

6.2.1	Definizione e proprietà . . . . .	65
6.2.2	V.c. equivalenti: l'identica distribuzione . . . . .	66
6.3	Prime leggi notevoli . . . . .	66
6.4	Le leggi discrete in generale . . . . .	67
6.4.1	Tabelle che rappresentano leggi discrete . . . . .	68
6.5	Le leggi uniformi discrete . . . . .	69

7.1	Definizione . . . . .	75
7.2	Le leggi marginali . . . . .	76
7.3	Le leggi condizionali . . . . .	77
7.4	La specificazione gerarchica . . . . .	78
7.5	V.c. discrete con componenti indipendenti . . . . .	80
7.6	La covarianza . . . . .	81
7.7	Esercizi risolti	82

9.1	Definizione . . . . .	103
9.2	Le probabilità di intervalli . . . . .	104
9.3	Relazioni tra f.r. e f.m.p. o f.d.p. . . . .	105
9.3.1	Da $p_X(x)$ a $F_X(x)$ . . . . .	105
9.3.2	Nel discreto: dalla f.r. alla f.m.p. . . . .	106
9.3.3	Nel continuo: dalla f.r. alla f.d.p. . . . .	106
9.4	Esercizi risolti	107

17.1	Genesi e definizione . . . . .	189
17.2	Chiusura sotto trasformazioni affini . . . . .	191
17.3	F.g.m. e momenti di una normale . . . . .	192
17.4	Additività delle normali indipendenti . . . . .	193
17.5	Esercizi risolti . . . . .	194
17.6	Esercizi . . . . .	196

**Definizione 1.1** Costituisce un **esperimento casuale**, indicato con  $\mathcal{E}$ , ogni attività, ripetibile essenzialmente nelle stesse condizioni secondo preassegnate regole, che, se eseguita, produce un risultato. Il risultato di  $\mathcal{E}$  (anch'esso rilevato secondo regole predeterminate) è incerto, perché generalmente varia al ripetersi delle prove.

**Definizione 1.2** Lo spazio campionario di un esperimento casuale  $\mathcal{E}$ , indicato con  $S$ , è l'insieme di tutti i possibili risultati  $s$  di  $\mathcal{E}$ ,  $S = \{s\}$ .

Quale  $s$  in  $S$  si realizzerà è incerto:  $\mathcal{E} \rightarrow ?$ . È tuttavia certo che  $? \in S$ :  $S$  comprende tutti i possibili risultati macroscopici dell'esperimento.

**Definizione 1.3** Si dice **evento**, indicato con  $E$  (ma anche con  $A$ ,  $B$ , o, usando pedici,  $A_1$ ,  $A_2$ , eccetera) un sottoinsieme di  $S$ ,  $E \subseteq S$ .

Ogni evento è dunque una collezione di possibili risultati di  $\mathcal{E}$ , da tutti a nessuno.

Ad  $E$  corrisponde la predizione che si osserverà un  $s \in E$ . Tradurre in un sottoinsieme di  $S$  una predizione espressa in forma verbale richiede molta

**Definizione 1.4** Dato un evento  $E \subseteq S$ , si dice che  $E$  è realizzato se è osservato un  $s \in E$ , altrimenti si dice che  $E$  non è realizzato.

Si capisce ora meglio perché l'evento  $S$  è certo, l'evento  $\emptyset$  impossibile:  $S$  si realizza sempre,  $\emptyset$  mai.

**Esempio 1.3** (Lancio di un dado) Si consideri l'esperimento casuale  $\mathcal{E} =$

**Definizione 2.1** Due eventi  $A$  e  $B$  per cui  $A \cap B = \emptyset$  sono detti **incompatibili**, o anche **mutuamente esclusivi**.

Se  $A$  e  $B$  non sono incompatibili, si dicono **compatibili**. Si ha allora  $A \cap B \supset \emptyset$ .  
Si può generalizzare l'incompatibilità a una successione di eventi.

**Definizione 2.2** Una successione di eventi  $A_i$ ,  $i \in I \subseteq \mathbb{N}$ , si dice costituita da eventi a due a due incompatibili, o anche mutuamente esclusivi, se

$$i, j \in I \text{ con } i \neq j \implies A_i \cap A_j = \emptyset.$$

**Definizione 2.3** La classe degli eventi è la collezione di tutti gli eventi di cui ha senso parlare nel contesto del dato esperimento casuale con spazio campionario  $S$ . È indicata con  $\mathcal{F}$  (o  $\mathcal{B}$ ).

Ovviamente  $\mathcal{F} \subseteq \mathcal{P}(S)$ , dove  $\mathcal{P}(S)$  è l'insieme delle parti di  $S$ . Se  $\mathcal{F} \subset \mathcal{P}(S)$ ,  $\mathcal{F}$  dovrà essere ‘simile’ a  $\mathcal{P}(S)$ : le usuali operazioni fatte su eventi in  $\mathcal{F}$  dovranno

un modo semplice per valutare  $P(E)$  è seguire la cosiddetta **definizione classica** della probabilità (Laplace, 1812). Per essa, la probabilità  $P(E)$  di un evento  $E$  si ottiene rapportando il numero di risultati dell'esperimento che comportano che  $E$  sia realizzato (casi favorevoli) al numero complessivo dei risultati possibili (casi possibili), supponendo che tutti i risultati siano ‘egualmente possibili’.

Quindi

$$P(E) = \frac{|E|}{|\Omega|}$$

La seguente definizione riassume i passi fin qui fatti per assiomatizzare il Calcolo delle Probabilità.

**Definizione 2.4** *La coppia ordinata  $(S, \mathcal{F})$ , dove  $S$  è non vuoto e  $\mathcal{F}$  è una  $\sigma$ -algebra di parti di  $S$ , è detta spazio probabilizzabile.*

**Definizione 2.4** La coppia ordinata  $(S, \mathcal{F})$ , dove  $S$  è non vuoto e  $\mathcal{F}$  è una  $\sigma$ -algebra di parti di  $S$ , è detta spazio probabilizzabile.

### 2.5.3 Misure di probabilità

Gli assiomi del Calcolo delle Probabilità sono esplicati nella seguente definizione.

Gli assiomi del Calcolo delle Probabilità sono espressi nella seguente definizione che vincola il comportamento delle applicazioni  $P : \mathcal{F} \rightarrow \mathbb{R}$  utilizzabili per probabilizzare gli eventi di una  $\sigma$ -algebra  $\mathcal{F}$ .

**Definizione 2.5** *Dato lo spazio probabilizzabile  $(S, \mathcal{F})$ , si dice misura di probabilità su  $\mathcal{F}$  una applicazione*

**Definizione 2.5** Dato lo spazio probabilizzabile  $(S, \mathcal{F})$ , si dice misura di probabilità su  $\mathcal{F}$  una applicazione

$$P : \mathcal{F} \rightarrow [0, 1]$$

che soddisfi le tre proprietà, dette assiomi di Kolmogorov:

$$\begin{aligned} A \in \mathcal{F}, P(A) > 0 \quad & \Rightarrow \exists A' \in \mathcal{T}_A \text{ s.t. } A \subseteq A' \\ A \in \mathcal{F}, P(A) = 0 \quad & \Rightarrow \forall A' \in \mathcal{T}_A \text{ s.t. } A \subseteq A' \end{aligned}$$

La seguente definizione esprime in forma sintetica le possibili modellazioni probabilistiche.

**Definizione 2.6** *Sia  $(S, \mathcal{F})$  uno spazio probabilizzabile e  $P$  una misura di probabilità su  $\mathcal{F}$ . La terna ordinata  $(S, \mathcal{F}, P)$  è detta spazio probabilizzato.*

Le incertezze in gioco nell'esperimento casuale  $\mathcal{E}$  sono completamente de-

**Definizione 2.6** Sia  $(S, \mathcal{F})$  uno spazio probabilizzabile e  $P$  una misura di probabilità su  $\mathcal{F}$ . La terna ordinata  $(S, \mathcal{F}, P)$  è detta **spazio probabilizzato**.

Le incertezze in gioco nell'esperimento casuale  $\mathcal{E}$  sono completamente descritte dalla modellazione con un particolare spazio probabilizzato. Quindi il processo di realizzazione  $\mathcal{E} \rightarrow ?$  è ora descritto da  $(S, \mathcal{F}, P)$  per opportuni  $S$ ,  $\mathcal{F}$ ,  $P$ . Questo è lo schema astratto. In pratica, la modellazione più delicata è

Definizione 2.5, e non solo per la probabilità classica per cui sono ovvi, valgono i seguenti teoremi.

