Matematica utile

definizioni, formule ed esempi

Pietro Barbiero

Quest'opera contiene informazioni tratte da wikipedia (http://www.wikipedia.en) e dalle dispense relative al corso di Metodi Matematici per l'Ingegneria tenuto dal professor Bazzanella Danilo del Dipartimento di Scienze Matematiche del Politecnico di Torino (IT).

Indice

Parte I

Formulario di Probabilità

Capitolo 1

Calcolo combinatorio

1.1 Sequenze ordinate

1.1.1 Numero di permutazioni senza ripetizione

Il numero delle permutazioni senza ripetizione di un insieme S di numerosità n è il fattoriale di n

$$P_n = n! = n(n-1)(n-2)\dots 1 \tag{1.1}$$

Il primo elemento può essere scelto in n modi, il secondo in $(n-1)$, il terzo in $(n-2)$, l'n-esimo in 1 modo solo: $n(n-1)(n-2)\dots 1$

1.1.2 Numero di permutazioni con ripetizione

Il numero di permutazioni con ripetizione di in insieme S di numerosità n di cui r elementi (con $r \leq n$) si ripetono $k_1, k_2, \dots, k_r$ volte è il rapporto tra il numero di permutazioni senza ripetizione P_n di S e il prodotto tra il numero di permutazioni di ciascun elemento ripetiuto $k_i!$

$$P_n^{k_1,k_2,\dots,k_r} = \frac{P_n}{P_{k_1}P_{k_2}\dots P_{k_r}} = \frac{n!}{k_1!k_2!\dots k_r!} \tag{1.2}$$

1.1.3 Numero di disposizioni senza ripetizione

Il numero di disposizioni senza ripetizione di lunghezza k di un insieme S di numerosità n (con $k \leq n$) è il rapporto tra il numero di permutazioni senza ripetizione P_n di S e il numero di permutazioni della differenza tra la numerosità n di S e la lunghezza di ogni disposizione k

$$D_{n,k} = \frac{P_n}{P_{(n-k)}} = \frac{n!}{(n-k)!} \tag{1.3}$$

1.1.4 Numero di disposizioni con ripetizione

Il numero di disposizioni con ripetizione di lunghezza k di un insieme S di numerosità n è la potenza k-esima della numerosità n

$$D'_{n,k} = n^k = n_1 \cdot n_2 \dots n_k \tag{1.4}$$

Il primo elemento può essere scelto in n modi, il secondo in n modi, il k-esimo in n modi: $n_1 \cdot n_2 \dots n_k$

1.2 Sequenze non ordinate

1.2.1 Numero di combinazioni senza ripetizione

Il numero di combinazioni senza ripetizione di lunghezza k di un insieme S di numerosità n è il rapporto tra il numero di disposizioni senza ripetizione di S e il numero di permutazioni della lunghezza k ($k!$ è la numerosità delle classi di combinazioni aventi gli stessi elementi)

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n!}{k!(n-k)!} = \binom{n}{k} \tag{1.5}$$

1.2.2 Numero di combinazioni con ripetizione

Il numero di combinazioni con ripetizione di lunghezza k di un insieme S di numerosità n è il numero di combinazioni senza ripetizione di lunghezza k di un insieme di numerosità $n+k-1$

$$C'_{n,k} = C_{n+k-1,k} = \binom{n+k-1}{k} \tag{1.6}$$

Capitolo 2

Probabilità

2.1 Probabilità

2.1.1 Probabilità assiomatica

La probabilità è una funzione che associa un numero reale compreso nell'intervallo $[0, 1]$ ad ogni evento A_i di un campo di Borel $\mathfrak{B}$ (costituito da sottoinsiemi di uno spazio campione Ω) in modo che: le probabilità dell'evento certo Ω sia 1; la probabilità che si verifichino due eventi incompatibili sia la somma delle rispettive probabilità

$$\mathbb{P}: \quad \mathfrak{B} \to [0, 1] \quad | \quad \begin{cases} \mathbb{P}(\Omega) = 1 \\ i \neq j \quad \wedge \quad A_i \cap A_j = \emptyset \iff \mathbb{P}(A_i \cup A_j) = \mathbb{P}(A_i) + \mathbb{P}(A_j) \end{cases} \tag{2.1}$$

2.1.2 Proprietà della funzione probabilità

$$\forall A \in \mathfrak{B} \quad 0 \leq \mathbb{P}(A) \leq 1 \tag{2.2}$$

$$\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A) \tag{2.3}$$

$$\mathbb{P}(\emptyset) = 0 \tag{2.4}$$

$$A_i \subset A_j \implies \mathbb{P}(A_i) \leq \mathbb{P}(A_j) \tag{2.5}$$

$$A_i \cap A_j \neq \emptyset \implies \mathbb{P}(A_i \cup A_j) = \mathbb{P}(A_i) + \mathbb{P}(A_j) - \mathbb{P}(A_i \cap A_j) \tag{2.6}$$

2.1.3 Probabilità di un evento

La probabilità di un evento A_i composto da N_A eventi elementari è il quoziente tra il numero di eventi elementari favorevoli ad A_i e il numero di eventi elementari che costituiscono lo spazio campione Ω

$$\mathbb{P}(A_i) = \frac{N_A}{N} \tag{2.7}$$

2.1.4 Probabilità di un evento condizionata

La probabilità di un evento A_i condizionata dal verificarsi dell'evento A_j è il quoziente tra la probabilità favorevole al verificarsi sia di A_i sia di A_j e la probabilità che si verifichi solo A_j

$$\mathbb{P}(A_i | A_j) = \frac{\mathbb{P}(A_i \cap A_j)}{\mathbb{P}(A_j)} \tag{2.8}$$

2.1.5 Formule per eventi condizionati

2.1.5.1 Eventi statisticamente indipendenti

Due eventi A_i e A_j sono statisticamente indipendenti se e solo se il verificarsi di A_i non condiziona la probabilità che si verifichi A_j

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \tag{2.9}$$

2.1.5.2 Formula di Bayes

La formula di Bayes permette di calcolare la probabilità che la causa A_i abbia provocato l'evento E (che sappiamo essersi verificato)

$$\mathbb{P}(A_i|E) = \frac{\mathbb{P}(A_i)\mathbb{P}(E|A_i)}{\sum_j \mathbb{P}(A_j)\mathbb{P}(E|A_j)} = \frac{\mathbb{P}(A_i \cap E)}{\mathbb{P}(E)} \tag{2.10}$$

2.1.5.3 Formula delle probabilità totali

La formula delle probabilità totali permette di calcolare la probabilità di E se è nota la sua probabilità condizionata dagli eventi A_j

$$\mathbb{P}(E) = \sum_j \mathbb{P}(A_j)\mathbb{P}(E|A_j) \tag{2.11}$$

2.2 Variabili aleatorie

2.2.1 Variabile aleatoria

Una variabile aleatoria è una funzione che associa ad ogni evento elementare ω un numero reale x e una probabilità associata ad x

$$X: \quad (\Omega, \mathfrak{B}, \mathbb{P}) \to (\mathbb{R}, \mathfrak{B}^*, \mathbb{P}^*) \tag{2.12}$$

2.2.2 Funzione di distribuzione cumulata (o di ripartizione)

La funzione di distribuzione cumulata è una misura della probabilità che una variabile aleatoria $X(\omega)$ assuma valori minori o uguali al numero reale x

$$F_X: \quad \mathbb{R} \to [0,1] \quad | \quad F_X(x) = \mathbb{P}(X \leq x) \tag{2.13}$$

