APPUNTI DI "PROBABILITÀ E STATISTICA"

Alexandru Bradatan Gabriel

1 PROBABILITÀ

1.1 Spazio di probabilità

La probabilità studia i modelli matematici per i fenomeni aleatori.

(ESPERIMENTI)

DEFINIZIONE - FENOMENO ALEATORIO Un fenomeno aleatorio è un fenomeno di cui non possiamo prevedere a priori il risultato.

Il fenomeno aleatorio è il contrario del fenomeno deterministico.

Supponiamo che ogni possibile risultato dell'esperimento aleatorio sia un elemento $\omega \in \Omega$, allora Ω è detto spazio degli eventi elementari

spazio campionatorio

ESEMPI
- Lancio un dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Lancio una moneta finché non esce testa: $\Omega = \mathbb{N} \cup \{\infty\}$

DEFINIZIONE - EVENTO Un evento E è un insieme $E \subseteq \Omega$ di esperimenti aleatori che soddisfano una certa proprietà.

Possiamo costruire eventi tramite le operazioni insiemistiche come unione e intersezione di altri eventi.

Un evento si è verificato se il risultato dell'esperimento ω sta in E.

Definiamo Ω evento certo poiché è sempre verificato e $\emptyset$ evento impossibile poiché nessun evento gli appartiene.

DEFINIZIONE - EVENTI INCOMPATIBILI Definiamo E, F eventi incompatibili se $E \cap F = \emptyset$.

Gli eventi E e $\neg E$ sono sempre incompatibili.

Ciò che manca è un modello matematico che descriva quella intuitiva di probabilità

DEFINIZIONE - PROBABILITÀ (ASSIOMATICA) Sia Ω lo spazio campionario e $\mathcal{F}$ la famiglia di $E \subseteq \Omega$ eventi di interesse. Una probabilità P su $(\Omega, \mathcal{F})$ è una funzione definita $\forall E \in \mathcal{F}$ tale che:

a1) $P(E) \geq 0 \quad \forall E \in \mathcal{F}$

a2) $P(\Omega) = 1$

a3) Se $E_1, \dots, E_n$ successione di eventi a due a due incompatibili, allora:

$$P\left(\bigcup_1^{\infty} E_i\right) = \sum_1^{\infty} P(E_i)$$ → assioma di σ-additività

→ assiomi di Kolmogorov

DEFINIZIONE - SPAZIO DI PROBABILITÀ La terna $(\Omega, \mathcal{F}, P)$ è detta spazio di probabilità.

1.2 Proprietà della probabilità

Ora che abbiamo definito lo spazio di probabilità ed enunciato gli assiomi, studiamone le proprietà.

PROPOSIZIONE - PROPRIETÀ PROBABILITÀ Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità, valgono:

1) $P(\emptyset) = 0$

2) se $E_1, \dots, E_n \in \mathcal{F}$ eventi a 2 a 2 incompatibili allora

$$P\left(\bigcup_{k=1}^{n} E_k\right) = \sum_{k=1}^{n} P(E_k)$$ ← additività finita

3) $P(E^c) = 1 - P(E)$

4) $P(E) \leq 1$

5) se $E \subseteq F$ allora $P(F \setminus E) = P(F) - P(E)$

6) se $E \subseteq F$ allora $P(E) \leq P(F)$ ← monotonia

7) $\forall E, F \in \mathcal{F}$ $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

DIMOSTRAZIONE

1) Poiché $\emptyset = \emptyset \cup \ldots \cup \emptyset$, unione disgiunta, possiamo scrivere per a3 che $P(\emptyset) = P(\emptyset) + P(\emptyset) + \cdots$ che converge solo se $P(\emptyset) = 0$.

2) Possiamo dire che

$$E_1 \cup \cdots \cup E_n = E_1 \cup \ldots \cup E_n \cup \emptyset \cup \ldots \cup \emptyset = \bigcup_{i=1}^{\infty} E_i \qquad E_j = \emptyset \text{ per } j \geq n+1$$

Per Hp gli eventi sono a 2 a 2 disgiunti e

$$P\left(\bigcup_{i=1}^{\infty} E_1\right) = P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{\infty} P(E_i) = \sum_{i=1}^{n} P(E_i)$$

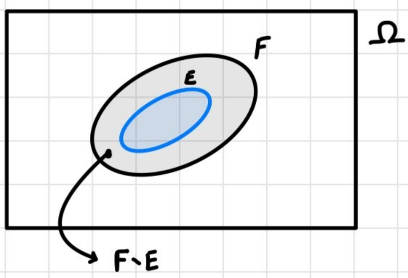
perché $P(E_j) = 0 \quad j \geq n+1$

3) $\Omega = E \cup E^c$ unione disgiunta. Per a2 sappiamo $P(\Omega) = 1$ e per la proprietà 2 dimostrata segue:

$$P(\Omega) = P(E \cup E^c) = P(E) + P(E^c) = 1 \implies P(E^c) = 1 - P(E)$$

4) Deriva da 3:

$$P(E) = 1 - P(E^c) \leq 1 \quad \text{poiché } P(E^c) \geq 0 \text{ per a1}$$

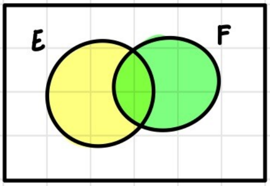
5) 

Possiamo scrivere l'insieme F come unione disgiunta tra $F \setminus E$ ed E. Quindi per la 2 si ha

$$P(F) = P(F \setminus E) + P(E) \implies P(F \setminus E) = P(F) - P(E)$$

6) Per a1 $P(F \setminus E) \geq 0$, quindi se $E \subseteq F$ per 5 si ha

$$P(F \setminus E) = P(F) - P(E) \geq 0 \implies P(F) \geq P(E)$$

7) 

Ω

Possiamo scrivere $E \cup F = E \cup F \setminus (E \cap F)$
e quindi per le precedenti proprietà

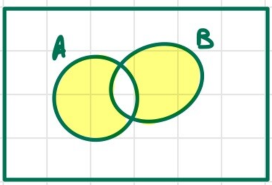
$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$ ■

ESERCIZIO

Siano A e B eventi in $(\Omega, \mathcal{F}, P)$:
1) Se $P(A) = \frac{1}{3}$ e $P(B^c) = \frac{1}{4}$, A e B possono essere incompatibili?
2) Se $P(A)$ e $P(B) = \frac{3}{8}$ può essere $P(A \cup B) = \frac{1}{4}$? E $P(A \cup B) = \frac{7}{8}$?
3) Dimostrare che la probabilità che si verifichi uno e uno solo tra A e B vale $P(A) + P(B) - 2P(A \cap B)$

1) $P(B) = \frac{3}{4}$. I due eventi non possono essere incompatibili poiché se lo fossero avremmo $P(A) + P(B) = \frac{1}{3} + \frac{3}{4} > 1$, che è impossibile.
2) $P(A) = P(B) = \frac{3}{8}$, allora $P(A \cup B)$ non può essere $\frac{1}{4}$ poiché $P(A) = P(B) \leq P(A \cup B)$ e ciò non è vero $\left(\frac{3}{8} > \frac{1}{4}\right)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$. $P(A \cup B)$ non può essere $\frac{7}{8}$ poiché $\frac{7}{8} > P(A) + P(B) = \frac{6}{8}$

3) 

$$A \,\Delta\, B = A \setminus (A \cap B) \cup B \setminus (A \cap B)$$

$$P(A \,\Delta\, B) = P(A) - P(A \cap B) + P(B) - P(A \cap B) = P(A) + P(B) - 2P(A \cap B)$$

1.3 Spazio finito o numerabile

Ora che abbiamo definito come operare con le probabilità, definiamo come possiamo assegnarne una su uno spazio finito o numerabile.

Sia uno spazio campione $\Omega = \{w_1, w_2, \dots, w_n\}$, i nostri eventi d'interesse sono $\mathcal{F} = \mathcal{P}(\Omega)$. Come possiamo assegnare una probabilità ad ogni sottoinsieme di Ω? Iniziamo con l'assegnare una successione di pesi $p_1, \dots, p_n$ tali che $p_k \geq 0$ $\forall k$ e $\sum_{i=1} p_i = 1$.
Se assegniamo ad ogni evento elementare $\{w_k\}$ $k = 1, 2, \dots$ una

probabilità $P(\{w_K\}) = p_K$ allora se voglio che P sia una probabilità necessariamente $\forall E \in \mathcal{P}(\Omega)$ $P(E) = \sum_{K: w_K \in E} p_K$. Infatti considerato un evento E qualsiasi, esso è unione disgiunta di eventi elementari e poiché ogni probabilità deve essere σ-additiva e additiva deve valere che:

$$P(E) = P\left(\bigcup_{K: w_K \in E} \{w_K\}\right) = \sum_{K: w_K \in E} P(\{w_K\}) = \sum_{K: w_K \in E} p_K$$

È facile verificare questa probabilità rispetta gli assiomi su $(\Omega, \mathcal{P}(\Omega))$:

a1) $P(E) \geq 0$ $\forall E$ perché $p_K > 0$ $\forall K$

a2) $P(\Omega) = \sum_{K: w_K \in E} p_K = \sum_{K=1}^{\infty} p_K = 1$ per definizione di p_K

a3) Da fare per esercizio

Se lo spazio è finito la costruzione è analoga.

Vediamo alcune assegnazioni di questi pesi:

1. MODELLO DI POISSON dipende da un parametro $\lambda > 0$:

$$\Omega = \{0, 1, \dots\} \quad p_K = P(\{K\}) = e^{-\lambda} \frac{\lambda^K}{K!}$$

$$p_K \geq 0 ; \quad \sum_{K=0}^{\infty} e^{-\lambda} \frac{\lambda^K}{K!} = e^{-\lambda} \sum_{K=0}^{\infty} \frac{\lambda^K}{K!} = e^{-\lambda} e^{\lambda} = 1$$

2. MODELLO GEOMETRICO dipende da un parametro $0 < p < 1$:

$$\Omega = \mathbb{N} \qquad p_K = P(\{K\}) = p(1-p)^{K-1} \quad K = 1, 2, \dots$$

$$p_K \geq 0 ; \quad \sum_{K=1}^{\infty} p(1-p)^{1-K} \overset{J = K-1}{=} p \sum_{J=0}^{\infty} (1-p)^J = p \frac{1}{1-(1-p)}$$

3. MODELLO BINOMIALE dipende da 2 parametri $n \in \mathbb{N}$ e $0 < p < 1$

$$\Omega = \{0, \dots, n\} \qquad p_K = P(\{K\}) = \overbrace{\binom{n}{K}}^{\frac{n!}{K!(n-K)!}} p^K (1-p)^{n-K}$$

$$p_K \geq 0 ; \quad \sum_{K=0}^{n} p_K = \sum_{K=0}^{n} \binom{n}{K} p^K (1-p)^{n-K} = (p + 1 - p)^n = 1$$

4. MODELLO UNIFORME (PROBABILITÀ CLASSICA)

$$\Omega = \{ w_1, \dots, w_n \} \qquad p_1 = p_2 = \dots = p_n = p$$

$$p \geq 0 \;;\; \sum_{i=0}^{n} p_i = \sum_{i=1}^{n} p = np = 1 \Rightarrow p = \frac{1}{n} = \frac{1}{|\Omega|}$$

$$P(E) = \sum_{K:\, w_K \in E} p_K = \frac{|E|}{|\Omega|}$$

In questa prima parte del corso utilizzeremo principalmente il modello uniforme.

1.4 Probabilità condizionata

Consideriamo il seguente esempio

ESEMPIO Consideriamo il lancio di due dadi regolari:

$$\Omega = \{ (i,j) \quad i = 1..6 \quad j = 1..6 \} \Rightarrow |\Omega| = 36 \quad ; \quad \mathcal{F} = \mathcal{P}(\Omega)$$

Se i dadi non sono truccati allora possiamo usare uno spazio di probabilità uniforme

$$P(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{36} \qquad \forall E \subseteq \Omega$$

Calcoliamo la probabilità che esca 12: $E_{12} = \frac{|E_{12}|}{|\Omega|} = \frac{1}{36}$. Supponiamo di avere una conoscenza parziale dell'evento: sappiamo che sul primo dado sia uscito 6, come varia la probabilità di E_{12}?

La soluzione al problema dell'esempio è la probabilità condizionata:

DEFINIZIONE - PROBABILITÀ CONDIZIONATA Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità e sia $F \in \mathcal{F}$ un evento tale che $P(F) > 0$. Dato un qualsiasi $E \in \mathcal{F}$, si definisce probabilità condizionata di E dato F:

$$P(E|F) := \frac{P(E \cap F)}{P(F)}$$

Tornando all'esempio:

ESEMPIO (CONT.) $P(E_{12}|F) = \frac{P(E_{12} \cap F)}{P(F)} = \frac{1}{6}$ con F = "il risultato del 2° dado è 6"

Nota bene: se $(\Omega, \mathcal{F}, P)$ è uno spazio di probabilità uniforme, cioè Ω è finito e $P(E) = |E|/|\Omega|$ $\forall E \in \Omega$, allora:
- $P(F) > 0 \iff F \neq \emptyset$
- $P(E|F) = |E \cap F|/|F|$

ESERCIZIO Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità e sia $F \in \mathcal{F}$ tale che $P(F) > 0$. Sia P_F la funzione definita da:

$$P_F(E) := P(E|F) \quad \forall E \in \mathcal{F}$$

Mostrare che
1) P_F è una probabilità su $(\Omega, \mathcal{F})$
2) $P_F(F) = 1$
3) Dedurre che $P(E^c|F) = 1 - P(E|F)$ e che $P_F(B \setminus A) = P_F(B) - P_F(A)$

1) P è una probabilità -> soddisfa assiomi di Kolmogorov. Devo dimostrare che anche P_F soddisfa gli assiomi.
   a1) $P_F(E) := \frac{P(E \cap F)}{P(F)} \geq 0 \Rightarrow$ vale a_1
   a2) $P_F(\Omega) := \frac{P(\Omega \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1 \Rightarrow$ vale a_2
   a3) prendiamo una successione di eventi $E_1 \dots E_n$ a 2 a 2 incompatibili; vale:

$$P_F\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P_F(E_i) = \frac{P((\bigcup_{i=1}^{\infty} E_i) \cap F)}{P(F)} = \frac{P((\bigcup_{i=1}^{\infty} E_i \cap F))}{P(F)}$$

poiché $E_i \cap E_j = \emptyset$ $i \neq j \Rightarrow (E_i \cap F) \cap (E_j \cap F) = \emptyset$ $i \neq j$. Quindi si ricava che:

$$\frac{P((\bigcup_{i=1}^{\infty} E_i \cap F))}{P(F)} = \frac{\sum_{i=1}^{\infty} P(E_i \cap F)}{P(F)} = \sum_{i=1}^{\infty} \frac{P(E_i \cap F)}{P(F)} = \sum_{i=1}^{\infty} P_F(E)$$

quindi vale a3.

