenti di X esistono finiti e possono essere calcolati mediante la formula

$$\mu_r = E(X^r) = \frac{d^r}{dt^r} M_X(t) \Big|_{t=0}$$

Abbiamo quindi che per calcolare valore atteso e varianza data la M_X possiamo eseguire questi passaggi:

$$\begin{aligned} E(X) &= \mu_1 = \frac{d}{dt} M_X(t) \Big|_{t=0} \\ E(X^2) &= \mu_2 = \frac{d^2}{dt^2} M_X(t) \Big|_{t=0} \\ V(X) &= E(X^2) - (E(X))^2 \end{aligned}$$

Inoltre se X ha funzione generatrice dei momenti propria, allora la successione la successione di tutti i suoi momenti $\mu_r = E(X^r)$ determina la legge di probabilità di X , cioè

$$M_X(t) = M_Y(t) \quad \forall t \in (-\epsilon, \epsilon) \quad \Rightarrow \quad X \sim Y$$

in altre parole

una funzione generatrice dei momenti propria determina la legge di probabilità della variabile casuale.

Trasformazioni di variabili casuali

Sia X **variabile casuale discreta** con supporto $S_X = \{x_i, i \in I \subseteq \mathbb{N}^+\}$ e funzione di massa di probabilità $p_X(x)$ tale che $\forall x_i \in S_X$

$$p_X(x_i) = P(X = x_i) = p_i$$

e Y variabile casuale definita come variabile trasformata da X

$$Y = g(X)$$

Allora anche Y sarà una variabile casuale discreta con supporto

$$S_Y = g(S_X) = \{g(x_i), i \in I \subseteq \mathbb{N}^+\} = \{y_j, j \in J \subseteq \mathbb{N}^+\}$$

e funzione di probabilità $p_Y(y)$ tale che $\forall y_j \in S_Y$

$$p_Y(y_j) = \sum_{i \in I: g(x_i) = y_j} p_i$$

Da ciò discende che

$$E(Y) = \sum_{j \in J} y_j \cdot p_Y(y_j) = \sum_{j \in J} y_j \cdot \sum_{i \in I: g(x_i) = y_j} p_i = \sum_{i \in I} g(x_i) \cdot p_X(x_i)$$

Ad esempio prendiamo una variabile casuale discreta X con supporto $S_X = \{-2, -1, 0, 1, 2\}$ e funzione di probabilità $p_X(x)$ tale che

$$\begin{aligned} p_X(-2) &= \frac{1}{5} \\ p_X(-1) &= p_X(0) = \frac{1}{6} \\ p_X(1) &= \frac{2}{15} \\ p_X(2) &= \frac{1}{3} \end{aligned}$$

Si consideri $Y = X^2$. Allora $S_Y = \{0, 1, 4\}$ e la funzione di probabilità è

$$\begin{aligned} p_Y(0) &= p_X(0) = \frac{1}{6} \\ p_Y(1) &= p_X(-1) + p_X(1) = \frac{1}{6} + \frac{2}{15} = \frac{5+4}{30} = \frac{9}{30} = \frac{3}{10} \\ p_Y(4) &= p_X(-2) + p_X(2) = \frac{1}{5} + \frac{1}{3} = \frac{3+5}{15} = \frac{8}{15} \end{aligned}$$

e il valore atteso è

$$\begin{aligned}
 E(Y) = E(X^2) &= \sum_{x_i \in S_X} g(x_i) \cdot p_X(x_i) \\
 &= (-2)^2 p_X(-2) + (-1)^2 p_X(-1) + (0)^2 p_X(0) + (1)^2 p_X(1) + (2)^2 p_X(2) \\
 &= 4 \frac{1}{5} + \frac{1}{6} + 0 + \frac{2}{15} + 4 \frac{1}{3} = 4 \frac{8}{15} + \frac{1}{6} + \frac{2}{15} = \frac{34}{15} + \frac{1}{6} = \frac{68+5}{30} = \frac{73}{30}
 \end{aligned}$$

calcolabile anche come

$$E(Y) = E(X^2) = \sum_{j \in J} y_j \cdot p_Y(y_j) = 0 + \frac{3}{10} + 4 \frac{8}{15} = \frac{9+64}{30} = \frac{73}{30}$$

Se invece X è **variabile casuale assolutamente continua** con funzione di ripartizione $F_X(x)$, funzione di densità di probabilità $p_X(x)$ e supporto S_X e $Y = g(X)$ con $g(\cdot)$ funzione misurabile, monotona in senso stretto e continua con inversa differenziabile, allora

$$\begin{aligned}
 S_Y &= \{y \in \mathbb{R} : y = g(x) \text{ e } x \in S_X\} \\
 F_Y(y) &= \begin{cases} F_X(g^{-1}(y)) & g(\cdot) \text{ crescente} \\ 1 - F_X(g^{-1}(y)) & g(\cdot) \text{ decrescente} \end{cases} \\
 p_Y(y) &= p_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \\
 E(Y) &= \int_{\mathbb{R}} y_j \cdot p_Y(y_j) dy = \int_{S_Y} y_j \cdot p_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| dy = \int_{S_X} g(x) \cdot p_X(x) dx
 \end{aligned}$$