2.2.3 Funzione di distribuzione cumulata (o di ripartizione) congiunta (o mista)

La funzione di distribuzione cumulata congiunta di una coppia di variabili aleatorie X e Y è una misura della probabilità che le due variabili aleatorie assumano valori minori o uguali ai numeri reali x e y

$$F_{XY}: \quad \mathbb{R}^2 \to [0,1] \quad | \quad F_{XY}(x,y) = \mathbb{P}((X \leq x) \cap (Y \leq y)) \tag{2.14}$$

2.2.4 Funzione di densità di probabilità

Una funzione di densità di probabilità $f_X(x)$ è una funzione a valori reali tale che l'integrale su un sottoinsieme del suo dominio (o supporto) $B \in S$ è una misura della probabilità che la variabilie aleatoria X assuma valori in B

$$\int_B f_X(x)dx = \mathbb{P}(X \in B) \tag{2.15}$$

La relazione tra distribuzione cumulata e densità di probabilità è:

$$\int_{-\infty}^{x} f_X(t)dt = \mathbb{P}(X \leq x) = F_X(x) \iff f_X(x) = \frac{dF_X(x)}{dx} \tag{2.16}$$

2.2.5 Funzione di densità di probabilità congiunta

Una funzione di densità di probabilità congiunta $f_{XY}(x)$ è una funzione a valori reali tale che l'integrale su un sottoinsieme del suo dominio (o supporto) $B \in S$ è una misura della probabilità che le variabili aleatorie X e Y assumano valori in B

$$\int\int_B f_{XY}(x,y)dxdy = \mathbb{P}((X,Y) \in B) \tag{2.17}$$

La relazione tra distribuzione cumulata congiunta e densità di probabilità congiunta è:

$$\int_{-\infty}^{x} \left(\int_{-\infty}^{y} f_{XY}(x,y)dy \right) dx = \mathbb{P}((X \leq x) \cap (Y \leq y)) = F_{XY}(x,y) \iff f_{XY}(x,y) = \frac{dF_{XY}(x,y)}{dxdy} \tag{2.18}$$

2.2.6 Funzione di densità di probabilità marginale

Le funzioni di densità di probabilità marginali $f_Y(y)$ e $f_X(x)$ sono gli integrali della densità congiunta su tutti i valori che può assumere l'altra variabile aleatoria

$$f_Y(y) = \int_{D_x(y)} f_{XY}(x,y)dx = \frac{dF_Y(y)}{dy} = \mathbb{P}(Y \in D_x(y)) \tag{2.19}$$

$$f_X(x) = \int_{D_y(x)} f_{XY}(x,y)dy = \frac{dF_X(x)}{dx} = \mathbb{P}(X \in D_y(x)) \tag{2.20}$$

2.2.7 Momenti di variabili aleatorie

2.2.7.1 Momento di ordine q

Il momento di ordine q di una variabile aleatoria X di densità $f_X(x)$ è l'integrale del prodotto tra la funzione di densità di probabilità e la potenza q-esima della variabile x

$$E\{X^q\} = \int_{-\infty}^{+\infty} x^q f_X(x)dx \tag{2.21}$$

2.2.7.2 Momento congiunto di ordine q

Il momento congiunto di ordine q di una coppia di variabili aleatorie X e Y di densità congiunta $f_{XY}(x,y)$ è l'integrale del prodotto tra la funzione di densità di probabilità congiunta e le potenze q-esime delle variabili x e y

$$E\{X^qY^q\} = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} x^q y^q f_{XY}(x,y)dx \right) dy \tag{2.22}$$

2.2.7.3 Valor medio (o atteso)

Il valor medio di una variabile aleatoria X è il momento del primo ordine di X

$$m_X = E\{X\} = \begin{cases} \int_{-\infty}^{\infty} x f_X(x)dx \\ \sum_i p_i x_i \end{cases} \tag{2.23}$$

2.2.7.4 Varianza

La varianza di una variabile aleatoria X è il momento del secondo ordine centrato nel valor medio

$$\sigma_X^2 = E\{(X - m_X)^2\} = \int_{-\infty}^{\infty} (x - m_X)^2 f_X(x)dx \tag{2.24}$$

Relazione tra la varianza e il momento di primo e secondo ordine non centrati

$$\sigma_X^2 = E\{X^2\} - E^2\{X\} \tag{2.25}$$

2.2.7.5 Covarianza

La covarianza di una coppia di variabili aleatorie X e Y è il momento del secondo ordine congiunto centrato nel valor medio di X e di Y

$$Cov(X,Y) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} (x - m_X)(y - m_Y) f_{XY}(x,y)dx \right) dy \tag{2.26}$$

Relazione tra la covarianza e il momento di primo ordine congiunto

$$Cov(X,Y) = E\{XY\} - m_X m_Y \tag{2.27}$$

2.2.8 Diseguaglianza di Tchebyshev

$$\mathbb{P}(|X - m_X| \geq k) \leq \frac{\sigma_X^2}{k^2} \tag{2.28}$$

2.3 Distribuzioni di probabilità notevoli (continue)

2.3.1 Distribuzione uniforme

2.3.1.1 Funzione di densità di probabilità

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \quad \vee \quad x > b \end{cases} \tag{2.29}$$

2.3.1.2 Funzione di distribuzione cumulata

$$F_X(x) = \int_a^x \frac{1}{b-a} dx = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases} \tag{2.30}$$

2.3.1.3 Valor medio

$$m_X = \frac{a+b}{2} \tag{2.31}$$

2.3.1.4 Varianza

$$\sigma_X^2 = \frac{(b-a)^2}{12} \tag{2.32}$$

2.3.2 Distribuzione normale $\mathcal{N}(m_X, \sigma_X^2)$

2.3.2.1 Funzione di densità di probabilità

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-m_X)^2}{2\sigma_X^2}} \tag{2.33}$$

2.3.2.2 Funzione di distribuzione cumulata

$$F_X(x) = f_X(x)dx = \frac{1}{2} + erf\left(\frac{x-m_X}{\sigma_X}\right) \tag{2.34}$$

2.3.3 Distribuzione normale standardizzata $\mathcal{N}(0,1)$

2.3.3.1 Funzione di densità di probabilità

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{2.35}$$

2.3.3.2 Funzione di distribuzione cumulata

$$F_Z(z) = f_Z(z)dz = \frac{1}{2} + erf(z) \tag{2.36}$$

2.3.3.3 Cambio di variabile

$$z = \frac{x-m_X}{\sigma_X} \iff x = \sigma_X z + m_X \tag{2.37}$$

2.3.3.4 Teorema del limite centrale

La densità di probabilità della somma di una successione di variabili aleatorie statisticamente indipendenti X_i con uguali densità di probabilità $f_i(x_i)$ converge per $n \to \infty$ alla distribuzione normale $\mathcal{N}(n\mu, n\sigma^2$

2.3.4 Distribuzione Gamma

2.3.4.1 Funzione di densità di probabilità

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \qquad \Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt \tag{2.38}$$

2.3.4.2 Valor medio

$$m_X = \frac{\alpha}{\lambda} \tag{2.39}$$

2.3.4.3 Varianza

$$\sigma_X^2 = \frac{\alpha}{\lambda^2} \tag{2.40}$$

2.3.5 Distribuzione esponenziale

2.3.5.1 Funzione di densità di probabilità

$$f_X(x) = \lambda e^{-\lambda x} \qquad \forall \lambda > 0 \tag{2.41}$$

2.3.5.2 Funzione di distribuzione cumulata

$$F_X(x) = 1 - e^{-\lambda x} \tag{2.42}$$

2.3.5.3 Valor medio

$$m_X = \frac{1}{\lambda} \tag{2.43}$$

2.3.5.4 Varianza

$$\sigma_X^2 = \frac{1}{\lambda^2} \tag{2.44}$$

2.3.6 Distribuzione di Maxwell

2.3.6.1 Funzione di densità di probabilità

$$f_X(x) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\sigma^3} e^{-\frac{x^2}{2\sigma^2}} \tag{2.45}$$

