Università di Pisa

Dipartimento di Informatica
Corso di Laurea Triennale in Informatica

Formulario di Statistica

Professore:
Prof. Francesco Grotto

Autore:
Filippo Ghirardini

Anno Accademico 2023/2024

Contents

Formulario di Statistica

Realizzato da: Ghirardini Filippo

A.A. 2023-2024

---

1 Formulario

1.1 Statistica descrittiva

1.1.1 Indici statistici

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{(Media campionaria)}$$

$$var(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{(Varianza campionaria)}$$

$$\sigma(x) = \sqrt{var(x)} \qquad \text{(Deviazione standard)}$$

$$b = \frac{1}{\sigma^3} \cdot \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 \qquad \text{(Sample skewness)}$$

Proposizione 1.1.1.

$$\frac{\#\{x_i : |x_i - \bar{x}| > d\}}{n-1} \leq \frac{var(x)}{d^2} \tag{1}$$

1.1.2 Quantili

$$F_e(t) = \frac{\#\{i|x_i \leq t\}}{n} \qquad \text{(Funzione di ripartizione empirica)}$$

1.1.3 Dati multivariati

$$cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} \qquad \text{(Covarianza campionaria)}$$

$$r(x,y) = \frac{cov(x,y)}{\sigma(x)\sigma(y)} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad \text{(Coefficiente di correlazione)}$$

1.2 Probabilità e indipendenza

1.2.1 Spazi di probabilità

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n) \qquad (\sigma\text{-addittività})$$

Proposizione 1.2.1. *Operazioni su spazi di probabilità:*

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(\emptyset) = 0$

- $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$

Proposizione 1.2.2. *Vale:*

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) \tag{2}$$

1.2.2 Probabilità discreta

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} \tag{Probabilità}$$

Proposizione 1.2.3. *Calcolo combinatorio:*

- *Sequenze ordinate con possibile ripetizione di k numeri da 1 a n: n^k*
- *Numero di modi in cui si può ordinare $\{1,\dots,n\}$: $n!$*
- *Numero di sequenze ordinate senza ripetizione di k numeri di $\{1,\dots,n\}$: $\frac{n!}{(n-k)!}$*
- *Numero di sottoinsiemi di $\{1,\dots,n\}$ formati da k elementi: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$*

Definizione 1.2.1 (Funzione di massa)**.** *Se $\Omega = \{x_1, x_2, \dots\} \subset \mathbb{R}$ è un sottoinsieme numerabile:*

$$\Omega \ni x_i \mapsto p(x_i) = \mathbb{P}(\{x_i\}) \in [0,1] \tag{3}$$

e valgono:

$$\mathbb{P}(A) = \sum_{i:x_i\in A} p(x_i) \qquad \forall A \subseteq \mathbb{R} \tag{4}$$

$$p(x_i) \geq 0 \tag{5}$$

$$\sum_{i=1,2,\dots} p(x_i) = 1 \tag{6}$$

1.2.3 Probabilità condizionata

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \tag{Probabilità condizionata}$$

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}) \tag{Condizionamento ripetuto}$$

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i) \tag{Formula di fattorizzazione}$$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \tag{Formula di Bayes}$$

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^{n}\mathbb{P}(A|B_j)\mathbb{P}(B_j)} \tag{Formula di Bayes - Sistema di alternative}$$

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_k}) \tag{Eventi indipendenti}$$

1.2.4 Entropia di Shannon

$$H^{(n)}(p_1,\ldots,p_n) = -\sum_{i=1}^{n} p_i \log(p_i) \qquad \text{(Entropia)}$$