Ad esempio si consideri una variabile casuale assolutamente continua con

$$S_X = [0, +\infty)$$

e

$$F_X(x) = \begin{cases} 1 - e^{-2x} & x \in S_X \\ 0 & x \notin S_X \end{cases}$$

Se $Y = e^X$, allora

$$X = g^{-1}(Y) = \ln Y$$

$$S_Y = \{g(x), x \in S_X\} = [1, +\infty)$$

$$F_Y(y) = F_X(g^{-1}(y)) = 1 - e^{-2 \ln y} = 1 - e^{\ln \frac{1}{y^2}} = 1 - \frac{1}{y^2} \quad \forall y \in S_Y \quad \text{perchè } g(y) = e^x \text{ è crescente.}$$

$$\begin{aligned}
 p_X(x) &= F'_X(x) = 2e^{-2x} \quad \text{se } x \in S_X \\
 E(Y) &= \int_{S_X} g(x) p_X(x) dx = \int_0^{+\infty} e^x \cdot 2e^{-2x} dx = 2 \int_0^{+\infty} e^{-x} dx = 2 \int_0^{+\infty} \frac{1}{e^x} dx = 2 \left[-\frac{1}{e^x} \right]_0^{+\infty} \\
 &= 2 \left(\lim_{x \rightarrow +\infty} \left(-\frac{1}{e^x} \right) + \frac{1}{e^0} \right) = 2(0 + 1) = 2 \\
 p_Y(y) &= p_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| = 2e^{-2 \ln y} \left| \frac{d}{dy} \ln y \right| = 2e^{\ln \frac{1}{y^2}} \left| \frac{1}{y} \right| = 2 \frac{1}{y^2} \left| \frac{1}{y} \right| = \frac{2}{|y^3|}
 \end{aligned}$$

Distribuzione uniforme discreta e distribuzione degenera

La distribuzione uniforme discreta è utile per descrivere esperimenti casuali con un numero finito $n \in \mathbb{N}^+$ di potenziali esiti numerici $x_1, \dots, x_n \in \mathbb{R}$ *supposti equiprobabili*. Un esempio può essere il lancio di un dado o moneta, oppure un'estrazione da un'urna con n palline.

Una variabile casuale ha distribuzione **uniforme discreta** con possibili valori $x_1, \dots, x_n \in \mathbb{R}$ con $n \in \mathbb{N}^+$ fissato e si indica con

$$X \sim Ud(x_1, \dots, x_n)$$

se ha supporto

$$S_X = \{x_1, \dots, x_n\}$$

e funzione di probabilità

$$f(x) = \begin{cases} \frac{1}{n} & \text{se } x = x_1, \dots, x_n \\ 0 & \text{altrimenti} \end{cases}$$

Se $X \sim Ud(x_1, \dots, x_n)$ si ha che

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

$$V(X) = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2$$

$$M_X(t) = \frac{1}{n} \sum_{i=1}^n e^{tx_i} \quad t \in \mathbb{R}$$

Se si considera una distribuzione uniforme discreta con parametro $n = 1$ si ottiene una distribuzione **degenere** e si scrive

$$X \sim D(x_0)$$

Una distribuzione degenere ha

$$S_X = \{x_0\}$$

$$p_X(x_0) = 1$$

$$E(X) = x_0$$

$$V(X) = 0$$

$$M_X(t) = e^{tx_0}$$

Distribuzione binomiale e distribuzione di Bernoulli (o binomiale elementare)

Si consideri un'urna con $N \in \mathbb{N}^+$ palline totali di cui $M \in \{0, \dots, N\}$ sono bianche e $N - M$ sono nere. Se si estraggono *con reinserimento* $n \in \mathbb{N}^+$ palline, la distribuzione della variabile casuale X che conta il numero di palline bianche ottenute in n estrazioni con reinserimento.