**Teorema 3.1**  $P(\emptyset) = 0$ .

*Dimostrazione.* Poiché  $S = S \cup \emptyset$ , con  $S$  e  $\emptyset$  eventi incompatibili ( $S \cap \emptyset = \emptyset$ ), segue da **A3** che  $P(S) = P(S) + P(\emptyset)$ , dove  $P(S) = 1$  per **A2**.  $\square$

anti  $A_i$ ,  $i \in I$

**Definizione 3.2** Una v.c.  $X$  con possibili realizzazioni nell'insieme  $S_X = \{x \in \mathbb{N} : \max(0, n - N + D) \leq x \leq \min(n, D)\}$ , come (3.1), tale che ogni realizzazione ha probabilità  $P(X = x) = P(E_x)$  come in (3.2), si dice che ha legge ipergeometrica con indice  $n$  e parametri  $D$  e  $N$ . Si scrive in breve  $X \sim IG(n; D, N)$ , dove  $0 < n < N$  e  $0 < D < N$ .

**Definizione 4.1** Sia  $A$  un evento non trascurabile, detto condizione,  $E$  un altro evento. Si dice probabilità di  $E$  condizionale ad  $A$ , e si indica con  $P(E | A)$ , il valore

$$P(E | A) = \frac{P(E \cap A)}{P(A)}. \quad (4.4)$$

Non conviene memorizzare la definizione con nomi specifici degli eventi, meglio il motto **probabilità dell'intersezione fratto probabilità della condizione**.

Se  $A$  è realizzato, o pensato realizzato, la probabilità di  $E$  pertinente alla nuova situazione, reale o ipotizzata, non è più  $P(E)$  ma  $P(E | A)$ . Il condizionamento ad  $A$  fa passare dall'esperimento  $\mathcal{E}$  all'esperimento  $\mathcal{E}_A$  che ha le stesse regole di  $\mathcal{E}$  ma in cui si ‘squalificano’ tutte le replicazioni che non realizzano  $A$ . Per l'esperimento  $\mathcal{E}_A$ , l'evento  $A$  assume il ruolo di evento certo. In tale

situazione,  $E$  si realizza se e solo se si realizza un  $s$  in  $E \cap A$ . La definizione (4.4) rinormalizza  $P(E \cap A)$  semplicemente riproporzionandola a  $P(A)$ , di modo che sia  $P(A | A) = 1$ , come pure, per  $E \supseteq A$ , anche  $P(E | A) = 1$ . Dal punto di vista frequentista, fatte  $R$  replicazioni di  $\mathcal{E}$ , la probabilità  $P(E | A)$  relativa all'esperimento  $\mathcal{E}_A$  è approssimata da

$$\frac{n^0 \text{ realizzazioni in } E \cap A}{n^0 \text{ realizzazioni in } E \cap A} = P(E \cap A)$$

disfatti. La non-negatività di  $P(E \mid A)$  è ovvia dalla definizione (4.4). La normalizzazione segue da  $P(S \mid A) = P(S \cap A)/P(A) = 1$ . Si verifica la  $\sigma$ -additività, per una successione finita o numerabile di eventi  $E_i \in \mathcal{F}$  a due a due incompatibili, considerando che

$$P(\cup_{i \in I} E_i \mid A) = \frac{P((\cup_{i \in I} E_i) \cap A)}{P(A)} = \frac{P(\cup_{i \in I} (E_i \cap A))}{P(A)} = \frac{\sum_{i \in I} P(E_i \cap A)}{P(A)}$$

pone la seguente definizione.

**Definizione 4.2** (Indipendenza di due eventi) *A e B in  $\mathcal{F}$  sono detti eventi indipendenti se  $P(A \cap B) = P(A)P(B)$ .*

**Definizione 4.2 (Indipendenza di due eventi)** *A e B in  $\mathcal{F}$  sono detti eventi indipendenti se  $P(A \cap B) = P(A)P(B)$ .*

Si noti che l'evento impossibile è indipendente da qualunque evento. Anche ogni evento trascurabile è indipendente da qualunque evento. Per più di due eventi si ha la seguente definizione.

eventi si ha la seguente definizione.

**Definizione 4.3 (Indipendenza degli eventi di una successione)** La successione  $A_i$ ,  $i \in I$  con  $|I| > 2$ , è una successione di eventi indipendenti se per ogni  $J \subseteq I$ ,  $J$  con cardinalità finita pari almeno a 2, vale

**Definizione 4.3 (Indipendenza degli eventi di una successione)** La successione  $A_i, i \in I$  con  $|I| > 2$ , è una successione di eventi indipendenti se per ogni  $J \subseteq I$ ,  $J$  con cardinalità finita pari almeno a 2, vale

$$P(\cap_{i \in J} A_i) = \prod_{i \in J} P(A_i).$$

**Definizione 4.4** La variabile casuale  $X$ , con possibili realizzazioni in

$$S_X = \{x \in \mathbb{N} : 0 \leq x \leq n\} = \{0, 1, \dots, n\}, \quad (4.8)$$

tale che ogni realizzazione ha probabilità  $P(X = x) = P(E_x)$  come in (4.7), viene detta **con legge binomiale con parametri  $n$  e  $p$** . Si scrive in breve  $X \sim Bi(n, p)$ , dove  $n \in \mathbb{N}^+ = \{1, 2, \dots\}$  e  $0 < p < 1$ .

*Dimostrazione.* Per definizione,  $P(A_i | E) = P(A_i \cap E) / P(E)$ . La formula della probabilità composta, Teorema 4.3, dà  $P(A_i \cap E) = P(A_i) P(E | A_i)$ . La formula della probabilità totale, Teorema 4.4, dà  $P(E) = \sum_{j \in I} P(A_j) P(E | A_j)$ .  $\square$

Oltre alle probabilità iniziali e alle probabilità aggiornate, compaiono nella formula i valori  $P(E | A_i)$ ,  $i \in I$ . Nel contesto della formula di Bayes  $E$  è supposto realizzato, per cui i valori  $P(E | A_i)$  appaiono come delle probabilità solo per

descrivere la nozione precisa di v.c.. Si dà la seguente definizione.

**Definizione 6.1** Sia  $(S, \mathcal{F}, P)$  uno spazio probabilizzato. Una variabile casuale  $d$ -variata è una applicazione

$$X : S \rightarrow \mathbb{R}^d$$

misurabile, ossia tale che per ogni  $B \in \mathcal{B}_d$  si ha  $X^{-1}(B) \in \mathcal{F}$ .

## 6.2.1 Definizione e proprietà

Le v.c. sono descritte dalla loro legge di probabilità.

**Definizione 6.2** *La legge di probabilità di una v.c.  $d$ -variata  $X$  è una applicazione  $P_X : \mathcal{B}_d \rightarrow [0, 1]$  definita, per ogni  $B \in \mathcal{B}_d$ , da*

$$P_X(B) = P(X^{-1}(B)). \quad (6.1)$$

**Definizione 6.2** La legge di probabilità di una v.c.  $d$ -variata  $X$  è una applicazione  $P_X : \mathcal{B}_d \rightarrow [0, 1]$  definita, per ogni  $B \in \mathcal{B}_d$ , da

$$P_X(B) = P(X^{-1}(B)). \quad (6.1)$$

**Osservazione 1.** La misurabilità di  $X$  è richiesta affinché la legge di probabilità sia ben definita. Definire i  $\mathcal{B}_d$  è l'argomento

**Definizione 6.3** Due v.c.  $X$  e  $Y$  si dicono **identicamente distribuite**, e si scrive  $X \sim Y$ , se  $P_X = P_Y$ , ossia per ogni  $B \in \mathcal{B}_d$  vale  $P_X(B) = P_Y(B)$ .

