

# Elementi di Probabilità e Statistica. Anno 2017-18

## Lista di definizioni ed enunciati

### 1 Nozioni fondamentali

**Definizione: Algebra di parti** Dato un insieme  $\Omega$ , si chiama algebra di parti una famiglia  $\mathcal{F}$  di sottinsiemi di  $\Omega$  tale che:

- a) l'insieme vuoto  $\emptyset$  e l'intero insieme  $\Omega$  sono elementi di  $\mathcal{F}$ ;
- b) se  $A \in \mathcal{F}$ , anche il suo complementare  $A^c \in \mathcal{F}$ ;
- c) se  $A$  e  $B$  sono elementi di  $\mathcal{F}$ , anche  $A \cup B \in \mathcal{F}$ .

**Definizione: Probabilità finitamente additiva** Data un'algebra  $\mathcal{F}$  di parti di un insieme  $\Omega$ , si chiama probabilità (finitamente additiva) una funzione  $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$  tale che

- a) se  $A, B \in \mathcal{F}$  e  $A \cap B = \emptyset$ , allora  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ ;
- b)  $\mathbf{P}(\Omega) = 1$ .

Gli elementi dell'algebra di parti  $\mathcal{F}$  sono chiamati **eventi**, si chiama **trascurabile** un evento  $A$  tale che  $\mathbf{P}(A) = 0$  e si chiama **quasi certo** un evento  $A$  tale che  $\mathbf{P}(A) = 1$ .

**Definizione:  $\sigma$ -algebra di parti** Dato un insieme  $\Omega$ , si chiama  $\sigma$ -algebra di parti una famiglia  $\mathcal{F}$  di sottinsiemi di  $\Omega$  tale che:

- a) l'insieme vuoto  $\emptyset$  e l'intero insieme  $\Omega$  sono elementi di  $\mathcal{F}$ ;
- b) se  $A \in \mathcal{F}$ , anche il suo complementare  $A^c \in \mathcal{F}$ ;
- c) se  $(A_n)_{n \geq 1}$  è una successione di elementi di  $\mathcal{F}$ , anche  $\bigcup_{n=1}^{+\infty} A_n \in \mathcal{F}$ .

**Definizione: Probabilità** Assegnato un insieme  $\Omega$  ed una  $\sigma$ -algebra  $\mathcal{F}$  di parti di  $\Omega$ , si chiama probabilità una funzione  $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$  tale che

- a) se  $(A_n)_{n=1,2,\dots}$  è una successione di elementi di  $\mathcal{F}$  a due a due disgiunti, si ha  $\mathbf{P}(\bigcup_{n=1}^{+\infty} A_n) = \sum_{n=1}^{+\infty} \mathbf{P}(A_n)$ ;
- b)  $\mathbf{P}(\Omega) = 1$ .

**Proposizione** Sia  $\mathcal{F}$  una  $\sigma$ -algebra di parti di un insieme  $\Omega$  e sia  $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$  semplicemente additiva (e tale che  $\mathbf{P}(\Omega) = 1$ ). Sono equivalenti le seguenti proprietà:

- 1)  $\mathbf{P}$  è  $\sigma$ -additiva;
- 2)  $A_n \uparrow A \implies \mathbf{P}(A_n) \rightarrow \mathbf{P}(A)$  (o anche  $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$ );
- 3)  $A_n \downarrow A \implies \mathbf{P}(A_n) \rightarrow \mathbf{P}(A)$  (o anche  $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$ );
- 4)  $A_n \uparrow \Omega \implies \mathbf{P}(A_n) \rightarrow 1$ ;
- 5)  $A_n \downarrow \emptyset \implies \mathbf{P}(A_n) \rightarrow 0$ .

Nel caso in cui  $\Omega$  sia un insieme finito e gli *eventi elementari*  $\omega_i$  siano equiprobabili, si parla di *distribuzione uniforme di probabilità su  $\Omega$* : in questo caso si ottiene la formula

$$\mathbf{P}(A) = \frac{\#A}{\#\Omega} = \frac{|A|}{|\Omega|}$$

dove con  $\#A$  o con  $|A|$  si indica la *cardinalità* (o numero degli elementi) dell'insieme  $A$ .

**Permutazioni** Il numero di modi in cui si possono ordinare gli elementi di  $\{1, \dots, n\}$  è  $n!$

**Coefficiente binomiale** Siano  $0 \leq k \leq n$ : il numero di sottinsiemi di  $\{1, \dots, n\}$  formati da  $k$  elementi è

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

**Definizione: Probabilità condizionata.** Assegnato uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbf{P})$  ed un evento  $B$  non trascurabile, si chiama *probabilità condizionata* di  $A$  rispetto a  $B$  il numero

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

**Proposizione.** Siano  $A_1, \dots, A_n$  eventi, e supponiamo che  $A_1 \cap \dots \cap A_{n-1}$  sia non trascurabile: vale la formula

$$\mathbf{P}(A_1 \cap \dots \cap A_n) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2|A_1) \dots \mathbf{P}(A_n|A_1 \cap \dots \cap A_{n-1})$$

**Definizione: Sistema di alternative** Si chiama *sistema di alternative* una partizione di  $\Omega$  in  $n$  eventi non trascurabili  $B_1, \dots, B_n$ .

**Proposizione: Formula di Bayes** Sia  $B_1, \dots, B_n$  un sistema di alternative: assegnato una qualunque evento  $A$  non trascurabile, valgono le formule

$$\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A|B_i) \mathbf{P}(B_i)$$

$$\mathbf{P}(B_i|A) = \frac{\mathbf{P}(A|B_i) \mathbf{P}(B_i)}{\sum_{j=1}^n \mathbf{P}(A|B_j) \mathbf{P}(B_j)}$$

**Definizione: Indipendenza stocastica** Due eventi  $A$  e  $B$  sono detti indipendenti se vale l'eguaglianza

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$$

**Definizione: Indipendenza di più eventi** Assegnati  $n$  eventi  $A_1, \dots, A_n$ , questi si dicono *indipendenti* se per ogni intero  $k$  con  $2 \leq k \leq n$  e per ogni scelta di interi  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ , vale l'eguaglianza

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

**Proposizione:** Gli eventi  $A_1, \dots, A_n$  sono indipendenti se e solo se, per ogni possibile scelta di  $B_i = A_i$  oppure  $B_i = A_i^c$ , vale l'eguaglianza

$$\mathbf{P}(B_1 \cap \dots \cap B_n) = \mathbf{P}(B_1) \dots \mathbf{P}(B_n)$$

## 2 Spazio di probabilità numerabile

Consideriamo un insieme numerabile  $E = \{e_1, e_2, \dots\}$  sul quale sia definita una misura  $\mathbf{m}$ : per ogni insieme  $A \subset E$  si ha

$$\mathbf{m}(A) = \sum_{e_i \in A} \mathbf{m}(e_i)$$

Consideriamo ora una funzione  $f : E \rightarrow \mathbb{R}$ .

**Definizione: Integrale** Si dice che la funzione  $f$  è integrabile se

$$\sum_i |f(e_i)| \mathbf{m}(e_i) < +\infty$$

ed in tal caso chiamiamo *integrale* di  $f$  il numero

$$\int f \, d\mathbf{m} = \sum_i f(e_i) \mathbf{m}(e_i)$$

Indichiamo con  $\mathcal{L}^1$  lo spazio delle funzioni integrabili. Osserviamo ancora che, se  $f$  è a valori positivi, ha sempre senso parlare di integrale di  $f$ , cioè  $\int f \, d\mathbf{m} = \sum_{i \geq 1} f(e_i) \mathbf{m}(e_i) \in [0, +\infty]$ .

Si chiama *trascurabile* un insieme che ha misura nulla; una proprietà verificata ovunque eccetto che su un insieme trascurabile è detta valere *quasi ovunque* (e si scrive q.o.), mentre in probabilità si preferisce dire *quasi certamente* (e si scrive q.c.).

**Teorema: Convergenza monotona** Sia  $(f_n)_{n \geq 1}$  una successione *crescente* di funzioni positive, convergente ad  $f$ : la successione degli integrali  $(\int f_n \, d\mathbf{m})_{n \geq 1}$  converge (crescendo) a  $\int f \, d\mathbf{m}$ .

In maniera più sintetica si scrive

$$0 \leq f_n \quad , \quad f_n \uparrow f \quad \implies \quad \int f_n \, d\mathbf{m} \uparrow \int f \, d\mathbf{m}$$

**Teorema: Convergenza dominata** Sia  $(f_n)_{n \geq 1}$  una successione di funzioni convergente *puntualmente* ad  $f$  e supponiamo che esista  $g$  positiva integrabile tale che si abbia  $|f_n| \leq g$  qualunque sia  $n$ : vale allora la relazione

$$\lim_{n \rightarrow \infty} \int f_n \, d\mathbf{m} = \int f \, d\mathbf{m}$$

**Teorema: Diseguaglianza di Schwartz.** Siano  $f, g$  tali che  $\int f^2 \, d\mathbf{m} < +\infty$  e  $\int g^2 \, d\mathbf{m} < +\infty$ : allora il prodotto  $fg$  è integrabile e vale la diseguaglianza

$$\left| \int fg \, d\mathbf{m} \right| \leq \sqrt{\int f^2 \, d\mathbf{m}} \sqrt{\int g^2 \, d\mathbf{m}}$$

Inoltre, se la diseguaglianza sopra scritta è una eguaglianza, le funzioni  $f$  e  $g$  coincidono a meno di una costante moltiplicativa (cioè esiste  $t$  reale tale che  $f(e_i) = t g(e_i)$  q.o.).