2.3.6.2 Valor medio

$$m_X = 2\sigma \sqrt{\frac{2}{\pi}} \tag{2.46}$$

2.3.6.3 Varianza

$$\sigma_X^2 = \sigma^2(3 - \frac{8}{\pi}) \tag{2.47}$$

2.3.7 Distribuzione t-Student a n gradi di libertà

2.3.7.1 Funzione di densità di probabilità

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \tag{2.48}$$

2.3.7.2 Valor medio

$$m_X = 0 \tag{2.49}$$

2.3.7.3 Varianza

$$\sigma_X^2 = \frac{n}{n-2} \qquad \forall n > 2 \tag{2.50}$$

Per $n \to \infty$ la distribuzione t-Student tende alla distribuzione normale standardizzata $\mathcal{N}(0,1)$

2.3.8 Distribuzione Chi-quadrato

2.3.8.1 Funzione di densità di probabilità

$$f_X(x) = \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \tag{2.51}$$

2.3.8.2 Valor medio

$$m_X = n \tag{2.52}$$

2.3.8.3 Varianza

$$\sigma_X^2 = 2n \tag{2.53}$$

Per $n \to \infty$ la distribuzione Chi-quadrato tende alla distribuzione normale standard $\mathcal{N}(n, 2n)$

2.3.9 Distribuzione F (o di Fisher)

2.3.9.1 Funzione di densità di probabilità

$$f_X(x) = \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{x^{\frac{m}{2}-1}}{\left(1 + \frac{mx}{n}\right)^{\frac{m+n}{2}}} \tag{2.54}$$

2.3.9.2 Valor medio

$$m_X = \frac{n}{n-2} \qquad \forall n > 2 \tag{2.55}$$

2.3.9.3 Varianza

$$\sigma_X^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \qquad \forall n > 4 \tag{2.56}$$

2.4 Distribuzioni di probabilità notevoli (discrete)

2.4.1 Distribuzione binomiale $B(n, p)$ (grandi numeri)

La distribuzione binomiale di prove di numerosità n ognuna con probabilità di successo p ha una funzione di densità di probabilità pari alla somma somma dei prodotti tra: il numero di combinazioni senza ripetizione di lunghezza k dell'insieme di numerosità n dei tentativi effettuati; la probabilità di successo p elevata al numero di successi k; la probabilità di insuccesso $1 - p$ elevata al numero di insuccessi $n - k$

2.4.1.1 Funzione di densità di probabilità

$$f_X(x) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta(x-k) \tag{2.57}$$

2.4.1.2 Funzione di distribuzione cumulata

$$F_X(x) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} U(x-k) \tag{2.58}$$

2.4.1.3 Valor medio

$$m_X = np \tag{2.59}$$

2.4.1.4 Varianza

$$\sigma_X^2 = npq \tag{2.60}$$

2.4.1.5 Teorema locale di asintoticità locale di Moivre-Laplace

Per $np, npq \to \infty$ la distribuzione binomiale tende alla distribuzione normale $\mathcal{N}(np, npq)$

2.4.2 Distribuzione di Poisson (eventi rari)

2.4.2.1 Funzione di densità di probabilità

$$f_X(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta(x-k) \tag{2.61}$$

2.4.2.2 Valor medio

$$m_X = \lambda \tag{2.62}$$

2.4.2.3 Varianza

$$\sigma_X^2 = \lambda \tag{2.63}$$

2.4.3 Distribuzione geometrica

2.4.3.1 Funzione di densità di probabilità

$$f_X(x) = \sum_{k=0}^{\infty} p(1-p)^k \delta(x-k) \qquad\qquad 0 < p < 1 \tag{2.64}$$

2.4.3.2 Funzione di distribuzione cumulata

$$F_X(x) = \sum_{k=0}^{\infty} p(1-p)^k U(x-k) \tag{2.65}$$

2.4.3.3 Valor medio

$$m_X = \frac{1-p}{p} \tag{2.66}$$

2.4.3.4 Varianza

$$\sigma_X^2 = \frac{1-p}{p^2} \tag{2.67}$$

2.4.4 Distribuzione ipergeometrica

2.4.4.1 Funzione di densità di probabilità

$$f_X(x) = \frac{\binom{N_A}{x}\binom{N_B}{n-x}}{\binom{N}{n}} \tag{2.68}$$

2.4.4.2 Valor medio

$$m_X = \frac{nN_a}{N} \tag{2.69}$$

2.4.4.3 Varianza

$$\sigma_X^2 = \frac{nN_A(N-N_A)(N-n)}{N^2(N-1)} \tag{2.70}$$

2.4.5 Distribuzione Beta

2.4.5.1 Funzione di densità di probabilità

$$f_X(x) = \begin{cases} Ax^b(1-x)^c & \forall x \in [0,1] \quad \wedge \quad b,c > -1 \\ 0 \end{cases} \tag{2.71}$$

$$A = \frac{\Gamma(b+c+2)}{\Gamma(b+1)\Gamma(c+1)}$$

2.4.5.2 Valor medio

$$m_X = \frac{b+1}{b+c+2} \tag{2.72}$$

2.4.5.3 Varianza

$$\sigma_X^2 = \frac{(b+1)(c+1)}{(b+c+2)^2(b+c+3)} \tag{2.73}$$

2.4.6 Distribuzione di Weibull $\mathcal{W}(\alpha, \beta)$

2.4.6.1 Funzione di densità di probabilità

$$f_X(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} & \forall x \in [0, +\infty) \quad \wedge \quad \alpha, \beta \in \mathbb{R} \\ 0 \end{cases} \tag{2.74}$$

2.4.6.2 Valor medio

$$m_X = \frac{1}{\alpha^\beta}\Gamma\left(1 + \frac{1}{\beta}\right) \tag{2.75}$$

2.4.6.3 Varianza

$$\sigma_X^2 = \alpha^{-\frac{2}{\beta}}\left(\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right) \tag{2.76}$$

Capitolo 3

Statistica

3.1 Statistica descrittiva

3.1.1 Numerosità del campione

La numerosità di un campione è il numero di elementi di una popolazione presi in considerazione per le analisi statistiche

$$n \tag{3.1}$$

3.1.2 Valori empirici (o determinazioni o realizzazioni di X)

I valori empirici sono i dati ricavati dalle osservazioni effettuate su un campione di una popolazione

$$\{x_1, x_2, \dots, x_n\} \tag{3.2}$$

3.1.3 Insieme delle modalità di un carattere X

L'insieme delle modalità di un carattere X è l'insieme dei valori che possono assumere le modalità del carattere

$$\Delta \subseteq \mathbb{R} \tag{3.3}$$

3.1.4 Classi di modalità di un carattere X

Le classi di modalità di un carattere X sono una possibile partizione dell'insieme delle modalità Δ del carattere

$$\Delta_i = [a_i, b_i) \in \Delta \quad | \quad \begin{cases} \Delta = \bigcup_i^m \Delta_i \\ \Delta_i \cap \Delta_j = \emptyset \quad \forall i \neq j \end{cases} \tag{3.4}$$