Nel Calcolo delle Probabilità le v.c. identicamente distribuite sono identificate. Quindi non importa come è definita una v.c. come applicazione, importa solo la sua legge di probabilità. È come dire che solo le probabilità importano nel

**Definizione 6.4** Si dice che una v.c.  $d$ -variata  $X$  ha legge degenera in  $x_0 \in \mathbb{R}^d$ , valore prefissato, e si scrive  $X \sim \mathcal{D}(x_0)$ , se per  $B \in \mathcal{B}_d$  la legge di probabilità di  $X$  è

$$P_X(B) = P(X \in B) = \begin{cases} 1 & \text{se } x_0 \in B \\ 0 & \text{se } x_0 \notin B. \end{cases}$$

**Definizione 6.5** Si dice che la v.c. univariata  $X$  ha legge binomiale con indice  $n \in \mathbb{N}^+$  e parametro  $p \in (0, 1)$ , e si scrive  $X \sim Bi(n, p)$ , se per ogni  $B \in \mathcal{B}_1$  vale

$$P_X(B) = P(X \in B) = \sum_{x \in B \cap S_X} \binom{n}{x} p^x (1-p)^{n-x},$$

**Definizione 6.6** Si dice che la v.c. univariata  $X$  ha legge ipergeometrica con indice  $n \in \mathbb{N}^+$  e parametri  $N$  e  $D$ , dove  $1 \leq n \leq N \in \mathbb{N}^+$  e  $D \in \mathbb{N}^+$  con  $1 \leq D \leq N$ , e si scrive in breve  $X \sim IG(n; D, N)$ , se vale per ogni  $B \in \mathcal{B}_1$

$$P_X(B) = P(X \in B) = \sum_{x \in B \cap S_X} \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}},$$

**Definizione 6.7** Dato il supporto discreto  $S_X$  e la f.m.p.,  $p_X(x)$  positiva sul supporto e normalizzata, la legge discreta di  $X$  è definita da

$$P_X(B) = P(X \in B) = \sum_{x \in B \cap S_X} p_X(x). \quad (6.2)$$

**Definizione 6.8** La v.c.  $d$ -variata  $X$  ha legge uniforme discreta in  $D = \cup_{i=1}^k \{x_i\}$ ,  $x_i \in \mathbb{R}^d$ , dove gli  $x_i$  sono  $k$  punti distinti, e si scrive in breve  $X \sim Ud(x_1, \dots, x_k)$ , se

- $S_X = D$
- $p_X(x) = \frac{1}{k}$  per ogni  $x \in S_X$ .

## 7.1 Definizione

Si adotta spesso per una v.c. bivariata, cioè per un punto in  $\mathbb{R}^2$  determinato dal caso, la notazione  $(X, Y)$ , ispirata alla notazione  $(x, y)$  per le coordinate di un generico punto del piano cartesiano. Si parla di supporto congiunto, legge congiunta, eccetera, della v.c. bivariata per sottolineare il fatto che viene modellata la variabilità simultanea delle v.c. univariate  $X$  e  $Y$ , componenti di  $(X, Y)$ .

$$= p_X(x)p_{Y|X=x}(y). \quad (\text{per definizione di f.m.p.})$$

discreto si ha la seguente definizione.

**Definizione 7.1** *La v.c. con legge discreta  $(X, Y)$  si dice **con componenti indipendenti** se*

$$S_{X,Y} = S_X \times S_Y \quad (7.2)$$

*e se inoltre per ogni  $(x, y) \in S_{X,Y}$  si ha*

**Definizione 7.1** La v.c. con legge discreta  $(X, Y)$  si dice **con componenti indipendenti** se

$$S_{X,Y} = S_X \times S_Y \quad (7.2)$$

e se inoltre per ogni  $(x, y) \in S_{X,Y}$  si ha

$$p_{X,Y}(x, y) = p_X(x)p_Y(y). \quad (7.3)$$

**Definizione 8.1** Si dice che  $X$  è una v.c. univariata con legge di tipo continuo (in breve: con legge continua; ancora più in breve: una v.c. univariata continua) se per ogni  $B \in \mathcal{B}_1$  si può esprimere  $P_X(B)$  in forma integrale come

$$P_X(B) = P(X \in B) = \int_B p_X(x) dx = \int_a^b p_X(x) dx \quad (8.1)$$

**Definizione 8.2** Si dice che  $X$  ha legge uniforme continua in  $(a, b)$ , dove  $a < b$ , e si scrive in breve  $X \sim U(a, b)$ , se la f.d.p. di  $X$  è

$$p_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in (a, b) \\ 0 & \text{se } x \notin (a, b). \end{cases}$$

- Definizione:

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n, \quad x \in \mathbb{R}. \quad (8.2)$$

- Proprietà delle potenze:

$$e^0 = 1$$

$$e^{a+b} = e^a e^b$$

Si è ora pronti ad esaminare la definizione delle leggi esponenziali.

**Definizione 8.3** Si dice che  $X$  ha legge esponenziale con parametro  $\lambda > 0$ , e si scrive in breve  $X \sim Esp(\lambda)$ , se la f.d.p. di  $X$  è

$$\pi(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

**Definizione 8.3** Si dice che  $X$  ha legge esponenziale con parametro  $\lambda > 0$ , e si scrive in breve  $X \sim Esp(\lambda)$ , se la f.d.p. di  $X$  è

$$p_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0. \end{cases}$$

**Osservazione.** Conviene memorizzare la definizione delle leggi esponenziali tramite la funzione di sopravvivenza:

$$X \sim Esp(\lambda), \lambda > 0, \iff \text{per ogni } x > 0 \quad P(X > x) = e^{-\lambda x}. \quad (8.5)$$

Il valore atteso di  $X \sim Esp(\lambda)$  è  $E[X] = 1/\lambda$ . Infatti

**Definizione 8.4** Sia  $X$  una v.c. univariata. Il supporto di  $X$ , indicato con  $S_X$ , è il più piccolo sottoinsieme chiuso di  $\mathbb{R}$  al quale  $P_X$  assegna probabilità 1.

Quindi  $S_X \in \mathcal{B}_1$  è tale che:

- i)  $S_X$  è chiuso

La Definizione 8.4 è la definizione generale di supporto per una v.c. univariata. Estende la definizione del caso discreto data nella Unità 8, Si generalizza poi la nozione di supporto a v.c. bivariate e multivariate considerando i sottoinsiemi chiusi di  $\mathbb{R}^d$ .

La Definizione 8.4 è la definizione generale di supporto per una v.c. univariata. Estende la definizione del caso discreto data nella Unità 8, Si generalizza poi la nozione di supporto a v.c. bivariate e multivariate considerando i sottoinsiemi chiusi di  $\mathbb{R}^d$ .

ta. Estende la definizione del caso discreto data nella Unità 8, Si generalizza poi la nozione di supporto a v.c. bivariate e multivariate considerando i sottoinsiemi chiusi di  $\mathbb{R}^d$ .

**Definizione 8.5** Una v.c. bivariata  $(X, Y)$  con legge continua ha componenti indipendenti se per ogni  $(x, y) \in \mathbb{R}^2$  vale

$$p_{X,Y}(x, y) = p_X(x) p_Y(y).$$

Vale allora anche che  $S_{X,Y} = S_X \times S_Y$ . Inoltre si ha

**Definizione 8.6** Una v.c. multivariata  $X = (X_1, \dots, X_d)$  con legge continua ha componenti indipendenti se per ogni  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  vale

$$p_X(x) = p_{X_1, \dots, X_d}(x_1, \dots, x_d) = p_{X_1}(x_1) \cdots p_{X_d}(x_d).$$

Si ha allora

## 9.1 Definizione

Cii si occuperà principalmente di funzioni di ripartizione (f.r.) di v.c. univariate. Poco costa tuttavia dare la definizione generale, per una v.c. multivariata con legge di probabilità  $P_X$ .

**Definizione 9.1** Si dice funzione di ripartizione di  $X = (X_1, \dots, X_d)$  la

riate. Poco costa tuttavia dare la definizione generale, per una v.c. multivariata con legge di probabilità  $P_X$ .