Proposizione 1.2.4. *L'entropia ha queste proprietà:*

- *È una funzione simmetrica*
- $H^{(n)}(1,0,\ldots,0) = 0$
- *È coerente tra n diversi:* $H^{(n)}(p_1 = 0, p_2,\ldots,p_n) = H^{(n-1)}(p_2,\ldots,p_n)$
- $H^{(n)}(p_1,\ldots,p_n) \leq H^{(n)}(\frac{1}{n},\ldots,\frac{1}{n})$
- *Data una probabilità su* $n \times m$ *oggetti,* $\Omega = \{x_{11},\ldots,x_{ij},\ldots,x_{nm}\}$, $\mathbb{P}(\{x_{ij}\}) = q_{ij}$, *considerando gli eventi* $A_i = \{x_{i,1},\ldots,x_{i,m}\}$, $\mathbb{P}(A_i) = p_i$*:*

$$H^{(nm)}(q_{11},\ldots,q_{ij},\ldots,q_{nm}) = H^{(n)}(p_1,\ldots,p_n) + \sum_{i=1}^{n} p_i H^{(m)}\left(\frac{q_{i1}}{p_i},\ldots,\frac{q_{im}}{p_i}\right) \qquad (7)$$

Teorema 1.2.1 (Teorema di Shannon)**.** Una funzione continua che soddisfa queste proprietà deve avere la forma:

$$cH^{(n)} \qquad c > 0 \qquad (8)$$

1.2.5 Densità di probabilità

$$\mathbb{P}(A) = \int_A f(x)dx \qquad A \subseteq \Omega \qquad \text{(Probabilità di una densità)}$$

$$\mathbb{P}(A \cup B) = \int_{A\cup B} f(x)dx = \int_A f(x)dx + \int_B f(x)dx = \mathbb{P}(A) + \mathbb{P}(B) \qquad A \cap B = \emptyset \qquad \text{(Somma di probabilità)}$$

$$\mathbb{P}(\{t\}) = \int_{\{t\}} f(x)dx = 0 \qquad \text{(Probabilità di un punto)}$$

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{altrove} \end{cases} \qquad \text{(Densità uniforme)}$$

$$\mathcal{X} : \mathbb{R} \to \{0,1\} \qquad \mathcal{X}_S(x) = \begin{cases} 1 & x \in S \\ 0 & x \notin S \end{cases} \qquad \text{(Funzione indicatrice)}$$

1.3 Variabili aleatorie

1.3.1 Legge di una variabile aleatoria

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \sum_{x_i \in A} p_X(x_i) \qquad \text{(Variabile aleatoria discreta)}$$

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \int_A f(x)dx \qquad \text{(Variabile aleatoria continua)}$$

$$\mathbb{P}(X \in A) = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx \qquad \text{(Variabile aleatoria continua con segmento)}$$

1.3.2 Funzione di ripartizione e quantili

$$F_X : \mathbb{R} \to [0,1] \qquad F_X(x) = \mathbb{P}\{X \le x\} \tag{Cumulative Distribution Function}$$

$$\mathbb{P}\{a < X \le b\} = F(b) - F(a) \tag{9}$$

$$F_X(t) = \sum_{x_i \le t} p(x_i) \tag{c.d.f. discreta}$$

$$\mathbb{P}\{X = x\} = F(x) - F_{\text{-}}(x) \tag{10}$$

$$f(x) = \frac{dF(x)}{dx} \tag{c.d.f. continua}$$

Proposizione 1.3.1. *Proprietà della c.d.f.:*

- *F non è decrescente ($x < y \Rightarrow F(x) \le F(y)$)*
- $\lim_{x\to-\infty} F(x) = 0 \qquad \lim_{x\to+\infty} F(x) = 1$
- *F è continua a destra ($\forall x \in \mathbb{R} \quad F(x_n) \to F(x)$ per ogni successione $x_n \to x \quad x_n \ge x$)*

$$r_\beta = inf\{r \in \mathbb{R} : F(r) \ge \beta\} \qquad \beta \in (0,1) \tag{β-quantile}$$

$$F^{\leftarrow} : (0,1) \to \mathbb{R} \qquad F^{\leftarrow}(t) = inf\{r \in \mathbb{R} : F(r) \ge t\} \tag{Inversa generalizzata}$$

Proposizione 1.3.2. *Proprietà dell'inversa generalizzata:*

- *Se F è strettamente crescente $F^{\leftarrow} = F^{-1}$*
- *$F^{\leftarrow}$ è sempre non decrescente*
- $F^{\leftarrow}(F(t)) \le t \quad \forall t \in \mathbb{R}$
- $F(F^{\leftarrow}(t)) \ge t \quad \forall t \in \mathbb{R}$
- $F^{\leftarrow}(t) \le s \Leftrightarrow F(s) \ge t$