Consideriamo ora  $(\Omega, \mathcal{F}, \mathbf{P})$  con  $\Omega$  numerabile.

**Definizione: Variabile aleatoria** Si chiama *variabile aleatoria* reale (discreta) una funzione  $X : \Omega \rightarrow \mathbb{R}$ .

**Definizione: Legge di Probabilità** Si chiama *legge di probabilità* (o anche *distribuzione di probabilità*) della v.a. reale  $X$  la probabilità definita sui sottinsiemi di  $\mathbb{R}$  dalla formula

$$\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A))$$

La probabilità  $\mathbf{P}_X$  viene anche chiamata la *probabilità immagine* (di  $\mathbf{P}$  mediante  $X$ ) e indicata  $X(\mathbf{P})$ . Quando due variabili aleatorie hanno la stessa legge di probabilità sono dette **equidistribuite**.

Anche l'immagine di  $X$  è un sottinsieme numerabile della retta, cioè  $(x_1, x_2, \dots)$ ; per ogni punto  $x_i$ , si consideri  $p(x_i) = \mathbf{P}\{X = x_i\} = \mathbf{P}(X^{-1}(x_i))$ . Vale la formula:

$$\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A)) = \sum_{x_i \in A} p(x_i)$$

Alla funzione  $x \rightarrow p(x) = \mathbf{P}\{X = x\}$  viene dato il nome di *funzione di probabilità* o anche *densità discreta*.

**Esempio: Variabile Binomiale** La variabile Binomiale di parametri  $n$  e  $p$  ha come valori gli interi  $\{0, 1, \dots, n\}$  e vale, per  $0 \leq k \leq n$ , la formula

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

Quando  $n = 1$  viene anche chiamata **di Bernoulli di parametro  $p$** .

**Esempio: Variabile di Poisson** La variabile di Poisson (di parametro  $\lambda$ ,  $\lambda > 0$ ) è una variabile che assume tutti i valori naturali con probabilità

$$p(n) = \mathbf{P}\{X = n\} = e^{-\lambda} \frac{\lambda^n}{n!}$$

**Esempio: Variabile Geometrica** La variabile Geometrica (di parametro  $p$ ,  $0 < p < 1$ ) ha come valori possibili gli interi strettamente positivi e si ha

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

**Teorema: Integrazione rispetto a una probabilità immagine** Sia  $X$  una v.a. discreta,  $\mathbf{P}_X = X(\mathbf{P})$  la sua *legge di probabilità* e  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ .  $\varphi$  è integrabile rispetto a  $\mathbf{P}_X$  se e solo se  $\varphi \circ X$  è integrabile rispetto a  $\mathbf{P}$ , e in tal caso vale l'eguaglianza

$$\int_{\mathbb{R}} \varphi(x) d\mathbf{P}_X(x) = \int_{\Omega} \varphi(X(\omega)) d\mathbf{P}(\omega) \quad (2.1)$$

**Definizione: Valore atteso** Data una v.a. reale discreta  $X$ , si dice che essa ha *valore atteso* se è integrabile rispetto a  $\mathbf{P}$ , e in tal caso si chiama *valore atteso* (o *speranza matematica*) l'integrale

$$\mathbf{E}[X] = \int_{\Omega} X(\omega) d\mathbf{P}(\omega) = \sum_i X(\omega_i) \mathbf{P}(\omega_i)$$

Se  $X$  è a valori positivi, ha senso scrivere  $\mathbf{E}[X] = \int_{\Omega} X(\omega) d\mathbf{P}(\omega) \in [0, +\infty[$ .

**Definizione: Momenti** Sia  $1 \leq p < +\infty$  e  $X$  una v.a.: si chiama *momento assoluto* di ordine  $p$  il numero

$$\mathbf{E}[|X|^p] = \sum_i |x_i|^p p(x_i) \in [0, +\infty]$$

e se questo numero risulta finito, si dice che  $X$  ammette momento di ordine  $p$ . Dato un intero positivo  $n$  si chiama *momento di ordine  $n$*  (se esiste) il numero  $\mathbf{E}[X^n]$ .

**Proposizione** Siano  $1 \leq p < q < +\infty$ : se  $X$  ha momento di ordine  $q$ , ammette anche momento di ordine  $p$ .

**Definizione: Varianza** Sia  $X$  una variabile aleatoria dotata di momento secondo: si chiama *Varianza* di  $X$  il numero

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

**Diseguaglianza di Markov** Sia  $X$  una v.a. a valori positivi e  $t$  una costante positiva: vale la diseguaglianza

$$t \mathbf{P}\{X \geq t\} \leq \mathbf{E}[X]$$

**Diseguaglianza di Chebishev** Sia  $X$  una v.a. dotata di momento secondo: vale la diseguaglianza

$$t^2 \mathbf{P}\{|X - \mathbf{E}[X]| \geq t\} \leq \text{Var}(X)$$

**Osservazione** La varianza di una v.a.  $X$  è eguale a 0 se e solo se  $X$  è costante q.c.

Consideriamo una variabile aleatoria *doppia* o *bidimensionale*, cioè una applicazione  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ : la sua *legge di probabilità*, denotata  $\mathbf{P}_{X,Y}$ , è una probabilità sui sottinsiemi di  $\mathbb{R}^2$ .

L'immagine di  $(X, Y)$  è un sottinsieme numerabile di  $\mathbb{R}^2$  cioè un insieme di punti  $\{(x_i, y_j) \mid i \geq 1, j \geq 1\}$  e la *funzione di probabilità* è definita da  $p(x_i, y_j) = \mathbf{P}\{X = x_i, Y = y_j\}$ . Per ogni sottinsieme  $B \subset \mathbb{R}^2$  si ha

$$\mathbf{P}_{X,Y}(B) = \mathbf{P}\{(X, Y) \in B\} = \sum_{(x_i, y_j) \in B} p(x_i, y_j)$$

Il *teorema di integrazione rispetto ad una misura immagine* si traduce nell'eguaglianza

$$\begin{aligned} \mathbf{E}[\varphi(X, Y)] &= \int_{\Omega} \varphi(X(\omega), Y(\omega)) d\mathbf{P}(\omega) = \iint_{\mathbb{R}^2} \varphi(x, y) d\mathbf{P}_{X,Y}(x, y) = \\ &= \sum_{x_i, y_j} \varphi(x_i, y_j) p(x_i, y_j) \end{aligned}$$

**Definizione: Covarianza** Supponiamo che  $X$  ed  $Y$  ammettano momento secondo: si chiama *covarianza* il numero

$$\text{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

Se  $\text{Cov}(X, Y) = 0$ , le due variabili sono dette *incorrelate*.

**Proposizione** Siano  $X, Y$  dotate di momento secondo: vale la diseguaglianza

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}$$

Se  $X, Y$  ammettono momento secondo e non sono costanti, si chiama *coefficiente di correlazione* il numero

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

**Matrice delle covarianze** Sia  $(X_1, \dots, X_n)$  una variabile aleatoria  $n$ -dimensionale, supponiamo che ogni componente  $X_i$  abbia momento secondo e indichiamo con  $C$  la *matrice delle covarianze* (cioè  $C_{ij} = Cov(X_i, X_j)$ ).

La matrice  $C$  è *simmetrica, semidefinita positiva*; inoltre vale la formula

$$Var\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i,j=1}^n C_{ij} a_i a_j$$

Sia  $(X, Y)$  una variabile doppia, la sua legge di probabilità è identificata dalla funzione di probabilità  $p(x_i, y_j)$ ; ognuna delle due componenti  $X$  ed  $Y$  è una v.a. reale, e indichiamo con  $p_X(x_i) = \mathbf{P}\{X = x_i\}$  (e analogamente per  $p_Y$ ) le relative funzioni di probabilità.

**Proposizione** Valgono le formule

$$p_X(x_i) = \sum_{y_j} p(x_i, y_j) \quad p_Y(y_j) = \sum_{x_i} p(x_i, y_j)$$

**Definizione** Due variabili aleatorie  $X$  ed  $Y$  si dicono *indipendenti* se, scelti comunque due sottinsiemi  $A$  e  $B$  di  $\mathbb{R}$ , gli eventi  $X^{-1}(A)$  e  $Y^{-1}(B)$  sono indipendenti, cioè se vale la formula

$$\mathbf{P}\{X \in A, Y \in B\} = \mathbf{P}\{X \in A\} \mathbf{P}\{Y \in B\}$$

**Proposizione** Due variabili discrete  $X$  ed  $Y$  sono indipendenti se e solo se le relative funzioni di probabilità sono legate dalla formula

$$p(x_i, y_j) = p_X(x_i) p_Y(y_j) \quad (2.2)$$

**Definizione: Probabilità prodotto** Siano  $\mathbf{P}_1$  e  $\mathbf{P}_2$  due probabilità sui sottinsiemi di  $\mathbb{R}$ : si chiama *probabilità prodotto* (e si indica  $\mathbf{P}_1 \otimes \mathbf{P}_2$ ) la probabilità definita sui sottinsiemi di  $\mathbb{R}^2$  tale che, se  $A, B$  sono sottinsiemi di  $\mathbb{R}$ , si abbia

$$\mathbf{P}_1 \otimes \mathbf{P}_2(A \times B) = \mathbf{P}_1(A) \mathbf{P}_2(B)$$

**Proposizione** Due variabili aleatorie  $X_1, X_2$  sono indipendenti se e solo se la legge di probabilità congiunta è il prodotto delle singole leggi, cioè se si ha

$$\mathbf{P}_{X_1, X_2} = \mathbf{P}_{X_1} \otimes \mathbf{P}_{X_2}$$

Di conseguenza si può dire, per **definizione**, che  $n$  v.a.  $X_1, \dots, X_n$  sono indipendenti se la legge congiunta è il prodotto delle singole leggi, cioè se si ha

$$\mathbf{P}_{X_1, \dots, X_n} = \mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_n}$$

**Proposizione** Siano  $X, Y$  due v.a. indipendenti e  $f, g$  due funzioni reali: le variabili  $f \circ X$  e  $g \circ Y$  sono indipendenti.