Poiché valgono tutti e 3 gli assiomi, P_F è una probabilità su $(\Omega, \mathcal{F})$

2) $P_F(F) = \frac{P(F \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$

3) Poiché abbiamo dimostrato che P_F è una probabilità valgono tutte le regole di calcolo per le probabilità e in particolare:

$$\forall E \in \mathcal{F} \quad P_F(E^c) = P(E^c|F) = 1 - P_F(E) = 1 - P(E|F)$$

Analogamente se $A \subseteq B$:

$$P_F(B \setminus A) = P(B \setminus A|F) = P_F(B) - P_F(A)$$

Spesso si ha a che fare con gli eventi le cui cause sono a loro volta aleatorie. La seguente proposizione ci viene in aiuto.

PROPOSIZIONE - FORMULA DELLE PROBABILITÀ TOTALI

Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità e sia $\{F_1, \dots, F_n\}$, $F_i \in \mathcal{F}$ una partizione di Ω, cioè $F_i \cap F_j = \emptyset$ $i \neq j$ e $\bigcup_{i=1}^n F_i = \Omega$, tale che $P(F_i) > 0$ per ogni $i = 1, n$. Allora per ogni $E \in \mathcal{F}$ vale

$$P(E) = \sum_{i=1}^n P(E|F_i) P(F_i)$$

DIMOSTRAZIONE Poiché $\Omega = \bigcup_{i=1}^\infty F_i$

$$E = E \cap \Omega = E \cap \bigcup_{i=1}^n F_i = \bigcup_{i=1}^n (E \cap F_i)$$

Poiché le partizioni sono disgiunte, l'unione sopra è anch'essa disgiunta e quindi possiamo dire:

$$P(E) = P\left(\bigcup_{i=1}^n (E \cap F_i)\right) = \sum_{i=1}^n P(E \cap F_i) = \sum_{i=1}^n P(E|F_i) P(F_i) \quad \blacksquare$$

Nel caso in cui ci viene data una informazione a posteriori se un evento aleatorio e ci viene chiesto in che modo si sia realizzato tale evento è utile la formula di Bayes.

PROPOSIZIONE - FORMULA DI BAYES Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità e $F_1 \dots F_n \in \mathcal{F}$ una partizione finita di Ω tale che $P(F_k) > 0 \quad \forall k = 1 \dots n$. Se $E \in \mathcal{F}$ è tale che $P(E) > 0$ allora si ha:

$$P(F_n|E) = \frac{P(E|F_n)P(F_n)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

DIMOSTRAZIONE Sia $h = 1..n$ fissato. Si ha:

$$P(F_h|E) = \frac{P(F_h \cap E)}{P(E)} = \frac{P(E|F_h)P(F_h)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \qquad \blacksquare$$

ESERCIZIO 1.5.15

Una impresa ha installato un sistema di controllo che rimuove un pezzo difettoso con probabilità 0,995 e sbaglia, rimuovendo un pezzo buono con probabilità 0,001. La probabilità che un pezzo sia difettoso è 0.2. Si calcoli la probabilità che un pezzo non eliminato dopo il controllo sia difettoso

E = "il pezzo viene eliminato"
D = "è difettoso"

$$\rightarrow \begin{cases} P(E|D) = 0,995 \Rightarrow P(E^c|D) = 0,005 \\ P(E|D^c) = 0,001 \Rightarrow P(E^c|D^c) = 0,999 \\ P(D) = 0,2 \Rightarrow P(D^c) = 0,8 \end{cases}$$

$P(D|E^c)$

$$\hookrightarrow P(D|E^c) = \frac{P(E^c|D)P(D)}{P(E^c|D)P(D) + P(E^c|D^c)P(D^c)} \approx 0,00125$$

Consideriamo l'esperimento di estrarre, senza reimettere, una pallina da un'urna con r palline rosse e b bianche. La probabilità che la prima pallina sia rossa e la seconda sia bianca si può calcolare così:

$$P(R_1 \cap B_2) = P(B_2 | R_1) P(R_1) = \frac{r}{r+b} \cdot \frac{b}{r+b-1}$$

- $R_K :=$ "la K-esima pallina è rossa"
- $B_K :=$ "la K-esima pallina è bianca"

Come possiamo estendere il precedente ragionamento al calcolo di $P(R_1 \cap B_2 \cap R_3 \cap B_4)$? Usiamo questa formula.

PROPOSIZIONE - FORMULA DI MOLTIPLICAZIONE Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità e $E_1, \dots, E_n$ eventi $(E_i \in \mathcal{F})$ tali che $P(E_1 \cap \dots \cap E_{n-1}) > 0$. Allora

$$P(E_1 \cap \dots \cap E_n) = P(E_1) P(E_2 | E_1) P(E_3 | E_1 \cap E_2) \dots P(E_n | E_1 \cap \dots \cap E_{n-1})$$

DIMOSTRAZIONE $E_1 \cap \dots \cap E_{n-1} \subseteq E_1 \cap \dots \cap E_{n-1} \subseteq \dots \subseteq E_1$ quindi per la proprietà di monotonia $0 < P(E_1 \cap \dots \cap E_{n-1}) < P(E_1 \cap \dots \cap E_{n-1}) < \dots < P(E_1)$, quindi le probabilità nella nostra tesi sono ben definite. Possiamo allora dire che:

$$P(E_1) P(E_2 | E_1) P(E_3 | E_1 \cap E_2) \dots P(E_n | E_1 \cap \dots \cap E_{n-1}) =$$

$$= \cancel{P(E_1)} \frac{\cancel{P(E_1 \cap E_2)}}{\cancel{P(E_1)}} \cdot \frac{P(E_3 \cap E_1 \cap E_2)}{\cancel{P(E_1 \cap E_2)}} \cdot \dots \cdot \frac{P(E_n \cap E_1 \cap \dots \cap E_{n-1})}{\cancel{P(E_1 \cap \dots \cap E_{n-1})}} =$$

$$= P(E_1 \cap \dots \cap E_n)$$

■

1.4.1 Indipendenza di eventi

Il concetto di indipendenza di eventi è un concetto strettamente collegato alla probabilità condizionata.

DEFINIZIONE - EVENTI INDIPENDENTI Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità. Gli eventi $E, F \in \mathcal{F}$ si dicono indipendenti se

$$P(E \cap F) = P(E) \cdot P(F)$$

Osservazioni 1) Se E, F indipendenti, allora $P(E | F) = P(E)$. Viceversa se $P(E) = P(E|F)$ allora E, F indipendenti. Quindi se $P(F) > 0$ allora E, F sono indipendenti se e solo se $P(E|F) = P(E)$. Inoltre scambiando E con F otteniamo lo stesso risultato.

2) Non confondere la nozione di incompatibilità con la nozione di indipendenza! Se E, F incompatibili allora $E \cap F = \emptyset$, diverso dalla nozione di indipendenza. Inoltre se E, F indipendenti con $P(E)P(F) \neq 0$ allora non sono indipendenti.

Definiamo ora l'indipendenza di più eventi.

DEFINIZIONE - INDIPENDENZA DI PIÙ EVENTI Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità. Gli eventi $E_1, \dots, E_n \in \mathcal{F}$ $n \geq 2$ si dicono indipendenti se per ogni $k \geq 2$ e per ogni scelta di k indici $\{h_1, \dots, h_k\} \subseteq \{1, \dots, n\}$ si ha:

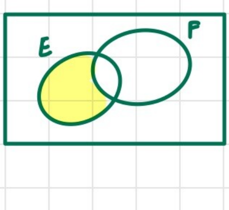
$$P(E_{h_1} \cap \dots \cap E_{h_k}) = \prod_{i=0}^{k} P(E_{h_i})$$

DEFINIZIONE - SUCCESSIONE DI EVENTI IND. Una successione $E_1, E_2, \dots$ in uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ è detta di eventi indipendenti se per ogni $n \geq 2$ gli eventi $E_1, \dots, E_n$ sono indipendenti secondo la definizione di eventi indipendenti

ESERCIZIO Mostrare che se E, F sono eventi indipendenti lo sono anche E e F^c, E^c e F, E^c e F^c.

Basta mostrare che da E, F indipendenti segue che E e F^c sono ancora indipendenti.

$$P(E \cap F^c) = P(E) - P(E \cap F) = P(E) - P(E)P(F) = P(E)(1 - P(F)) = P(E)P(F^c)$$



Si può dimostrare che se ho n eventi $E_1, \dots, E_n$ indipendenti, allora sono indipendenti anche le famiglie di eventi che ottengo sostituendo a uno o più E_i i loro complementari (generalizzazione esercizio sopra).

ESEMPIO Supponiamo di avere un'urna con r palline rosse e b palline bianche. Estraiamo n palline in sequenza. Consideriamo gli eventi:

$B_K :=$ "La K-esima pallina estratta è bianca"
$R_K :=$ "La K-esima pallina estratta è rossa" $= B_K^c$

Consideriamo i 2 casi:

1) RIMPIAZZO Posso estrarre n qualsiasi palline e per $K \leq n$ $P(B_2) = P(B_1) = \dots = P(B_K) = \frac{b}{r+b}$. In particolare $P(B_2 | B_1) = \frac{b}{r+b} = P(B_2)$ ciò significa che gli eventi B_1, B_2 sono indipendenti e facilmente si può estendere il risultato a B_K e R_K.
2) CON REIMMISSIONE Posso estrarre $n \leq r+b$ palline. Per $K \leq n$ vale ancora $P(B_2) = P(B_1) = P(B_K) = \frac{b}{r+b}$ ma $P(B_2 | B_1) \neq P(B_2)$ quindi i B_K eventi non sono indipendenti

1.4.2 Prove di Bernoulli

Consideriamo un esperimento aleatorio "PROVA" con due possibili risultati: "successo" (S) e "insuccesso" (I). Ripeto n volte l'esperimento con le seguenti ipotesi:
1) condizioni identiche
2) il risultato di ciascuna prova non influenza il risultato delle altre

Costruiamo uno spazio di probabilità partendo dalle caratteristiche sopra indicate. Prendiamo:

$$\Omega = \{ \omega = (a_1, \dots, a_n) \quad a_i = 0,1 \quad i = 1..n \} \rightarrow |\Omega| = 2^n$$

$$a_i = \begin{cases} 1 & \text{se il risultato della i-esima prova è un successo} \\ 0 & \text{" " " " " " " insuccesso} \end{cases}$$

$\mathcal{F} = \mathcal{P}(\Omega)$

Si tratta ora di assegnare $P(\{\omega\})$. Consideriamo gli eventi:

$E_1 :=$ "La prima prova è un successo"
$E_2 :=$ "La seconda prova è un successo"
$E_n :=$ "L'n-esima prova è un successo"

Le nostre due ipotesi si traducono in:

1) $P(E_1) = P(E_2) = \dots = P(E_n) = p$
2) $E_1, \dots, E_n$ sono eventi indipendenti

La probabilità $P(\{\omega\})$ di un $\omega = (a_1, \dots, a_n)$ qualsiasi sarà:

$$\{\omega\} = \bigcap_{h: a_h = 1} E_h \cap \bigcap_{h: a_h = 0} E_h^c$$

$$P(\{\omega\}) = \prod_{h: a_h = 1} P(E_h) \cdot \prod_{h: a_h = 0} P(E_h)$$
$$= \prod_{h: a_h = 1} p \cdot \prod_{h: a_h = 0} (1-p) = p^{|\{h: a_h = 1\}|} (1-p)^{|\{h: a_h = 0\}|}$$
$$= p^{\sum_{i=1}^n a_i} (1-p)^{(n - \sum_{i=1}^n a_i)}$$

Dopo aver ~~provato~~ verificato gli assiomi di Kolmogorov (non lo facciamo) possiamo definire lo spazio di n prove di Bernoulli

DEFINIZIONE - SPAZIO DI N PROVE DI BERNOULLI Sia $n \in \mathbb{N}$ e $p \in (0,1)$. Lo spazio di probabilità $(\Omega, \mathcal{F}, P)$, dove $\Omega = \{\omega = (a_1 \dots a_n)\ a_i = 0,1\ i = 1 \dots n\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ e $P(\{\omega\}) = p^{\sum_{i=1}^n a_i} (1-p)^{n - \sum_{i=1}^n a_i}$ per ogni $\omega \in \Omega$ è detto spazio di n prove di Bernoulli con probabilità di successo in ogni prova uguale a p.