1.3.3 Variabili aleatorie notevoli discrete

$$\mathbb{P}(X = h) = \binom{n}{h} p^h (1-p)^{n-h} \tag{Binomiale $B(n,p)$}$$

$$\mathbb{P}(X = h) = (1-p)^{h-1} p \qquad h \in \mathbb{N}_0 \tag{Geometriche $G(p)$}$$

$$\mathbb{P}(X = k) = \frac{\binom{h}{k}\binom{n-h}{r-k}}{\binom{n}{r}} \qquad k = 0, \ldots, h \tag{Ipergeometriche $I(n,h,r)$}$$

$$\mathbb{P}(X = h) = e^{-\lambda} \frac{\lambda^h}{h!} \qquad h \in \mathbb{N} \tag{Poisson $P(\lambda)$}$$

$$\sum_{k=0}^{h} \binom{h}{k} \binom{n-h}{r-k} = \binom{n}{r} \tag{Identità di Vandermonde}$$

1.3.4 Variabili aleatorie notevoli con densità

$$f(t) = \begin{cases} \frac{1}{b-a} & a < t < b \\ 0 & \text{altrove} \end{cases} \qquad F(t) = \begin{cases} 0 & t \leq a \\ \frac{t}{b-a} & 0 < t \leq b \\ 1 & t > b \end{cases} \qquad \text{(Uniformi su intervalli)}$$

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases} \qquad F(t) = \begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases} \qquad \text{(Esponenziali)}$$

$$f(t) = \begin{cases} \alpha x_m^\alpha t^{-1-\alpha} & t > x_m \\ 0 & t \leq x_m \end{cases} \qquad F(t) = \begin{cases} 1 & t < x_m \\ 1 - (\frac{x_m}{t})^\alpha & t \geq x_m \end{cases} \qquad \text{(Pareto)}$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} \qquad \text{(Gaussiane standard } N(0,1))$$

$$f_Y(t) = \frac{1}{\sigma} f_X\left(\frac{t-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-m)^2}{2\sigma^2}} \qquad \text{(Gaussiane non standard } N(m,\sigma^2))$$

$$F_Y(t) = \mathbb{P}\{Y \leq t\} = \mathbb{P}\{\sigma X + m \leq t\} = \mathbb{P}\left(X \leq \frac{t-m}{\sigma}\right) = \Phi\left(\frac{t-m}{\sigma}\right) \qquad \text{(Gaussiane non standard 2)}$$

Proposizione 1.3.3. *Proprietà delle Gaussiane standard:*

$$\mathbb{P}\{-t \leq Z \leq t\} = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1 \tag{11}$$

$$\Phi(0) = \mathbb{P}\{X \geq 0\} = \mathbb{P}\{X \leq 0\} = \frac{1}{2} \tag{12}$$

Definizione 1.3.1 (Variabili Gaussiane Non Standard -)**.** *Data X una v.a. $N(0,1)$ e Y una v.a. del tipo $Y = \sigma X + m$.*

1.3.5 Trasformazioni di variabili con densità

$$Y : \Omega \to \mathbb{R} \qquad Y = h \circ X \qquad F_Y(y) = \mathbb{P}\{Y \leq y\} = \mathbb{P}\{h(X) \leq y\} \tag{13}$$

Proposizione 1.3.4 (Cambio di variabile)**.** *Sia $h : A \to B$ biunivoca, differenziabile e con inversa differenziabile:*

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \cdot \left| \frac{dh^{-1}(y)}{dy} \right| & y \in B \\ 0 & y \notin B \end{cases} \tag{14}$$

Se h è crescente:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(h(X) \leq y) = \mathbb{P}(X \leq h^{-1}(y)) = F_X(h^{-1}(y)) \tag{15}$$

$$f_Y(y) = f_X(h^{-1}(y)) \cdot \frac{dh^{-1}(y)}{dy} \tag{16}$$

Se h è decrescente:

$$\mathbb{P}(h(X) \leq y) = \mathbb{P}(X \leq h^{-1}(y)) = 1 - F_X(h^{-1}(y)) \tag{17}$$