Poiché in n estrazioni si possono ottenere da 0 a n palline bianche, è immediato concludere che il supporto di X è $S_X = \{0, \dots, n\}$

Per determinare la $p_X(x)$ invece si consideri il caso in cui le prime x palline siano bianche e le rimanenti $n - x$ nere. Dal momento che le estrazioni sono indipendenti l'una dall'altra (per il reinserimento) la probabilità di questo evento è

$$\underbrace{p \cdots p}_{x \text{ volte}} \underbrace{(1-p) \cdots (1-p)}_{n-x \text{ volte}} = p^x \cdot (1-p)^{n-x}$$

dove $p = \frac{M}{N}$ è la probabilità che venga estratta una bianca in una singola estrazione. In n estrazioni abbiamo che le configurazioni totali che possiamo avere sono pari alle combinazioni semplici di n elementi in gruppi di x , ossia ci sono $\binom{n}{x}$ configurazioni.

Abbiamo quindi che per ogni $x \in S_X$, la $p_X(x)$ vale

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

In generale, una distribuzione con

$$S_X = \{0, \dots, n\}$$

e

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

viene definita distribuzione **binomiale** con parametri n e p ,

$$X \sim Bi(n, p)$$

dove $n \in \mathbb{N}^+$ e $p \in (0, 1)$.

La distribuzione di probabilità binomiale descrive, come già detto, esperimenti aleatori che possono essere rappresentati come estrazioni con reinserimento da un'urna di configurazione nota. In sintesi, una

variabile casuale $X \sim Bi(n, p)$ descrive il numero di successi in n prove *indipendenti* con probabilità di successo costante pari a p .

La funzione generatrice dei momenti per la distribuzione binomiale si ottiene mediante la formula del binomio di Newton secondo la quale $(a + b)^k = \sum_{i=0}^k \binom{k}{i} a^i b^{k-i}$:

$$M_X(t) = \sum_{x=0}^n e^{tx} p_X(x) = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x} = (1-p + pe^t)^n \quad \forall t \in \mathbb{R}$$

A questo punto si possono trovare banalmente valore atteso e varianza

$$E(X) = \left. \frac{d}{dt} M_X(t) \right|_{t=0} = npe^t(1-p + pe^t)^{n-1} \Big|_{t=0} = np$$

$$V(X) = \left. \frac{d^2}{dt^2} M_X(t) \right|_{t=0} - (E(X))^2 = np[1 + p(n-1)] - (np)^2 = np(1-p)$$

Nel caso che il parametro $n = 1$ si ha la distribuzione **binomiale elementare** (o di **Bernoulli**), in simboli $X \sim Ber(p)$.

In tal caso avremo

$$S_X = \{0, 1\}$$

$$p_X(x) = p^x (1-p)^{1-x}$$

$$M_X(t) = \sum_{x=0}^n e^{tx} p_X(x) = \sum_{x=0}^1 e^{tx} p_X(x) = e^{t \cdot 0} p_X(0) + e^{t \cdot 1} p_X(1) = 1 - p + pe^t$$

$$E(X) = \left. \frac{d}{dt} M_X(t) \right|_{t=0} = pe^t(1-p + pe^t) \Big|_{t=0} = p$$

$$V(X) = \left. \frac{d^2}{dt^2} M_X(t) \right|_{t=0} - (E(X))^2 = p(1-p)$$

Distribuzione ipergeometrica

Una variabile casuale X con supporto

$$S_X = \{x \in \mathbb{N} : \max\{0, n - (N - M)\} \leq x \leq \min\{n, M\}\}$$

e funzione di massa di probabilità

$$p_X(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

ha distribuzione di probabilità **ipergeometrica** di parametri $M, N, n \in \mathbb{N}^+$, $M \leq N$ e $n \leq N$, in simboli

$$X \sim Ig(M, N, n)$$

È utile per rappresentare esperimenti casuali di estrazione di n palline *senza reinserimento* da un'urna con N palline totali, di cui M bianche. X conta il numero di bianche estratte.

Si ha che

$$E(X) = n \frac{M}{N}$$

$$V(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{(N-n)}{N-1}$$

Distribuzione geometrica

Una variabile casuale X con supporto

$$S_X = \mathbb{N}^+$$

e funzione di probabilità

$$p_X(x) = (1-p)^{x-1} p$$

ha distribuzione di probabilità **geometrica** con parametro p , in simboli

$$X \sim Ge(p)$$

In generale, una variabile casuale con distribuzione geometrica, descrive *il tempo di attesa* espresso come numero di replicazioni dell'esperimento (bernoulliano) di base, con probabilità p di successo, per osservare per la prima volta un successo.

Ad esempio si può considerare l'esperimento che consiste nel lanciare ripetutamente una moneta fino a che non si visualizza testa per la prima volta. In tal caso $p = \frac{1}{2}$ e il supporto è formato da un'infinità numerabile di elementi x_i dove $x_i =$ esce testa all' i -esimo lancio $= \{C \times C \times \dots \times T\}$, quindi $S_X = \mathbb{N}^+$, perché almeno un lancio è necessario e potenzialmente potrei vedere per la prima volta testa a numeri molto grandi.