La numerosità m delle classi dovrebbe avvicinarsi alla parte intera di: $m' = 1 + \frac{10}{3} log_{10} n$

3.1.5 Distribuzioni di frequenze

3.1.5.1 Frequenza assoluta

La frequenza assoluta di una classe Δ_i è il numero di elementi della serie di dati che appartengono a Δ_i

$$n_i \quad | \quad x_i \in \Delta_i \implies \sum_{i=1}^{n} n_i = n \tag{3.5}$$

3.1.5.2 Frequenza relativa (o probabilità empirica)

La frequenza relativa di una classe Δ_i è il rapporto tra il numero di elementi della serie di dati che appartengono alla classe Δ_i e il numero di dati totale

$$f_i = \frac{n_i}{n} \tag{3.6}$$

3.1.5.3 Frequenza cumulata assoluta

La frequenza cumulata assoluta è la somma delle prime i frequenze assolute delle prime i classi

$$N_i = \sum_{j=1}^{i} n_j \tag{3.7}$$

3.1.5.4 Frequenza cumulata relativa

La frequenza cumulata relativa è la somma delle prime i probabilità empiriche delle prime i classi

$$F_i = \sum_{j=1}^{i} f_j = \frac{N_i}{n} \tag{3.8}$$

3.1.6 Funzione di distribuzione delle frequenze

La funzione di distribuzione delle frequenze associa ad ogni classe Δ_i il rapporto tra la propria frequenza assoluta e la propria ampiezza $\Delta_i = b_i - a_i$

$$\phi_n(x): \quad \Delta \to \mathbb{R} \quad | \quad \phi_n(x) = \frac{n_i}{\Delta_i} = \frac{n_i}{b_i - a_i} \tag{3.9}$$

3.1.7 Funzione di distribuzione delle probabilità empiriche

La funzione di distribuzione delle probabilità empiriche associa ad ogni classe Δ_i il rapporto tra la propria probabilità empirica e la propria ampiezza $\Delta_i = b_i - a_i$

$$f_n(x) = \frac{n_i}{n\Delta_i} = \frac{f_i}{\Delta_i} \tag{3.10}$$

Per $n \to \infty \quad \wedge \quad \Delta_i \to 0$ la distribuzione delle probabilità empiriche $f_n(x)$ tende alla densità di probabilità $f_X(x)$

3.1.8 Funzione di distribuzione delle frequenze cumulate relative

La funzione di distribuzione delle frequenze cumulate relative associa ad ogni classe Δ_i la probabilità che X assuma valori minori o uguali al valore centrale della classe

$$F_i(x) = \int_{\Delta_i} f_n(x)dx = \sum_{i=1}^{n} \frac{f_i}{\Delta_i} \tag{3.11}$$

Per $n \to \infty \quad \wedge \quad \Delta_i \to 0$ la distribuzione delle frequenze cumulate relative $F_i(x)$ tende alla distribuzione cumulata $F_X(x)$

3.1.9 Indici di posizione

3.1.9.1 Media pesata

La media pesata è il rapporto tra: la somma dei valori centrali $\bar{x}_i$ di ciascuna classe moltiplicati per la frequenza assoluta della classe e il numero di dati totale

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{m} n_i \bar{x}_i \tag{3.12}$$

Nel caso in cui la serie di dati non sia raggruppata, e quindi $n_i = 1$, la media pesata viene detta media aritmetica

3.1.9.2 Media spuntata ($\pm\delta$)

La media spuntata è la media pesata calcolata considerando solo una percentuale dei dati centrali che si trovano nell'intervallo $[1+\delta, n-\delta]$

$$\bar{x}_t = \frac{1}{n} \sum_{i=1+\delta}^{n-\delta} x_i \tag{3.13}$$

3.1.9.3 Media mobile

La media mobile è la media pesata a intervalli regolari di periodo k; è data dalla somma del prodotto tra k dati x_{j+k} presi in considerazione e i pesi p_j

$$\bar{x}_M(t+h) = \sum_{j=1}^{k} p_j x_{t+j} \qquad t = 0, 1, \ldots, n-k \tag{3.14}$$

3.1.9.4 Moda

La moda è il valore centrale della classe Δ_i avente la frequenza maggiore

$$Mo = \bar{x}_i \in \Delta_i \quad | \quad f_i = max\{f_1, f_2, \ldots, f_m\} \tag{3.15}$$

3.1.9.5 Mediana

La mediana è il valore dell'insieme delle modalità di un carattere Δ che divide in parti uguali la superficie coperta dalla funzione delle frequenze relative $f_n(x)$

$$\tilde{x} \in \Delta \quad | \quad \mathbb{P}(X \leq \tilde{x}) = \mathbb{P}(X \geq \tilde{x}) = \frac{1}{2} \implies \tilde{x} = a_k + \frac{\Delta_k}{f_k}\left(\frac{1}{2} - F_{k-1}\right) \tag{3.16}$$

3.1.9.6 Quantile q-esimo

Il quantile q-esimo è il valore massimo delle modalità per cui la frequenza cumulata relativa è minore o uguale al valore q

$$x_q \in \Delta_q \quad | \quad \mathbb{P}(X \leq x_q) = F_X(x_q) = q \tag{3.17}$$

3.1.10 Indici di dispersione

3.1.10.1 Varianza

La varianza è il rapporto tra: la somma pesata dei quadrati degli scarti tra ogni valore centrale $\bar{x}_i$ e la media pesata della distribuzione $\bar{x}$; il numero di dati totale.

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(\bar{x}_i - \bar{x})^2 \tag{3.18}$$

La varianza si può esprimere come: $\sigma_X^2 = \sigma_W^2 + \sigma_B^2$

3.1.10.2 Scarto quadratico medio (o deviazione standard)

Lo scarto quadratico medio è la radice quadrata positiva della varianza

$$\sigma_X = \sqrt{\sigma_X^2} \tag{3.19}$$

3.1.10.3 Media pesata delle varianze

La media pesata delle varianze è il rapporto tra la somma pesata delle varianze calcolate all'interno di ciascuna classe Δ_i e il numero di dati totale

$$\sigma_W^2 = \frac{1}{2}\sum_{i=1}^{m} n_i\sigma_i^2 \tag{3.20}$$

La media pesata delle varianze si avvicina alla varianza σ_X^2 se le classi Δ_i hanno una varianza omogenea (cioè hanno un grado di dispersione simile)

3.1.10.4 Varianza delle medie

La varianza delle medie è il rapporto tra la somma pesata dei quadrati degli scarti tra: le medie $\bar{x}_{\Delta_i}$ all'interno di ciascuna classe e la media aritmetica della serie di dati; il numero di dati totale

$$\sigma_B^2 = \frac{1}{n}\sum_{i=1}^{m} n_i(\bar{x}_{\Delta_i} - \bar{x})^2 \tag{3.21}$$

La varianza delle medie misura l'eterogeneità delle classi Δ_i

3.1.10.5 Coefficiente di asimmetria

Il coefficiente di asimmetria è il rapporto tra il momento centrale di ordine 3 di una distribuzione e il cubo della deviazione standard

$$\alpha_3 = \frac{m_3}{\sigma^3} \qquad m_3 = \frac{1}{n}\sum_{i=1}^{m} n_i(\bar{x}_i - \bar{x})^3 \tag{3.22}$$

Se il coefficiente di asimmetria è nullo la distribuzione è simmetrica; se è negativo la distribuzione è asimmetrica verso sinistra; se è positivo la distribuzione è asimmetrica verso destra