1.3.6 Valore atteso, varianza e momenti

Definizione 1.3.2 (Valore atteso)**.**

$$\mathbb{E}[X] = \sum_i x_i p_X(x_i) \qquad \mathbb{E}[X] = \int_{-\infty}^{+\infty} t f_X(t) dt \tag{18}$$

Proposizione 1.3.5. *Sia X discreta, la variabile $g(x)$ ammette valore atteso se $\sum_i |g(x_i)| p(x_i) < +\infty$. In quel caso vale:*

$$\mathbb{E}[g(X)] = \sum_i g(x_i) p(x_i) \tag{19}$$

Sia X con densità, la variabile $g(x)$ ammette valore atteso se $\int_{-\infty} *{+\infty}|g(x)|f(x)dx < +\infty$. In quel caso vale:

$$\mathbb{E}[g(X)] = \int_{-\infty} *{+\infty} g(x)f(x)dx \tag{20}$$

Proposizione 1.3.6. *Se X ha valore atteso, valgono:*

- $\forall a, b \in \mathbb{R} \quad \mathbb{E}[aX + b] = a\mathbb{E}[X] + b \quad \mathbb{E}[b] = b$
- $|\mathbb{E}[X]| \leq E[|X|]$
- $\mathbb{P}(X \geq 0) = 1 \Longrightarrow \mathbb{E}[X] \geq 0$

Definizione 1.3.3 (Momento di ordine n)**.**

$$\mathbb{E}[X^n] \qquad \mathbb{E}[|X|^n] < +\infty \tag{21}$$

Definizione 1.3.4 (Disuguaglianza di Jensen)**.**

$$\mathbb{E}[|X|^m]^{\frac{1}{m}} \leq \mathbb{E}[|X|^n]^{\frac{1}{n}} \qquad 1 \leq m \leq n \tag{22}$$

Definizione 1.3.5 (Disuguaglianza di Markov)**.**

$$a\mathbb{P}\{X \geq a\} \leq \mathbb{E}[X] \qquad a > 0 \tag{23}$$

Definizione 1.3.6 (Varianza)**.**

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \tag{24}$$

Definizione 1.3.7 (Scarto quadratico medio)**.**

$$\sigma(X) = \sqrt{Var(X)} \tag{25}$$

Definizione 1.3.8 (Disuguaglianza di Chebyshev)**.**

$$\mathbb{P}\{|X - \mathbb{E}[X]| > d\} \leq \frac{Var(X)}{d^2} \qquad d > 0 \tag{26}$$

1.3.7 Momenti di variabili aleatorie notevoli

Proposizione 1.3.7 (Variabili di Bernoulli - $B(1, p)$)**.**

$$\mathbb{E}[X^k] = p \qquad Var(X) = p - p^2 = p(1-p) \qquad k \geq 1 \tag{27}$$

Proposizione 1.3.8 (Variabili Binomiali)**.**

$$\mathbb{E}[X] = np \qquad Var(X) = np(1-p) \tag{28}$$

Proposizione 1.3.9 (Variabili di Poisson)**.**

$$\mathbb{E}[X] = \sum_{h=0}^{+\infty} he^{-\lambda}\frac{\lambda^h}{h!} = \lambda \tag{29}$$

$$\mathbb{E}[X^2] = \lambda + \lambda^2 \tag{30}$$

$$Var(X) = \lambda \tag{31}$$

Proposizione 1.3.10 (Variabili uniformi su intervalli finiti)**.**

$$\mathbb{E}[X] = \int_a^b \frac{x}{b-a}dx = \frac{a^2 + ab + b^2}{3} \qquad Var(X) = \frac{(b-a)^2}{12} \tag{32}$$

Proposizione 1.3.11 (Variabili Esponenziali)**.**

$$\mathbb{E}[X^n] = \frac{n!}{\lambda^n} \qquad Var(X) = \frac{1}{\lambda^2} \tag{33}$$

Proposizione 1.3.12 (Variabili Gaussiane Standard).

$$\mathbb{E}[X^{2h+1}] = 0 \tag{34}$$

$$\mathbb{E}[X^{2h+2}] = (2h+1)\mathbb{E}[X^{2h}] \tag{35}$$

$$Var(X) = 1 \tag{36}$$

Proposizione 1.3.13 (Variabili Gaussiane).