La funzione di probabilità, per l'evento $x_i \in S_X$, invece è data dal prodotto di tutti i primi $i - 1$ lanci in cui è uscito croce e del i -esimo in cui è uscito testa, quindi

$$(1 - p)^{x-1}p$$

Se $X \sim Ge(p)$, allora, per $e^t < 1$, ovvero per $t < -\log(1 - p)$, si ha che

$$M_X(t) = \sum_{x \in S_X} e^{tx} p_X(x) = \sum_{x \in S_X} e^{tx} (1 - p)^{x-1} p = p e^t \sum_{x \in S_X} [e^{tx} (1 - p)]^{x-1} = \frac{p e^t}{1 - e^t (1 - p)}$$

Da ciò segue che

$$E(X) = \left. \frac{d}{dt} M_X(t) \right|_{t=0} = \frac{1}{p}$$

$$V(X) = \left. \frac{d^2}{dt^2} M_X(t) \right|_{t=0} - (E(X))^2 = \frac{1 - p}{p^2}$$

La distribuzione geometrica è caratterizzata dalla proprietà di assenza di memoria, definita dal seguente teorema:

Se $X \sim Ge(p)$, $p \in (0, 1)$, allora per ogni $s, t \in S_X$

$$P(X > s + t | X > s) = P(X > t)$$

Ossia: verificato l'evento $X > s$, la probabilità che successivamente si verifichi anche l'evento $X > t$ non è dipendente dal fatto che si sia verificato $X > s$, infatti la probabilità condizionale $P_{X>s}(X > t)$ è uguale alla semplice probabilità di $P(X > t)$, in pratica come se $X > s$ non fosse mai stato visualizzato, da cui assenza di memoria.

Distribuzione di Poisson

Una variabile casuale X con supporto e funzione di probabilità pari rispettivamente a

$$S_X = \mathbb{N}$$

$$p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

con $\lambda > 0$, ha distribuzione di probabilità di **Poisson** con parametro λ , in simboli

$$X \sim P(\lambda)$$

Questa distribuzione di probabilità è utile per descrivere *problemi di conteggio* senza una limitazione superiore per il supporto o quando la limitazione risulta irrilevante (cioè $n \gg E(X) \rightarrow n \geq 10E(X)$).

Per fare un esempio si consideri un numero molto grande $n \in \mathbb{N}^+$ di esperimenti *bernoulliani indipendenti* con una piccola probabilità p di successo uguale per tutti. La variabile casuale X conta il numero complessivo di successi, quindi si potrebbe assumere $X \sim Bi(n, p)$. In questo ambito, però si può approssimare la distribuzione binomiale con una distribuzione di Poisson. Infatti, se $p = \frac{\lambda}{n}$ con λ costante reale positiva, si verifica che il limite per $x \rightarrow +\infty$ della funzione di massa di probabilità di una variabile casuale con legge binomiale esiste finito e per $x \in \mathbb{N}$ corrisponde a

$$\lim_{x \rightarrow +\infty} p_X(x) = \lim_{x \rightarrow +\infty} \binom{n}{x} p^x (1 - p)^{n-x} = \frac{\lambda^x}{x!} e^{-\lambda}$$

Si ha che $\frac{\lambda^x}{x!} e^{-\lambda} > 0$ per ogni $x \in \mathbb{N}$ e che $\sum_{x=0}^{+\infty} \frac{\lambda^x}{x!} e^{-\lambda} = 1$.

L'approssimazione di una binomiale con la distribuzione di Poisson è possibile grazie al **teorema di Poisson** secondo il quale

se $X_n \sim Bi\left(n, \frac{\lambda}{n}\right)$, $x \in \mathbb{N}$ allora

$$p_{X_n}(x) = P(X_n = x) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \rightarrow \frac{\lambda^x}{x!} e^{-\lambda}$$

A parole, per n sufficientemente grande (ossia per $n \geq 10E(X)$) si può sostituire la probabilità binomiale calcolata dalla legge *esatta*

$$Bi\left(n, \frac{\lambda}{n}\right)$$

con la probabilità di Poisson, data da $P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$ poiché per $x \in \mathbb{N}$

$$\lim_{x \rightarrow +\infty} p_{X_n}(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Se una variabile casuale $X \sim P(\lambda)$ allora

$$M_X(t) = \sum_{x \in S_X} e^{xt} p_X(x) = \sum_{x \in S_X} e^{xt} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x \in S_X} \frac{[e^t \lambda]^x}{x!} = e^{-\lambda} \sum_{x \in S_X} \frac{[e^t \lambda]^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

$$E(X) = \left. \frac{d}{dt} M_X(t) \right|_{t=0} = (e^{\lambda(e^t - 1)} \lambda e^t) \Big|_{t=0} = \lambda$$

$$V(X) = \left. \frac{d^2}{dt^2} M_X(t) \right|_{t=0} - (E(X))^2 = (e^{\lambda(e^t - 1) + t} (\lambda e^t + 1)) \Big|_{t=0} - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Distribuzione uniforme continua