3.1.10.6 Covarianza

La covarianza è il rapporto tra: le somme pesate del prodotto tra gli scarti tra i valori centrali delle classi (di X e di Y) e la media aritmetica (di X e di Y); il numero di dati totale

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{l} n_{ij}(\bar{x}_i - \bar{x})(\bar{y}_j - \bar{y}) \tag{3.23}$$

3.1.11 Indice di connessione χ^2 di Pearson

$$\chi^2 = n\left(\sum_{i=1}^{m}\sum_{j=1}^{l}\frac{n_{ij}}{n_i n_j}\right) \tag{3.24}$$

L'indice di connessione χ^2 di Pearson è nullo se e solo se i dati del campione sono statisticamente indipendenti

3.1.12 Regressione lineare

3.1.12.1 Coefficiente di traslazione

Il coefficiente di traslazione è la differenza tra la media $\bar{y}$ e il prodotto tra la media $\bar{x}$ e il rapporto tra la covarianza s_{xy} e la varianza σ_x^2

$$a = \bar{y} - \frac{s_{xy}}{\sigma_x^2}\bar{x} \tag{3.25}$$

3.1.12.2 Coefficiente di regressione lineare

Il coefficiente di regressione lineare è il rapporto tra la covarianza s_{xy} e la varianza σ_x^2

$$b = \frac{s_{xy}}{\sigma_x^2} \tag{3.26}$$

3.1.12.3 Retta di regressione lineare

La retta di regressione lineare di Y su X è una retta il cui coefficiente di traslazione è a e il cui coefficiente angolare è il coefficiente di regressione lineare b

$$y = a + bx \tag{3.27}$$

3.1.12.4 Coefficiente di determinazione lineare

Il coefficiente di determinazione lineare è il rapporto tra la covarianza s_{xy} e il prodotto delle deviazioni standard di X e di Y

$$\rho(X,Y) = \frac{s_{xy}}{\sigma_x\sigma_y} \in [-1,1] \tag{3.28}$$

Se il valore assoluto del coefficiente di determinazione lineare è circa a 1 X e Y hanno una forte correlazione lineare; se è circa 0 X e Y hanno una scarsa correlazione lineare

3.1.12.5 Errore standard

L'errore standard è il prodotto tra la deviazione standard di Y e la radice quadrata della differenza tra 1 e il quadrato del coefficiente di determinazione lineare ρ

$$\sigma_{xy} = \sigma_y\sqrt{1-\rho^2(X,Y)} \tag{3.29}$$

3.2 Modelli statistici

3.2.1 Modello uniforme $\mathcal{R}(\theta_1, \theta_2)$

Il modello uniforme descrive caratteri che possono assumere casualmente qualsiasi modalità in un intervallo limitato incognito

3.2.1.1 Famiglia di densità

$$f(x, \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1} \qquad x \in [\theta_1, \theta_2] \tag{3.30}$$

3.2.1.2 Parametri incogniti

$$(\theta_1, \theta_2) \mid \quad -\infty < \theta_1 < \theta_2 < +\infty \tag{3.31}$$

3.2.2 Modello Normale-1 $\mathcal{N}(\theta, \sigma^2)$

Il modello Normale-1 descrive caratteri che hanno una distribuzione normale con valor medio incognito

3.2.2.1 Famiglia di densità

$$f(x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \tag{3.32}$$

3.2.2.2 Parametri incogniti

$$\theta \in \Theta = \mathbb{R} \tag{3.33}$$

3.2.3 Modello Normale-2 $\mathcal{N}(\mu, \theta)$

Il modello Normale-2 descrive caratteri che hanno una distribuzione normale con varianza incognita

3.2.3.1 Famiglia di densità

$$f(x, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-\mu)^2}{2\theta}} \tag{3.34}$$

3.2.3.2 Parametri incogniti

$$\theta \in \Theta = \mathbb{R}^+ \tag{3.35}$$

3.2.4 Modello Normale generale $\mathcal{N}(\theta_1, \theta_2)$

Il modello Normale generale descrive caratteri che hanno una distribuzione normale con varianza e valor medio incogniti

3.2.4.1 Famiglia di densità

$$f(x, \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x-\theta_1)^2}{2\theta_2}} \tag{3.36}$$

3.2.4.2 Parametri incogniti

$$(\theta_1, \theta_2) \mid \quad -\infty < \theta_1 < +\infty \quad \wedge \quad \theta_2 > 0 \tag{3.37}$$

3.2.5 Modello Binomiale $\mathcal{B}i(n, \theta)$

Il modello Binomiale descrive i dati risultanti da una sequenza di n (grande) prove ripetute e indipendenti con probabilità di successo incognita

3.2.5.1 Famiglia di densità

$$f(x, \theta) = \sum_{k=0}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} \delta(x-k) \tag{3.38}$$

3.2.5.2 Parametri incogniti

$$\theta \in \Theta = (0, 1) \tag{3.39}$$

3.2.6 Modello di Poisson $\Pi(\theta)$

Il modello di Poisson descrive i dati risultanti da una sequenza di n prove ripetute e indipendenti con probabilità di successo incognita (piccola)

3.2.6.1 Famiglia di densità

$$f(x, \theta) = e^{-\theta} \sum_{k=0}^{\infty} \frac{\theta^k}{k!} \delta(x-k) \tag{3.40}$$

3.2.6.2 Parametri incogniti

$$\theta \in \Theta = \mathbb{R}^+ \tag{3.41}$$

3.2.7 Modello esponenziale $\mathcal{E}(\theta)$

3.2.7.1 Famiglia di densità

$$f(x, \theta) = \theta e^{\theta x} \qquad x \geq 0 \tag{3.42}$$

3.2.7.2 Parametri incogniti

$$\theta \in \Theta = \mathbb{R}^+ \tag{3.43}$$

3.3 Statistiche campionarie

3.3.1 Statistica campionaria (o riassunto campionario)

La statistica campionaria è una variabile casuale $g(\mathbf{X})$ che si esprime per mezzo delle n variabili casuali $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$

3.3.2 Momenti campionari

3.3.2.1 Momento campionario di ordine q

Il momento campionario di ordine q è una statistica campionaria data dal rapporto tra la somma delle q-esime potenze delle variabili aleatorie X_i e la numerosità del campione

$$G_n = \frac{1}{n}\sum_{i=1}^{n} X_i^q \tag{3.44}$$

3.3.2.2 Media campionaria

La media campionaria è il rapporto tra la somma delle n variabili aleatorie X_i e la numerosità del campione

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3.45}$$

3.3.2.3 Varianza campionaria

La varianza campionaria è il rapporto tra il quadrato degli scarti tra la i-esima variabile aleatoria X_i e la media campionaria μ_n e la numerosità del campione

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \mu_n)^2 \tag{3.46}$$

3.3.2.4 Varianza campionaria corretta

La varianza campionaria corretta è il rapporto tra il quadrato degli scarti tra la i-esima variabile aleatoria X_i e la media campionaria μ_n e la numerosità del campione meno 1

$$\hat{S}_n^2 = \frac{n}{n-1} S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \mu_n)^2 \tag{3.47}$$

3.3.2.5 Deviazione standard campionaria corretta (campionamento con ripetizione)

La deviazione standard campionaria corretta è la radice quadrata della varianza campionaria corretta

$$\hat{S}_n = \sqrt{\hat{S}_n^2} = \sqrt{\frac{n}{n-1}} S_n \tag{3.48}$$

3.3.2.6 Deviazione standard campionaria corretta (campionamento senza ripetizione)

La deviazione standard campionaria corretta è la radice quadrata della varianza campionaria corretta

$$\hat{S}_n = \sqrt{\hat{S}_n^2} = \sqrt{\frac{N-1}{N}\frac{n}{n-1}} S_n \tag{3.49}$$