$$\mathbb{E}[Y] = \mathbb{E}[\sigma X + m] \qquad Var(Y) = Var(\sigma X + m) = \sigma^2 Var(X) \tag{37}$$

1.4 Distribuzioni multivariate

1.4.1 Variabili doppie

$$\Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2 \qquad \mathbb{P}_{(X,Y)}(A) = \mathbb{P}((X,Y) \in A) = \mathbb{P}\{\omega \in \Omega : (X(\omega), Y(\Omega)) \in A\} \tag{38}$$

Proposizione 1.4.1 (Distribuzione di probabilità di variabile doppia discreta).

$$p(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j) \tag{39}$$

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}\{(X,Y) \in A\} = \sum_{(x_i,y_j)\in A} p(x_i, y_j) \tag{40}$$

$$p_X(x_i) = \sum_{j=1}^{\infty} p(x_i, y_j) \tag{41}$$

$$p_Y(y_j) = \sum_{i=1}^{\infty} p(x_i, y_j) \tag{42}$$

Proposizione 1.4.2 (Distribuzione di probabilità di variabile doppia con densità).

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}\{(X,Y) \in A\} = \int\int_A f(x,y)dxdy \tag{43}$$

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy \tag{44}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx \tag{45}$$

1.4.2 Indipendenza di variabili aleatorie

Definizione 1.4.1 (Variabili aleatorie indipendenti).

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdot \ldots \cdot \mathbb{P}(X_n \in A_n) \tag{46}$$

$$p(x_i, y_j) = p_X(x_i) \cdot p_Y(y_j) \qquad \forall (x_i, y_j) \tag{47}$$

$$f(x,y) = f_X(x) \cdot f_Y(y) \qquad \forall (x,y) \tag{48}$$

1.4.3 Funzioni di variabili indipendenti

Proposizione 1.4.3 (Somma di Binomiali).

$$X \to B(n,p) \quad Y \to B(m,p) \Longrightarrow Z = X + Y \to B(n+m, p) \tag{49}$$

Proposizione 1.4.4 (Funzione di massa di somma di variabili discrete).

$$Z = X + Y \Longrightarrow p_Z(n) = \sum_{h=0}^{n} p_X(h) \cdot p_Y(n-h) \tag{50}$$

Proposizione 1.4.5 (Funzione di massa di somma di variabili con densità o formula della convoluzione).

$$Z = X + Y \Longrightarrow p_Z(n) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)dx = \int_{-\infty}^{+\infty} f_Y(y) f_X(z-y)dy \tag{51}$$

1.4.4 Covarianza e correlazione

Proposizione 1.4.6 (Valore atteso di somma di variabili). *Dati X e Y con valore atteso:*

- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

- $X \geq Y \Longrightarrow \mathbb{E}[X] \geq \mathbb{E}[Y]$

Se sono anche indipendenti:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \tag{52}$$

$$\mathbb{E}[h(X)k(Y)] = \mathbb{E}[h(X)] \cdot \mathbb{E}[k(Y)] \tag{53}$$

Proposizione 1.4.7 (Disuguaglianza di Schwartz).

$$\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]} \cdot \sqrt{\mathbb{E}[Y^2]} \tag{54}$$

Definizione 1.4.2 (Covarianza).

$$Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{55}$$

Definizione 1.4.3 (Coefficiente di correlazione).

$$\rhd(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} \tag{56}$$

1.5 Variabili indipendenti e teoremi limite

Definizione 1.5.1 (Convergenza in probabilità).

$$\forall \epsilon > 0 \qquad \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0 \tag{57}$$

Proposizione 1.5.1 (Convergenza in probabilità ad una costante).

$$\lim_{n\to\infty} \mathbb{E}[X_n] = c \in \mathbb{R} \wedge \lim_{n\to\infty} Var(X_n) = 0 \Longrightarrow (X_n)_{n\geq 1} \longrightarrow c \tag{58}$$

Proposizione 1.5.2.

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|S_n^2 - \sigma^2| > \epsilon) = 0 \tag{59}$$

Definizione 1.5.2 (Convergenza in distribuzione).

$$\lim_{n\to\infty} F_n(t) = F(t) \quad \forall t \in \mathbb{R} \Longrightarrow (X_n)_{n\geq 1} \longrightarrow X \tag{60}$$

Definizione 1.5.3 (Teorema centrale del limite).