Una variabile casuale continua X ha distribuzione **uniforme continua** o **rettangolare**, in simboli

$$X \sim U(a, b)$$

se ha supporto

$$S_X = [a, b] \quad a, b \in \mathbb{R}, \quad a < b$$

e funzione di densità di probabilità

$$p_X(x) = \frac{1}{b - a}$$

La funzione di ripartizione vale

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x - a}{b - a} & a \leq x < b \\ 1 & x \geq b \end{cases}$$

Il valore atteso e la varianza sono

$$\begin{aligned} E(X) &= \int_a^b x \cdot p_X(x) dx = \int_a^b x \cdot \frac{1}{b - a} dx = \frac{1}{b - a} \int_a^b x dx = \frac{1}{b - a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b - a} \frac{b^2 - a^2}{2} \\ &= \frac{1}{b - a} \frac{(b - a)(b + a)}{2} = \frac{b + a}{2} \end{aligned}$$

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 = \frac{1}{b - a} \int_a^b x^2 dx - \left(\frac{b + a}{2} \right)^2 = \frac{1}{b - a} \left[\frac{x^3}{3} \right]_a^b - \left(\frac{b + a}{2} \right)^2 = \frac{1}{3} \frac{b^3 - a^3}{b - a} - \frac{(b + a)^2}{4} \\ &= \frac{4(b^3 - a^3) - 3(b + a)^2(b - a)}{12(b - a)} = \frac{4(b - a)(b^2 + a^2 + ab) - 3(b + a)^2(b - a)}{12(b - a)} \\ &= \frac{4(b^2 + a^2 + ab) - 3(b + a)^2}{12} = \frac{4a^2 + 4b^2 + 4ab - 3b^2 - 3a^2 - 6ab}{12} \\ &= \frac{a^2 - 2ab + b^2}{12} = \frac{(b - a)^2}{12} \end{aligned}$$

mentre la funzione generatrice dei momenti è

$$M_X(t) = \int_a^b e^{xt} p_X(x) dx = \int_a^b e^{xt} \frac{1}{b-a} dx = \frac{e^t}{b-a} \int_a^b e^x dx = \frac{[e^{tx}]_a^b}{t(b-a)} = \frac{e^{bt} - e^{at}}{t(b-a)}$$

Distribuzione esponenziale

Una variabile casuale X ha legge esponenziale

$$X \sim \text{Esp}(\lambda)$$

di parametro $\lambda > 0$ se

$$S_X = [0, +\infty)$$

e, per $x \in S_X$

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

La funzione di ripartizione è

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

La funzione generatrice dei momenti è

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} = \frac{\lambda}{\lambda - t} \quad \text{se } t < \lambda$$

Il valore atteso e la varianza invece sono

$$E(X) = \frac{1}{\lambda}$$

$$V(X) = \frac{1}{\lambda^2}$$

Come per la distribuzione geometrica, la distribuzione esponenziale ha la proprietà di assenza di memoria. Quindi:

Se X è una variabile casuale tale che $S_X = [0, +\infty)$ e $P(X = 0) < 1$, allora

$$X \sim \text{Esp}(\lambda) \Leftrightarrow \forall s, t \in S_X \quad P_{X>s}(s < X \leq s+t) = P(0 < X \leq t)$$

La funzione tasso di guasto e le leggi di Weibull

Sia X una variabile casuale *assolutamente continua* non negativa, cioè tale che $P(X \geq 0) = 1$, con supporto S_X . La funzione reale di variabile reale

$$r_X(x) = \frac{p_X(x)}{1 - F_X(x)}, \quad x \in S_X$$

funzione **tasso di guasto**.

Se la funzione $r_X(x)$ è:

- Crescente viene descritta una situazione con presenza di usura;
- Decrescente viene descritta una situazione con presenza di selezione infantile;
- Costante viene descritta una situazione con assenza di memoria.

Una variabile casuale X ha distribuzione di **Weibull**

$$X \sim \text{We}(\beta, \gamma)$$

con parametri $\beta, \gamma > 0$ se

$$S_X = [0, +\infty)$$

$$p_X(x) = \begin{cases} \beta \gamma x^{\gamma-1} \exp(-\beta x^\gamma) & x \in S_X \\ 0 & \text{altrimenti} \end{cases}$$

La distribuzione normale

Una variabile casuale X ha distribuzione **normale** con parametro di posizione $\mu \in \mathbb{R}$ e parametro di scala $\sigma > 0$, in simboli

$$X \sim N(\mu, \sigma^2)$$

se il supporto

$$S_X = \mathbb{R}$$

e la funzione di densità di probabilità è per ogni $x \in \mathbb{R}$

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

La distribuzione normale è utile per descrivere gli errori accidentali associati ad una misurazione. È stata studiata da Gauss, difatti è anche chiamata **gaussiana**.