**Definizione 9.1** Si dice funzione di ripartizione di  $X = (X_1, \dots, X_d)$  la funzione

$$F_{\mathbf{X}} : \mathbb{R}^d \rightarrow [0, 1]$$

**Definizione 9.1** Si dice funzione di ripartizione di  $X = (X_1, \dots, X_d)$  la funzione

$$F_X : \mathbb{R}^d \rightarrow [0, 1]$$

che a ciascun punto  $x = (x_1, \dots, x_d)$  di  $\mathbb{R}^d$  fa corrispondere il valore d'immagine

completi la definizione della funzione di densità di  $X$ , determinando il valore della costante di normalizzazione  $c$ . Si calcoli la funzione di ripartizione di  $X$ , esplicitandola in tutti i suoi tratti. Si ottengano valore atteso e varianza di  $X$ . La costante di normalizzazione è determinata dalla condizione di normalizzazione:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad \Rightarrow \quad \int_0^c x^2 dx = 1$$

*completi la definizione della funzione di densità di  $X$ , determinando il valore della costante di normalizzazione  $c$ . Si calcoli la funzione di ripartizione di  $X$ , esplicitandola in tutti i suoi tratti.*

La costante di normalizzazione è determinata dalla condizione di normalizzazione:

$$1 = \int_{-\infty}^{+\infty} f(x) dx = \int_{-x_1}^1 c x^{-\frac{1}{2}} dx = c \left[ -x^{\frac{1}{2}} \right]_{-x_1}^1 = c \left( -\sqrt{1} + \sqrt{-x_1} \right)$$

Si completi la definizione della funzione di densità di  $X$ , determinando il valore della costante  $c$ . Si calcoli la funzione di ripartizione di  $X$ , esplicitandola in tutti i suoi tratti. Si ottenga il valore atteso di  $X$ . Si calcoli  $P(X = 0)$ .

completi la definizione della funzione di densità di  $X$ , determinando il valore della costante di normalizzazione  $c$ . Si calcoli la funzione di ripartizione di  $X$ , esplicitandola in tutti i suoi tratti. Si calcoli  $P(0.4 < X < 0.6)$ . Si ottengano valore atteso e la varianza di  $X$ .

**Esercizio 9.6** Sia  $X$  una variabile casuale con supporto  $S_X = [-1, 1]$  e funzione di densità di probabilità di forma  $p_v(x) = c|x|$  per  $x \in S_X$  e 0 altrove. Si

completi la definizione della funzione di densità di  $X$ , determinando il valore della costante di normalizzazione  $c$ . Si calcoli la funzione di ripartizione di  $X$ , esplicitandola in tutti i suoi tratti. Si calcoli poi  $P(X > 0.5)$ . Si ottenga  $E[X]$ .

**Esercizio 9.7** Sia  $X$  una variabile casuale con supporto  $S_X = [0, 2]$  e funzione di densità di probabilità di forma  $p_x(x) = ce^x$  per  $x \in S_X$  e 0 altrove. Si completi la definizione della funzione di densità di  $X$ , determinando il valore

completi la definizione della funzione di densità di  $X$ , determinando il valore della costante di normalizzazione  $c$ . Si calcoli la funzione di ripartizione di  $X$ , esplicitandola in tutti i suoi tratti. Si ottenga  $P(X < 2)$ .

**Esercizio 9.8** Sia  $X$  una variabile casuale con supporto  $S_X = [-1, 1]$  e funzione di densità di probabilità di forma  $p_x(x) = cx^2$  per  $x \in S_X$  e 0 altrove. Si completi la definizione della funzione di densità di  $X$ , determinando il valore

completi la definizione della funzione di densità di  $X$ , determinando il valore della costante di normalizzazione  $c$ . Si calcoli la funzione di ripartizione di  $X$ , esplicitandola in tutti i suoi tratti. Si calcoli poi  $P(-0.5 < X < 0.5)$ . Si ottenga il valore atteso di  $X$ .

**Esercizio 9.9** Sia  $X$  una v.c. con supporto  $S_X = [0, +\infty)$  e f.r.  $F_X(x) = 1 - e^{-x}$  per  $x > 0$ ,  $F_X(x) = 0$  per  $x < 0$ . Si ottengano supporto e f.r. della legge di

Nella Unità precedente si è vista la definizione della funzione di ripartizione di una v.c.. Si è anche iniziato ad usare la f.r. per calcolare probabilità di eventi relativi a una v.c. univariata. In questa Unità si studiano le proprietà che distinguono le funzioni di ripartizione delle v.c. univariate dalle altre funzioni reali di variabile reale. Si enuncia subito il teorema che esplicita le proprietà strutturali delle funzioni di ripartizione delle v.c. univariate, ossia quelle proprietà che ogni f.r. necessariamente ha

I risultati del paragrafo precedente portano alla seguente definizione.

**Definizione 11.1** Si dice che  $X$  ha legge di Poisson con parametro  $\lambda$ ,  $\lambda > 0$ , e in breve si scriverà  $X \sim P(\lambda)$ , se è una v.c. univariata con legge discreta che ha supporto  $S_X = \mathbb{N}$  e f.m.p., per  $x \in S_X$ , pari a

**Definizione 11.1** Si dice che  $X$  ha legge di Poisson con parametro  $\lambda$ ,  $\lambda > 0$ , e in breve si scriverà  $X \sim P(\lambda)$ , se è una v.c. univariata con legge discreta che ha supporto  $S_X = \mathbb{N}$  e f.m.p., per  $x \in S_X$ , pari a

$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}. \quad (11.1)$$

Per verificare che si tratta di una buona definizione occorre controllare due condizioni:

- positività di  $p_X(x)$  sul supporto  $S_X = \mathbb{N}$ :  
banale perché  $p_X(x)$  è prodotto di tre fattori positivi;
- normalizzazione:

considerazioni svolte sono sintetizzate dalla seguente definizione.

**Definizione 11.2** Una v.c. univariata  $X$  con legge discreta che ha supporto  $S_X = \mathbb{N}^+$  e f.m.p. per  $x \in S_X$  pari a

$$p_X(x) = p(1-p)^{x-1} \quad (11.2)$$

**Definizione 11.2** Una v.c. univariata  $X$  con legge discreta che ha supporto  $S_X = \mathbb{N}^+$  e f.m.p. per  $x \in S_X$  pari a

$$p_X(x) = p(1 - p)^{x-1} \quad (11.2)$$

si dice che ha legge geometrica con parametro  $p$ ,  $p \in (0, 1)$ . In breve si scriverà  $X \sim Ge(p)$ .

Per verificare che si tratta di una buona definizione occorre controllare due condizioni:

- positività di  $p_X(x)$  sul supporto  $S_X = \mathbb{N}^+$ :  
banale perché  $p_X(x)$  è prodotto di fattori positivi;
- normalizzazione:

questa definizione il supporto è traslato di una unità verso sinistra, non è  $\mathbb{N}^+$   
ma  $\mathbb{N}$ . Esempi d'uso:

```
> set.seed(12345)
> rgeom(10,prob=1/2)
[1] 1 1 1 3 1 0 1 1 4 1
> round(dgeom(0:6,prob=1/2),digits=5)
```

**Definizione 12.1** Si dice funzione tasso di guasto di  $T$  (*hazard rate, failure rate*) la funzione  $r_T(\cdot)$ , definita per i valori di  $t$  per cui  $F_T(t) < 1$ , da

$$r_T(t) = \frac{p_T(t)}{\bar{F}_T(t)} = -\frac{d}{dt} \log \bar{F}_T(t).$$

**Definizione 12.2** La v.c.  $T$  è detta con legge di Weibull biparametrica, con parametro di forma  $c > 0$  e di scala  $\lambda > 0$ , in breve si scrive  $T \sim W(c, \lambda)$ , se la f.d.p. di  $T$  è, per  $t > 0$ ,

$$p_T(t) = \lambda c(\lambda t)^{c-1} \exp\{-(\lambda t)^c\}. \quad (12.1)$$

**Definizione 12.3** La v.c.  $T$  è detta con legge gamma biparametrica, con parametro di forma  $\alpha > 0$  e di scala  $\lambda > 0$ , in breve si scrive  $T \sim Ga(\alpha, \lambda)$ , se la f.d.p. di  $T$  è, per  $t > 0$ ,

$$p_T(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}. \quad (12.2)$$

**Definizione 13.1** Sia  $X = (X_1, \dots, X_d)$  una v.c.  $d$ -variata. Si considerino poi un dominio di definizione  $D \subseteq \mathbb{R}^d$ , tale che  $D \in \mathcal{B}_d$  e sia  $P(X \in D) = 1$ , e una funzione

$$g : D \rightarrow \mathbb{R}^{d'} \quad \text{dove } d' \leq d$$

misurabile, ossia tale che per ogni  $B' \in \mathcal{B}_{d'}$  si abbia  $g^{-1}(B') \in \mathcal{B}_d$ , dove  $\mathcal{B}_{d'}$  e  $\mathcal{B}_d$  sono le  $\sigma$ -algebre di Borel associate a  $\mathbb{R}^{d'}$  e a  $\mathbb{R}^d$ , rispettivamente. Allora

poi un dominio di definizione  $D \subseteq \mathbb{R}^d$ , tale che  $D \in \mathcal{B}_d$  e sia  $P(X \in D) = 1$ ,  
e una funzione

$$g : D \rightarrow \mathbb{R}^{d'} \quad \text{dove } d' \leq d$$

misurabile, ossia tale che per ogni  $B' \in \mathcal{B}_{d'}$  si abbia  $g^{-1}(B') \in \mathcal{B}_d$ , dove  $\mathcal{B}_{d'}$  e  
 $\mathcal{B}_d$  sono le  $\sigma$ -algebre di Borel associate a  $\mathbb{R}^{d'}$  e a  $\mathbb{R}^d$ , rispettivamente. Allora  
**la v.c. trasformata**  $Y = g(X)$  è una v.c.  $d'$ -variata la cui legge di probabilità

La condizione di misurabilità che appare nella Definizione 13.1 è puramente tecnica. In pratica, tutte le funzioni che si riesce a scrivere esplicitamente sono misurabili.