Osservazioni

- $\forall E \subseteq \Omega$ allora $P(E) = \sum_{\omega \in E} P(\{\omega\})$

- se $p = \frac{1}{2}$, allora $1-p = \frac{1}{2}$ quindi $P(\{\omega\}) = \dots = \frac{1}{2^n} = \frac{1}{|\Omega|}$ quindi se $p = 1-p$ lo spazio di n prove di Bernoulli è uno spazio uniforme.

1.4.3 Indipendenza condizionale

Sia uno spazio campionario $(\Omega, \mathcal{F}, P)$, un evento $F \in \mathcal{F}$ con $P(F) > 0$ definiamo $P_F(\cdot) := P(\cdot | F)$ cioè $\forall E \in \mathcal{F}$ $P_F(E) = P(E|F)$. Abbiamo visto un esercizio che $P_F(\cdot)$ definisce una probabilità. Definiamo l'indipendenza condizionata di due eventi.

DEFINIZIONE - EVENTI CONDIZIONATAMENTE INDIPENDENTI Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità e siano $E_1, \dots, E_n, F$ eventi e $P(F) > 0$. Allora $E_1, \dots, E_n$ si dicono condizionatamente indipendenti dato F se sono indipendenti rispetto alla probabilità $P_F(E) = P(E|F)$.

2 VARIABILI ALEATORIE

Partiamo da un esempio.

ESEMPIO Supponiamo di lanciare 3 volte una moneta e sia X il numero di teste. X è un numero casuale $x \in \{0, 1, 2, 3\}$. Però poiché questo esperimento è descritto da uno spazio di probabilità uniforme possiamo dire $P(E) = \frac{|E|}{2^3}$. Se però conosco il valore di un esperimento $\omega = (a_1, a_2, a_3)$ posso determinare anche il valore di x. Posso, quindi, dire che X è una funzione di ω:

$$X(\omega) = X(\{a_1, a_2, a_3\}) = a_1 + a_2 + a_3 \rightarrow \begin{cases} a_n = 1 \text{ se testa} \\ a_n = 0 \text{ se croce} \end{cases}$$

Quindi scrivere $P(X=2)$ equivale a scrivere:

$$P(\{\omega = (a_1, a_2, a_3) : a_1 + a_2 + a_3 = 2\})$$

Diamo, quindi, la definizione formale di variabile aleatoria.

DEFINIZIONE - VARIABILE ALEATORIA Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità. Una variabile aleatoria X è una funzione $X : \Omega \to \mathbb{R}$ tale che per $t \in \mathbb{R}$
$\{X \le t\} := \{\omega : X(\omega) \le t\} \in \mathcal{F}$

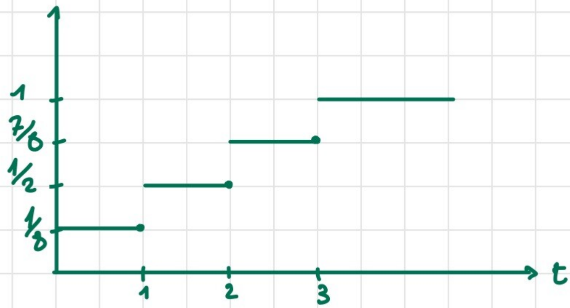
DEFINIZIONE - FUNZIONE DI RIPARTIZIONE Si definisce funzione di ripartizione della v.a. X la funzione $F_X : \mathbb{R} \to [0,1]$ così definita

$$F_X(t) := P(X \le t) = P(\underbrace{\{\omega : X(\omega) \le t\}}_{X^{-1}((-\infty;t])})$$

$X^{-1}((-\infty;t]) \to$ controimmagine

ESEMPIO (cont.) Calcoliamo la funzione di ripartizione dell'esperimento considerato:

$$F_X(t) := P(X \le t) \ \forall t \quad \begin{cases} t<0 & F_X(t) = P(\varnothing) = 0 \\ t \ge 3 & F_X(t) = P(\Omega) = 1 \\ t = 0 & F_X(0) = P(\{0,0,0\}) = \frac{1}{8} \\ 0 \le t < 1 & F_X(t) = P(\{0,0,0\}) = 0 \\ 1 \le t < 2 & F_X(t) = P(\{(0,0,0),(1,0,0) \dots\}) = \frac{1}{2} \\ 2 \le t < 3 & F_X(t) = P(\{(0,0,0),(1,1,0) \dots\}) = \frac{7}{8} \end{cases}$$



Enunciamo le proprietà di ogni funzione di ripartizione:
- F_X è monotona non decrescente
- F_X è continua a destra: $\lim_{s \to t^-} F_X(s) = F_X(t)$
- $\lim_{t \to +\infty} F_X(t) = 1$ e $\lim_{t \to -\infty} F_X(t) = 0$

2.1 Variabili aleatorie discrete

DEFINIZIONE - V.A. DISCRETA Una v.a. X su $(\Omega, \mathcal{F}, P)$ è una v.a discreta se è possibile determinare $S \subseteq \mathbb{R}$ al più numerabile tale che $P(X \in S) = 1$

DEFINIZIONE - DENSITÀ DISCRETA DI PROBABILITÀ Sia X una v.a.d. su $(\Omega, \mathcal{F}, P)$. Si definisce densità discreta di probabilità di X la funzione $p_X(x) := P(X = x)$

ESEMPIO (cont.) Consideriamo la nostra variabile aleatoria X pari al numero di teste ottenute in 3 lanci consecutivi di una moneta possiamo calcolare la seguente densità discreta:

$$p_X(x) = P(X = x) = \begin{cases} 1/8 & \text{se } x \in \{0,3\} \\ 3/8 & \text{se } x \in \{1,2\} \\ 0 & \text{se } x \notin \{0,1,2,3\} \end{cases}$$

PROPOSIZIONE - PROPRIETÀ DENSITÀ DI PROB. Sia X una v.a. discreta a valori in $S = \{x_k \ \ k \in I\}$ $I \subseteq \mathbb{Z}$ e con densità p_X. Allora

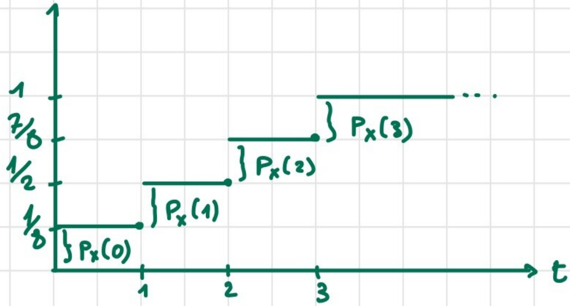
1) $0 \leq p_X(x) \leq 1 \quad \forall x \in \mathbb{R}$ e $p_X(x) = 0 \quad x \notin S$
2) $\sum_{k \in I} p_X(x_k) = 1$
3) se F_X è la f.d.r di X, allora $F_X(t) = \sum_{k : x_k \leq t} p_X(x_k) \quad \forall t \in \mathbb{R}$
4) se $\dots x_{k-1} < x_k < x_{k+1} \dots$ allora $p_X(x_k) = F_X(x_k) - F_X(x_{k-1})$
5) se $B \subseteq \mathbb{R}$ $P(X \in B) = \sum_{k : x_k \in B} p_X(x_k)$

DIMOSTRAZIONE (5) Poiché la probabilità $P(X \in S) = 1$, $S = \{x_k \ \ k \in I\}$, $I \subseteq \mathbb{Z}$:

$$P(X \in B) = P(X \in B \cap S) = P(\{w : X(w) \in B \cap S\}) =$$
$$= P\Big(\underbrace{\bigcup_{k : x_k \in B} \{w : X(w) = x_k\}}_{\text{unione disgiunta}}\Big) = \sum_{k : x_k \in B} P(\{w : X(w) = x_k\}) =$$
$$= \sum_{k : x_k \in B} p_X(x_k)$$

$\blacksquare$

ESEMPIO (cont.) Consideriamo di nuovo la nostra X pari al numero di teste in 3 lanci. Per la proprietà della densità possiamo dire che:



$$\sum_{k=0}^{3} p_k = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

Definiamo un altro concetto utile: la media di una v.a.d

DEFINIZIONE - MEDIA DI UNA V.A.D Sia X una v.a.d a valori $S = \{x_k \ k \in I\}$ con densità p_X. Allora diciamo che X ammette media (o valore atteso) se $\sum |x_k| p_X(x_k) < +\infty$. In tal caso si definisce media, o valore atteso, di X il numero

$$E(X) = \sum_{k \in I} x_k \, p_X(x_k)$$

ESEMPIO (cont.) Calcoliamo la media della nostra X: n° di teste in 3 lanci a valori in $S = \{0,1,2,3\}$.

$$E(X) = \sum_{k=0}^{3} k \, p_X(k) = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) + 3 \cdot p_X(3)$$
$$= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}$$

Possiamo vedere la media come un punto di equilibrio. Inoltre la media non è un valore assunto dalla variabile aleatoria.

ESEMPIO Consideriamo l'esperimento costituito da lanci successivi di una moneta equa fino all'uscita della prima testa. Sia X la v.a.d così definita:

$X :=$ "n° di volte (inclusa l'ultima) che la moneta viene lanciata"

1) Calcolare la densità e la media di X
2) Consideriamo una successione di prove di Bernoulli con probabilità di successo in ogni singola prova pari a p. Sia Y la v.a che conta il numero di prove fino ad ottenere il primo successo (compreso). Determinare la densità e la media di Y.

Si può notare che il punto 1 è un caso particolare del punto 2 dove $p = 1/2$.

$K \in \mathbb{N}$, $p_Y(K) = P(Y=K)$,
$E_i =$ "i-esima prova è un successo" $\to P(E_i) = p = 1 - P(E_i^c)$

$$p_Y(1) = P(Y=1) = P(E_1) = p$$
$$p_Y(2) = P(Y=2) = P(E_1^c \cap E_2) = (1-p)p$$
$$\vdots \; K \geq 2$$
$$p_Y(K) = P(Y=K) = P(E_1^c \cap \dots \cap E_{K-1}^c \cap E_K) = (1-p)^{K-1} p$$

$$P(Y \in \mathbb{N}) = \sum_{K=1}^{\infty} p_Y(K) = \sum_{K=1}^{\infty} p(1-p)^{K-1} = p \sum_{J=0}^{\infty} (1-p)^J$$
$$= p \frac{1}{1-(1-p)} = 1 \to Y \text{ è discreta con } p_Y(x) = 0 \text{ se } x \notin \mathbb{N}$$

$$p_Y(k) = \begin{cases} p(1-p)^{K-1} & K = 1, 2, \dots \\ 0 & K \notin \mathbb{N} \end{cases}$$

$$E(Y) = \sum_{K=1}^{\infty} K p_Y(K) = \sum_{K=1}^{\infty} K p (1-p)^{K-1} = p \sum_{K=1}^{\infty} K (1-p)^{K-1}$$

$$\to f(p) := \sum_{K=1}^{\infty} (1-p)^{K-1} = \sum_{J=0}^{\infty} (1-p)^J = \frac{1}{p}$$

$$\to \frac{df}{dp} = -\sum_{J=1}^{\infty} J(1-p)^{J-1} = -\frac{1}{p^2} \to \sum_{J=1}^{\infty} J(1-p)^{J-1} = \frac{1}{p^2}$$

$$= p \frac{1}{p^2} = \frac{1}{p}$$

Quindi il punto 1: $p_X(K) = \begin{cases} 1/2^K & K \in \mathbb{N} \\ 0 & K \notin \mathbb{N} \end{cases}$ e $E(X) = 2$

2.2 Variabili aleatorie discrete notevoli

2.2.1 Densità geometrica

DEFINIZIONE - V.A.D GEOMETRICA Una v.a.d X è detta geometrica di parametro p $(0<p<1)$ se ha densità:

$$p_X(K) = \begin{cases} p(1-p)^{K-1} & K = 1,2\dots \\ 0 & \end{cases}$$

Scriveremo $X \sim Geom(p)$ e $E(X) = 1/p$.

La v.a.d geometrica l'abbiamo appena incontrata nell'esempio precedente. Enunciamo la proprietà fondamentale di cui godono le v.a.d geometriche: l'assenza di memoria.

PROPOSIZIONE - MANCANZA DI MEMORIA Se $X \sim Geom(p)$ allora vale:

$$\forall K = 0,1,2,\dots \quad m = 1,2\dots \quad P(X > K+m \mid X > K) = P(X > m)$$

DIMOSTRAZIONE Sia $j = 0,1,\dots$, vale

$$P(X>j) = \sum_{K=j+1}^{\infty} p_X(K) = \sum_{K=j+1}^{\infty} p(1-p)^{K-1} = p(1-p)^j \sum_{K=j+1}^{\infty} (1-p)^{K-j-1}$$
$$= p(1-p)^j \sum_{K'=0}^{\infty} (1-p)^{K'} = \cancel{p}(1-p)^j \frac{1}{\cancel{1-(1-p)}} = (1-p)^j$$

$K' = K-j-1$

Quindi:

$$P(X>K+m \mid X>K) = \frac{P(\{X>K+m\} \cap \{X>K\})}{P(X>K)} = \frac{P(X>K+m)}{P(X>K)}$$
$$= \frac{(1-p)^{K+m}}{\cancel{(1-p)^K}} = (1-p)^m = P(X>m)$$

Si può dimostrare anche il viceversa: se X è una v.a.d gode della mancanza di memoria allora X è una v.a.d geometrica con $p = P(X \leq 1) = P(X=1)$. La proprietà sopra può anche essere enunciata come:

$$P(X = K+m \mid X>K) = P(X=m) \quad \forall m = 1,2,\dots \text{ e } K = 0,1,2,\dots$$

■

2.2.2 Densità binomiale e bernoulliana

DEFINIZIONE - V.A.D BINOMIALE Consideriamo n prove di Bernoulli $p \in (0,1)$, $(\Omega, \mathcal{F}, P)$ spazio (o schema) di Bernoulli.
Sia X = n° dei successi, $X(\omega) = \sum_{i=1}^{n} a_i$, v.a.d a valori in $S = \{0, 1, \dots, n\}$ con densità:

$$P_X(K) = \begin{cases} P(X = K) = \binom{n}{K} p^K (1-p)^{n-K} & K \in S \\ 0 \end{cases}$$

La v.a.d con queste proprietà è detta binomiale di parametri n, p.
Scriveremo $X \sim Bi(n, p)$.