3.3.3 Distribuzione campionaria delle medie

3.3.3.1 Valor medio

Il valor medio della distribuzione campionaria della media campionaria μ_n coincide con il valor medio μ della distribuzione teorica del carattere X

$$\mu = \mu_n \tag{3.50}$$

3.3.3.2 Varianza(campionamento con ripetizione)

La varianza della distribuzione campionaria delle medie è uguale al rapporto tra la varianza teorica della distribuzione del carattere X e la numerosità del campione

$$\sigma_n^2 = \frac{\sigma^2}{n} \tag{3.51}$$

3.3.3.3 Varianza (campionamento senza ripetizione)

La varianza della distribuzione campionaria delle medie è uguale al prodotto tra: il rapporto tra la varianza teorica σ^2 e la numerosità del campione n; il rapporto tra: la differenza tra il numero di elementi della popolazione N e la numerosità del campione n; il numero di elementi della popolazione meno 1

$$\sigma_n^2 = \frac{\sigma^2}{n}\frac{N-n}{N-1} \tag{3.52}$$

3.3.3.4 Deviazione standard

La deviazione standard della distribuzione campionaria delle medie è la radice quadrata della varianza

$$\sigma_n = \sqrt{\sigma_n^2} \tag{3.53}$$

3.3.4 Distribuzione campionaria delle varianze

3.3.4.1 Teorema χ^2

Se da una popolazione normalmente distribuita con varianza σ^2 si estraggono n campioni casuali, la variabile aleatoria

$$Q_n(\omega) = \frac{nS_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu_n)^2 \tag{3.54}$$

ha distribuzione χ^2_{n-1} con $n-1$ gradi di libertà

3.3.4.2 Valor medio corretto (campionamento con ripetizione)

Il valor medio della distribuzione campionaria delle varianze corretta coincide con la varianza teorica σ^2

$$E\{\hat{S}_n^2\} = \sigma^2 \tag{3.55}$$

3.3.4.3 Varianza corretta (campionamento con ripetizione)

La varianza della distribuzione campionaria delle varianze corretta è il rapporto tra 2 volte il quadrato della varianza teorica σ^2 e la numerosità del campione n meno 1

$$\sigma^2(\hat{S}_n^2) = \frac{2}{n-1}\sigma^4 \tag{3.56}$$

3.3.4.4 Valor medio corretto (campionamento senza ripetizione)

Il valor medio della distribuzione campionaria delle varianze è il prodotto tra: il rapporto tra la numerosità degli elementi della popolazione N e la numerosità N meno 1; il rapporto tra la numerosità del campione n meno 1 e la numerosità n; la varianza teorica σ^2

$$E\{\hat{S}_n^2\} = \frac{N}{N-1}\frac{n-1}{n}\sigma^2 \tag{3.57}$$

3.3.4.5 Varianza corretta (campionamento senza ripetizione)

La varianza della distribuzione campionaria delle varianze corretta è il prodotto tra: la numerosità degli elementi della popolazione N meno 1 e la numerosità N; il rapporto tra il quadrato dello scarto tra la i-esima variabile aleatoria e la media campionaria μ_n e la numerosità del campione n meno 1

$$\sigma^2(\hat{S}_n^2) = \frac{N-1}{N}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu_n)^2 \tag{3.58}$$

3.3.5 Deviazione standard corretta (campionamento senza ripetizione)

La deviazione standard della distribuzione campionaria delle varianze corretta è la radice quadrata della varianza

$$\sigma_n = \sqrt{\sigma^2(\hat{S}_n^2)} \tag{3.59}$$

3.3.6 Distribuzione campionaria delle frequenze

La distribuzione campionaria delle frequenze per $n \to \infty$ tende al modello Normale generale $\mathcal{N}(\theta_1, \theta_2)$

3.4 Stime di statistiche campionarie

3.4.1 Stimatore puntuale

Avendo a disposizione le realizzazioni $(x_1, x_2, \ldots, x_n)$ di un campione $\mathbf{X}$, lo stimatore puntuale di un parametro incognito θ di un modello statistico è una statistica campionaria G_n

$$G_n = g(\mathbf{X}) = g(X_1, X_2, \ldots, X_n) \tag{3.60}$$

3.4.2 Proprietà degli stimatori

3.4.2.1 Stimatore corretto (o imparziale o non distorto)

Lo stimatore corretto di θ è una statistica campionaria il cui valore atteso $E\{G_n\}$ coincide con il valore teorico del parametro da stimare

$$G_n \quad | \quad E\{G_n\} = \theta \tag{3.61}$$

3.4.2.2 Stimatore consistente in probabilità

Lo stimatore consistente in probabilità di θ è una statistica campionaria che converge al valore teorico θ con probabilità 1 quando la numerosità del campione tende ad ∞

$$G_n \quad | \quad \lim_{n\to\infty} \mathbb{P}(|G_n - \theta| \leq \epsilon) = 1 \qquad \forall \epsilon > 0 \tag{3.62}$$

3.4.2.3 Stimatore consistente in media quadratica

Lo stimatore consistente in media quadratica di θ è una statistica campionaria il cui scarto al quadrato con il valore teorico θ si annulla quando la numerosità del campione tende a ∞

$$G_n \quad | \quad \lim_{n\to\infty} (G_n - \theta)^2 = 0 \tag{3.63}$$

La consistenza in media quadratica implica la convergenza in probabilità (non vale il viceversa)
Uno stimatore è consistente in media quadratica se: $\lim_{n\to\infty} \sigma^2(G_n) = 0$

3.4.2.4 Stimatore efficiente

Lo stimatore efficiente di θ è uno stimatore corretto che ha una distribuzione campionaria avente varianza inferiore rispetto a quella degli altri stimatori di θ

$$G_n \quad | \quad \forall G_n^i, \quad \sigma^2(G_n) \leq \sigma^2(G_n^i) \tag{3.64}$$

3.4.3 Criterio di stima puntuale

Il miglior stimatore di θ è il più efficiente tra gli stimatori corretti e consistenti di θ

$$G_n \quad | \quad \begin{cases} \sigma^2(G_n) = \min\{\sigma^2(G_n^i)\} \\ \lim_{n\to\infty} \mathbb{P}(|G_n - \theta| \leq \epsilon) = 1 \qquad \forall \epsilon > 0 \end{cases} \tag{3.65}$$

3.4.4 Stima ottima di un parametro

La stima ottima di un parametro θ incognito è il valore del miglior stimatore di θ calcolato attraverso le n realizzazioni x_i del campione $\mathbf{X}$

$$\hat{\theta} = g_n(x_1, x_2, \ldots, x_n) \tag{3.66}$$

3.4.5 Stima ottima del valor medio

La stima ottima del valor medio è la media campionaria μ_n

$$\hat{\mu} = \mu_n = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.67}$$

3.4.6 Stima ottima della varianza

La stima ottima della varianza (effettuata con campioni estratti con ripetizione) è la varianza campionaria corretta $\hat{S}_n^2$

$$\hat{\sigma}^2 = \hat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \mu_n)^2 \tag{3.68}$$

3.4.7 Densità di probabilità congiunta di n realizzazioni x_i

La densità di probabilità congiunta di n variabili aleatorie indipendenti X_i aventi realizzazioni $x_1, x_2, \ldots, x_n$ e densità di probabilità $f(x_i, \theta)$ è il prodotto delle densità di probabilità delle variabili aleatorie