In linea di principio, la formula (13.1) determina la legge di probabilità della v.c. trasformata  $Y$  per ogni assegnata legge di probabilità della  $X$  di partenza. Tuttavia, molto spesso la legge di probabilità di  $X$  è definita in modo indiretto tramite strumenti complici quali supporto,  $S_X$ , e f.d.p.,  $p_X(x)$ , e

considerare  $D = (0, +\infty)$  come dominio di definizione della trasformazione.  
Ovvio che è  $P_X(D) = 1$ . Si ha  $g(D) = \mathbb{R}$ . Quindi  $g(D)$  è chiuso:  
ha come elementi tutti i suoi punti di accumulazione. Pertanto  $S_Y = \mathbb{R}$ .  $\triangle$

### 13.3 La f.m.p. di una trasformata

**Definizione 14.1** Sia  $X$  una v.c. con legge discreta o continua, con f.m.p. o f.d.p.  $p_X(x)$ . Si dice moda di  $X$ , indicata con  $x_{\text{mod}}$ , un valore del supporto di  $X$  per cui

$$p_X(x_{\text{mod}}) \geq p_X(x) \quad \text{per ogni } x \in S_X.$$

Nel caso continuo, si richiede inoltre che la densità di  $X$  sia continua almeno da destra o da sinistra in  $x_{\text{mod}}$ .

contenenti  $x_{mod}$ . Si osservi che la definizione ha senso anche se  $X$  è multivariata.  
Conviene considerare alcuni esempi.

- $X \sim Bi(2, 0.5)$  ha supporto  $S_X = \{0, 1, 2\}$  e f.m.p.  $p_X(0) = 1/4$ ,  $p_X(1) = 1/2$ ,  $p_X(2) = 1/4$ , per cui  $x_{mod} = 1$ .

**Definizione 14.2** Sia  $X$  una v.c. univariata con f.r.  $F_X(x)$ . Si dice **mediana** di  $X$ , indicata con  $x_{0.5}$ , un valore reale tale che valgano simultaneamente

$$P(X \leq x_{0.5}) \geq 0.5$$

$$e \quad P(X \geq x_{0.5}) \geq 0.5.$$

le diseguaglianze nella definizione.

Anche la mediana non è necessariamente unica. Tutte le soluzioni dell'equazione  $F_X(x) = 0.5$  sono mediana di  $X$ . D'altro canto, se l'equazione  $F_X(x) = 0.5$  non ha soluzioni, la mediana di  $X$  è unica ed è il più piccolo valore di  $x$  per cui  $F_X(x) \geq 0.5$ . Conviene considerare alcuni esempi.

- $X \sim Bi(2, 0.5)$  ha supporto  $S_X = \{0, 1, 2\}$ , f.m.p.  $p_X(0) = 1/4$ ,  $p_X(1) =$

**Definizione 14.3** Sia  $X$  una v.c. univariata con f.r.  $F_X(x)$  e  $p \in (0, 1)$ . Si dice **quantile- $p$**  di  $X$ , indicato con  $x_p$ , un valore reale tale che valgano simultaneamente

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

Tra i quantili si spiccano alcuni che per il loro uso frequente meritano un nome:

Conviene anzitutto richiamare la definizione data nei paragrafi 6.7 e 8.2.

**Definizione 14.4** Sia  $X$  una v.c. univariata con legge discreta o continua, con supporto  $S_X$  e f.m.p./f.d.p.  $p_X(x)$ . Si dice **valore atteso** di  $X$ , indicato con  $E[X]$ , il valore reale

**Definizione 14.4** Sia  $X$  una v.c. univariata con legge discreta o continua, con supporto  $S_X$  e f.m.p./f.d.p.  $p_X(x)$ . Si dice **valore atteso** di  $X$ , indicato con  $E[X]$ , il valore reale

$$E[X] = \begin{cases} \sum_{x \in S_X} x p_X(x) & \text{se } X \text{ ha legge discreta} \\ +\infty & \text{se } S_X \text{ è un insieme numerabile non finito} \end{cases}$$

Si osservi che, con opportune modificazioni, la definizione funziona anche se  $X$  è bivariata o multivariata. Diversamente da moda, mediana e quantili, l'esistenza del valore atteso non è garantita. In compenso, il valore atteso di una v.c., se esiste finito, è unico. Si considerano nel seguito solamente variabili casuali con valore atteso finito. Quindi se si scrive  $E[X]$  si sottintende che la v.c.  $X$  ha valore atteso finito.

Siano  $X$  e  $Y$  v.c. univariate con  $Y = g(X)$ . Allora si ha, seguendo la definizione,

$$\mathbb{E}[Y] = \begin{cases} \sum_{y \in S_Y} y p_Y(y) & \text{se } Y \text{ ha legge discreta} \\ \int_{-\infty}^{+\infty} y p_Y(y) dy & \text{se } Y \text{ ha legge continua} \end{cases}$$

Si richiama la definizione della varianza e si presenta l'indice di variabilità ad essa strettamente collegato, lo scarto quadratico medio.

La proprietà dei minimi quadrati del valore atteso, formula (14.6), mostra che il rischio quadratico, o media della perdita quadratica, associato a prevedere  $X$  con  $E[X]$ , ossia

$$E[(X - E[X])^2],$$

**Definizione 15.1** Sia  $X$  una v.c. univariata con legge discreta o continua, con supporto  $S_X$  e f.m.p. o f.d.p.  $p_X(x)$ . Si dice **varianza** di  $X$ , indicata con  $\text{Var}(X)$ , o anche  $V(X)$ ,  $\sigma_X^2$  o semplicemente  $\sigma^2$ , il valore reale

$$\text{Var}[X] = E[(X - E(X))^2]. \quad (15.1)$$

propria media. Esplicitando la definizione (15.1) si ha:

$$\text{Var}[X] = \begin{cases} \sum_{x \in S_X} (x - E[X])^2 p_X(x) & \text{se } X \text{ ha legge discreta} \\ \int_{-\infty}^{+\infty} (x - E[X])^2 p_X(x) dx & \text{se } X \text{ ha legge continua.} \end{cases} \quad (15.2)$$

La varianza è il principale indice di variabilità per una v.c. univariata, dal mo-

**Definizione 15.2** Lo scarto quadratico medio o deviazione standard, indicato con  $\sigma_X$  o semplicemente  $\sigma$ , è la radice quadrata aritmetica della varianza (l'unica non negativa). In simboli,

$$\sigma_X = \sqrt{\text{Var}[X]}.$$

La dimostrazione parte dalla definizione della varianza:

$$\begin{aligned}\text{Var}[X] &= E[(X - E[X])^2] \\ &= E[X^2 + (E[X])^2 - 2E[X]X] \quad \text{sviluppando il quadrato} \\ &= E[X^2] + (E[X])^2 - 2E[X]E[X] \quad \text{linearità del valore atteso} \\ &= E[X^2] - (E[X])^2.\end{aligned}$$

Di nuovo la dimostrazione parte dalla definizione della varianza:

$$\begin{aligned}\text{Var}[a + X] &= \mathbb{E}[(a + X - \mathbb{E}[a + X])^2] \\ &= \mathbb{E}[(a + X - a - \mathbb{E}[X])^2] \quad \text{perché } \mathbb{E}[a + X] = a + \mathbb{E}[X] \\ &= \mathbb{E}[(X - \mathbb{E}[X])^2] \quad \text{semplificando} \\ &= \text{Var}[X].\end{aligned}$$