**Definizione** Data una famiglia qualsiasi di variabili aleatorie  $(X_i)_{i \in I}$ , queste si dicono *indipendenti* se ogni sottofamiglia finita  $(X_{i_1}, \dots, X_{i_n})$  è formata da variabili indipendenti.

**Teorema: Formula del prodotto** Siano  $X, Y$  due variabili indipendenti dotate di momento primo: anche  $XY$  ammette momento primo e vale la formula

$$\mathbf{E}[XY] = \mathbf{E}[X] \mathbf{E}[Y]$$

**Corollario** Due variabili indipendenti dotate di momento secondo sono *incorrelate*

**Proposizione: Formula della convoluzione discreta** Siano  $X, Y$  due v.a. indipendenti a valori interi (relativi) e sia  $Z = X + Y$ : vale la formula

$$p_Z(n) = \mathbf{P}\{Z = n\} = \sum_{h=-\infty}^{+\infty} p_X(h)p_Y(n-h)$$

**Definizione: Funzione generatrice** Data una variabile aleatoria  $X$  a valori interi positivi, si chiama *funzione generatrice delle probabilità* la funzione  $G_X(\cdot)$  definita da

$$G_X(t) = \sum_{n=0}^{+\infty} t^n p(n) = \mathbf{E}[t^X]$$

**Proposizione.** Valgono le seguenti proprietà:

1.  $G_X(t) = G_Y(t) \iff X$  e  $Y$  sono equidistribuite;
2.  $X$  e  $Y$  indipendenti  $\implies G_{X+Y}(t) = G_X(t) \cdot G_Y(t)$ .

**Proposizione.** Sia  $X$  una v.a. a valori interi positivi: valgono le seguenti eguaglianze

1.  $\mathbf{E}[X] = \lim_{t \rightarrow 1^-} G'_X(t)$
2.  $\mathbf{E}[X(X-1)] = \lim_{t \rightarrow 1^-} G''_X(t)$

Tabella delle funzioni generatrici delle più usuali variabili aleatorie a valori interi:

- $X \sim B(n, p) \implies G_X(t) = [1 + p(t-1)]^n$ ;
- $X$  Geometrica di parametro  $p \implies G_X(t) = \frac{tp}{1-t(1-p)}$ ;
- $X$  di Poisson di parametro  $\lambda \implies G_X(t) = e^{\lambda(t-1)}$ .



### 3 Spazio di probabilità generale

**Definizione** Sia  $\mathcal{A}$  una famiglia di parti di un insieme  $E$ : si chiama  $\sigma$ -algebra generata da  $\mathcal{A}$  la più piccola  $\sigma$ -algebra contenente  $\mathcal{A}$ : essa coincide con l'intersezione di tutte le  $\sigma$ -algre contenenti  $\mathcal{A}$ .

**Proposizione: I boreliani** Sulla retta reale  $\mathbb{R}$  coincidono le  $\sigma$ -algre generate, ad esempio, da queste famiglie di insiemi:

1. le semirette del tipo  $] - \infty, x]$ , al variare di  $x \in \mathbb{R}$ ;
2. gli intervalli semiaperti  $]a, b[$  (oppure  $[a, b[$ ), con  $-\infty < a < b < +\infty$ ;
3. gli aperti di  $\mathbb{R}$ ;
4. i chiusi di  $\mathbb{R}$ .

La  $\sigma$ -algebra da essi generata è chiamata  $\sigma$ -algebra di Borel su  $\mathbb{R}$  (e indicata  $\mathcal{B}(\mathbb{R})$ ) ed i relativi elementi sono detti *boreliani*.

Analoga è la definizione della  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R}^n)$  dei boreliani di  $\mathbb{R}^n$  che è generata, ad esempio, dalle seguenti famiglie di insiemi:

1. gli aperti di  $\mathbb{R}^n$ ;
2. i prodotti cartesiani  $A_1 \times \dots \times A_n$ , dove ogni  $A_i$  è un boreliano di  $\mathbb{R}$ ;
3. i prodotti cartesiani della forma  $] - \infty, x_1] \times \dots \times ] - \infty, x_n]$ .

**Teorema: Unicità di Probabilità** Siano  $\mathbf{P}$  e  $\mathbf{Q}$  due probabilità definite su una  $\sigma$ -algebra  $\mathcal{F}$  di parti di un insieme  $E$  e supponiamo che  $\mathbf{P}$  e  $\mathbf{Q}$  coincidano su una famiglia  $\mathcal{I}$  di parti tale che:

- 1)  $\mathcal{I}$  genera  $\mathcal{F}$ ;
- 2)  $\mathcal{I}$  è stabile per l'intersezione (finita).

Allora  $\mathbf{P}$  e  $\mathbf{Q}$  coincidono su tutto  $\mathcal{F}$ .

**Teorema: Esistenza di Probabilità** Sia  $\mathcal{A}$  un'algebra di parti di un insieme  $E$  e sia  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$  una funzione  $\sigma$ -additiva (tale che  $\mathbf{P}(E) = 1$ ):  $\mathbf{P}$  si prolunga (in un sol modo) alla  $\sigma$ -algebra  $\mathcal{F}$  generata da  $\mathcal{A}$ .

**Definizione: Funzione di ripartizione** Sia  $\mathbf{P}$  una probabilità definita su  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ : si chiama *funzione di ripartizione* la funzione  $F : \mathbb{R} \rightarrow [0, 1]$  definita da  $F(x) = \mathbf{P}(] - \infty, x])$ .

La funzione di ripartizione sopra definita gode delle seguenti proprietà:

1. è crescente;
2. è continua a destra;
3.  $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$  e  $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ .

**Teorema: Esistenza di una Probabilità su  $\mathcal{B}(\mathbb{R})$**  Assegnata una funzione  $F : \mathbb{R} \rightarrow [0, 1]$  con le proprietà 1), 2) e 3) sopra scritte, esiste una ed una sola probabilità  $\mathbf{P}$  su  $\mathcal{B}(\mathbb{R})$  tale che, per ogni  $x \in \mathbb{R}$ , si abbia  $F(x) = \mathbf{P}([-\infty, x])$ .

**Probabilità discreta.**  $\mathbf{P}$  è concentrata su una successione di punti  $(x_1, x_2, \dots)$  e, per ogni  $A \in \mathcal{B}(\mathbb{R})$ , vale l'eguaglianza  $\mathbf{P}(A) = \sum_{x_i \in A} p(x_i)$  essendo  $p(x_i) = \mathbf{P}(\{x_i\})$ . In particolare la funzione di ripartizione soddisfa l'eguaglianza  $F(x) = \sum_{x_i \leq x} p(x_i)$ .

**Probabilità diffusa** Ogni punto è trascurabile per la probabilità  $\mathbf{P}$  associata alla funzione di ripartizione  $F$  se e solo se  $F$  è continua: questo è una conseguenza della formula  $\mathbf{P}(\{x\}) = \Delta F(x)$ . Le probabilità che godono di questa proprietà sono dette *diffuse*.

**Definizione: Spazio e applicazione misurabile** Si chiama *spazio misurabile* una coppia  $(E, \mathcal{E})$  dove  $E$  è un insieme e  $\mathcal{E}$  una  $\sigma$ -algebra di parti di  $E$ . Dati due spazi misurabili  $(E, \mathcal{E})$  e  $(F, \mathcal{F})$ , una applicazione  $f : E \rightarrow F$  è detta *misurabile* se, per ogni  $A \in \mathcal{F}$ ,  $f^{-1}(A) \in \mathcal{E}$ .

**Condizione sufficiente** Se  $\mathcal{A}$  è una famiglia di parti di  $F$  che genera la  $\sigma$ -algebra  $\mathcal{F}$ , affinché una funzione  $f : E \rightarrow F$  sia misurabile, è sufficiente che, per ogni  $A \in \mathcal{A}$ ,  $f^{-1}(A) \in \mathcal{E}$ .

Una funzione misurabile da  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  su  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  è detta *boreliana*.

**Definizione: Funzione semplice** Dato uno spazio misurabile  $(E, \mathcal{E})$ , si chiama *semplice* una funzione misurabile  $\varphi : E \rightarrow \mathbb{R}$  che prende un numero finito di valori (cioè la cui immagine è un insieme finito).