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta) \tag{3.69}$$

3.4.8 Stima di massima verosimiglianza

La stima di massima verosimiglianza per il parametro θ è il valore del parametro θ per cui è massima la densità di probabilità congiunta delle realizzazioni del carattere $\mathbf{X}$

$$\hat{\theta} = MLE(\theta) \quad | \quad L(\hat{\theta}, \mathbf{x}) = \max_{\theta \in \Theta}\{L(\theta_i, \mathbf{x})\} \tag{3.70}$$

3.4.9 Equazione di verosimiglianza

L'equazione di verosimiglianza è un'equazione differenziale che individua i punti stazionari di massimo della di densità di probabilità $L(\theta_i, \mathbf{x})$, cioè la stima di massima verosimiglianza di θ

$$\sum_{i=1}^{n} \frac{\partial f(\theta, x_i)}{\partial \theta} \frac{1}{f(\theta, x_i)} = 0 \tag{3.71}$$

3.4.10 Livello fiduciario (o probabilità fiduciaria)

Il livello fiduciario è valore di probabilità scelto a priori che quantifica il grado di accettazione dell'errore commesso nella stima di un parametro θ

$$\gamma \in [0,1] \tag{3.72}$$

3.4.11 Intervallo di confidenza (o fiduciario)

L'intervallo di confidenza è l'intervallo di valori di Θ per cui la probabilità che θ cada nell'intervallo è maggiore o uguale al livello fiduciario γ

$$[\theta_{min}, \theta_{max}] \in \Theta \quad | \quad \mathbb{P}(\theta \in [\theta_{min}, \theta_{max}]) \geq \gamma \qquad \forall \theta \in \Theta \tag{3.73}$$

3.4.11.1 Intervallo fiduciario simmetrico (o a due code)

L'intervallo fiduciario simmetrico è un intervallo fiduciario centrato sul valore calcolato dello stimatore puntuale $\hat{\theta}$

$$[\hat{\theta} - \delta, \hat{\theta} + \delta] \in \Theta \quad | \quad \mathbb{P}(\theta \in [\hat{\theta} - \delta, \hat{\theta} + \delta]) \geq \gamma \qquad \forall \theta \in \Theta \tag{3.74}$$

3.4.11.2 Intervallo fiduciario asimmetrico (o a una coda)

L'intervallo fiduciario asimmetrico è un intervallo fiduciario in cui si stima che θ sia minore o uguale oppure maggiore o uguale a determinati valori

$$(-\infty, \theta_{max}] \in \Theta \quad | \quad \mathbb{P}(\theta \in (-\infty, \theta_{max}]) \geq \gamma \qquad \forall \theta \in \Theta \tag{3.75}$$

$$[\theta_{min}, +\infty) \in \Theta \quad | \quad \mathbb{P}(\theta \in [\theta_{min}, +\infty)) \geq \gamma \qquad \forall \theta \in \Theta \tag{3.76}$$

3.4.12 Coefficiente fiduciario di una popolazione con varianza nota

Il coefficiente fiduciario è il quantile di ordine $q = \frac{1+\gamma}{2}$ della legge Normale standard $\mathcal{N}(0,1)$

$$z_{\frac{1+\gamma}{2}} \quad | \quad F_Z(z_{\frac{1+\gamma}{2}}) = \mathbb{P}(Z \leq z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2} \tag{3.77}$$

3.4.13 Coefficiente fiduciario di una popolazione con varianza ignota

Il coefficiente fiduciario è il quantile di ordine $q = \frac{1+\gamma}{2}$ della legge t-Student

$$t_{\frac{1+\gamma}{2}} \quad | \quad F_T(t_{\frac{1+\gamma}{2}}) = \mathbb{P}(T \leq t_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2} \tag{3.78}$$

3.4.14 Intervalli di confidenza per la media

3.4.14.1 Intervallo di confidenza simmetrico di una popolazione con varianza nota

L'intervallo di confidenza simmetrico per la media per una stima al livello fiduciario γ ha come estremi i limiti fiduciari dati dalla somma (differenza) tra il valore calcolato della media campionaria μ_n e il prodotto tra la variazione standard $\frac{\sigma}{\sqrt{n}}$ e il coefficiente fiduciario $z_{\frac{1+\gamma}{2}}$

$$\mu \in [\mu_n - \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}, \mu_n + \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}] \tag{3.79}$$

3.4.14.2 Intervallo di confidenza asimmetrico di una popolazione con varianza nota

L'intervallo di confidenza asimmetrico per la media per una stima al livello fiduciario γ ha come estremi i limiti fiduciari dati: da $-\infty$ e la somma tra il valore calcolato della media campionaria μ_n e il prodotto tra la variazione standard $\frac{\sigma}{\sqrt{n}}$ e il coefficiente fiduciario z_γ (superiore); oppure dalla differenza tra il valore calcolato della media campionaria μ_n e il prodotto tra la variazione standard $\frac{\sigma}{\sqrt{n}}$ e il coefficiente fiduciario z_γ e $+\infty$ (inferiore)

$$\mu \in (-\infty, \mu_n + \frac{\sigma}{\sqrt{n}} z_\gamma] \tag{3.80}$$

$$\mu \in [\mu_n - \frac{\sigma}{\sqrt{n}} z_\gamma, +\infty) \tag{3.81}$$

3.4.14.3 Intervallo di confidenza simmetrico di una popolazione con varianza ignota

L'intervallo di confidenza simmetrico per la media per una stima al livello fiduciario γ ha come estremi i limiti fiduciari dati dalla somma (differenza) tra il valore calcolato della media campionaria μ_n e il prodotto tra il valore calcolato della deviazione standard campionaria $\frac{\hat{S}_n^2}{\sqrt{n}}$ e il coefficiente fiduciario $t_{\frac{1+\gamma}{2}}$

$$\mu \in [\mu_n - \frac{\hat{S}_n^2}{\sqrt{n}} t_{\frac{1+\gamma}{2}}, \mu_n + \frac{\hat{S}_n^2}{\sqrt{n}} t_{\frac{1+\gamma}{2}}] \tag{3.82}$$

3.4.14.4 Intervallo di confidenza asimmetrico di una popolazione con varianza ignota

L'intervallo di confidenza asimmetrico per la media per una stima al livello fiduciario γ ha come estremi i limiti fiduciari dati: da $-\infty$ e la somma tra il valore calcolato della media campionaria μ_n e il prodotto tra il valore calcolato della deviazione standard campionaria $\frac{\hat{S}_n^2}{\sqrt{n}}$ e il coefficiente fiduciario $t_{\frac{1+\gamma}{2}}$ (superiore); oppure dalla differenza tra il valore calcolato della media campionaria μ_n e il prodotto tra il valore calcolato della deviazione standard campionaria $\frac{\hat{S}_n^2}{\sqrt{n}}$ e il coefficiente fiduciario $t_{\frac{1+\gamma}{2}}$ e $+\infty$ (inferiore)

$$\mu \in (-\infty, \mu_n + \frac{\hat{S}_n^2}{\sqrt{n}} t_\gamma] \tag{3.83}$$

$$\mu \in [\mu_n - \frac{\hat{S}_n^2}{\sqrt{n}} t_\gamma, +\infty) \tag{3.84}$$

3.4.15 Determinazione del livello fiduciario

Il livello fiduciario γ si può determinare, fissato uno scarto $\delta = \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}$, come 2 volte il valore che la funzione degli errori $erf(z_{\frac{1+\gamma}{2}})$ assume in $\frac{\sqrt{n}}{\sigma}\delta$

$$\gamma = 2erf\left(\frac{\sqrt{n}}{\sigma}\delta\right) \tag{3.85}$$

3.4.16 Determinazione della numerosità del campione

La numerosità del campione n necessaria per sostenere che il valor medio rientri nell'intervallo di confidenza fissato un livello fiduciario γ è dato dal quadrato del prodotto tra il coefficiente fiduciario $z_{\frac{1+\gamma}{2}}$ e il rapporto tra la varianza σ e lo scarto $\delta = |\mu_n - \mu|$

$$n \geq \left(z_{\frac{1+\gamma}{2}} \frac{\sigma}{\delta} \right)^2 \tag{3.86}$$