Ancora una volta, la dimostrazione parte dalla definizione della varianza:

$$\text{Var}[bX] = E[(bX - E[bX])^2]$$

$$= E[(bX - bE[X])^2]$$

$$= E[(b(X - E[X]))^2]$$

$$= E[b^2(X - E[X])^2]$$

perché  $E[bX] = bE[X]$

raccogliendo

proprietà delle potenze

$$= \mathbb{E} [(aX + bY - \mathbb{E}[aX + bY])^2]$$

definizione di  $T$

$$= \mathbb{E} [(aX + bY - a\mathbb{E}[X] - b\mathbb{E}[Y])^2]$$

linearità di  $\mathbb{E}[\cdot]$

$$= \mathbb{E} [(a(X - \mathbb{E}[X]) + b(Y - \mathbb{E}[Y]))^2]$$

raccogliendo

$$= a^2 \mathbb{E} [(X - \mathbb{E}[X])^2] + b^2 \mathbb{E} [(Y - \mathbb{E}[Y])^2]$$

$$+ 2ab \mathbb{E} [(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

**Definizione 16.1** Sia  $X$  una v.c. univariata con f.m.p. o f.d.p.  $p_X(x)$ . La funzione generatrice dei momenti di  $X$ , indicata con  $M_X(t)$ , è una funzione reale di variabile reale definita da

$$M_X(t) = \mathbb{E}[e^{tX}] = \begin{cases} \sum_{x \in S_X} e^{tx} p_X(x) & \text{se } X \text{ ha legge discreta} \\ \int_{-\infty}^{+\infty} e^{tx} F_X(x) dx & \text{se } X \text{ ha legge continua} \end{cases}$$

**Definizione 16.2** Sia  $X$  una v.c. univariata. Si dice che la v.c. univariata  $X$  ha funzione generatrice dei momenti propria se il dominio di finitezza di  $M_X(t)$  include l'origine come punto interno, ossia se esiste  $\varepsilon > 0$  tale che  $(-\varepsilon, \varepsilon) \subseteq D_X$ , e pertanto se per ogni  $t \in (-\varepsilon, \varepsilon)$  vale  $M_X(t) < +\infty$ .

$$M_S(t) = \mathbb{E}[e^{tS}]$$

$$= \mathbb{E} \left[ e^{t \sum_{i=1}^d X_i} \right]$$

$$= \mathbb{E} \left[ \prod_{i=1}^d e^{tX_i} \right]$$

$$d$$

definizione di f.g.m.

definizione di  $S$

proprietà di  $\exp(\cdot)$

$$= \mathbb{E} \left[ e^{t \sum_{i=1}^d X_i} \right] \quad \text{definizione di } S$$

$$= \mathbb{E} \left[ \prod_{i=1}^d e^{t X_i} \right] \quad \text{proprietà di } \exp(\cdot)$$

$$= \prod_{i=1}^d \mathbb{E}[e^{t X_i}] \quad \text{per l'indipendenza delle } X_i$$

$$= \prod_{i=1} M_{X_i}(t) \quad \text{definizione di f.g.m. .}$$

□

Il risultato permette di calcolare molto facilmente  $M_S(t)$ . Tuttavia di solito ciò che interessa è la legge di probabilità di  $S$ . Per individuarla, si dovrà controllare se  $M_S(t)$  appare nella lista delle f.g.m. già disponibili.

$$\begin{aligned}
 M_W(t) &= \mathbb{E}[e^{tW}] && \text{definizione di f.g.m.} \\
 &= \mathbb{E}[e^{t(-Y_1 - Y_2)}] && \text{definizione di } W \\
 &= \mathbb{E}[e^{-tY_1} e^{-tY_2}] && \text{proprietà di } \exp(\cdot) \\
 &= \mathbb{E}[e^{-tY_1}] \mathbb{E}[e^{-tY_2}] && Y_1 \text{ e } Y_2 \text{ indipendenti} \\
 &= M_{Y_1}(-t) M_{Y_2}(-t) && \text{definizione di f.g.m.}
 \end{aligned}$$

$$(-t)^2/2, (-t)^2/2$$

$$= \mathbb{E}[e^{t(-Y_1 - Y_2)}]$$

definizione di  $W$   
=  $\mathbb{E}[e^{-tY_1} e^{-tY_2}]$

proprietà di  $\exp(\cdot)$   
=  $\mathbb{E}[e^{-tY_1}] \mathbb{E}[e^{-tY_2}]$

$Y_1$  e  $Y_2$  indipendenti  
=  $M_{Y_1}(-t) M_{Y_2}(-t)$

definizione di f.g.m.  
=  $e^{(-t)^2/2} e^{(-t)^2/2}$

f.g.m. di  $Y \sim Y_1 \sim Y_2$

$$\begin{aligned}
 &= M_{Y_1}(-t) M_{Y_2}(-t) \quad \text{definizione di f.g.m.} \\
 &= e^{(-t)^2/2} e^{(-t)^2/2} \quad \text{f.g.m. di } Y \sim Y_1 \sim Y_2 \\
 &= e^{t^2} \quad \text{proprietà di } \exp(\cdot).
 \end{aligned}$$

Poiché le derivate di  $M_W(t)$  sono

$$M'(t) = 2t e^{t^2} \quad \text{e} \quad M''(t) = 2e^{t^2} + (2t)^2 e^{t^2}$$

$$\begin{aligned} M_W(t) &= \mathbb{E}[e^{tW}] && \text{definizione di f.g.m.} \\ &= \mathbb{E}[e^{t(Y_1+Y_2+Y_3)}] && \text{definizione di } W \\ &= \mathbb{E}[e^{tY_1} e^{tY_2} e^{tY_3}] && \text{proprietà di } \exp(\cdot) \\ &= \mathbb{E}[e^{tY_1}] \mathbb{E}[e^{tY_2}] \mathbb{E}[e^{tY_3}] && Y_1, Y_2, Y_3 \text{ indipendenti} \\ &= M_{Y_1}(t) M_{Y_2}(t) M_{Y_3}(t) && \text{definizione di f.g.m.} \\ &= M_{Y_1}(t)^3 && M_{Y_1}(t) = V_1 - V_1^2 \end{aligned}$$

definizione di  $W$   
proprietà di  $\exp(\cdot)$   
 $Y_1, Y_2, Y_3$  indipendenti  
definizione di f.g.m.  
 $Y \sim Y_1 \sim Y_2 \sim Y_3$

$$\begin{aligned} &= \mathbb{E}[e^{t(Y_1 + Y_2 + Y_3)}] \\ &= \mathbb{E}[e^{tY_1} e^{tY_2} e^{tY_3}] \\ &= \mathbb{E}[e^{tY_1}] \mathbb{E}[e^{tY_2}] \mathbb{E}[e^{tY_3}] \\ &= M_{Y_1}(t) M_{Y_2}(t) M_{Y_3}(t) \\ &= M_Y(t)^3 \\ &\quad (1 - e^{-t})^3 \end{aligned}$$

$= M_{Y_1}(t) M_{Y_2}(t) M_{Y_3}(t)$  definizione di f.g.m.

$= M_Y(t)^3$   $Y \sim Y_1 \sim Y_2 \sim Y_3$

$= \frac{(1 + e^t)^3}{8}$ .

## 17.1 Genesi e definizione

Si considerino  $n$  misure ripetute di una quantità  $\mu$ , effettuate tutte con lo stesso strumento di misura. Lo strumento è affetto da errore, per cui le misure  $x_i$  saranno realizzazioni di una v.c.  $X$ , in breve  $X \rightarrow x_i$ . Siano  $z_i$  gli errori di misura, espressi in una scala standard, realizzazioni di una v.c.  $Z$ , quindi  $Z \rightarrow z_i$ . La relazione fra misure ed errori di misura espressi su una scala standard è

**Definizione 17.1** Una v.c. univariata  $Z$  con supporto  $S_Z = \mathbb{R}$  e f.d.p.

$$p_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad (17.2)$$

è detta con legge normale standard, in breve  $Z \sim N(0, 1)$ .

per una v.c.  $X$  con legge normale generale.

**Definizione 17.2** Una v.c. univariata  $X$  con supporto  $S_X = \mathbb{R}$  e f.d.p.

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right\} \quad (17.3)$$

**Definizione 17.2** Una v.c. univariata  $X$  con supporto  $S_X = \mathbb{R}$  e f.d.p.

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \quad (17.3)$$

è detta **con legge normale con parametri  $\mu \in \mathbb{R}$  e  $\sigma^2 > 0$** , in breve si scrive  
 $X \sim N(\mu, \sigma^2)$

$\sim a + b(\mu + \sigma Z)$  definizione di  $X$

$\sim a + b\mu + b\sigma Z$

$\sim a + b\mu + |b|\sigma Z$   $Z \sim -Z$

$\sim N(a + b\mu, b^2\sigma^2)$ .