ESEMPIO Consideriamo $X :=$ "n° di T in $n = 3$ lanci di una moneta equa" $\Rightarrow X \sim Bi(3, 1/2)$

Se $n = 1$ le binomiali prendono il nome di bernoulliana di parametro p ($X \sim Be(p)$):

$$X \in \{0,1\} \qquad P_X(1) = P(X=1) = p = 1 - P(X=0) = 1 - P_X(0)$$
$$P_X(x) = 0 \qquad x \notin \{0,1\}$$

ESERCIZIO Sia $X \sim Bi(n,p)$. Mostrare che $E(X) = np$

$$X \in S = \{0, 1, 2, \dots, n\}, \quad P_X(K) = \binom{n}{K} p^K (1-p)^{n-K} \quad K \in S$$

$$E(X) = \sum_{K=0}^{n} K P_X(K) = \sum_{K=0}^{n} K \binom{n}{K} p^K (1-p)^{n-K} =$$
$$= \sum_{K=1}^{n} K \frac{n!}{K!(n-K)!} p^K (1-p)^{n-K} = np \sum_{K=1}^{n} \frac{(n-1)!}{(K-1)!(n-K)!} p^{K-1} (1-p)^{n-K} =$$
$$= np \sum_{K=1}^{n} \frac{(n-1)!}{(K-1)!(n-K)!} p^{K-1} (1-p)^{n-K} = np \sum_{K=1}^{n} \frac{(n-1)!}{(K-1)![n-1-(K-1)]!} p^{K-1} (1-p)^{n-K} =$$
$$= np \sum_{J=0}^{n-1} \frac{(n-1)!}{J!(n-1-J)!} p^J (1-p)^{n-1-J} = np(\cancel{p} + 1 - \cancel{p})^{n-1} = np$$

La media di $X \sim Bi(n,p)$ è $E(X) = np$. Nel caso della bernoulliana essa diventa $E(X) = p$.

2.2.3 Densità di Poisson

ESEMPIO Consideriamo il centralino di un numero verde. In genere è costituito da un certo numero di linee alle chiamate. Sia X il numero di chiamate che arrivano ad un certo operatore in un'ora. Pensiamo a numero n di utenti molto grande ognuno dei quali ha un piccola probabilità p di chiamare il centralino. Se assumiamo che ogni utente agisca indipendentemente, allora $X \sim Bi(n,p)$. Se il centralino è organizzato razionalmente il numero di linee è commisurato al bacino d'utente. Una condizione affinché ciò accada $\lambda = np$ sia un numero fissato non eccessivamente grande. Possiamo scrivere $X \sim Bi(n, \lambda/n)$, ossia:

$$P(X=K) = \binom{n}{K}\left(\frac{\lambda}{n}\right)^K\left(1-\frac{\lambda}{n}\right)^{n-K} = \frac{n!}{(n-K)!\,K!}\left(\frac{\lambda}{n}\right)^K\left(1-\frac{\lambda}{n}\right)^{n-K} =$$
$$= \frac{n!}{(n-K)!\,n^K}\left(1-\frac{\lambda}{n}\right)^{-K}\frac{\lambda^K}{K!}\left(1-\frac{\lambda}{n}\right)^n$$

$$\lim_{n\to+\infty}\frac{n!}{(n-K)\,n^K} = 1, \quad \lim_{n\to+\infty}\left(1-\frac{\lambda}{n}\right)^{-K} = 1, \quad \lim_{n\to+\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$\hookrightarrow P(X=K) \simeq \frac{\lambda^K}{K!}e^{-\lambda}$$

Se $X \sim Bi(n,p)$ con n grande e p piccola, possiamo approssimare la binomiale con la distribuzione di Poisson:

DEFINIZIONE - VAD DI POISSON Una v.a.d è detta di Poisson di parametro $\lambda > 0$ se ha densità:

$$p_X(K) = \begin{cases} e^{-\lambda}\frac{\lambda^K}{K!} & K = 0,1,2,\dots \\ 0 \end{cases}$$

e scriviamo $X \sim Poiss(\lambda)$. La media sarà:

$$E(X) = \sum_{K=0}^{\infty} K\,p_X(K) = \sum_{K=0}^{+\infty} K e^{-\lambda}\frac{\lambda^K}{K!} = \sum_{K=1}^{+\infty} e^{-\lambda}\frac{\lambda^K}{(K-1)!} =$$
$$= e^{-\lambda}\lambda\sum_{K=1}^{\infty}\frac{\lambda^{K-1}}{(K-1)!} = e^{-\lambda}\lambda\sum_{J=0}^{\infty}\frac{\lambda^J}{J!} = \cancel{e^{-\lambda}}\lambda\,\cancel{e^{\lambda}} = \lambda$$

ESEMPIO Consideriamo un esperimento che consiste nel contare il numero di particelle alfa emesse in un secondo da un grammo di un certo materiale radioattivo. Sappiamo dall'esperienza passata che il valore medio di questa v.a è 3.2. Qual è una buona approssimazione della probabilità che vengano emesse più di 3 particelle?

Possiamo pensare alla sorgente come un n° n molto grande di atomi radioattivi ciascuno dei quali ha una probabilità molto piccola di emettere una particella α in un secondo: $np = 3,2 \Rightarrow p = 3.2/n$. La probabilità richiesta è:

$$P(X>3) = 1 - P(X \leq 3) = 1 - e^{-3.2}\left(\frac{3.2^0}{0!} + \frac{3.2^1}{1!} + \frac{3.2^2}{2!} + \frac{3.2^3}{3!}\right) = 0,39748$$

2.2.4 Densità ipergeometrica

ESEMPIO Consideriamo un'urna con r palline rosse e b bianche. Consideriamo l'estrazione di n palline in sequenza. Consideriamo X definita come "numero di palline rosse estratte"

1) CON RIMPIAZZO: n può essere qualsiasi; i risultati delle estrazioni sono indipendenti. Se chiamo S l'estrazione di una pallina rossa con $p = P(S) = r/r+b$ allora $X \sim Bi(n, p = r/r+b)$
2) SENZA RIMPIAZZO: n deve essere minore di $r+b$; i risultati delle estrazioni NON sono indipendenti. Calcoliamo la densità di X:

$X \leq \min(n, r)$

se $n > b \Rightarrow n - b$ palline rosse

se $n \leq b \Rightarrow$ posso avere un'estrazione di sole palline bianche

$\Rightarrow X \geq \max(0, n-b)$

$\Rightarrow X$ va a valori in $S = \{\max(0, n-b), \max(0, n-b)+1, \dots\}$

$$p_X(\overset{K \in S}{K}) = P(X=K) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{r+b}{n}}$$

La densità ottenuta nel secondo esempio è detta ipergeometrica:

DEFINIZIONE - VAD IPERGEOMETRICA Una v.a.d. è detta ipergeometrica di parametri $r+b$, r, n se ha densità:

$$P_X(K) = \begin{cases} \dfrac{\binom{r}{K}\binom{b}{n-K}}{\binom{r+b}{n}} & K \in \{\max(0, n-b), \max(0, n-b)+1, \dots\} \\ 0 & \end{cases}$$

E scriveremo $X \sim \text{Iperg}(r+b, r, n)$.

2.3 Funzioni di v.a.d

ESERCIZIO Un dado equo viene lanciato. Indichiamo con Y il prodotto del punteggio sulla faccia superiore per quello sulla faccia inferiore del dado:

1) Si esprima Y in funzione della v.a. X che rappresenta il valore della faccia superiore.
2) Mostrare che Y è una v.a.d. e determinare la densità, la media e la f.d.r di Y.

Consideriamo $X :=$ "punteggio del dado sulla faccia superiore", allora:

$$Y = X(7-X) \rightarrow Y = g(X)$$

Ovviamente $X \in \{1,2,3,4,5,6\}$ e $P_X(i) = \frac{1}{6}$ per $i = 1..6$ e 0 altrimenti. Quindi $Y \in \{6, 10, 12\}$, quindi Y è v.a.d con densità, media e f.d.r:

$$P_Y(6) = P(X(7-X)=6) = P(X \in \{1,6\}) = P(X=1) + P(X=6) = \tfrac{1}{3}$$
$$P_Y(10) = \dots = P_Y(12) = \tfrac{1}{3}$$
$$P_Y(K) = 0 \quad \text{per } K \notin \{6, 10, 12\}$$

$$E(Y) = 6 \cdot \tfrac{1}{3} + 10 \cdot \tfrac{1}{3} + 12 \cdot \tfrac{1}{3} = \tfrac{28}{3}$$

$$F_Y(t) = \begin{cases} 0 & t < 6 \\ \frac{1}{3} & 6 \le t < 10 \\ \frac{2}{3} & 10 \le t \le 12 \\ 1 & t \ge 12 \end{cases}$$

In generale: sia X una v.a.d a valori in $S = \{x_K \ K \in I\}$ $I \subseteq \mathbb{Z}$ e densità p_X. Sia $Y = g(X)$ $g: S \to \mathbb{R}$ e $Y(\omega) = g(X(\omega))$ $\forall \omega$. Poiché $P(X \in S) = 1$, allora $P(Y \in g(S)) = 1$ quindi Y è anch'essa una v.a.d e $p_Y(y) = 0$ per $y \notin g(S)$. Se invece $y \in g(S)$ allora

$$p_Y(y) = P(g(X) = y) = P(X \in g^{-1}(\{y\})) = \sum_{K: g(x_K) = y} p_X(x_K))$$

2.4 Variabili aleatorie assolutamente continue

DEFINIZIONE - V.A. ASSOLUTAMENTE CONTINUA Una variabile aleatoria X su uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ è detta assolutamente continua se esiste una funzione $f_X : \mathbb{R} \mapsto \mathbb{R}_+$ integrabile tale che la f.d.r F_X di X si può scrivere:

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\,dx$$

$f_X(x)$ si dice densità di X.

OSSERVAZIONE Sappiamo che per ogni v.a. X vale $\forall x \in \mathbb{R}$ $P(X = x) = F_X(x) - F_X(x^-)$. Se X è ass. cont. allora F_X è continua e $F_X(x) = F_X(x^-)$, quindi $P(X = x) = 0$ $\forall x \in \mathbb{R}$.

ESEMPIO Sia X il tempo di vita di un componente non soggetto ad usura. Vedremo che un modello adeguato è:

$$F_X(t) = (1 - e^{-\lambda t})\, I_{[0;+\infty)}(t) = \begin{cases} 1 - e^{-\lambda t} & t \in [0;+\infty) \\ 0 & t \in (-\infty; 0) \end{cases}$$

Infatti se prendo $f_X(x) = \lambda e^{-\lambda x}$ $\Rightarrow$ $F_X = \int_{-\infty}^{t} f_X(x)\,dx$

PROPOSIZIONE - COND. SUFF. PER L'ASS. CONT. Sia X una v.a. e sia F_X la sua f.d.r. Se F_X è continua ovunque ed è derivabile con continuità tranne al più in un numero finito di punti $B = \{x_1, \dots, x_n\}$, allora X è una v.a.a.c e la funzione $f_X(x) = F_X'(x)$ $\forall x \notin B$ e definita in modo arbitrario per $x \in B$ è una densità per X

DEFINIZIONE - MEDIA DI UNA V.A.A.C Sia X una v.a.a.c con la la sua densità f_X. Se $\int_{-\infty}^{+\infty} |x| f_X(x)dx < +\infty$ allora X ammette media (o valore atteso) e si definisce media di X il numero:

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

PROPOSIZIONE - PROPRIETÀ VAAC Sia X una vaac con densità f_X, allora:

1) $\int_{-\infty}^{+\infty} f_X(x)dx = 1$
2) Se F_X è la f.d.r di X, allora $f_X(x) = F'_X(x)$ $\forall x$ in cui tale derivata esiste
3) Se $-\infty < a < b < +\infty$, allora:

$$P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

DIMOSTRAZIONE

1) Sia $F_X(t) = \int_{-\infty}^{t} f_X(x)dx$, allora $1 = \lim_{t\to+\infty} \int_{-\infty}^{t} f_X(x)dx = \int_{-\infty}^{+\infty} f_X(x)dx$.
2) Conseguenza del teorema fondamentale del calcolo integrale
3) $P(a < X \leq b) = F_X(b) - F_X(a)$. Se X a.c allora

$$\int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx = \int_a^b f_X(x)dx$$

Poichè l'integrale non varia per un insieme di di misura nullo, la presenza o no dell'uguale non cambia nulla.