**Definizione: Integrale delle funzioni semplici** Sia  $\varphi$  una funzione semplice della forma  $\varphi = \sum_{i=1}^n a_i I_{A_i}$ : definiamo *integrale* di  $\varphi$  il numero

$$\int_E \varphi(x) \, d\mathbf{m}(x) = \sum_{i=1}^n a_i \mathbf{m}(A_i)$$

**Teorema: Approssimazione con funzioni semplici** Sia  $f$  una funzione misurabile a valori positivi: esiste una successione di funzioni semplici  $(\varphi_n)_{n \geq 1}$  tale che

$$\varphi_n \uparrow f$$

**Definizione: Integrale delle funzioni a valori positivi** Sia  $f$  una funzione misurabile a valori positivi e consideriamo una successione di funzioni semplici  $(\varphi_n)_{n \geq 1}$  tale che  $\varphi_n \uparrow f$ : si definisce *integrale* di  $f$  il numero

$$\int f \, d\mathbf{m} = \lim_{n \geq 1} \int \varphi_n \, d\mathbf{m}$$

**Teorema: Convergenza monotona** Se  $(f_n)_{n \geq 1}$  è una successione di funzioni misurabili a valori positivi, si ha

$$f_n \uparrow f \implies \int f_n \, d\mathbf{m} \uparrow \int f \, d\mathbf{m}$$

Consideriamo ora una generica funzione misurabile  $f$ , e poniamo  $f^+ = f \vee 0 = \max(f, 0)$  e  $f^- = -(f \wedge 0) = -\min(f, 0)$ : entrambe sono funzioni misurabili e si ha  $|f| = f^+ + f^-$  e  $f = f^+ - f^-$ .

**Definizione: Funzione integrabile e integrale** Si dice che la funzione misurabile  $f$  è *integrabile* se  $\int |f| \, d\mathbf{m} < +\infty$ , e in tal caso si chiama *integrale* di  $f$  il numero

$$\int f \, d\mathbf{m} = \int f^+ \, d\mathbf{m} - \int f^- \, d\mathbf{m}.$$

**Teorema: Convergenza dominata** Sia  $(f_n)_{n \geq 1}$  una successione di funzioni misurabili convergente puntualmente ad  $f$  e supponiamo che esista  $g$  integrabile a valori positivi tale che si abbia, per ogni  $x \in E$ ,  $|f_n(x)| \leq g(x)$ : allora si ha

$$\lim_{n \rightarrow \infty} \int f_n \, d\mathbf{m} = \int f \, d\mathbf{m}.$$

Vale la *diseguaglianza di Schwartz*: se  $f^2$  e  $g^2$  sono integrabili, il prodotto  $fg \in \mathcal{L}^1$  e si ha

$$\left| \int fg \, d\mathbf{m} \right| \leq \sqrt{\int f^2 \, d\mathbf{m}} \sqrt{\int g^2 \, d\mathbf{m}}.$$

**Definizione: Densità di probabilità** Si chiama *densità di probabilità* su  $\mathbb{R}$  una funzione reale  $f$  definita su  $\mathbb{R}$ , misurabile e a valori positivi, integrabile (secondo Lebesgue) e tale che  $\int_{-\infty}^{+\infty} f(x) \, dx = 1$ .

Ad una densità  $f$  è associata una probabilità  $\mathbf{P}$  su  $\mathcal{B}(\mathbb{R})$  mediante la formula

$$\mathbf{P}(A) = \int_A f(x) \, dx$$

**Teorema: Integrazione rispetto a una misura definita da una densità** Una funzione misurabile  $g$  definita su  $\mathbb{R}$  è integrabile rispetto a  $\mathbf{P}$  se e solo se il prodotto  $gf$  è integrabile rispetto alla misura di Lebesgue, e in tal caso si ha

$$\int g(x) \, d\mathbf{P}(x) = \int g(x)f(x) \, dx.$$

Analoga è la definizione di *probabilità definita da una densità* su  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ , ed il relativo teorema di integrazione.

**Proposizione: Funzioni assolutamente continue** La probabilità associata ad una funzione di ripartizione  $F$  è definita da una densità se e solo se  $F$  è *assolutamente continua*, cioè per ogni  $\varepsilon > 0$ , esiste  $\delta > 0$  tale che, prese delle coppie di punti  $(x_i, y_i)$ ,

$$\sum_{i \leq n} |x_i - y_i| < \delta \implies \sum_{i \leq n} |F(x_i) - F(y_i)| < \varepsilon$$

**Definizione: Variabile aleatoria reale** Assegnato uno spazio di Probabilità  $(\Omega, \mathcal{F}, \mathbf{P})$ , si chiama *variabile aleatoria reale* una applicazione misurabile  $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

**Definizione: Legge di Probabilità** Si chiama *legge di probabilità* (o anche *distribuzione di probabilità*) di una variabile aleatoria reale  $X$  l'immagine di  $\mathbf{P}$  mediante  $X$ , cioè la probabilità  $\mathbf{P}_X$  definita sui boreliani dalla formula  $\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A))$ ; si chiama *funzione di ripartizione* di  $X$  la funzione di ripartizione della sua legge di probabilità.

**Teorema: Integrazione rispetto ad una probabilità immagine** Sia  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  boreliana:  $\varphi$  è integrabile rispetto a  $\mathbf{P}_X$  se e solo se  $\varphi \circ X$  è integrabile rispetto a  $\mathbf{P}$  e in tal caso vale la formula

$$\int_{\mathbb{R}} \varphi(x) d\mathbf{P}_X(x) = \int_{\Omega} \varphi(X(\omega)) d\mathbf{P}(\omega).$$

*Per definizione*, si chiama *variabile aleatoria doppia* una applicazione misurabile  $(X, Y) : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ ; le componenti  $X$  e  $Y$  sono due variabili aleatorie reali.

La *legge di probabilità* della coppia  $(X, Y)$  è l'immagine di  $\mathbf{P}$  mediante l'applicazione  $(X, Y)$ : il teorema di integrazione rispetto a una probabilità immagine si estende al caso vettoriale, in particolare presa  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  boreliana e limitata, vale la formula

$$\int_{\Omega} \varphi(X(\omega), Y(\omega)) d\mathbf{P}(\omega) = \iint_{\mathbb{R}^2} \varphi(x, y) d\mathbf{P}_{X,Y}(x, y)$$

**Definizione: Probabilità prodotto** Siano  $\mathbf{P}$  e  $\mathbf{Q}$  due probabilità su  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ : si chiama *probabilità prodotto* (e si indica  $\mathbf{P} \otimes \mathbf{Q}$ ) la probabilità su  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$  tale che, presi comunque due sottinsiemi boreliani  $A$  e  $B$  di  $\mathbb{R}$ , si abbia

$$\mathbf{P} \otimes \mathbf{Q}(A \times B) = \mathbf{P}(A) \cdot \mathbf{Q}(B)$$

Se  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  è boreliana e limitata (oppure a valori positivi) vale la formula di integrazione

$$\iint_{\mathbb{R}^2} \varphi(x, y) d\mathbf{P} \otimes \mathbf{Q}(x, y) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \varphi(x, y) d\mathbf{Q}(y) \right] d\mathbf{P}(x)$$

Vale l'estensione al caso generale della caratterizzazione provata nel caso delle variabili discrete: più precisamente  $X$  e  $Y$  sono indipendenti se e solo se  $\mathbf{P}_{X,Y} = \mathbf{P}_X \otimes \mathbf{P}_Y$ .

**Teorema: Formula del prodotto** Supponiamo che  $X$  ed  $Y$  siano indipendenti e dotate di momento primo: anche  $XY$  ha valore atteso e vale la formula

$$\mathbf{E}[XY] = \mathbf{E}[X] \mathbf{E}[Y]$$

**Definizione: Variabile con densità** Si dice che la v.a. reale  $X$  ha densità  $f$  se la sua legge di probabilità  $\mathbf{P}_X$  ha densità  $f$ , cioè se per ogni boreliano  $A$  vale la formula

$$\mathbf{P}\{X \in A\} = \mathbf{P}_X(A) = \int_A f(x) dx$$

**Proposizione** Sia  $X$  una variabile aleatoria reale. Sono equivalenti le due seguenti affermazioni:

1.  $X$  ha densità  $f$ ;
2. per ogni funzione reale  $\varphi$  boreliana e limitata, vale la formula

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x) dx$$

**Proposizione** Sia  $(X, Y)$  una variabile doppia con densità  $f(x, y)$ : anche le componenti  $X$  ed  $Y$  ammettono densità  $f_1$  ed  $f_2$  che soddisfano le formule

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

**Proposizione** Sia  $(X, Y)$  una variabile doppia con densità: le variabili  $X$  e  $Y$  sono indipendenti se e solo se tra le densità vale la seguente relazione (quasi ovunque)

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

**Proposizione: Formula della convoluzione** Siano  $X, Y$  due variabili indipendenti con densità rispettivamente  $f_1$  ed  $f_2$ : la somma  $(X + Y)$  ha densità  $g$  data dalla formula

$$g(x) = \int_{-\infty}^{+\infty} f_1(x - y) f_2(y) dy$$

**Proposizione** Sia  $X$  una v.a. reale con densità  $f$  diversa da 0 su un aperto  $A \subseteq \mathbb{R}$  e sia  $h : A \rightarrow B$  un diffeomorfismo. Consideriamo la variabile  $Y = h(X)$  : essa ha densità  $g$  data da

$$g(y) = \begin{cases} 0 & \text{se } y \notin B \\ f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| = f(x(y)) \left| \frac{dx(y)}{dy} \right| & \text{se } y \in B \end{cases}$$

Vediamo in concreto l'estensione di questa formula al caso bidimensionale, prendendo una variabile doppia  $(X, Y)$  con densità  $f$  diversa da 0 sull'aperto  $A$  di  $\mathbb{R}^2$ , considerando un diffeomorfismo  $h$  da  $A$  su  $B$  e chiamando  $(U, V) = h(X, Y)$ . La coppia  $(U, V)$  ha una densità  $g$  che si annulla fuori di  $B$ , mentre su  $B$  soddisfa la formula

$$g(u, v) = f(x(u, v), y(u, v)) \cdot \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

dove con  $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$  si intende il *valore assoluto del determinante* della matrice  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ .