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-1\frac{x^2}{2}} dx = \Phi(b) - \phi(a) \tag{61}$$

Definizione 1.5.4 (Approssimazione a variabile Gaussiana).

$$\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \qquad n \geq 50 \tag{62}$$

1.5.1 Variabili Chi-Quadro e di Student

Definizione 1.5.5 (Gamma di Eulero)**.**

$$\Gamma(r) = \int_0^{+\infty} x^{r-1}e^{-x}dx = (r-1)\Gamma(r-1) \qquad r > 0 \tag{63}$$

Definizione 1.5.6 (Densità Gamma - $\Gamma(r,\lambda)$)**.**

$$f(x) = \begin{cases} \frac{1}{\Gamma(r)}\lambda^r x^{r-1}e^{-\lambda x} & x > 0 \\ 0 & z \leq 0 \end{cases} \tag{64}$$

$$\mathbb{E}[X^\beta] = \frac{\Gamma(r+\beta)}{\Gamma(r)\lambda^\beta} \tag{65}$$

$$Var(X) = \frac{r}{\lambda^2} \tag{66}$$

Proposizione 1.5.3 (Somma di densità gamma)**.**

$$X \to \Gamma(r,\lambda) \quad Y \to \Gamma(s,\lambda) \Longrightarrow (X+Y) \to \Gamma(r+s,\lambda) \tag{67}$$

Definizione 1.5.7 (Densità Chi-quadro)**.**

$$X_1,\ldots,X_n \to N(0,1) \Longrightarrow (X_1^2,\ldots,X_n^") \to \Gamma(\frac{n}{2},\frac{1}{2}) = \mathcal{X}^2(n) \tag{68}$$

$$F_{X^2}(t) = \mathbb{P}(X^2 \leq t) = \begin{cases} 0 & t < 0 \\ \mathbb{P}(-\sqrt{t} \leq X \leq \sqrt{t}) = 2F_X(\sqrt{t}) - 1 & t \geq 1 \end{cases} \tag{69}$$

Proposizione 1.5.4 (Approssimazioni Chi-quadro)**.** *Data C_n variabile con densità $\mathcal{X}^2(n)$:*

- $\lim_{n\to\infty} \frac{C_2}{n} \approx 1$
- $\lim_{n\to\infty} \frac{C_n - n}{\sqrt{2n}} \approx N(0,1)$

Definizione 1.5.8 (Densità Student)**.**

$$T_n = \sqrt{n}\frac{X}{\sqrt{C_n}} \qquad X \to N(0,1) \quad C_n \to \mathcal{X}^2(n) \tag{70}$$

$$f_{T_n}(t) = \frac{\Gamma(\frac{n}{2}+\frac{1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-\frac{n}{2}-\frac{1}{2}} \tag{71}$$

Proposizione 1.5.5.

$$(T_n)_{n\geq 1} \longrightarrow N(0,1) \tag{72}$$

Proposizione 1.5.6. *Date $X_1,\ldots,X_n$ v.a. $N(m,\sigma^2)$:*

- *$\bar{X}_n$ e S_n^2 sono indipendenti*
- $\bar{X} \to N(m, \frac{\sigma^2}{n})$
- $\frac{n-1}{\sigma^2}S_n^2 \to \mathcal{X}^2(n-1)$
- $T = \sqrt{n}\frac{(\bar{X}_n - m)}{S} \to T(n-1)$

1.6 Campioni statistici e stimatori

1.6.1 Stima parametrica

Definizione 1.6.1 (Stimatore corretto). *Dato un parametro θ:*

$$\mathbb{E}_\theta[g(X_1,\dots,X_n)] = \theta \tag{73}$$

Proposizione 1.6.1.

$$\mathbb{E}[\bar{X}] = \mathbb{E}[X_i] \tag{74}$$

$$Var(\bar{X}) = \frac{Var(X_i)}{n} \tag{75}$$

$$\mathbb{E}[S_n^2] = Var(X_i) \tag{76}$$

Definizione 1.6.2 (Stimatore consistente).