Se $X \sim N(\mu, \sigma^2)$ la funzione generatrice dei momenti è

$$M_X(t) = \int_{-\infty}^{+\infty} e^{xt} p_X(x) dx = \int_{-\infty}^{+\infty} e^{xt} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

posto $z = \frac{x-\mu}{\sigma}$

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{t(z\sigma+\mu)-\frac{z^2}{2}} dz = e^{t\mu} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{tz\sigma-\frac{z^2}{2}} dz \\ &= e^{t\mu} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{z^2-2t\sigma z+(t\sigma)^2-(t\sigma)^2}{2}} dz = e^{t\mu} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\sigma t)^2-\sigma^2 t^2}{2}} dz \\ &= e^{t\mu} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\sigma t)^2}{2}} e^{\frac{\sigma^2 t^2}{2}} dz = e^{t\mu+\frac{\sigma^2 t^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\sigma t)^2}{2}} dz = e^{t\mu+\frac{\sigma^2 t^2}{2}} \cdot 1 \\ &= e^{t\mu+\frac{\sigma^2 t^2}{2}} \end{aligned}$$

Da ciò si possono ricavare il valore atteso e la varianza:

$$\begin{aligned} E(X) &= \frac{d}{dt} M_X(t) \Big|_{t=0} = \mu \\ V(X) &= \frac{d^2}{dt^2} M_X(t) \Big|_{t=0} - (E(X))^2 = \sigma^2 \end{aligned}$$

Dunque si ha che μ è il valor medio, la moda e la mediana di X , mentre σ^2 è la varianza e σ lo scarto quadratico medio di X .

Se $X \sim N(\mu, \sigma^2)$ si ha che $\frac{X-\mu}{\sigma} = Z \sim N(0, 1)$, detta anche distribuzione **normale standard**.

Normalmente, la funzione di ripartizione di Z si indica con $\Phi(z)$, mentre la funzione di densità di probabilità si indica con $\phi(z)$. Poiché $\phi(z)$ è simmetrica rispetto all'origine, si ha che

$$\Phi(-z) = 1 - \Phi(z) \quad \forall z \geq 0$$

In generale la funzione di ripartizione e densità di una variabile casuale $X \sim N(\mu, \sigma^2)$ sono indicate con $\Phi_{\mu, \sigma^2}(x)$ e $\phi_{\mu, \sigma^2}(x)$. Si conclude che

$$\begin{aligned} \Phi_{\mu, \sigma^2}(x) &= \Phi\left(\frac{x-\mu}{\sigma}\right) \\ \phi_{\mu, \sigma^2}(x) &= \sigma^{-1} \phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

Abbiamo quindi che

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Variabili casuali bivariate e multivariate con componenti indipendenti

Definizione generale di variabile casuale bivariata

Sia (Ω, \mathcal{F}, P) uno spazio probabilizzato e $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$ uno spazio probabilizzabile. Si definisce **variabile casuale bivariata** definita su Ω a valori in \mathbb{R}^2 , un'applicazione $(X, Y): \Omega \rightarrow \mathbb{R}^2$ misurabile rispetto alle σ -algebre \mathcal{F} e $\mathcal{B}_{\mathbb{R}^2}$ tale che per ogni $B \in \mathcal{B}_{\mathbb{R}^2}$ si ha che $(X, Y)^{-1}(B) \in \mathcal{F}$ dove

$$(X, Y)^{-1}(B) = \{\omega \in \Omega : (X, Y)(\omega) \in B\}$$

Definizione generale di variabile casuale multivariata

Sia (Ω, \mathcal{F}, P) uno spazio probabilizzato e $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ uno spazio probabilizzabile. Si definisce **variabile casuale multivariata** definita su Ω a valori in \mathbb{R}^n , un'applicazione $X: \Omega \rightarrow \mathbb{R}^n$ misurabile rispetto alle σ -algebre \mathcal{F} e $\mathcal{B}_{\mathbb{R}^n}$ tale che per ogni $B \in \mathcal{B}_{\mathbb{R}^n}$ si ha che $X^{-1}(B) \in \mathcal{F}$

Poiché il caso bivariato è un caso particolare del multivariato, verrà visto principalmente quest'ultimo.

Data una variabile casuale multivariata (X_1, \dots, X_n) , il supporto S_{X_1, \dots, X_n} è chiamato **supporto congiunto** ed è l'insieme dei vettori (punti, nel bivariato) $(x_1, \dots, x_n) \in \mathbb{R}^n$ i cui intorno sono eventi di probabilità strettamente positiva rispetto a P_{X_1, \dots, X_n} .