□

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] && \text{definizione di f.g.m.} \\ &= \mathbb{E}[e^{t(\mu+\sigma Z)}] && \text{definizione di } X \\ &= \mathbb{E}[e^{t\mu} e^{t\sigma Z}] && \text{proprietà di } \exp(\cdot) \\ &= e^{t\mu} \mathbb{E}[e^{t\sigma Z}] && \text{linearità del valore atteso} \\ &= e^{t\mu} M_Z(t\sigma) && \text{definizione di f.g.m..} \end{aligned}$$

$$\begin{aligned} &= \mathbb{E} [e^{t(\mu + \sigma Z)}] && \text{definizione di } X \\ &= \mathbb{E} [e^{t\mu} e^{t\sigma Z}] && \text{proprietà di } \exp(\cdot) \\ &= e^{t\mu} \mathbb{E} [e^{t\sigma Z}] && \text{linearità del valore atteso} \\ &= e^{t\mu} M_Z(t\sigma) && \text{definizione di f.g.m..} \end{aligned}$$

Si ha poi

$$= e^{t\mu} M_Z(t\sigma) \quad \text{definizione di f.g.m.}$$

Si ha poi

$$\begin{aligned} M_Z(t) &= \mathbb{E}[e^{tZ}] \\ &= \int_{-\infty}^{+\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \end{aligned}$$

$$\begin{aligned} M_S(t) &= \mathbb{E}[e^{tS}] && \text{definizione di f.g.m.} \\ &= \mathbb{E}[e^{t(X+Y)}] && \text{definizione di } S \\ &= \mathbb{E}[e^{tX} e^{tY}] && \text{proprietà di } \exp(\cdot) \\ &= \mathbb{E}[e^{tX}] \mathbb{E}[e^{tY}] && X \text{ e } Y \text{ indipendenti} \\ &= M_X(t) M_Y(t) && \text{definizione di f.g.m.} \end{aligned}$$

$$= \mathbb{E} [e^{t(X+Y)}]$$

definizione di  $S$

$$= \mathbb{E} [e^{tX} e^{tY}]$$

proprietà di  $\exp(\cdot)$

$$= \mathbb{E} [e^{tX}] \mathbb{E} [e^{tY}]$$

$X$  e  $Y$  indipendenti

$$= M_X(t) M_Y(t)$$

definizione di f.g.m.

$$= e^{t\mu_X + \frac{1}{2}t^2\sigma_X^2} e^{t\mu_Y + \frac{1}{2}t^2\sigma_Y^2}$$

$X$  e  $Y$  normali

$= M_X(t) M_Y(t)$  definizione di f.g.m.

$= e^{t\mu_X + \frac{1}{2}t^2\sigma_X^2} e^{t\mu_Y + \frac{1}{2}t^2\sigma_Y^2}$   $X$  e  $Y$  normali

$= e^{t(\mu_X + \mu_Y) + \frac{1}{2}t^2(\sigma_X^2 + \sigma_Y^2)}$  proprietà di  $\exp(\cdot)$  e raccoglimenti.

□

La somma di due normali indipendenti ha la stessa distribuzione di una normale.

$$= \prod_{i=1}^d M_{X_i}(ta_i) \quad \text{definizione di f.g.m.}$$

$$= \prod_{i=1}^d \exp \left\{ \mu_i(ta_i) + \frac{1}{2} \sigma_i^2(ta_i)^2 \right\}$$

$$= \exp \left\{ t \sum_{i=1}^d a_i \mu_i + \frac{t^2}{2} \sum_{i=1}^d a_i^2 \sigma_i^2 \right\}$$

$$\begin{aligned} F_X(x) &= P(X \leq x) && \text{definizione di f.r.} \\ &= P(\mu + \sigma Z \leq x) && \text{definizione di } X \\ &= P\left(Z \leq \frac{x - \mu}{\sigma}\right) && \text{algebra} \\ &= F_Z\left(\frac{x - \mu}{\sigma}\right) && \text{definizione di f.r. .} \end{aligned}$$

$= P(\mu + \sigma Z \leq x)$  definizione di  $X$

$= P\left(Z \leq \frac{x - \mu}{\sigma}\right)$  algebra

$= F_Z\left(\frac{x - \mu}{\sigma}\right)$  definizione di f.r. .

La trasformazione

$$= F_Z \left( \frac{x - \mu}{\sigma} \right)$$

definizione di f.r. .

La trasformazione

$$\text{da } x \text{ a } \frac{x - \mu}{\sigma}$$

è detta **standardizzazione**. Esprime  $x$  come deviazione da  $\mu$  in unità di scarto

**Definizione 19.1** Si dice che la successione di v.c. univariate  $X_n$  converge in probabilità alla costante  $c \in \mathbb{R}$  se per ogni  $\varepsilon > 0$  si ha

$$\lim_{n \rightarrow +\infty} P(|X_n - c| > \varepsilon) = 0.$$

Si scrive allora  $X_n \xrightarrow{p} c$ .

**Definizione 19.2** Si dice che la successione di v.c. univariate  $X_n$  converge in distribuzione alla v.c. univariata  $X$ , e si scrive  $X_n \xrightarrow{d} X$ , se vale

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

per ogni  $x \in \mathbb{R}$  in cui  $F_X(x)$  è continua.

**Esercizio 19.3** Sia  $X_n \sim Ge(1/n)$ . Usando la Definizione 19.2, si mostri che la successione  $T_n = X_n/n$  converge in distribuzione a  $T \sim Esp(1)$ .

**Esercizio 19.4** Sia  $X \sim Bi(10^4, 0.5)$ . Si calcolino in via approssimata  $P(X > 4900)$ ,  $P(X \leq 5200)$ , il novantanovesimo percentile di  $X$ .

**Esercizio 19.5** Sia  $X \sim Bi(20, 0.5)$ . Si calcoli in via approssimata  $P(X \leq$

**Definizione 20.1** Una v.c.  $Y = (Y_1, \dots, Y_n)$  esprime un esperimento di campionamento casuale semplice con numerosità  $n$  se

- il supporto di  $Y$ , detto anche spazio campionario e indicato con  $\mathcal{Y}$ , è

$$\mathcal{Y} = S_{Y_1} \times \cdots \times S_{Y_n} = S_{Y_1}^n;$$

**Definizione 20.2** Una statistica  $t = t(y)$  è una applicazione che ha come dominio lo spazio campionario  $\mathcal{Y}$  e assume valori in uno spazio  $\mathbb{R}^P$ .

Esempi di statistiche sono la somma,  $s_n = \sum_{i=1}^n y_i$ , la media campionaria,  $\bar{y}_n = \sum_{i=1}^n y_i/n$ , il minimo,  $\min_{i=1}^n y_i$ , il massimo,  $\max_{i=1}^n y_i$ . Quali statistiche riassuntive utilizzare è spesso suggerito dalla forma di  $p_Y(y; \theta)$ .

**Definizione 20.3** La distribuzione campionaria di una statistica  $t = t(y)$  è la legge di probabilità della v.c. trasformata  $T = t(Y)$  quando la legge di  $Y$  è descritta da  $p_Y(y; \theta)$ .

Conviene considerare subito due esempi.

**Esempio 20.4 (Stima di una probabilità)** Si supponga di aver osservato

**Definizione 20.4** Siano  $Z_1, Z_2, \dots, Z_\nu$  v.c. i.i.d. con legge  $N(0, 1)$ . Si dice che la v.c.  $T_\nu$  ha legge chi-quadrato con  $\nu$  gradi di libertà, e si scrive  $T_\nu \sim \chi_\nu^2$ , se  $T_\nu$  ha la legge di probabilità di  $\sum_{i=1}^\nu Z_i^2$ .