3.4.17 Intervalli di confidenza per la varianza

3.4.17.1 Intervallo di confidenza simmetrico

L'intervallo di confidenza simmetrico per la varianza per una stima al livello fiduciario γ ha come estremi i limiti fiduciari dati dal prodotto tra il valore calcolato della varianza campionaria corretta $\hat{S}_n^2$ e il rapporto tra la numerosità del campione n meno 1 e il coefficiente fiduciario sinistro $\chi_2 = \frac{1+\gamma}{2}$ e destro $\chi_1 = \frac{1-\gamma}{2}$

$$\mu \in [\frac{n-1}{\chi_2}\hat{S}_n^2, \frac{n-1}{\chi_1}\hat{S}_n^2] \tag{3.87}$$

3.4.17.2 Intervallo di confidenza asimmetrico

L'intervallo di confidenza asimmetrico per la varianza per una stima al livello fiduciario γ ha come estremi i limiti fiduciari dati: da $-\infty$ e il prodotto tra il valore calcolato della varianza campionaria corretta $\hat{S}_n^2$ e il rapporto tra la numerosità del campione n meno 1 e il coefficiente fiduciario destro $\chi_1 = 1-\gamma$ (superiore); oppure dal prodotto tra il valore calcolato della varianza campionaria corretta $\hat{S}_n^2$ e il rapporto tra la numerosità del campione n meno 1 e il coefficiente fiduciario sinistro $\chi_2 = \gamma$ e $+\infty$ (inferiore)

$$\mu \in (-\infty, \frac{n-1}{\chi_1}\hat{S}_n^2] \tag{3.88}$$

$$\mu \in [\frac{n-1}{\chi_2}\hat{S}_n^2, +\infty) \tag{3.89}$$

3.5 Test parametrici

3.5.1 Ipotesi statistiche

3.5.1.1 Ipotesi nulla

L'ipotesi nulla è un'ipotesi statistica formulata su un generico fenomeno aleatorio

$$H_0 \tag{3.90}$$

3.5.1.2 Ipotesi parametrica

L'ipotesi parametrica è un'ipotesi nulla che si riferisce a certi parametri θ di una distribuzione teorica appartenenti ad un determinato sottoinsieme Θ_0 del dominio Θ compatibile con il modello statistico adottato

$$H_0 : \quad \theta \in \Theta_0 \subset \Theta \tag{3.91}$$

3.5.1.3 Ipotesi non parametrica

L'ipotesi non parametrica è un'ipotesi nulla che si riferisce alla natura della funzione di distribuzione $F_X(x)$ appartenente ad un determinato sottoinsieme $\mathcal{F}_0$ del dominio delle funzioni di distribuzione $\mathcal{F}$ compatibili con il modello statistico adottato

$$H_0 : \quad F_X(x) \in \mathcal{F}_0 \subset \mathcal{F} \tag{3.92}$$

3.5.1.4 Ipotesi semplice

L'ipotesi semplice è un'ipotesi statistica che si riferisce ad un solo elemento

3.5.1.5 Ipotesi composta

L'ipotesi semplice è un'ipotesi statistica che si riferisce a più elementi

3.5.2 Test statistico

Il test statistico è la regola che permette di decidere se e in quale misura accettare o respingere un'ipotesi statistica esaminando le osservazioni effettuate su una statistica campionaria

3.5.3 Impostazione di un test statistico

L'impostazione di un test statistico avviene attraverso le seguenti operazioni:

- si definisce la legge probabilistica per il carattere aleatorio osservato compatibile con il modello statistico della popolazione
- si definisce l'ipotesi nulla H_0 da verificare
- si definisce l'ipotesi alternativa H_1, valida se e solo se si rifiuta H_0
- si definisce la statistica campionaria $G_n(\mathbf{X})$ avente distribuzione nota quando H_0 è valida
- si suddivide lo spazio delle possibili osservazioni campionarie $\mathcal{G}$ in due sottoinsiemi disgiunti: la regione di accettazione di H_0 ($\mathcal{A}$) e la regione di rifiuto (o regione critica) di H_0 ($\mathcal{C}$)
- si assume come criterio decisionale la regola: si accetta H_0 se la realizzazione osservata della statistica campionaria di $G_n(\mathbf{X})$ appartiene ad $\mathcal{A}$; si accetta H_1 se la realizzazione osservata della statistica campionaria di $G_n(\mathbf{X})$ appartiene ad $\mathcal{C}$

3.5.4 Errore di prima specie

L'errore di prima specie α è l'errore che si commette se si rigetta l'ipotesi nulla H_0 quando essa è vera (falso negativo)

$$\alpha = \mathbb{P}(H_1|H_0) = \mathbb{P}(G_n \in \mathcal{C}|H_0 \text{ vera}) \tag{3.93}$$

3.5.5 Errore di seconda specie

L'errore di seconda specie β è l'errore che si commette se si accetta l'ipotesi nulla H_0 quando essa non è vera (falso positivo)

$$\beta = \mathbb{P}(H_0|H_1) = \mathbb{P}(G_n \in \mathcal{A}|H_1 \text{ vera}) \tag{3.94}$$

Fissato α bisogna cercare il modo di minimizzare β

3.5.6 Livello di fiducia (o di significatività) di un test statistico

Il livello di fiducia di un test statistico è la probabilità di respingere l'ipotesi alternativa H_1 quando è vera l'ipotesi nulla H_0

$$1 - \alpha \tag{3.95}$$

3.5.7 Potenza di un test statistico

La potenza di un test statistico è la probabilità di respingere l'ipotesi nulla H_0 quando è vera l'ipotesi alternativa H_1

$$W = 1 - \beta \tag{3.96}$$

A parità di α maggiore è la potenza di un test minore è l'errore di seconda specie β

3.5.8 Test di Neyman-Pearson tra ipotesi semplici

Data una statistica campionaria $G_n(\mathbf{X})$ con valore empirico noto attraverso le n osservazioni $\mathbf{x}$ su un campione $\mathbf{X}$ estratto da una popolazione con funzione di distribuzione $F_X(x, \theta)$ assolutamente continua di densità $f_X(x, \theta)$, un numero reale arbitrario $c > 0$ e un errore di prima specie fissato, il test più potente (quello che minimizza β) è quello per cui la regione critica è l'insieme delle osservazioni $\mathbf{x}$ per cui il rapporto di verosimiglianza $l(\mathbf{x})$ tra le funzioni di densità di probabilità congiunta, calcolate in θ_0 e in θ_1, è minore di c

$$\mathcal{C} = \left\{ \mathbf{x} : \quad l(\mathbf{x}) = \frac{L(\theta_0, \mathbf{x})}{L(\theta_1, \mathbf{x})} < c \right\} \tag{3.97}$$

3.5.9 Test parametrici con ipotesi composte

3.5.9.1 Test sul valor medio per il modello normale