La funzione *Gamma* è definita, per  $r > 0$ , da  $\Gamma(r) = \int_0^{+\infty} x^{r-1} e^{-x} dx$ .

**Densità Gamma** Si chiama densità Gamma di parametri  $r$  e  $\lambda$ , ( $r > 0$ ,  $\lambda > 0$ ), (e si indica  $\Gamma(r, \lambda)$ ) la funzione definita da

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

Quando  $r = 1$ , la densità  $\Gamma(1, \lambda)$  si chiama più semplicemente *esponenziale di parametro  $\lambda$* .

Se  $X \sim \Gamma(r, \lambda)$  e  $\beta > 0$ , vale la formula

$$\mathbf{E}[X^\beta] = \frac{\Gamma(r + \beta)}{\Gamma(r) \lambda^\beta}$$

Se  $X \sim \Gamma(r_1, \lambda)$ ,  $Y \sim \Gamma(r_2, \lambda)$  e sono indipendenti, allora  $(X + Y) \sim \Gamma(r_1 + r_2, \lambda)$

Laplace ha calcolato che  $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$  : ne segue che la funzione  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  è una densità di probabilità, detta densità **Normale** o **Gaussiana**  $N(0, 1)$ , e la funzione  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$  è la relativa *funzione di ripartizione*.

Per una variabile  $X \sim N(0, 1)$  si ha  $\mathbf{E}[X] = 0$  e  $Var(X) = \mathbf{E}[X^2] = 1$ .

**Densità Gaussiana** Si dice che la variabile  $X$  ha legge gaussiana  $N(m, \sigma^2)$  ( $m \in \mathbb{R}, \sigma > 0$ ) se  $\frac{X-m}{\sigma}$  ha legge  $N(0, 1)$ . La densità di  $X$  è di conseguenza la funzione  $g$  definita da

$$g(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

inoltre  $\mathbf{E}[X] = m$ ,  $\text{Var}(X) = \sigma^2$ .

Se  $X \sim N(m_1, \sigma_1^2)$ ,  $Y \sim N(m_2, \sigma_2^2)$  e sono indipendenti, allora  $(X + Y) \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ .

Se  $X \sim N(0, 1)$ , allora  $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$ .

**Definizione: Quantile** Data una funzione di ripartizione  $F$  ed un numero  $0 < \alpha < 1$ , si chiama  $\alpha$ -quantile di  $F$  il numero così definito

$$r_\alpha = \inf \{x \in \mathbb{R} \mid F(x) > \alpha\}.$$

Il quantile della Legge Gaussiana  $N(0, 1)$  è indicato  $q_\alpha$ .

**Definizione: Legge chi-quadro** Si chiama *legge chi-quadro a n gradi di libertà* (e si indica  $\chi^2(n)$ ) la legge  $\Gamma(\frac{n}{2}, \frac{1}{2})$ .

Se  $(X_1, \dots, X_n)$  sono indipendenti gaussiane  $N(0, 1)$ , allora  $X_1^2 + \dots + X_n^2$  ha legge  $\chi^2(n)$ .

**Definizione: Legge di Student** Siano  $X \sim N(0, 1)$ ,  $Y \sim \chi^2(n)$  indipendenti: si chiama *legge di Student a n gradi di libertà* (e si indica  $T(n)$ ) la legge di

$$\frac{\sqrt{n} X}{\sqrt{Y}}$$

**Definizione: Legge di Fisher** Siano  $C_n$  e  $C_m$  due variabili indipendenti con legge rispettivamente  $\chi^2(n)$  e  $\chi^2(m)$ : si chiama *legge di Fisher*  $F_{n,m}$  la legge di

$$\frac{C_n/n}{C_m/m}$$

## 4 Convergenza e teoremi limite

**Definizione. Convergenza in probabilità** Si dice che la successione di variabili aleatorie  $(X_n)_{n \geq 1}$  converge in probabilità alla v.a.  $X$  se, per ogni  $\varepsilon > 0$ , si ha

$$\lim_{n \rightarrow \infty} \mathbf{P}\{|X_n - X| > \varepsilon\} = 0$$

**Teorema. Legge dei grandi numeri** Sia  $X_1, X_2, \dots$  una successione di variabili aleatorie dotate di momento secondo, incorrelate, e supponiamo

che  $\mathbf{E}[X_i] = m$  per ogni  $i$  (cioè hanno tutte lo stesso valore atteso) e che esista una costante  $K$  tale che si abbia  $\text{Var}(X_i) \leq K$  qualunque sia  $i$  (cioè le varianze sono equilimitate). Allora, posto  $S_n = X_1 + \dots + X_n$ , la successione  $\left(\frac{S_n}{n}\right)_{n \geq 1}$  converge in probabilità ad  $m$ .

**Proposizione** Sia  $(X_n)_{n \geq 1}$  una successione convergente in probabilità a  $c$  e sia  $g$  una funzione boreliana continua nel punto  $c$ : allora  $Y_n = g(X_n)$  converge in probabilità a  $g(c)$ .

**Definizione. Convergenza in legge** Si dice che la successione di v.a.  $(X_n)_{n \geq 1}$  converge in legge (o anche in distribuzione) alla v.a.  $X$  se per ogni  $f: \mathbb{R} \rightarrow \mathbb{R}$  continua e limitata, si ha

$$\lim_{n \rightarrow \infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)]$$

**Proposizione** Siano  $X_n$  e  $X$  variabili aleatorie,  $F_n$  ed  $F$  le relative funzioni di ripartizione; supponiamo inoltre che  $F$  sia continua (cioè la legge di  $X$  sia diffusa). Allora sono equivalenti le seguenti affermazioni:

- a) la successione  $(X_n)_{n \geq 1}$  converge a  $X$  in legge;
- b) per ogni  $x \in \mathbb{R}$ , si ha  $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ .

**Teorema. Limite Centrale per Variabili Binomiali** Sia  $X_1, X_2, \dots$  una successione di variabili indipendenti di Bernoulli di parametro  $p$  e sia  $S_n = X_1 + \dots + X_n$ : presi due numeri  $a, b$  con  $-\infty \leq a < b \leq +\infty$ , si ha

$$\lim_{n \rightarrow \infty} \mathbf{P}\left\{a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

**Teorema. Limite Centrale di Paul Lévy** Sia  $X_1, X_2, \dots$  una successione di variabili indipendenti equidistribuite, dotate di momento primo  $\mu$  e di varianza  $\sigma^2$  (diversa da 0): posto  $S_n = X_1 + \dots + X_n$ , la successione

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} = \sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right)$$

converge in legge alla variabile gaussiana  $N(0, 1)$ .

## 5 Inferenza statistica

**Definizione: Modello statistico** Si chiama *modello statistico* una terna  $(\Omega, \mathcal{F}, (\mathbf{P}^\theta, \theta \in \Theta))$  dove  $\Omega$  è un insieme,  $\mathcal{F}$  una  $\sigma$ -algebra di parti di  $\Omega$  e, per ogni  $\theta \in \Theta$ ,  $\mathbf{P}^\theta$  è una probabilità su  $(\Omega, \mathcal{F})$ .



A due parametri diversi  $\theta_1$  e  $\theta_2$  corrispondono due probabilità diverse (il modello è **identificabile**). Si chiama **trascurabile** un evento  $A \in \mathcal{F}$  trascurabile per ogni probabilità  $\mathbf{P}^\theta$ .

**Definizione. Verosimiglianza in un modello statistico discreto**  
Assegnato un modello statistico  $(\Omega, \mathcal{F}, (\mathbf{P}^\theta, \theta \in \Theta))$  con  $\Omega$  numerabile, si chiama *verosimiglianza* la funzione  $L : \Theta \times \Omega \rightarrow \mathbb{R}^+$  definita da

$$L(\theta, \omega) = \mathbf{P}^\theta(\{\omega\})$$

**Definizione. Modello con densità** Il modello statistico è detto *con densità* se soddisfa le seguenti condizioni:

- a)  $\Omega$  è uno spazio euclideo  $\mathbb{R}^n$  (o un sottinsieme misurabile di uno spazio euclideo);
- b)  $\mathcal{F}$  è la  $\sigma$ -algebra di Borel su  $\Omega$ ;
- c) le probabilità  $\mathbf{P}^\theta$  ammettono *densità* rispetto alla misura di Lebesgue  $n$ -dimensionale  $\lambda$ .