$$\lim_{n\to\infty} \mathbb{P}_\theta\{|g_n(X-1,\dots,X_n) - \theta| > \epsilon\} = 0 \tag{77}$$

Definizione 1.6.3 (Stimatore efficiente).

$$Var_\theta(g(X_1,\dots,X_n)) \leq Var_\theta(h(X_1,\dots,X_n)) \tag{78}$$

1.6.2 Massima verosimiglianza e metodo dei momenti

Definizione 1.6.4 (Funzione di verosimiglianza).

$$L(\theta, x_1,\dots,x_n) = \prod_{i=1}^n p_\theta(x_i) \tag{79}$$

$$L(\theta, x_1,\dots,x_n) = \prod_{i=1}^n f_\theta(x_i) \tag{80}$$

Definizione 1.6.5 (Stima di massima verosimiglianza).

$$L(\hat{\theta}; x_1,\dots,x_n) = \max_{\theta\in\Theta} L(\theta; x_1,\dots,x_n) \tag{81}$$

Definizione 1.6.6 (Metodo dei momenti).

$$\mathbb{E}_{\bar{\theta}}[X^k] = \frac{1}{n}\sum_{i=1}^n x_i^k \qquad \forall (x_1,\dots,x_n) \tag{82}$$

1.7 Intervalli di fiducia

Definizione 1.7.1 (Intervallo di fiducia per la media).

$$\left[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right] \tag{83}$$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 \tag{84}$$

$$\left[\bar{X}_n \pm \frac{S_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right] \tag{85}$$

Definizione 1.7.2 (Intervallo di fiducia per la media - Bernoulli).

$$\left[\bar{X}_n \pm \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} q_{1-\frac{\alpha}{2}}\right] \tag{86}$$

Definizione 1.7.3 (Intervallo di fiducia per la varianza - Gaussiani).

$$\left(0, \frac{(n-1)S_n^2}{\mathcal{X}^2_{\alpha,n-1}}\right] \qquad \left[\frac{(n-1)S_n^2}{\mathcal{X}^2_{1-\alpha,n-1}}, +\infty\right) \tag{87}$$

1.8 Test statistici

Definizione 1.8.1 (Regione critica di livello α)**.**

$$C = \left\{ \sqrt{n}\frac{|\bar{X} - m_0|}{\sigma} > q_{1-\frac{\alpha}{2}} \right\} = \left\{ |\bar{X}_n - m_0| > \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right\} \tag{88}$$

Definizione 1.8.2 (p-value)**.**

$$\bar{\alpha} = \mathbb{P}_{m_0}\left(\sqrt{n}\frac{\bar{X}_n - m_0|}{\sigma} > \frac{\sqrt{n}}{\sigma}|\bar{X}_n - m_0| \right) = 2\left[1 - \Phi\left(\frac{\sqrt{n}}{\sigma}|\bar{X}_n - m_0| \right) \right] \tag{89}$$

Definizione 1.8.3 (Curva operativa)**.**

$$\beta(m) = \Phi\left(\sqrt{n}\frac{|m_0 - m|}{\sigma} + q_{1-\alpha} \right) \tag{90}$$

1.8.1 Campione di Bernoulli

$$C = \left\{ \frac{\sqrt{n}|\bar{X} - m_0|}{\sqrt{p_0(1-p_0)}} > q_{1-\frac{\alpha}{2}} \right\} \tag{91}$$

$$\bar{\alpha} = 2\left[1 - \Phi\left(\frac{\sqrt{n}|\bar{X}_n - p_0|}{\sqrt{p_o(1-p_0)}} \right) \right] \tag{92}$$

1.8.2 Campione Gaussiano

$$C = \left\{ \frac{n-1}{\sigma_0^2} S_n^2 > \mathcal{X}^2_{2-\alpha, n-1} \right\} \tag{93}$$

$$\bar{\alpha} = 1 - G_{n-1}\left(\frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{\sigma_0^2} \right) \tag{94}$$

$$\beta(\sigma) = G_{n-1}\left(\frac{\sigma_0^2}{\sigma^2} \mathcal{X}^2_{1-\alpha, n-1} \right) \tag{95}$$