La **funzione di ripartizione (congiunta)** di una variabile casuale multivariata è

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

Data una variabile casuale multivariata X , le sue componenti marginali sono dette **indipendenti** se e solo se per ogni $(x_1, \dots, x_n) \in \mathbb{R}^n$

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$$

Se invece esiste almeno un vettore $(x_1, \dots, x_n) \in \mathbb{R}^n$ tale che $F_{X_1, \dots, X_n}(x_1, \dots, x_n) \neq F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$ allora le componenti di X vengono dette **dipendenti**.

Un *esempio di variabile casuale bivariata (multivariata) a componenti indipendenti* è la variabile che calcola la somma dei punteggi ottenuti nel lancio di 2 (n) dadi indipendenti.

Come conseguenze dell'indipendenza abbiamo che

- $S_X = S_{X_1, \dots, X_n} = S_{X_1} \times S_{X_2} \times \dots \times S_{X_n}$
- $p_X(x) = p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i)$
- $E(X_1 \cdot \dots \cdot X_n) = \prod_{i=1}^n E(X_i)$

Dimostrazione:

$$E(X_1 \cdot \dots \cdot X_n) = \sum_{x \in S_X} \prod_{i=1}^n x_i p_X(x)$$

se vale l'indipendenza abbiamo che

$$E(X_1 \cdot \dots \cdot X_n) = \prod_{i=1}^n \sum_{x \in S_X} x_i p_{X_i}(x_i) = \prod_{i=1}^n E(X_i)$$

Legge di probabilità di massimo, minimo e somma di variabili casuali indipendenti

Sia, $T = \max_{i=1 \dots n} x_i$, $U = \min_{i=1 \dots n} x_i$ e $S_n = \sum_{i=1}^n x_i$ e sia *propria* la funzione generatrice dei momenti delle componenti marginali $M_{X_i}(t) = E(e^{tx_i})$

$$F_T(t) = P(T \leq t) = P(\max(x_1 \dots x_n) \leq t) = P(X_1 \leq t \dots X_n \leq t) = F_{X_1, \dots, X_n}(t \dots t) = \prod_{i=1}^n F_{X_i}(t)$$

$$F_U(t) = 1 - F_T(t) = 1 - \prod_{i=1}^n F_{X_i}(t)$$

$$M_{S_n}(t) = E(e^{tS_n}) = E(e^{t \sum_{i=1}^n x_i}) = E(e^{tx_1 + \dots + tx_n}) = E(e^{tx_1} \dots e^{tx_n}) = \prod_{i=1}^n E(e^{tx_i}) = \prod_{i=1}^n M_{X_i}(t)$$

La distribuzione della media campionaria di normali indipendenti e identicamente distribuite

Si dice che una variabile casuale n -variata $X = (X_1 \dots X_n)$ descrive un esperimento di campionamento casuale semplice con numerosità n da una variabile casuale con legge assegnata (quella di X_i) se le componenti X_i sono indipendenti e identicamente distribuite (con la stessa distribuzione).

Ad esempio si consideri un campionamento casuale semplice dalla variabile n -variata $X \sim N(\mu, \sigma^2)$, $X = (X_1 \dots X_n)$ tale che

$$X_i \sim N(\mu, \sigma^2) \quad i = 1 \dots n$$

e che X_i siano indipendenti.

Se la variabile casuale *media* campionaria $\bar{X}_n = \frac{1}{n} S_n$ dove $S_n = \sum_{i=1}^n x_i \sim N(n\mu, n\sigma^2)$ è la variabile *somma*, allora la funzione generatrice dei momenti di \bar{X}_n è

$$M_{\bar{X}_n}(t) = E(e^{t\bar{X}_n}) = E\left(e^{\left(\frac{t}{n}\right)S_n}\right) = M_{S_n}\left(\frac{t}{n}\right)$$

Si consideri che $M_{S_n}(t) = e^{tn\mu + \frac{1}{2}t^2n\sigma^2}$, quindi

$$M_{\bar{X}_n}(t) = e^{t\mu + \frac{t^2}{2n}\sigma^2} \rightarrow \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Distribuzioni marginali di una variabile casuale con legge discreta