**Definizione. Verosimiglianza in un modello con densità** Si chiama *verosimiglianza* una funzione  $L : \Theta \times \Omega \rightarrow \mathbb{R}^+$  tale che, fissato  $\theta$ ,  $L(\theta, \cdot)$  sia una versione della *densità* di  $\mathbf{P}^\theta$  (rispetto alla misura di Lebesgue  $\lambda$ ).

**Definizione: Campione** Sia  $(\mathbf{m}^\theta, \theta \in \Theta)$  una famiglia parametrizzata di leggi di probabilità su  $\mathbb{R}$ : si chiama *campione di taglia  $n$  e legge  $\mathbf{m}^\theta$*  una famiglia  $(X_1, \dots, X_n)$  di  $n$  variabili aleatorie indipendenti ciascuna con legge  $\mathbf{m}^\theta$ .

Cominciamo col caso in cui ogni probabilità  $\mathbf{m}^\theta$  è discreta: il modo canonico per rappresentare come modello statistico un campione di legge  $(\mathbf{m}^\theta, \theta \in \Theta)$  è il seguente. Sia  $C$  l'insieme su cui sono concentrate le probabilità  $\mathbf{m}^\theta$ , e poniamo (per  $\theta \in \Theta$  e  $x_i \in C$ ),  $p(\theta, x_i) = \mathbf{m}^\theta(\{x_i\})$ .

Poniamo poi  $\Omega = C^n$ ,  $\mathcal{F} = \mathcal{P}(\Omega)$  e scegliamo come verosimiglianza

$$L(\theta; x_1, \dots, x_n) = p(\theta, x_1) \cdots p(\theta, x_n)$$

Considerando come  $X_i$  la proiezione canonica di indice  $i$  da  $\Omega$  su  $C$ , le variabili  $X_1, \dots, X_n$  sono effettivamente indipendenti e ciascuna con legge  $\mathbf{m}^\theta$  (se si considera su  $\Omega$  la probabilità  $\mathbf{P}^\theta$ ).

Vediamo ora il caso in cui le probabilità  $\mathbf{m}^\theta$  sono definite da una densità. Sia  $(f(\theta, \cdot), \theta \in \Theta)$  una famiglia parametrizzata di densità di probabilità su  $\mathbb{R}$ : si chiama *campione di taglia  $n$  e densità  $f(\theta, \cdot)$*  una famiglia di variabili aleatorie indipendenti, equidistribuite, aventi densità  $f(\theta, \cdot)$  (sotto  $\mathbf{P}^\theta$ ).

La costruzione canonica del modello è la seguente: si prende  $\Omega = \mathbb{R}^n$  e si considera come *verosimiglianza* la funzione

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(\theta, x_i)$$

Si definiscono inoltre come variabili  $X_i$  le *proiezioni canoniche* di indice  $i$ : è immediato verificare che ponendo su  $\Omega$  la probabilità  $\mathbf{P}^\theta$  definita dalla densità  $L(\theta, \cdot)$  queste variabili risultano indipendenti ciascuna con densità  $f(\theta, \cdot)$ .

Se ogni densità  $f(\theta, \cdot)$  si annulla fuori di un intervallo  $\mathcal{I} \subseteq \mathbb{R}$ , conviene considerare come spazio  $\Omega = \mathcal{I}^n$  anzichè  $\mathbb{R}^n$ .

**Definizione: Stima** Assegnato un modello statistico  $(\Omega, \mathcal{F}, (\mathbf{P}^\theta, \theta \in \Theta))$ , si chiama *stima* una variabile aleatoria  $U : \Omega \rightarrow \mathbb{R}$ .

In genere una stima è accoppiata ad una funzione  $g : \Theta \rightarrow \mathbb{R}$  e lo scopo di  $U$  è appunto valutare  $g(\theta)$ .

**Definizione: Stima corretta** Assegnata una funzione  $g : \Theta \rightarrow \mathbb{R}$ , la stima  $U$  di  $g(\theta)$  è detta *corretta* se, per ogni  $\theta$ ,  $U$  è  $\mathbf{P}^\theta$ -integrabile e si ha  $\mathbf{E}^\theta[U] = g(\theta)$ .

**Definizione: Stima consistente** Sia  $(\mathbf{m}^\theta, \theta \in \Theta)$  una famiglia di leggi di probabilità discrete su  $\mathbb{R}$  e consideriamo, per ogni  $n$ , un campione  $X_1, \dots, X_n$  di legge  $\mathbf{m}^\theta$ ; sia poi  $U_n = h_n(X_1, \dots, X_n)$  una stima di  $g(\theta)$  basata sulle osservazioni del campione  $n$ -simo. Si dice che la successione di stime  $(U_n)_{n \geq 1}$  è *consistente* se, scelti comunque  $\theta \in \Theta$  ed  $\varepsilon > 0$ , si ha

$$\lim_{n \rightarrow \infty} \mathbf{P}^\theta \{|U_n - g(\theta)| > \varepsilon\} = 0$$

**Definizione: Rischio** Sia  $U$  una stima della funzione  $g(\theta)$ : si chiama *Rischio* (quadratico) il numero

$$R(\theta, U) = \mathbf{E}^\theta[(U - g(\theta))^2]$$

La definizione di rischio introduce un criterio di *ordinamento parziale* tra le stime, più precisamente diremo che

- $U$  è *preferibile* a  $V$  se, per ogni  $\theta$ ,  $R(\theta, U) \leq R(\theta, V)$ ;
- $U$  è *strettamente preferibile* a  $V$  se è preferibile e, per almeno un parametro  $\bar{\theta}$ ,  $R(\bar{\theta}, U) < R(\bar{\theta}, V)$ ;
- $U$  è *ammissibile* se non esistono stime strettamente preferibili a  $U$ ;
- $U$  è *ottimale* se è preferibile a ogni altra stima.

**Definizione: Riassunto esaustivo** Sia  $T : \Omega \rightarrow E$  una variabile aleatoria: si dice che  $T$  è un *riassunto esaustivo* se si può scrivere la verosimiglianza nella forma

$$L(\theta, \omega) = h(\theta, T(\omega)) k(\omega)$$

**Teorema: Stima ed esaustività** Sia  $T$  un riassunto esaustivo,  $U$  una stima di  $g(\theta)$  e supponiamo che  $U$  sia di quadrato integrabile per ogni probabilità  $\mathbf{P}^\theta$ . Esiste una stima  $V$  della forma  $V(\omega) = f(T(\omega))$  preferibile a  $U$ , inoltre  $V$  è *strettamente* preferibile a meno che  $U$  non sia già nella forma  $f \circ T$ . Infine, se  $U$  è corretta, anche  $V$  è corretta.

**Definizione: Stima di massima verosimiglianza** Sia assegnato un modello statistico  $(\Omega, \mathcal{F}, (\mathbf{P}^\theta, \theta \in \Theta))$  tale che  $\Theta \subset \mathbb{R}$ : si dice che  $U$  è una stima di massima verosimiglianza del parametro  $\theta$  se, per ogni  $\omega \in \Omega$ , si ha

$$L(U(\omega), \omega) = \sup_{\theta \in \Theta} L(\theta, \omega)$$

Usualmente la stima di massima verosimiglianza, se esiste, viene indicata  $\hat{\theta}(\omega)$ .

**Teorema** Sia  $(\mathbf{m}^\theta, \theta \in \Theta)$  una famiglia di leggi di probabilità concentrate sugli interi positivi, e supponiamo che  $\Theta$  sia un *intervallo* di  $\mathbb{R}$  e che, ponendo  $p(\theta, k) = \mathbf{m}^\theta(\{k\})$ , questa si possa scrivere nella forma

$$p(\theta, k) = c(\theta) \exp(\theta T(k)) g(k)$$

dove  $T : \mathbb{N} \rightarrow \mathbb{R}$ . Consideriamo un campione infinito  $X_1, X_2, \dots$  di legge  $\mathbf{m}^\theta$  e supponiamo che esista, per ogni  $n$ , la stima di massima verosimiglianza  $\hat{\theta}_n$  relativa al campione di taglia  $n$ : allora la successione di stime  $(\hat{\theta}_n)_{n \geq 1}$  è consistente.

**Teorema** Supponiamo che  $\Theta$  sia un *intervallo* di  $\mathbb{R}$  e sia assegnata una famiglia di densità  $(f(\theta, x), \theta \in \Theta)$  che si possano scrivere nella forma

$$f(\theta, x) = c(\theta) \cdot \exp(\theta T(x)) \cdot g(x)$$

con una opportuna applicazione  $T : \mathbb{R} \rightarrow \mathbb{R}$ . Consideriamo un campione infinito  $X_1, X_2, \dots$  con densità  $f(\theta, \cdot)$  e supponiamo che esista, per ogni  $n$ , la stima di massima verosimiglianza  $\hat{\theta}_n$  relativa al campione di taglia  $n$ : allora la successione di stime  $(\hat{\theta}_n)_{n \geq 1}$  è consistente.

**Definizione: Regione di Fiducia** Sia assegnato, per ogni  $\omega \in \Omega$ , un sottoinsieme dei parametri  $C(\omega) \subset \Theta$ : si dice che  $C(\omega)$  è una *regione di fiducia* per il parametro  $\theta$  al livello  $(1 - \alpha)$  se, qualunque sia  $\theta$ , si ha

$$\mathbf{P}^\theta \{\omega \mid \theta \in C(\omega)\} \geq 1 - \alpha$$

o (ciò che è lo stesso)  $\mathbf{P}^\theta \{\omega \mid \theta \notin C(\omega)\} \leq \alpha$ .