Possiamo enunciare delle regole di calcolo per le vaac:

1) $B = B_1 \cup B_2 \cup \ldots \cup B_i$ con B_i intervalli disgiunti $(a_i, b_i]$, allora:

$$P(X \in B) = P\left(\bigcup_{i=1}^{n} \{X \in (a_i, b_i]\}\right) = \sum_{i=1}^{n} P(a < X \leq b_i)$$
$$= \int_{\cup (a_i; b_i]} f_X(x)dx = \int_B f_X(x)dx$$

In generale: $P(X \in B) = \int_B f_X(x)dx$ ■

2.5 Variabili aleatorie assolutamente continue notevoli

2.5.1 Densità uniforme continua

DEFINIZIONE - VAAC UNIFORME Una vaac X è detta uniforme sull'intervallo (a,b) se ha densità costante in (a,b) e nulla al di fuori:

$$f_X(x) = \begin{cases} K & x \in (a,b) \\ 0 & x \notin (a,b) \end{cases} = K\, I_{(a;b)}(x)$$

Scriviamo che $X \sim U(a,b)$. Nota che la costante deve avere valore appropriato:

$$1 = \int_{-\infty}^{+\infty} f_X(x)\,dx = \int_a^b K\,dx = K(b-a) \Rightarrow K = \frac{1}{b-a}$$

PROPOSIZIONE - PROPRIETÀ VAAC UNIFORME Se $X \sim U(a,b)$ allora:

1) $P(X \in (a,b)) = \int_a^b f_X(x)\,dx = \int_a^b \frac{1}{b-a}\,dx = 1 \Rightarrow X$ assume valori in (a,b) con probabilità 1
2) Sia $a < c < d < b$, $P(X \in (c,d)) = \int_c^d f_X(x)\,dx = \frac{d-c}{b-a}$. La probabilità, quindi, dipende dalla lunghezza di $(c;d)$ non dai punti. Quindi ad intervalli di pari lunghezza corrispondono pari probabilità.
3) La f.d.r di X sarà:

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\,dx = \int_{-\infty}^{t} \frac{1}{b-a} I_{(a,b)}(x)\,dx = \begin{cases} 0 & t \le a \\ \frac{t-a}{b-a} & a < t < b \\ 1 & t \ge b \end{cases}$$

4) La media di X sarà:

$$E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\,dx = \int_{-\infty}^{+\infty} x \frac{1}{b-a} I_{(a,b)}(x)\,dx = \int_a^b x \frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = (a+b)/2$$

2.5.2 Densità esponenziale

Consideriamo un componente non soggetto ad usura che si può guastare per motivi contingenti. Sia X il tempo che intercorre dall'inizio del funzionamento a quando si guasta, ossia il tempo di vita.

$s \leq 0 \quad P(X \geq s) = 1$

$s > 0 \rightarrow$ per mancanza di usura possiamo dire che se $X > t$ allora la probabilità che non si guasti per un ulteriore tempo s è uguale alla probabilità che non si guasti nell'intervallo $(0; s] \rightarrow$ MANCANZA DI MEMORIA

$$\rightarrow P(X > t+s \mid X > t) = P(X > s)$$

$$\Updownarrow$$

$$\frac{P(\{X > t+s\} \cap \{X > t\})}{P(\{X > t\})} = \frac{P(\{X > t+s\})}{P(\{X > t\})} = P(X > s)$$

$$\Updownarrow$$

$$\circledast \; P(X > t+s) = P(X > s)\, P(X > t)$$

Se indichiamo con $\bar{F}_X(t) = P(X > t) = 1 - F_X(t)$ (funzione di sopravvivenza) allora cerchiamo $\bar{F}_X$ tali che $\bar{F}_X(t+s) = \bar{F}_X(t)\bar{F}_X(s)$ $\forall t, s > 0$. Si può dimostrare che l'unica soluzione continua della nostra equazione è $\bar{F}_X(t) = e^{\alpha t}$ e poiché $\forall t \; \bar{F}_X(t) < 1$ necessariam. $\alpha < 0$ e scriviamo $\alpha = -\lambda$ con λ positiva. Quindi l'unica f.d.r che soddisfa i nostri criteri è

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

Abbiamo già visto che tale X è assolutamente continua con densità $f_X(t) = F_X'(t) = \lambda e^{-\lambda t} I_{(0,+\infty)}(t)$. Il modello descritto è l'unico che traduce la relazione $\circledast$ e il tempo di vita senza usura. Questo tipo di variabile aleatoria è detta esponenziale.

DEFINIZIONE - VAAC ESPONENZIALE Una v.a.a.c X è detta esponenziale di parametro λ se ha densità:

$$f_X(x) = \lambda e^{-\lambda} I_{(0,+\infty)}(x)$$

Scriveremo $X \sim \mathcal{E}(\lambda)$. La media sarà $E(X) = 1/\lambda$. Inoltre ogni v.a.a.c che gode di assenza memoria è esponenziale e viceversa.

Come modelliamo i casi di usura e rodaggio? Enunciamo le condizioni da rispettare:

- USURA: $P(X > t+s \mid X > t) < P(X > s)$
- RODAGGIO: $P(X > t+s \mid X > t) > P(X > s)$

Introduciamo degli strumenti che ci permettono di generalizzare ciò visto precedentemente.

DEFINIZIONE - INTENSITÀ DI ROTTURA (HAZARD FUNCTION) Sia X una v.a.a.c tale che $P(X > 0) = 1$ con f.d.r F_X e densità f_X. Si definisce intensità di rottura la funzione

$$\lambda(t) = f_X(t) / (1 - F_X(t))$$

L'intensità di rottura ha il seguente significato:

$$P(t < X \leq t + \Delta t \mid X > t) = \frac{P(\{t < X \leq t+\Delta t\} \cap \{X > t\})}{P(X > t)} =$$

$$= \frac{P(t < X \leq t + \Delta t)}{P(X > t)} = \frac{F_X(t+\Delta t) - F_X(t)}{1 - F_X(t)}$$

$$\lim_{\Delta t \to 0} \frac{P(t < X \leq t + \Delta t \mid X > t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{F_X(t+\Delta t) - F_X(t)}{\Delta t\,(1 - F_X(t))} = \frac{f_X(t)}{1 - F_X(t)} = \lambda(t)$$

Nel caso della distribuzione esponenziale otteniamo che la hazard function risulta:

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{1 - (1 - e^{-\lambda t})} = \lambda \rightarrow \text{costante}$$

Come ricaviamo la funzione di ripartizione di una v.a X dalla hazard function? Si noti che se considero $\lambda(t)$, ciò equivale a

$$\lambda(s) = f_X(s) / (1 - F_X(s)) = -\frac{d}{ds} \ln(1 - F_X(s))$$

Quindi integrando ottengo:

$$\int_0^t \lambda(s)ds = -\left[\ln(1-F_X(t)) - \cancel{\ln(1-F_X(0))}^{\,\to 0}\right] = -\ln(1-F_X(t))$$

$$\Rightarrow 1-F_X(t) = e^{-\int_0^t \lambda(s)ds} \to F_X(t) = e^{-\int_0^t \lambda(s)ds} - 1$$

Grazie all'intensità di rottura possiamo modellizzare le condizioni di usura e rodaggio:

$$P(X>t+s \mid X>t) = \frac{P(X>t+s)}{P(X>t)} = \frac{e^{-\int_0^{t+s}\lambda(u)du}}{e^{-\int_0^t \lambda(u)du}}$$

1) se λ è crescente, otteniamo che per monotonia dell'integrale $P(X>t+s) < P(X>t)$ e quindi avremo un modello per l'usura
2) se λ è decrescente, otteniamo per lo stesso motivo di 1 che $P(X>t+s) > P(X>t)$ e quindi un modello per il rodaggio
3) se λ è costante, $X \sim \mathcal{E}(\lambda)$.

ESEMPIO (Distribuzioni di Weibull) Sia $\lambda = \alpha\beta t^{\beta-1}$, $t>0$, $\alpha, \beta > 0$.
Notiamo che:
- se $\beta < 1$ $\lambda(t)$ è decrescente $\to$ rodaggio
- se $\beta = 1$ $\lambda(t)$ è costante $\to$ $X \sim \mathcal{E}(\lambda)$
- se $\beta > 1$ $\lambda(t)$ è crescente $\to$ usura

Inoltre:

$$F_X(t) = 1 - e^{-\int_0^t \lambda(u)du} = 1 - e^{-\alpha t^\beta} \qquad f_X(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^\beta}$$

2.5.3 Distribuzione gaussiana

DEFINIZIONE - VAAC GAUSSIANA STANDARD Sia una v.a.a.c Z è detta gaussiana standard se ha densità

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad x \in \mathbb{R}$$

Scriviamo che $Z \sim \mathcal{N}(0,1)$.

Si può dimostrare che $\int_{-\infty}^{+\infty} \varphi(x)dx = 1$. La funzione di ripartizione è:

$$\Phi(t) = \int_{-\infty}^{t} \varphi(x)dx \qquad t > 0 \rightarrow \text{calcolata numericamente!!}$$

Per $-t$ possiamo usare la simmetria di φ calcolando:

SIMMETRIA

$$\Phi(-t) = P(Z \leq -t) = \int_{-\infty}^{-t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = P(Z \geq t) = P(Z > t) = 1 - P(Z \leq t)$$
$$= 1 - \Phi(t)$$

La densità della gaussiana è pari a $E(X) = 0$. Infine definiamo:

DEFINIZIONE - QUANTILE DI CODA DESTRA DI ORDINE α Si definisce quantile di coda destra di ordine α l'unico numero z_α tale che:

$$P(Z > z_\alpha) = \alpha \qquad 0 < \alpha < 1$$

Inoltre valgono:

- $P(Z > -z_\alpha) = 1 - \alpha \Rightarrow -z_\alpha = z_{1-\alpha}$
- $z_{0,5} = 0 \rightarrow \begin{cases} \alpha < 0 \Rightarrow z_\alpha > 0 \\ \alpha > 0 \Rightarrow z_\alpha < 0 \end{cases}$

2.6 Funzioni di v.a.a.c

Sia X una v.a.a.c con densità f_X, allora non è detto che $Y = g(X)$ sia anch'essa a.c. Vedremo che però $Y = X^2$ e $Y = aX + b$ $a, b \in \mathbb{R}$ $a \neq 0$ sono ancora a.c. Ma per esempio se $X \sim \mathcal{N}(0,1)$ e $Y := \mathbb{I}_{(0;+\infty)}(x)$ allora $Y \sim Be(p)$ e $p = P(Y=1) = P(X>0)$. Enunciamo dei criteri per determinare la continuità di Y.

2.6.1 Criteri per l'assoluta continuità

METODO DELLA F.D.R Sia $Y = g(X)$, $X \sim f_X$ densità a.c. Calcoliamo la f.d.r per ogni $t \in \mathbb{R}$:

$F_Y(t) := P(Y \le t) = P(g(X) \le t) = P(X \in g^{-1}\{(-\infty; t)\}) = \int_{g^{-1}\{(-\infty;t)\}} f_X(x)\,dx$

A questo punto se F_Y così ottenuta è continua ovunque e differenziabile con continuità tranne al più in un insieme finito $B = \{t_1, \dots, t_n\}$ allora Y è assolutamente continua con densità $f_Y(t) = F_Y'(t)$ $t \notin B$ e $f_Y(t)$ definita arbitrariamente per $t \in B$.

ESEMPIO Sia $Y = g(X) = X^2$, $g(x) = x$ $X \sim f_X$. Applichiamo il metodo della f.d.r.:

$$F_Y(t) = P(Y \le t) = P(X^2 \le t) = 0 \qquad t < 0$$
$$F_Y(0) = P(X^2 \le 0) = P(X^2 = 0) = P(X = 0) = 0 \quad \text{poiché } X \text{ a.c.}$$

Sia $t > 0$:
$$F_Y(t) = P(X^2 \le t) = P(-\sqrt{t} \le X \le \sqrt{t}) = \int_{-\sqrt{t}}^{\sqrt{t}} f_X(x)\,dx = F_X(\sqrt{t}) - F_X(-\sqrt{t})$$

Se $F_Y(t)$ rispetta i criteri di derivabilità/continuità allora Y a.c con $f_Y(t) = F_Y'(t) = f_X(\sqrt{t})\frac{1}{2\sqrt{t}} + f_X(-\sqrt{t})\frac{1}{2\sqrt{t}}$ per $t > 0$.