Si supponga una variabile casuale bivariata (X, Y) di cui si conosce il *supporto congiunto* $S_{X,Y}$, la *funzione di ripartizione congiunta* $F_{X,Y}$ e la *funzione di probabilità congiunta* $p_{X,Y}$. Il **supporto marginale** S_X si ricava dal supporto congiunto considerando tutti i possibili valori della componente X , al variare delle potenziali determinazioni della componente Y , ossia

$$S_X = \{x \in \mathbb{R} : \exists y \in \mathbb{R} : (x, y) \in S_{X,Y}\}$$

$$S_Y = \{y \in \mathbb{R} : \exists x \in \mathbb{R} : (x, y) \in S_{X,Y}\}$$

La **funzione di ripartizione marginale** F_X si ottiene a partire dalla funzione di ripartizione congiunta utilizzando la relazione

$$F_X(x) = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y)$$

Se la variabile è discreta con supporto $S_{X,Y} = \{(x_i, y_j) \in \mathbb{R}^2 : (i, j) \in K\}$ $K \subseteq \mathbb{N}^+ \times \mathbb{N}^+$, allora anche le distribuzioni marginali sono discrete con supporti marginali

$$S_X = \{x_i \in \mathbb{R} : i \in I\}$$

$$S_Y = \{y_j \in \mathbb{R} : j \in J\}$$

$$I, J \subseteq \mathbb{N}^+$$

In questo caso, si può vedere che per ogni $x_i \in S_X$ e $y_j \in S_Y$ fissati si ha

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

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

Quindi $P(X = x_i)$ e $P(Y = y_j)$ sono pari alla somma delle probabilità associate alle coppie $(x, y) \in S_{X,Y}$ avendo fissato $x = x_i$ (nella prima) e $y = y_j$ (nella seconda). p_{i+} indica la somma fissato l'indice fissato i (idem per p_{+j}). Grazie a queste formule è immediato definire le **funzioni di probabilità marginali**.

Le leggi condizionali

Leggi di Y condizionali a $X = x$

$$P(X = x) > 0$$

Leggi di X condizionali a $Y = y$

$$P(Y = y) > 0$$

Per $x \in S_X$ la legge di $Y|X = x$ ha

- Supporto $S_{Y|X=x} = \{y \in \mathbb{R} : (x, y) \in S_{X,Y}\}$
- Funzione di massa di probabilità $p_{Y|X=x}(y) = \frac{p_{XY}(x,y)}{p_X(x)}$

Per $y \in S_Y$ la legge di $X|Y = y$ ha

- Supporto $S_{X|Y=y} = \{x \in \mathbb{R} : (x, y) \in S_{X,Y}\}$

- Funzione di massa di probabilità $p_{X|Y=y}(x) = \frac{p_{XY}(x,y)}{p_Y(y)}$

Ad esempio se il supporto congiunto fosse

$$S_{XY} = \{(0,0), (0,1), (1,1)\}$$

$$S_{Y|X=0} = \{0,1\}$$

$$S_{Y|X=1} = \{1\}$$

$$S_{X|Y=0} = \{0\}$$

$$S_{X|Y=1} = \{0,1\}$$

Proprietà additive

(X, Y) variabile casuale bivariata con componenti indipendenti

$$X \sim D(x_0) \quad Y \sim D(y_0)$$

$$S = X + Y \sim D(x_0 + y_0)$$

Infatti:

$$M_S(t) = M_X(t)M_Y(t) = e^{tx_0}e^{ty_0} = e^{t(x_0+y_0)}$$

funzione generatrice dei momenti propria della variabile casuale $S \sim D(x_0 + y_0)$

$$X \sim Bi(m, p) \quad Y \sim Bi(n, p)$$

$$S = X + Y \sim Bi(m + n, p)$$

Infatti:

$$M_S(t) = M_X(t)M_Y(t) = (1 - p + pe^t)^m(1 - p + pe^t)^n = (1 - p + pe^t)^{m+n}$$

funzione generatrice dei momenti propria della variabile casuale $S \sim Bi(m + n, p)$

$$X \sim P(\lambda) \quad Y \sim P(\mu)$$

$$S = X + Y \sim P(\lambda + \mu)$$

Infatti:

$$M_S(t) = M_X(t)M_Y(t) = e^{\lambda(e^t-1)}e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$$

funzione generatrice dei momenti propria della variabile casuale $S \sim P(\lambda + \mu)$

$$X \sim N(\mu_x, \sigma_x^2) \quad Y \sim N(\mu_y, \sigma_y^2)$$

$$S = X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

Infatti:

$$M_S(t) = M_X(t)M_Y(t) = e^{t\mu_x + \frac{\sigma_x^2 t^2}{2}} e^{t\mu_y + \frac{\sigma_y^2 t^2}{2}} = e^{t\mu_x + \frac{\sigma_x^2 t^2}{2} + t\mu_y + \frac{\sigma_y^2 t^2}{2}} = e^{t(\mu_x + \mu_y) + \frac{(\sigma_x^2 + \sigma_y^2)t^2}{2}}$$

funzione generatrice dei momenti propria della variabile casuale $S \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$