Se  $\Theta \subseteq \mathbb{R}$  e  $C(\omega)$  è un intervallo, si parla di *intervallo di fiducia*.

**Esempio: Intervallo di fiducia per il controllo di qualità** Consideriamo un campione  $X_1, \dots, X_n$  di legge di Bernoulli di parametro  $\theta$  e vogliamo individuare un intervallo di fiducia per il parametro  $\theta$ : partiamo dal fatto che  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  è una stima corretta di  $\theta$  e che  $Var^\theta(\bar{X}) = \frac{\theta(1-\theta)}{n}$ .

Ci aspettiamo un intervallo di fiducia per  $\theta$  intorno alla sua stima, più precisamente della forma  $I = [\bar{X}(\omega) - d, \bar{X}(\omega) + d]$  (con  $d$  da determinare).

Utilizzando la disuguaglianza di Chebiscev si ottiene, al livello  $(1 - \alpha)$ , l'intervallo di fiducia  $[\bar{X}(\omega) - \frac{1}{\sqrt{4n\alpha}}, \bar{X}(\omega) + \frac{1}{\sqrt{4n\alpha}}]$ , o (come si scrive più sinteticamente)  $\bar{X}(\omega) \pm \frac{1}{\sqrt{4n\alpha}}$ .

Utilizzando invece il *Teorema limite di De Moivre-Laplace* si ottiene invece l'intervallo di fiducia (approssimato)  $\bar{X}(\omega) \pm \frac{q_{1-\frac{\alpha}{2}}}{2\sqrt{n}}$ .

Per **pianificare un test** bisogna per prima cosa *formulare un'ipotesi* e questo si ottiene stabilendo una partizione dell'insieme dei parametri  $\Theta$  in due sottinsiemi non vuoti  $\Theta_0$  e  $\Theta_1$ . L'ipotesi e l'alternativa sono indicate rispettivamente  $\mathcal{H}_0$ ) e  $\mathcal{H}_1$ ) e si usa dire, ad esempio: consideriamo un test dell'ipotesi  $\mathcal{H}_0) \theta \in \Theta_0$  contro l'alternativa  $\mathcal{H}_1) \theta \in \Theta_1$ .

Il secondo passo è *pianificare un esperimento*, cioè stabilire una regola che, secondo il risultato dell'esperienza  $\omega$ , permetta di decidere se accettare o rifiutare l'ipotesi. Questo equivale a scegliere un *evento*  $D \in \mathcal{F}$  che consiste nell'insieme dei risultati  $\omega$  che portano a rifiutare l'ipotesi: tale insieme  $D$  viene chiamato *regione critica*.

**Definizione: Livello e potenza** Si chiama *taglia* di un test di regione critica  $D$  il numero

$$\sup_{\theta \in \Theta_0} \mathbf{P}^\theta(D)$$

Si dice che il test è di *livello*  $\alpha$  se la sua taglia è minore o eguale ad  $\alpha$ .

Si chiama *potenza* del test la funzione  $\pi_D : \Theta_1 \rightarrow [0, 1]$  definita da  $\theta \rightarrow \mathbf{P}^\theta(D)$ .

Diremo che il test di regione critica  $D$  è *più potente* del test di regione critica  $D^*$  se, per ogni  $\theta \in \Theta_1$ , si ha  $\mathbf{P}^\theta(D) \geq \mathbf{P}^\theta(D^*)$ .

Quando  $\Theta_0$  è ridotto a un solo punto (cioè  $\Theta_0 = \{\theta_0\}$ ) si dice che *l'ipotesi è semplice*.

**Lemma di Neyman-Pearson** Supponiamo assegnato un modello statistico nel quale l'insieme  $\Theta$  dei parametri è ridotto a due punti ( $\Theta = \{\theta_0, \theta_1\}$ ) e sia dato il test dell'ipotesi  $\mathcal{H}_0) \theta = \theta_0$  contro  $\mathcal{H}_1) \theta = \theta_1$ . Consideriamo l'insieme  $D$  così definito

$$D = \{ \omega \in \Omega \mid L(\theta_0, \omega) \leq c L(\theta_1, \omega) \}$$

dove  $c$  è una costante positiva. Allora

1.  $D$  è la regione critica di un test più potente di ogni altro test di livello  $\mathbf{P}^{\theta_0}(D)$  ;
2. vale la disuguaglianza  $\mathbf{P}^{\theta_1}(D) \geq \mathbf{P}^{\theta_0}(D)$ .

**Definizione: Rapporto di verosimiglianza crescente** Supponiamo assegnato un modello statistico nel quale l'insieme dei parametri  $\Theta$  è un intervallo di  $\mathbb{R}$  e sia  $T$  una variabile aleatoria reale definita su  $\Omega$ : si dice che il modello è *a rapporto di verosimiglianza crescente* rispetto a  $T$  se, scelti comunque  $\theta_1 < \theta_2$ , esiste una funzione reale (strettamente) crescente a valori positivi  $f_{\theta_1, \theta_2}$  tale che valga l'eguaglianza

$$\frac{L(\theta_2, \omega)}{L(\theta_1, \omega)} = f_{\theta_1, \theta_2}(T(\omega))$$

**Teorema: Test unilatero** Supponiamo che il modello sia a rapporto di verosimiglianza crescente rispetto a  $T$  e consideriamo il test unilatero  $\mathcal{H}_0) \theta \leq \theta_0$  contro l'alternativa  $\mathcal{H}_1) \theta > \theta_0$ ; consideriamo poi l'insieme  $D = \{\omega \mid T(\omega) \geq d\}$  dove  $d$  è un opportuno numero. Il test di regione critica  $D$  è tale che:

1. vale l'eguaglianza  $\sup_{\theta \leq \theta_0} \mathbf{P}^\theta(D) = \mathbf{P}^{\theta_0}(D)$ ;
2.  $D$  è più potente di qualsiasi altro test  $D^*$  con *livello*  $\mathbf{P}^{\theta_0}(D)$ .

**Soglia di accettazione** Spesso ci si trova in questa situazione: per ogni numero  $0 < \alpha < 1$ , è assegnata una regione critica  $D_\alpha$  di livello  $\alpha$  in modo tale che, se  $\alpha_1 \leq \alpha_2$ , allora  $D_{\alpha_1} \subseteq D_{\alpha_2}$ . Inoltre  $\cup_{0 < \alpha < 1} D_\alpha = \Omega$  e  $\cap_{0 < \alpha < 1} D_\alpha = \emptyset$ .

Allora, per ogni  $\bar{\omega} \in \Omega$  (cioè per ogni *risultato dell'indagine statistica*) è assegnato un numero  $\bar{\alpha}$  tale che, se  $\alpha < \bar{\alpha}$ ,  $\bar{\omega} \notin D_\alpha$  e se  $\alpha > \bar{\alpha}$ ,  $\bar{\omega} \in D_\alpha$ . Tale numero  $\bar{\alpha}$  sarà chiamato *soglia di accettazione*.

## 6 Statistica sui modelli gaussiani

**Lemma** Sia  $\mathbf{X} = (X_1, \dots, X_n)$  un vettore aleatorio formato da  $n$  v.a. indipendenti con densità  $N(0, 1)$ , sia  $A$  una matrice  $n \times n$  *ortogonale* (cioè la matrice di un cambio di base) e sia  $\mathbf{Y} = A\mathbf{X}$ . Anche le componenti  $(Y_1, \dots, Y_n)$  sono indipendenti con densità  $N(0, 1)$ .

Se  $(X_1, \dots, X_n)$  è un campione di  $n$  variabili aleatorie, indichiamo con  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  la media empirica, e con

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

(e naturalmente  $S$  ne è la radice quadrata).

**Teorema** Siano  $X_1, \dots, X_n$  indipendenti con densità  $N(m, \sigma^2)$ . Si hanno i seguenti risultati:

- a) le variabili  $\bar{X}$  e  $S^2$  sono indipendenti;

b)  $\bar{X}$  ha densità  $N(m, \frac{\sigma^2}{n})$  e  $\frac{\sum_{i \leq n} (X_i - \bar{X})^2}{\sigma^2}$  ha densità  $\chi^2(n-1)$ ;

c) la variabile

$$\frac{\sqrt{n}(\bar{X} - m)}{S}$$

ha densità di Student  $T(n-1)$ .

Consideriamo come modello statistico un **campione di taglia n e densità**  $N(m, \sigma^2)$ .

Valgono le seguenti *stime di massima verosimiglianza* per i parametri:

1)  $\hat{m} = \bar{X}$  sempre;

2)  $\hat{\sigma}^2 = \frac{\sum_i (X_i - m)^2}{n}$  se  $m$  è nota;

3)  $\hat{\sigma}^2 = \frac{\sum_i (X_i - \bar{X})^2}{n}$  se  $m$  è sconosciuta.

Una stima corretta della varianza è data da

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

**Esempio: Intervallo di fiducia per la media con varianza nota**

Vogliamo trovare un intervallo di fiducia al livello  $(1-\alpha)$  per la media di un campione gaussiano, con varianza nota.

Si ottiene l'intervallo di fiducia  $\bar{X}(\omega) \pm \frac{q_{1-\alpha/2} \sigma}{\sqrt{n}}$ .