ESERCIZIO Sia $Z \sim N(0,1)$. Determinare la densità di $Y = Z^2$

Se $Z \sim N(0,1)$ allora Z è a.c con $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Avremo, quindi, che Y sarà a.c. con

$$f_Y(t) = \frac{1}{2\sqrt{t}}\left[\varphi(\sqrt{t}) + \varphi(-\sqrt{t})\right] = \frac{1}{\sqrt{t}}\varphi(\sqrt{t}) = \frac{1}{\sqrt{t\,2\pi}} e^{-\frac{t}{2}} \quad \text{per } t > 0.$$

Questa densità si chiama chi-quadro con 1 grado di libertà, ossia $Y \sim \chi^2(1)$

ESERCIZIO

1) Sia X a.c. con densità f_X. Mostrare che $Y = aX + b$ con $a, b \in \mathbb{R}$ $a \neq 0$ è una v.a assolutamente continua con densità $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$
2) Sia $X \sim \mathcal{E}(\lambda)$. Determinare e riconoscere la densità di

$Y = aX$ con $a > 0$

3) Sia $X \sim \mathcal{U}(a,b)$. Mostrare che $Y = \frac{X-a}{b-a} \sim \mathcal{U}(0,1)$

1) Calcoliamo la f.d.r di $Y = aX + b$:

$$F_Y(t) = P(Y \leq t) = P(aX + b \leq t) = \begin{cases} P\left(X \leq \frac{t-b}{a}\right) = F_X\left(\frac{t-b}{a}\right) & a > 0 \\ P\left(X \geq \frac{t-b}{a}\right) = 1 - F_X\left(\frac{t-b}{a}\right) & a < 0 \end{cases}$$

$$f_Y(t) = F_Y'(t) = \begin{cases} \frac{1}{a} f_X\left(\frac{t-b}{a}\right) & a > 0 \\ -\frac{1}{a} f_X\left(\frac{t-b}{a}\right) & a < 0 \end{cases}$$

Quindi per $a \neq 0$ Y è a.c con densità $f_Y = \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right)$

2. $X \sim \mathcal{E}(\lambda)$ e $Y = aX$ $a > 0$. Calcoliamone la densità:

$$f_Y(t) = \frac{1}{a} f_X\left(\frac{t}{a}\right) = \frac{\lambda}{a} e^{-\frac{\lambda}{a}t} I_{(0;+\infty)}(y) \implies Y \sim \mathcal{E}\left(\frac{\lambda}{a}\right)$$

3. Come esercizio

PROPOSIZIONE - CONDIZIONE SUFF. ASS. CONTINUITÀ $Y = g(X)$, X ASS. CONT. Sia X una v.a.a.c con densità f_X e sia Y una v.a funzione di X $Y = g(X)$. Supponiamo sia possibile determinare un intervallo aperto $S \subseteq \mathbb{R}$ tale che:

1) $P(X \in S) = 1$
2) g è differenziabile con continuità su S
3) la derivata di g non si annulla in S

Sia $g^{-1} : g(S) \to S$ la funzione inversa di g ristretta a S. Allora $Y = g(X)$ è ass. cont. con densità:

$$f_Y(y) = f_X(g^{-1}(y)) \, |(g^{-1})'(y)| \, I_{g(S)}(y)$$

ESERCIZIO Sia Y l'area di un cerchio con raggio lunghezza scelta a caso in $(0;1)$. Y è a.c.? Qual è la densità?

Sia X la lunghezza del raggio. $X \sim \mathcal{U}(0,1)$ con $f_X(t) = I_{(0,1)}(t)$
Avremo $Y = \pi X^2$. Se prendiamo $S = (0,1)$ abbiamo che $P(X \in S) = 1$, $g \in C^2(S)$ con $g' = 2\pi x \neq 0$ $\forall x \in S$. Le ipotesi della condizione suff.

sono soddisfatte e possiamo applicare la formula:

$$g(S) = (0,\pi) \quad , \quad g^{-1}(x) = \sqrt{\frac{x}{\pi}}$$

$$f_Y(y) = \mathbb{I}_{(0,1)}\left(\sqrt{y/\pi}\right) \frac{1}{2\sqrt{\pi y}} \mathbb{I}_{(0,\pi)}(y) = \frac{1}{2\sqrt{\pi y}} \mathbb{I}_{(0,\pi)}(y)$$

2.7 Media di funzioni di v.a

PROPOSIZIONE - MEDIA DI FUNZIONI DI V.A

1) Siano X una v.a.a.c con densità f_X e g una funzione tale che $Y := g(X)$ sia un v.a funzione di X. Allora se $\int_{\mathbb{R}} |g(x)| f(x)\, dx < +\infty$ allora Y ammette media pari a:

$$E(Y) = \int_{\mathbb{R}} g(x) f_X(x)\, dx$$

2) Siano X una v.a.d a valori in $S = \{x_k \ k \in I\}$, $I \subseteq \mathbb{Z}$ e densità p_X e g una funzione reale definita almeno su S e $Y := g(X)$. Se $\sum_{k\in I} |g(x_k)|\, p_X(x_k) < +\infty$, allora Y ammette media e

$$E(Y) = \sum_{k\in I} g(x_k)\, p_X(x_k)$$

PROPOSIZIONE – PROPRIETÀ DELLA MEDIA Sia X una v.a

1. Se $P(X=c)=1$ $c\in\mathbb{R}$ $\Rightarrow$ $E(X)=c$
2. Se $B \subseteq \mathbb{R}$ $Y = \mathbb{I}_{x\in B}(x)$ v.a. $\Rightarrow$ $E(Y) = P(X\in B)$
3. Se X ammette media e $\alpha,\beta\in\mathbb{R}$, allora $\alpha X+\beta$ ammette media e vale: $E(\alpha X+\beta) = \alpha E(X) + \beta$

DIMOSTRAZIONE

1. $p_X(c) = 1$ e $p_X(x) = 0$ $\forall x \neq c$ $\Rightarrow$ $E(X) = c \cdot p_X(c) = c$
2.

$$Y = \begin{cases} 1 & x\in B \\ 0 & x\notin B \end{cases} \Rightarrow Y \sim Be(p) \quad p = P(Y=1) = P(X\in B)$$

$$E(Y) = 1\cdot p + 0\cdot(1-p) = p = P(X\in B)$$

3. Dimostriamo la proposizione nel caso discreto. Per il caso ass. cont si vedano le dispense.

Se X ammette media $\Rightarrow \sum_{k\in I} |x_k| p_X(x_k) < +\infty$ dove $S = \{x_k \; k\in I\}$ è tale che $P(X\in S)=1$ e p_X è la densità di X. Sia $Y = \alpha X + \beta$

$$\sum_{k\in I} |g(x_k)| p_X(x_k) = \sum_{k\in I} |\alpha x_k + \beta| p_X(x_k) \leq$$
$$\sum_{k\in I} (|\alpha||x_k| + |\beta|) p_X(x_k) = |\alpha| \sum_{k\in I} |x_k| p_X(x_k) + |\beta| \sum_{k\in I} p_X(x_k)$$
$$= |\alpha| \sum_{k\in I} |x_k| p_X(x_k) + \beta < +\infty \Rightarrow Y \text{ ammette media}$$
$$E(Y) = \sum_{k\in I} (\alpha x_k + \beta) p_X(x_k) = \alpha \sum_{k\in I} x_k p_X(x_k) + \beta \sum_{k\in I} p_X(x_k) =$$
$$= \alpha E(X) + \beta$$

■

PROPOSIZIONE - ALTRE PROPRIETÀ DELLA MEDIA

4. Se X è una v.a e g e h sono funzioni tali che $g(X)$ e $h(X)$ ammettono media allora $E(g(X)+h(X)) = E(g(X)) + E(h(X))$.
5. Se X è una v.a tale che $P(X\geq 0)=1$ e X ammette media allora $E(X)\geq 0$. Se in aggiunta $E(X)=0$ allora $P(X=0)=1$
6. Se a,b sono numeri tali che $P(a\leq X\leq b)=1$ allora $a\leq E(X)\leq b$
7. Se $g(X)$ e $h(X)$ ammettono media e $P(g(X)\geq h(X))=1$ allora $E(g(X))\geq E(h(X))$

2.7.1 Varianza

Supponiamo di giocare a testa e croce con un amico. Indichiamo con X il guadagno netto puntando 1€ su T e Y il guadagno netto puntando 1000€ su T:

$$X = \begin{cases} 1 & 1/2 \\ -1 & 1/2 \end{cases} \qquad Y = \begin{cases} 1000 & 1/2 \\ -1000 & 1/2 \end{cases}$$

Osserviamo le medie: $E(X)=0$ e $E(Y)=0$. Entrambi sono, quindi, giochi equi. La distanza dei valori di Y dalla media è però molto maggiore rispetto a X. Questa distanza è la varianza.

DEFINIZIONE - VARIANZA Sia X una v.a. che ammette media. Se $(X-E(X))^2$ ammette media, allora diciamo che X ammette varianza e si definisce varianza di X il numero:

$$\operatorname{var}(X) := E\left((X-E(X))^2\right)$$

$\sqrt{var(x)}$ è detta deviazione standard di X. Usando le proprietà della media calcoliamo var(X) come:

$$var(x) = \begin{cases} \sum_{k \in I} (x_k - E(x))^2 p_x(x_k) & X \text{ v.a.d} \\ \int_{-\infty}^{+\infty} (x - E(x))^2 f_x(x)\, dx & X \text{ v.a.a.c} \end{cases}$$

Usando ancora le proprietà della media si ricavano quelle della varianza:

PROPOSIZIONE - PROPRIETÀ VARIANZA Sia X una v.a che ammette varianza

1. $var(X) \geq 0$
2. $var(X) = 0 \iff P(X=c) = 1$ e $E(X) = c$
3. $var(\alpha X + \beta) = \alpha^2 var(X) \quad \forall \alpha, \beta \in \mathbb{R}$
4. X^2 ammette media e $var(X) = E(X^2) - E(X)^2$

DIMOSTRAZIONE

1. $var(x) = E((x - E(x))^2) \geq 0$ per proprietà 5 della media
2. Se $P(X=c) = 1 \Rightarrow var(X) = E((X - E(X))^2) = E((c-c)^2) = 0$
   viceversa sia X tale che $0 = var(X) = E((X - E(X))^2)$
   $\Rightarrow P((X - E(X))^2 = 0) = 1 = P(X = E(X)) \Rightarrow P(X=c) = 1$
3. $var(\alpha X + \beta) = E((\alpha X + \beta - E(\alpha X + \beta))^2) = E((\alpha X + \cancel{\beta} - \alpha E(x) - \cancel{\beta})^2) =$
   $= E(\alpha^2 (X - E(X))^2) = \alpha^2 E((x - E(x))^2) = \alpha^2 var(X)$
4. $E(X^2) = E((X - E(X) + E(X))^2) \leq E((X - E(X))^2 + E(X)^2) \leq E((X - E(X))^2) + E(X)^2$
   $\leq +\infty \Rightarrow X^2$ ammette media
   $var(X) = E(X^2 + E(X)^2 - 2XE(X)) = E(X^2) + E(E(X)^2) + E(-2XE(X))$
   $= E(X^2) + E(X)^2 - 2E(X)E(X) = E(X^2) - E(X)^2$ ∎

ESEMPI

1. $X \sim Be(p) \quad f_X(x) = p^x (1-p)^{1-x} I_{(0;1)}(x) \quad E(X) = p$
   $E(X^2) = 1^2 p_X(1) + 0\, p_X(0) = p \quad var(X) = p(1-p)$

2. $X \sim Poiss(\lambda) \quad p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} I_{k \geq 0}(k) \quad E(X) = \lambda$
   $E(X^2) = \sum_{k=0}^{\infty} k^2 p_X(k) = \sum_{k=1}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{(k-1)!} =$
   $= \lambda e^{-\lambda} \left[\sum_{k=1}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right] = \lambda e^{-\lambda} \left[\underbrace{\sum_{j=0}^{\infty} j \frac{\lambda^j}{j!}}_{E(X)} + \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}}_{1} \right] =$
   $= \lambda^2 + \lambda$

$var(X) = E(X^2) - E(X) = \dots = \lambda$

3. $X \sim \mathcal{E}(\lambda)$ $\quad f_X(x) = \lambda e^{-\lambda x} I_{(0,+\infty)}(x)$ $\quad E(X) = 1/\lambda$
$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x)dx = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = \dots = 2/\lambda^2$
$var(X) = 1/\lambda^2$

4. $Z \sim \mathcal{N}(0,1)$ $\quad \varphi_x = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $\quad E(Z) = 0$ — indicati in $\mathcal{N}(0,1)$
$var(Z) = E(X^2) = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \dots = 1$

ESERCIZIO Mostrare che
1) Se $X \sim Bi(n,p)$ allora $var(X) = np(1-p)$
2) Se $X \sim Geom(p)$ allora $var(X) = 1-p/p^2$
3) Se $X \sim U(a,b)$ allora $var(X) = (b-a)^2/12$

2.7.2 Standardizzazione

ESEMPIO Consideriamo la misurazione di una grandezza fisica μ. La misuriamo con uno strumento che non ha errori sistematici e con precisione $1/\sigma$. Definiamo $X :=$ "misurazione di μ" $= \mu + \varepsilon$ con ε l'errore di misura. ε è una v.a con $E(\varepsilon) = 0$ e $\frac{1}{\sqrt{var(\varepsilon)}} = \frac{1}{\sigma}$ cioè $var(\varepsilon) = \sigma^2$. Otteniamo che $E(X) = \mu$ e $var(X) = var(\mu + \varepsilon) = \sigma^2$. Supponiamo di poter dire che $\varepsilon = \sigma Z$ con $Z \sim \mathcal{N}(0,1)$ con $E(\varepsilon) = 0$ e $var(\varepsilon) = \sigma^2$. Possiamo modellizzare X come $X = \mu + \sigma Z$. Qual è la densità di X?

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

se $\mu = 0, \sigma = 1 \Rightarrow X = Z$, $f_X(x) = \varphi$. La trasformazione affine della gaussiana standard prende il nome di gaussiana.

DEFINIZIONE - GAUSSIANA Una v.a.a.c è detta gaussiana di media μ e varianza σ^2 se ha densità

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

e scriveremo $X \sim \mathcal{N}(\mu, \sigma^2)$.

Nell'esempio abbiamo mostrato che se $X = \mu + \sigma Z$ con $Z \sim \mathcal{N}(0,1)$ allora $X \sim \mathcal{N}(\mu, \sigma^2)$. Dimostriamo ora il viceversa.

$$X \sim \mathcal{N}(\mu, \sigma) \Rightarrow Y = \frac{X-\mu}{\sigma} \text{ è anch'essa una trasformazione affine}$$
$$\Rightarrow f_Y(y) = \sigma f_X(\sigma y + \mu) = \dots = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \Rightarrow Y \sim \mathcal{N}(0,1)$$

L'operazione $Y = \frac{X - E(X)}{\sqrt{var(X)}}$ è detta standardizzazione.