Conviene indagare in primo luogo la legge chi-quadrato con un grado di libertà,  $\chi_1^2$ . È la legge di  $T = Z^2$  dove  $Z \sim N(0, 1)$ . Il supporto di  $T$  è

legge di probabilità di  $\hat{\sigma}_{\mu_0}^2$  si trova facilmente sostituendo nella sua definizione  $Y_i$  con  $\mu_0 + \sigma Z_i$ , dove le  $Z_i$  sono i.i.d.  $N(0, 1)$ . Si ottiene

$$\begin{aligned}\hat{\sigma}_{\mu_0}^2 &= \frac{1}{n} \sum_{i=1}^n (\mu_0 + \sigma Z_i - \mu_0)^2 \\ &= \sigma^2 \sum_{i=1}^n Z_i^2\end{aligned}$$

**Definizione 21.1** Un modello statistico, indicato con  $\mathcal{F}$ , è una collezione di leggi di probabilità ritenute possibili per la v.c.  $Y$  generatrice dei dati  $y$ .

**Definizione 21.2** Quando in (21.1)  $\Theta \subseteq \mathbb{R}^p$ , il modello  $\mathcal{F}$  è detto **modello statistico parametrico**.

Sono modelli statistici parametrici:

- il modello di c.c.s. con numerosità  $n$  da  $Bi(1, p)$ ,  $p \in (0, 1)$  ignoto valore atteso del conteggio 0/1

**Definizione 21.3** La funzione di ripartizione empirica è la f.r. della distribuzione empirica, ossia

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(y_i), \quad x \in \mathbb{R}. \quad (21.2)$$

Sopra  $I_{(-\infty, x]}(\cdot)$  è la funzione indicatrice

**Definizione 21.4** Una procedura di stima puntuale di  $\theta$  (intendendo di  $\theta^*$ ) è una applicazione (misurabile)

$$\hat{\theta} : \mathcal{Y} \rightarrow \Theta.$$

In corrispondenza al valore  $y$  osservato, la procedura produce  $\hat{\theta}(y)$ , detta stima di  $\theta$ .

**Definizione 21.5** La v.c.  $\hat{\theta} = \hat{\theta}(Y)$  è detta stimatore.

L'informazione su  $\theta$  offerta dallo stimatore  $\hat{\theta}$  è espressa dalla distribuzione campionaria della statistica  $\hat{\theta}(Y)$ .

**Definizione 21.6** La distribuzione campionaria dello stimatore  $\hat{\theta}$  è la legge di probabilità della v.c. trasformata  $\hat{\theta}(Y)$  sotto  $\theta$ , ossia quando la legge di  $Y$  è descritta da  $p_Y(y; \theta)$ .

**Esempio 21.2 (C.c.s. con numerosità  $n$  da  $P(\lambda)$ )** Si considerino dati

**Definizione 21.7** Uno stimatore  $\hat{\theta}_n$  è detto non distorto per  $\theta$  se, sotto  $\theta$ ,

$$\mathbb{E}[\hat{\theta}_n] = \theta$$

per ogni  $\theta \in \Theta$ .

Per interpretare la non distorsione, si tenga presente che una proprietà che vale per ogni  $\theta \in \Theta$  vale anche per il vero valore del parametro. Dunque uno stimatore non distorto pone il vero valore del parametro al centro della propria

**Definizione 21.8** Uno stimatore  $\hat{\theta}_n$  è detto **asintoticamente non distorto per  $\theta$**  se, sotto  $\theta$ ,

$$\lim_{n \rightarrow +\infty} E[\hat{\theta}_n] = \theta \quad \text{per ogni } \theta \in \Theta.$$

Quando  $n$  è sufficientemente grande, uno stimatore asintoticamente non distorto pone il vero valore del parametro al centro della propria distribuzione di

**Definizione 21.9** Uno stimatore  $\hat{\theta}_n$  è detto **consistente per  $\theta$**  se, sotto  $\theta$ ,

$$\hat{\theta}_n \xrightarrow{P} \theta$$

per ogni  $\theta \in \Theta$ .

Se lo stimatore è consistente, quando vi è moltissima informazione nel campione, si avrà

**Definizione 21.10** Uno stimatore  $\hat{\theta}_n$  è detto **asintoticamente normale** se, sotto  $\theta$ ,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$$

per ogni  $\theta \in \Theta$ . La funzione  $\sigma^2(\theta)$  è detta **varianza asintotica dello stimatore**.

**Definizione 21.11** La funzione di verosimiglianza (likelihood function) per i dati  $y$  e il modello statistico parametrico  $\mathcal{F} = \{p_Y(y; \theta), y \in \mathcal{Y}, \theta \in \Theta \subseteq \mathbb{R}^p\}$  è

$$L(\theta) = L(\theta; y) = p_Y(y; \theta) \quad \text{dove } \theta \in \Theta.$$

È quindi la f.m.p. o f.d.p. del modello calcolata ai dati. Per un modello di c.c.s. la funzione di verosimiglianza è semplicemente  $L(\theta) = \prod_{i=1}^n p_{Y_i}(y_i; \theta)$ .

**Definizione 21.12** Data una funzione di verosimiglianza  $L(\theta) = L(\theta; y)$ , si dice **stima di massima verosimiglianza** di  $\theta$  un valore  $\hat{\theta} \in \Theta$  tale che

$$L(\hat{\theta}) \geq L(\theta) \quad \text{per ogni } \theta \in \Theta.$$

funzione di  $\sum_{i=1}^n y_i$  e  $\sum_{i=1}^n y_i^2$  informativa su  $\sigma^2$ . La sua definizione è

$$s_n^2 = \frac{n}{n-1} \left( \frac{1}{n} \sum_{i=1}^n y_i^2 - \left( \frac{1}{n} \sum_{i=1}^n y_i \right)^2 \right) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_n)^2$$

dove  $\bar{y}_n = \sum_{i=1}^n y_i/n$  è la media campionaria.

**Definizione 22.1** Si dice regione di confidenza per  $\theta$  una applicazione con dominio  $\mathcal{Y}$  che mappa  $y \rightarrow \hat{\Theta}(y) \subseteq \Theta$ .

Quando  $p = 1$  e la regione risulta un intervallo si parlerà di intervallo di confidenza. Spesso sarà interessante seguire il comportamento di una regione o intervallo di confidenza al crescere della numerosità campionaria  $n$ . Si userà allora la notazione  $\hat{\Theta}_n(y)$ . Per valutare l'efficacia di una procedura che definisce

**Definizione 22.2** Si dice che  $\Theta(y)$  è una regione di confidenza per  $\theta$  con livello  $1 - \alpha$  se

$$P_\theta(\theta \in \hat{\Theta}(Y)) = 1 - \alpha, \quad \text{per ogni } \theta \in \Theta.$$

Nella definizione data il livello è esatto. Di solito si otterranno regioni di confidenza con livello  $1 - \alpha$  approssimato, ad esempio circa  $1 - \alpha$ , quando  $\alpha$  è piccolo.

Nella definizione data il livello è esatto. Di solito si otterranno regioni di confidenza con livello  $1 - \alpha$  approssimato, ad esempio circa  $1 - \alpha$  quando  $n$  è grande, valendo il limite

$$\lim_{n \rightarrow \infty} P_\theta(\theta \in \hat{\Theta}_n(Y)) = 1 - \alpha, \quad \text{per ogni } \theta \in \Theta.$$

Per studiare la distribuzione campionaria di  $T_n$  è utile la seguente definizione.

Definizione 22.3 Siano  $Z \sim N(0, 1)$  e  $W_\nu \sim \chi_\nu^2$  due v.c. indipendenti.  
Allora la legge di probabilità di

$$S = \frac{Z}{\sqrt{W_\nu / \nu}}$$

**Definizione 22.3** Siano  $Z \sim N(0, 1)$  e  $W_\nu \sim \chi_\nu^2$  due v.c. indipendenti.  
Allora la legge di probabilità di

$$S_\nu = \frac{Z}{\sqrt{\frac{W_\nu}{\nu}}}$$

**Definizione 23.1** Il test decisionale  $t_n = t_n(y)$  per  $H_0 : \theta = \theta_0$  contro  $H_1 : \theta \neq \theta_0$  con regione di rifiuto  $R$  è detto con livello di significatività  $\alpha$  se

$$P_{\theta_0}(Y \in R) = \alpha,$$

esattamente o, sovente, come approssimazione asintotica.

**Definizione 23.2** Sia  $C_y$  l'insieme dei valori  $y^*$  almeno altrettanto critici di  $y$  contro  $H_0 : \theta \in \Theta_0$ . Allora il  $P$ -value associato ai dati  $y$  è

$$P = \sup_{\theta \in \Theta_0} P_\theta(Y \in C_y).$$

Se  $H_0$  è semplice, di forma  $\theta = \theta_0$ , si ha  $P = P_{\theta_0}(Y \in C_y)$ . Se  $H_0$  è com-