**Esempio: Test unilatero** Consideriamo il test della forma  $\mathcal{H}_0) m \leq m_0$  contro  $\mathcal{H}_1) m > m_0$ , con varianza nota, al livello  $\alpha$

La regione critica è della forma  $D = \{\bar{X} \geq c\}$ ; è più comodo scrivere la regione critica nella forma  $\{\bar{X} - m_0 \geq d\}$ , e ricordando che (sotto  $\mathbf{P}^{m_0}$ )  $\frac{\sqrt{n}}{\sigma}(\bar{X} - m_0)$  ha densità  $N(0, 1)$ , si ottiene  $\frac{\sqrt{n}}{\sigma} d = q_{1-\alpha}$ .

Esaminiamo ora il caso di test sulla media di un campione gaussiano *con varianza sconosciuta*, che è noto col nome di **test di Student**.

**Definizione: Legge di Student decentrata** Si chiama legge di Student a  $n$  gradi di libertà decentrata di  $a$  (indicata anche  $T(n)$  *decentrata di a*) la legge di

$$\frac{\sqrt{n} X}{\sqrt{Y}}$$

dove  $X \sim N(a, 1)$ ,  $Y \sim \chi^2(n)$  e sono indipendenti. Le densità di Student decentrate di  $a$ , al variare di  $a$ , sono a rapporto di verosimiglianza crescente.

**Osservazione** La variabile aleatoria  $\frac{\sqrt{n} \bar{X}}{S}$  (sotto  $\mathbf{P}^{m, \sigma^2}$ ) ha legge di Student  $T(n-1)$  *decentrata di*  $\frac{m \sqrt{n}}{\sigma}$ .

**Esempio: Test di Student unilatero** Consideriamo, al livello  $\alpha$ , la regione critica di un test dell'ipotesi  $\mathcal{H}_0) m \leq m_0, \sigma$  qualsiasi, contro l'alternativa  $\mathcal{H}_1) m > m_0, \sigma$  qualsiasi.

Si ottiene una regione critica della forma

$$D = \left\{ \frac{\sqrt{n}(\bar{X} - m_0)}{S} \geq t_{(1-\alpha, n-1)} \right\}$$

**Esempio: Test di Student** Consideriamo il test

$$\mathcal{H}_0) m = m_0, \sigma \text{ qualsiasi} \quad \mathcal{H}_1) m \neq m_0, \sigma \text{ qualsiasi}$$

al livello  $\alpha$ . La regione critica è della forma

$$D = \left\{ \frac{\sqrt{n}|\bar{X} - m_0|}{S} \geq t_{(1-\frac{\alpha}{2}, n-1)} \right\}$$

**Intervallo di fiducia per la media, con varianza sconosciuta**

L'intervallo di fiducia per la media al livello  $(1-\alpha)$ , con varianza sconosciuta, è della forma

$$\bar{X}(\omega) \pm \frac{t_{(1-\frac{\alpha}{2}, n-1)} S(\omega)}{\sqrt{n}}.$$

Prima di affrontare i test sulla varianza, osserviamo che valgono le seguenti proprietà:

- se  $m$  è noto,  $\frac{\sum_i (X_i - m)^2}{\sigma^2}$  ha densità  $\chi^2(n)$ ;
- se  $m$  è sconosciuto,  $\frac{\sum_i (X_i - \bar{X})^2}{\sigma^2}$  ha densità  $\chi^2(n-1)$ .

**Esempio: Test sulla varianza con media sconosciuta** Consideriamo il test

$$\mathcal{H}_0) \sigma^2 \leq \sigma_0^2, m \text{ qualsiasi} \quad \text{contro} \quad \mathcal{H}_1) \sigma^2 > \sigma_0^2, m \text{ qualsiasi}$$

al livello  $\alpha$ . Si ottiene la regione critica

$$D = \left\{ \sum_i (X_i - \bar{X})^2 \geq c \right\}$$

con  $c$  dato da  $\frac{c}{\sigma_0^2} = \chi_{(1-\alpha, n-1)}^2$ .

Ci occupiamo ora del caso in cui l'osservazione statistica sia formata da due campioni indipendenti  $X_1, \dots, X_n$  (di legge  $N(m_1, \sigma_1^2)$ ) e  $Y_1, \dots, Y_k$  (di legge  $N(m_2, \sigma_2^2)$ ).

**Esempio: Confronto tra due varianze** Identifichiamo il test

$$\mathcal{H}_0) \sigma_1^2 \leq \sigma_2^2 \quad \text{contro} \quad \mathcal{H}_1) \sigma_1^2 > \sigma_2^2$$

al livello  $\alpha$ .

Se chiamiamo  $F_{(1-\alpha, n, k)}$  lo  $(1 - \alpha)$ -quantile della legge  $F_{n, k}$ , la regione critica del test richiesto è data da

$$D = \left\{ \frac{\sum_{i \leq n} (X_i - \bar{X})^2 / (n - 1)}{\sum_{j \leq k} (Y_j - \bar{Y})^2 / (k - 1)} \geq F_{(1-\alpha, n-1, k-1)} \right\}$$

**Definizione: Problema di Behrens-Fisher** Si chiama *problema di Behrens-Fisher* l'individuazione della regione critica del test dell'ipotesi

$$\mathcal{H}_0) m_1 = m_2 \quad \text{contro} \quad \mathcal{H}_1) m_1 \neq m_2 .$$

Noi ci limitiamo al caso più semplice nel quale si abbia  $\sigma_1^2 = \sigma_2^2$ .

**Lemma** Se  $m_1 = m_2$  e  $\sigma_1^2 = \sigma_2^2$ , la variabile

$$Z_{n,k} = \frac{\bar{X} - \bar{Y}}{\sqrt{\sum_{i \leq n} (X_i - \bar{X})^2 + \sum_{j \leq k} (Y_j - \bar{Y})^2}} \frac{\sqrt{n+k-2}}{\sqrt{\frac{1}{n} + \frac{1}{k}}}$$

ha densità di Student  $T(n+k-2)$ .

Se consideriamo l'ipotesi  $\mathcal{H}_0) m_1 = m_2$ , si prende come regione critica (al livello  $\alpha$ )

$$D = \left\{ |Z_{n,k}| \geq t_{(1-\frac{\alpha}{2}, n+k-2)} \right\}$$

mentre il test dell'ipotesi  $\mathcal{H}_0) m_1 \leq m_2$  avrà regione critica

$$D = \left\{ Z_{n,k} \geq t_{(1-\alpha, n+k-2)} \right\} .$$

**Definizione: Modelli lineari** Si chiama *modello statistico lineare* un modello nel quale l'osservazione è data da  $n$  variabili aleatorie  $X_1, \dots, X_n$  che si possano scrivere nella forma

$$X_i = \sum_{j=1}^k a_{ij} \theta_j + \sigma W_i$$

con le seguenti proprietà:

- a)  $k < n$ ,  $(\theta_1, \dots, \theta_k) \in \mathbb{R}^k$  e  $\sigma > 0$ ;
- b) la matrice  $n \times k$ ,  $A = [a_{ij}]$  è di rango massimo (e quindi l'applicazione lineare ad essa associata  $A : \mathbb{R}^k \rightarrow \mathbb{R}^n$  è iniettiva;



c) le variabili  $W_1, \dots, W_n$  sono gaussiane  $N(0, 1)$  indipendenti.

**Definizione: Modello di regressione** Il modello è detto *di regressione* quando è della forma

$$X_i = \theta_1 + \theta_2 z_i + \dots + \theta_k z_i^{k-1} + \sigma W_i$$

con  $z_1 \neq z_2 \neq \dots \neq z_n$  (e  $k < n$ ).

Per i modelli lineari useremo anche la notazione vettoriale  $\mathbf{X} = A\boldsymbol{\theta} + \sigma\mathbf{W}$ .

**Lemma** Sia  $A : \mathbb{R}^k \rightarrow \mathbb{R}^n$  una applicazione lineare *iniettiva*. Dato  $\mathbf{x} \in \mathbb{R}^n$ , il punto  $\mathbf{y} \in \mathbb{R}^k$  che minimizza  $\|\mathbf{x} - A\mathbf{y}\|^2$  è dato da  $\mathbf{y} = U\mathbf{x}$ , essendo  $U = (A^t A)^{-1} A^t$ .

L'espressione della verosimiglianza del modello in forma vettoriale si scrive

$$L(\boldsymbol{\theta}, \sigma^2; \mathbf{x}) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{\|\mathbf{x} - A\boldsymbol{\theta}\|^2}{2\sigma^2} - n \log \sigma\right)$$

Vediamo le *stime di massima verosimiglianza*: la stima di  $\boldsymbol{\theta}$  è  $\hat{\boldsymbol{\theta}} = U\mathbf{X}$ , e la stima di  $\sigma^2$  è

$$\hat{\sigma}^2 = \frac{\|\mathbf{X} - A\hat{\boldsymbol{\theta}}\|^2}{n} = \frac{\|\mathbf{X} - AU\mathbf{X}\|^2}{n}.$$

**Teorema di Gauss Markov**  $U\mathbf{X}$  è una stima corretta di  $\boldsymbol{\theta}$ , di rischio minimo tra tutte le stime lineari corrette. Inoltre

$$\frac{\|\mathbf{X} - AU\mathbf{X}\|^2}{n - k}$$

è una stima corretta di  $\sigma^2$ .