DEFINIZIONE - STANDARDIZZATA Sia X una v.a. che ammette media e varianza. Indichiamo con μ la media di X e con σ^2 la varianza di X. Supponiamo $\sigma^2 > 0$. Si chiama standardizzata di X la v.a. Y così definita:

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

Nota bene: $E(Y) = 0$ e $var(Y) = 1$. Inoltre $F_X(t) = F_Y\left(\frac{t-\mu}{\sigma}\right)$.

ESERCIZIO Determinare per $\alpha \in (0,1)$ il quantile q_α di ordine α di una densità $\mathcal{N}(\mu, \sigma^2)$.

$$P(X \geq q_\alpha) = P\left(\frac{X-\mu}{\sigma} \geq \frac{q_\alpha - \mu}{\sigma}\right) = P\left(Z \geq \frac{q_\alpha - \mu}{\sigma}\right) = \alpha$$
$$\Leftrightarrow \frac{q_\alpha - \mu}{\sigma} = z_\alpha \text{ è un quantile di una gaussiana standard}$$
$$\Leftrightarrow q_\alpha = \sigma z_\alpha + \mu$$

2.7.3 Varianza come indice di dispersione

PROPOSIZIONE - DISUGUAGLIANZA DI CHEBYSHEV Sia X una v.a. che ammette media e varianza. Allora vale:

$$\forall \varepsilon > 0 \quad P(|X - E(X)| > \varepsilon) \leq \frac{var(X)}{\varepsilon^2}$$

La disuguaglianza di Chebyshev è un caso particolare della disuguaglianza di Markov:

PROPOSIZIONE - DISUGUAGLIANZA DI MARKOV Sia Y una v.a positiva tale che Y^K ammette media per K intero positivo.

Allora vale:

$$\forall \varepsilon > 0 \quad P(Y > \varepsilon) \leq \frac{E(Y^K)}{\varepsilon^K}$$

DIMOSTRAZIONE (Disuguaglianza di Markov)

$$Y^K \geq Y^K I_{Y>\varepsilon} \Rightarrow E(Y^K) \geq E(Y^K I_{Y>\varepsilon})$$
$$\geq \varepsilon^K E(I_{Y>\varepsilon}) = \varepsilon^K P(Y > \varepsilon)$$
$$\Rightarrow P(Y > \varepsilon) \leq E(Y^K)/\varepsilon^K$$

■

2.7.4 Momenti e funzioni generatrici di momenti

DEFINIZIONE - MOMENTO K-ESIMO Sia X una v.a tale che $|X|^K$ ammette media. Allora $E(X^K)$ è detto momento K-esimo, o momento di ordine K, della v.a X

PROPOSIZIONE - NUMERO DI MOMENTI Se X è una v.a che ammette momento K-esimo per qualche $K \geq 2$, allora X ammette anche momento h-esimo per ogni h con $1 < h < K$

ESEMPIO Sia X una v.a.a.c con densità $f_X(x) = 5x^{-6} I_{(1;+\infty)}(x)$

$$E(X^4) = \int_{-\infty}^{+\infty} x^4 f_X(x)\,dx = \int_1^{+\infty} x^4\, 5x^{-6}\,dx = \dots = 5$$
$$\Rightarrow X \text{ ammette momento } 4$$

Per la proposizione precedente esisteranno anche $E(x)$, $E(x^2)$ e $E(x^4)$

DEFINIZIONE - FUNZIONI GENERATRICI DI MOMENTI Sia X una v.a per la quale esiste un intervallo aperto O contenente lo 0 tale che per ogni $t \in O$ e^{tX} ammette media allora la funzione

$$m_X(t) := E(e^{tX})$$

definita per almeno $t \in O$ è detta funzione generatrice di momenti.

PROPOSIZIONE - LEGAME MEDIA-MOMENTO Se X ammette f.g.m allora esistono i momenti di ordine K di X, per ogni intero K, e vale:

$$E(X) = m_X'(0), \quad E(X^2) = m_X''(0), \quad \dots, \quad E(X^K) = m_X^{(K)}(0)$$

DIMOSTRAZIONE PARZIALE

$$m_X'(t) = E(e^{tX})' = E((e^{tX})') = E(Xe^{tX}) \to m_X'(0) = E(X)$$
$$m_X''(t) = E(Xe^{tX})'' = \dots = E(X^2 e^{tX}) \to m_X''(0) = E(X^2)$$
$$\vdots$$
$$m_X^{(K)}(t) = E(X^K e^{tX}) \to m_X^{(K)}(0) = E(X^K)$$

ESERCIZIO Calcolare se esiste la f.g.m di
1) $X \sim \text{Poiss}(\lambda)$
2) $Z \sim \mathcal{N}(0,1)$
3) Dedurre dal punto precedente la f.g.m di $Y \sim \mathcal{N}(\mu, \sigma^2)$

1) $X \sim \text{Poiss}(\lambda) \Rightarrow p_X(K) = e^{-\lambda}\frac{\lambda^K}{K!} I_{\{0,1,\dots\}}(K)$

$$m_X = E(e^{tX}) = \sum_{K=0}^{+\infty} e^{tK} e^{-\lambda}\frac{\lambda^K}{K!} = e^{-\lambda}\sum_{K=0}^{+\infty}\frac{(e^t\lambda)^K}{K!} = e^{-\lambda}e^{e^t\lambda}$$
$$= e^{-\lambda(1-e^t)}$$

2) $Z \sim \mathcal{N}(0,1) \Rightarrow$ densità $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$$m_Z = \int_{-\infty}^{+\infty} e^{tx}\varphi(x)\,dx = \int_{-\infty}^{+\infty} e^{tx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx = \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-t)^2+\frac{t^2}{2}}\,dx$$
$$= e^{t^2/2}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-t)^2}\,dx = e^{t^2/2}$$

3) $Y \sim \mathcal{N}(\mu,\sigma^2) \iff \frac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1) \iff Y = \sigma Z + \mu$

$$m_Y = E(e^{(\sigma Z+\mu)t}) = e^{\mu t}E(e^{t\sigma Z}) = e^{\mu t}m_Z(t\sigma) = e^{\mu t}e^{\frac{t^2\sigma^2}{2}}$$

Osservazione Se X e Y sono due v.a che ammettono f.g.m e hanno la stessa densità, allora le f.g.m coincidono $\forall t$ per le quindi sono definite. Vale anche il viceversa!

PROPOSIZIONE - LEGAME MOMENTO-FDR Siano X e Y v.a che ammettono f.g.m m_X e m_Y rispettivamente e siano F_X e F_Y le f.d.r di X e Y rispettivamente. Allora $F_X = F_Y \iff m_X = m_Y$

3 VETTORI ALEATORI

Introduciamo l'idea di vettore aleatorio con il seguente esempio:

ESEMPIO Lanciamo 10 volte 2 monete eque. Possiamo considerare diverse v.a:

X := "numero di teste della 1° moneta" $\rightarrow$ $X \sim Bi(10, \frac{1}{2})$
Y := "numero di teste della 2ª moneta" $\rightarrow$ $Y \sim Bi(10, \frac{1}{2})$
Z := "numero di teste totali" $\rightarrow$ $Z \sim Bi(20, \frac{1}{2})$

Consideriamo la seguente probabilità:

$$P(X=3, Y\leq 5) = P(\{w: X(w)=3\} \cap \{w: Y(w)\leq 5\}) \overset{\text{INDIPENDENTI}}{=} P(X=3)P(X\leq 5)$$

Cosa accadrebbe se volessimo calcolare

$$P(X=3, Z\leq 5) = P(\{w: X(w)=3\} \cap \{w: Z(w)\leq 5\}) \overset{\text{NON INDIPENDENTI}}{\neq} P(X=3)(Z\leq 5)$$
$$= P(X=3, Y\leq 2) = P(X=3)P(Y\leq 2) = \cdots \neq P(X=3)(Z\leq 5)$$

Diamo della notazione: $B_1, \dots, B_n$ $B_i \subseteq \mathbb{R}$ $i=1..n$ e $X_1, \dots, X_n$ delle v.a definite nello stesso $(\Omega, \mathcal{F}, P)$. Diciamo che:

$$\{X_1 \in B_1, \dots, X_n \in B_n\} := \{w: X_1(w) \in B_1\} \cap \cdots \cap \{w: X_n(w) \in B_n\}$$

DEFINIZIONE - VETTORE ALEATORIO Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità. Un vettore aleatorio n-dimensionale è una funzione $\vec{X} = (X_1, \dots, X_n): \Omega \rightarrow \mathbb{R}^n$ tale che ciascuna X_j $j=1..n$ è una v.a. su $(\Omega, \mathcal{F}, P)$.

DEFINIZIONE - F. DI RIPARTIZIONE Sia $\vec{X} = (X_1, \dots, X_n)$ un vettore aleatorio n-dimensionale su uno spazio $(\Omega, \mathcal{F}, P)$. Si chiama funzione di ripartizione di $\vec{X}$ la funzione $F_{\vec{X}}: \mathbb{R}^n \to [0,1]$ definita da

$$F_{\vec{X}}(t_1, \dots, t_n) := P(X_1 \leq t_1, \dots, X_n \leq t_n)$$

DEFINIZIONE - V.A. INDIPENDENTI Siano $X_1, \dots, X_n$ n v.a definite su un medesimo $(\Omega, \mathcal{F}, P)$. Si dicono indipendenti se per ogni $B_1, \dots, B_n \subseteq \mathbb{R}$ vale

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{j=1}^{n} P(X_j \in B_j)$$

Osservazione Se $X_1, \dots, X_n$ sono indipendenti, allora se applichiamo la definizione a $B_j = (-\infty; t_j]$ otteniamo

$$P(X_1 \leq t, \dots, X_n \leq t_n) = F_{\vec{X}}(t_1, \dots, t_n) = \prod_{j=1}^{n} P(X_j \leq t_j) = \prod_{j=1}^{n} F_{X_j}(t_j)$$

PROPOSIZIONE - CNS DI INDIPENDENZA Le componenti di un vettore aleatorio sono v.a. indipendenti se e solo se la f.d.r $F_{\vec{X}}$ coincide con il prodotto delle f.d.r marginali F_{X_j}:

$$F_X(t_1, \dots, t_n) = \prod_{j=1}^{n} F_{X_j}(t_j) \quad \forall t_1, \dots, t_n \in \mathbb{R}^n$$

PROPOSIZIONE - FDR MARGINALE DA FDR Sia $\vec{X}$ un vettore aleatorio con f.d.r $F_{\vec{X}}$. Allora per $j = 1..n$ vale:

$$\forall t \quad F_{X_j} = \lim_{t_k \to +\infty, k \neq j} F_{\vec{X}}(t_1, \dots, t_{k-1}, t, t_{k+1}, \dots, t_n)$$

3.1 Vettori aleatori discreti

DEFINIZIONE - VETT. A. DISCRETO Un vettore aleatorio $\vec{X}$ si dice discreto se le sue componenti X_i sono v.a discrete.

Osservazione Se $\vec{X}$ discreto, allora assumerà valori in $S \subseteq S_1 \times S_2 \times \dots \times S_n$. Poiché S_i è al più numerabile, allora anche S sarà al più numerabile.

DEFINIZIONE - DENSITÀ DISCRETA Sia $\vec{X}$ un vettore a.d.. La funzione $p_{\vec{X}} : \mathbb{R} \to [0,1]$ definita per ogni $\vec{x} \in \mathbb{R}^n$ da:

$$p_{\vec{X}}(\vec{x}) := P(\vec{X} = \vec{x})$$

è detta densità discreta di $\vec{X}$ o anche densità congiunta di $X_1, \dots, X_n$.

__Osservazione__ Poiché $\vec{X}$ ha valori in S numerabile, la densità $p_{\vec{X}}(\vec{x}) = 0$ per un numero numerabile di elementi.

PROPOSIZIONE - PROPRIETÀ DENSITÀ Sia $p_{\vec{X}}$ la densità di un vettore aleatorio n-dimensionale discreto a valori in S al più numerabile allora:

1) $0 \le p_{\vec{X}}(\vec{x}) \le 1$ e $p_{\vec{X}}(\vec{x}) = 0$ se $\vec{x} \notin S$
2) $\sum_{\vec{x} \in S} p_{\vec{X}}(\vec{x}) = 1$
3) Se $B \subseteq \mathbb{R}^n$ allora $P(\vec{X} \in B) = \sum_{\vec{x} \in B \cap S} p_{\vec{X}}(\vec{x})$

Poniamoci ora le seguenti domande:

1) È possibile calcolare le densità marginali p_{X_i} a partire dalla densità di $\vec{X}$?

$$p_{X_1}(x) = P(X_1 = x) = P(X_1 = x, X_2 \in \mathbb{R}, \dots, X_n \in \mathbb{R}) = P(X \in \overbrace{\{x\} \times \mathbb{R} \times \dots \times \mathbb{R}}^{B})$$

$$= \sum_{\vec{x} \in B \cap S} p_{\vec{X}}(\vec{x}) = \sum_{\substack{(x_2, \dots, x_n): \\ (x, x_2, \dots x_n) \in S}} p_{\vec{X}}(x, x_2, \dots, x_n) \leftarrow \text{MARGINALIZZAZIONE}$$

2) Viceversa, se conosco p_{X_i} per ogni $i = 1..n$, posso calcolare la densità di $\vec{X}$? È possibile farlo quando so che le v.a X_i sono indipendenti. Vale, infatti, la proposizione:

PROPOSIZIONE - COSTRUZIONE DENSITÀ CONGIUNTA Le componenti di vettore a.d. $\vec{X}$ sono v.a. indipendenti se e solo se $p_{\vec{X}}$ si può calcolare come $p_{\vec{X}}(\vec{x}) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n)$

ESERCIZIO In una scatola vi sono 6 pile di cui 3 nuove, 2 parzialmente usate e 1 completamente esaurita. Se ne scelgono 3 a caso senza reimmissione. Sia X la v.a che conta il numero di pile nuove e Y quella che conta il numero di pile parzialmente usate tra quelle estratte.

1) Determinare la densità congiunta di (X,Y) e le rispettive densità marginali
2) X e Y sono indipendenti?
3) Quanto vale la probabilità che tra le pile scelte non ci sia quella esaurita?

1) $X :=$ n° pile nuove $\in \{0,1,2,3\} = S_1$
$Y :=$ n° delle pile parzialmente usate $\in \{0,1,2\} = S_2$
$\vec{X} = (X,Y) \in S \subseteq S_1 \times S_2$

$P_{\vec{X}}(0,0) = P_{\vec{X}}(0,1) = P_{\vec{X}}(1,0) = 0$

$$P_{\vec{X}}(0,2) = \frac{\binom{3}{0}\binom{2}{2}\binom{1}{1}}{\binom{6}{3}} = 1/20$$

$$P_{\vec{X}}(1,2) = \frac{\binom{3}{1}\binom{2}{2}\binom{1}{0}}{\binom{6}{3}} = 3/20$$

$$P_{\vec{X}}(1,1) = \frac{\binom{3}{1}\binom{2}{1}\binom{1}{1}}{\binom{6}{3}} = 6/20$$

...

X \ Y	0	1	2	P_X
0	0	0	1/20	1/20
1	0	6/20	3/20	9/20
2	3/20	6/20	0	9/20
3	1/20	0	0	1/20
P_Y	4/20	12/20	4/20	1

2) X e Y sono indipendenti se e solo se $P_{\vec{X}}(x,y) = P_X(x)P_Y(y)$.
Ciò è falso perché $0 = P_{\vec{X}}(3,1) \neq P_X(3)P_Y(1) = 12/144$

3) $P(X+Y=3) = P((X,Y) \in \{(3,0),(2,1),(1,2)\}) = \sum_{(x,y)\in B} P_{\vec{X}}(x,y) =$
$P_{\vec{X}}(3,0) + P_{\vec{X}}(2,1) + P_{\vec{X}}(1,2) = 1/2$

Possiamo notare che $X+Y=3$ equivale a $g(X,Y)=3$ con g una funzione del vettore a. e ho calcolato il valore della sua densità in 3.

3.2 Funzioni di vettori aleatori discreti

Abbiamo $\vec{X}$ un vettore a.d a valori in $S \subseteq \mathbb{R}^n$ al più numerabile e con densità $p_{\vec{X}}$. Consideriamo la funzione $\vec{g} : (g_1, \dots, g_m) : \mathbb{R}^n \to \mathbb{R}^m$ tale che $\vec{Y} = (Y_1, \dots, Y_m) = \vec{g}(\vec{X})$. Che proprietà ha $\vec{Y}$? Innanzitutto:

$$P(\vec{X} \in S) = 1 \Rightarrow P(\vec{Y} \in \vec{g}(S)) = P(\vec{g}(\vec{X}) \in \vec{g}(S)) = 1$$

E poiché $\vec{g}(S)$ rimane al più numerabile $\vec{Y}$ è un vettore a.d. Quindi varrà $p_{\vec{Y}}(\vec{y}) = P_{\vec{Y}}(\vec{Y} = \vec{y}) = 0$ per $\vec{y} \notin \vec{g}(S)$. Se invece $\vec{y} \in \vec{g}(S)$ allora:

$$p_{\vec{Y}}(\vec{y}) = P(g(\vec{x}) = \vec{y}) = P(\vec{X} \in \vec{g}^{-1}\{\vec{y}\}) = \sum_{\vec{x} \in S \cap \vec{g}^{-1}\{\vec{y}\}} p_{\vec{X}}(\vec{x}) =$$
$$= \sum_{\vec{x} \in S:\ \vec{g}(\vec{x}) = \vec{y}} p_{\vec{X}}(\vec{x})$$

ESEMPIO Consideriamo il vettore aleatorio discreto $\vec{X} \sim p_{\vec{X}}$ con $P(\vec{X} \in S) = 1$. Analizziamo uno a uno i seguenti casi:

1) $Y = X_1 + X_2$ con $\vec{X} = (X_1, X_2)$. Sia S al più che numerabile, allora $g(S) = \{x_1 + x_2 \quad (x_1, x_2) \in \}$ è anch'esso numerabile e sarà $P(Y = y) = 0$ per $y \notin g(S)$. Avremo che la densità di Y sarà:

$$p_Y = \sum_{\vec{x} \in S:\ x_1 + x_2 = y} p_{\vec{X}}(x_1, x_2) = \sum_{x_1 : (x_1, x_1 - y) \in S} p_{\vec{X}}(x_1, x_1 - y)$$

detta convoluzione discreta.

2) $\vec{Y} = A\vec{X} + \vec{b}$ con $A \in Mat(n,n)$ con $\det A \neq 0$. Possiamo ricavare $p_{\vec{Y}}$ facendo

$$p_{\vec{Y}}(\vec{y}) = P(\vec{Y} = \vec{y}) = P(A\vec{X} + \vec{b} = \vec{y}) = P(\vec{X} = A^{-1}(\vec{y} - \vec{b})) =$$
$$= p_{\vec{X}}(A^{-1}(\vec{y} - \vec{b}))$$

3) Consideriamo il caso speciale di 2) in cui abbiamo

$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ con $\vec{b} = 0$. Otteniamo così che $\vec{y} = \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix}$.
Applicando la formula trovata in 2) otteniamo:

$$A^{-1}\vec{y} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \frac{y_1 + y_2}{2} \\ \frac{y_1 - y_2}{2} \end{pmatrix} \Rightarrow p_{\vec{y}}(\vec{y}) = p_{\vec{x}}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right)$$

ESERCIZIO Siano X_1 e X_2 v.a indipendenti con $X_i \sim Poiss(\lambda_i)$ $i = 1,2$ $(\lambda_i > 0)$. Determinare la densità di $Y = X_1 + X_2$

$$\begin{aligned} p_Y(y) &= P(X_1 + X_2 = y) = \sum_{K:(K,y-K)\in S} p_{X_1,X_2}(K, y-K) = \\ &= \sum_{K:(K,y-K)\in S} p_{X_1}(K)\, p_{X_2}(y-K) = \sum_{K:(K,y-K)\in S} e^{-\lambda_1}\frac{\lambda_1^K}{K!} e^{-\lambda_2}\frac{\lambda_2^{y-K}}{(y-K)!} = \\ &= e^{-(\lambda_1+\lambda_2)}/y! \sum_{K:(K,y-K)\in S} y!/K!(y-K)!\ \lambda_1^K \lambda_2^{y-K} = \\ &= e^{-(\lambda_1+\lambda_2)}/y! \sum_{K=0}^{y} \binom{y}{K} \lambda_1^K \lambda_2^{y-K} = e^{-(\lambda_1+\lambda_2)}/y!\ (\lambda_1 + \lambda_2)^y \end{aligned}$$

$Y \in g(S) = \mathbb{N}_0 \Rightarrow$ $p_Y(y) = 0$ $y \notin g(S)$

COROLLARIO - RIPRODUCIBILITÀ DI POISSON Se $X_1, \dots, X_n$ sono n v.a. definite sul medesimo $(\Omega, \mathcal{F}, P)$, indipendenti e tali che $X_i \sim Poiss(\lambda_i)$ $i = 1..n$, allora $X_1 + \dots + X_n \sim Poiss(\lambda_1 + \dots + \lambda_n)$

3.3 Vettori aleatori assolutamente continui

DEFINIZIONE - VETTORE ALEATORIO ASS. CONT. Un vettore aleatorio $\vec{X} = (X_1, \dots, X_n)$ è detto assolutamente continuo se esiste una funzione $f_{\vec{X}}: \mathbb{R}^n \to \mathbb{R}_+$ integrabile tale che la f.d.r di $\vec{X}$ $F_{\vec{X}}$ si può scrivere come:

$$F_{\vec{X}}(t_1, \dots, t_n) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} f_{\vec{X}}(x_1, \dots, x_n)\, dx_1 \cdots dx_n$$

per ogni $(t_1, \dots, t_n) \in \mathbb{R}^n$; $f_{\vec{X}}$ prende il nome di densità del vettore aleatorio $\vec{X}$ (o densità congiunta).

PROPOSIZIONE - PROPRIETÀ VETT. A.C Sia $f_{\vec{X}}$ una densità di un vettore aleatorio assolutamente continuo $\vec{X}$. Allora:

- $\int_{\mathbb{R}^n} f_{\vec{X}}\, dx_1 \dots dx_n = 1$
- Se $F_{\vec{X}}$ è la fdr di $\vec{X}$, allora $\frac{\partial^n}{\partial x_1 \dots \partial x_n} F_{\vec{X}} = f_{\vec{X}}(\vec{x})$ per ogni $\vec{x} \in \mathbb{R}$

per cui esiste tale derivata parziale
- Se $B \subset \mathbb{R}^n$ è un dominio regolare allora $P(\vec{x} \in B) = \int_B f_{\vec{x}} \, dx_1 \dots dx_n$

Se $\vec{x}$ è un vettore aleatorio a.c le componenti X_i sono assolutamente continue? Se sì, come le calcolo? Per rispondere a questa domanda consideriamo $\vec{x}$ un vettore aleatorio a.c. Mostriamo che ogni sua componente è una v.a.a.c:

$i=1$: $F_{X_1} := P(X_1 \le t) = P(X_1 \le t, X_2 \in \mathbb{R}, \dots, X \in \mathbb{R}) =$

$$= P((X_1, \dots, X_n) \in \underbrace{\boxed{(-\infty, t] \times \mathbb{R}^{n-1}}}_{B \subset \mathbb{R}^n})$$

$$= \int_{-\infty}^{t} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\vec{x}}(x_1, \dots, x_n) \, dx_1 \cdots dx_n =$$

$$= \int_{-\infty}^{t} \left(\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \dots, x_n) \, dx_2 \cdots dx_n \right) dx_1 =$$

$$= \int_{-\infty}^{t} g(x_1) \, dx_1 \implies X_1 \text{ è a.c con densità } g(x_1)$$

$\vdots$

Questa è la dimostrazione della proposizione:

PROPOSIZIONE - CONTINUITÀ DELLE COMPONENTI Sia $\vec{x}$ un vettore aleatorio a.c di densità $f_{\vec{x}}$. Allora per ogni $i = 1..n$ X_i è una v.a.a.c e

$$f_{X_i}(x) = \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n-1 \text{ volte}} f_{\vec{x}}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) \, dx_1 \cdots dx_{i-1} \, dx_{i+1} \cdots dx_n$$

Attenzione: il viceversa in generale non vale!

Se $\vec{x}$ è, invece, un vettore le cui componenti X_i sono v.a.a.c., sotto quali condizioni posso affermare la continuità del vettore $\vec{x}$ formato dagli X_i? Come nel caso discreto è l'indipendenza:

PROPOSIZIONE - INDIPENDENZA Sia $\vec{x}$ un vettore aleatorio tale che le componenti sono v.a.a.c e indipendenti. Allora anche $\vec{x}$ è a.c. e la densità di $\vec{x}$ è:

$$f_{\vec{x}}(x_1, \dots, x_n) = \prod_{j=1}^{n} f_{X_j}(x_j)$$

Con f_{X_j} la densità marginale di X_j. Viceversa se $\vec{X}$ è assolutamente continuo di densità $f_{\vec{X}}$ pari a $\prod_{j=1}^{n} f_{X_j}(x_j)$, f_{X_j} densità marginale di X_j, allora $X_1, \dots, X_n$ sono v.a. indipendenti.

3.4 Funzioni di vettori aleatori assolutamente continui

Siano $\vec{X}$ un vettore aleatorio a.c. con densità $f_{\vec{X}}$, la funzione $\vec{g}: \mathbb{R}^n \to \mathbb{R}^m$ e la variabile aleatoria $\vec{Y} = \vec{g}(\vec{X})$. Per semplificare la trattazione considereremo 3 esempi:

1. $n=2$, $m=1$ $\quad g(x_1, x_2) = x_1 + x_2 \implies Y = X_1 + X_2$
2. $n=m$, A matrice $n \times n$ non singolare, $\vec{b} = (b_1, \dots, b_n)^T \in \mathbb{R}^n$

   $\implies \vec{Y} = (Y_1, \dots, Y_n)^T = A\vec{X} + \vec{b}$ con $\vec{X} = (X_1, \dots, X_n)^T$ e $\vec{g}(\vec{x}) = A\vec{x} + \vec{b}$

3. caso particolare del punto precedente in cui $n=m=2$ e $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ con $\vec{b} = 0$.

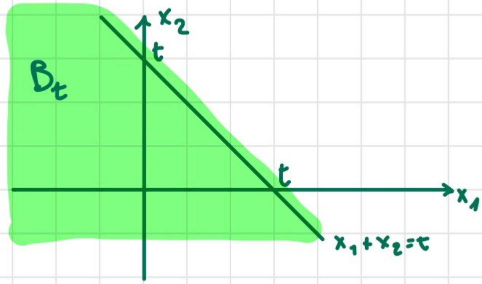
Il secondo ed il terzo esempio sono analoghi al caso discreto e otteniamo rispettivamente:

$$f_{\vec{Y}}(\vec{y}) = \frac{f_{\vec{X}}(A^{-1}(\vec{y} - \vec{b}))}{|\det A|} \quad ; \quad f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}\left(\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}\right)$$

Per quanto riguarda il primo punto, lo affrontiamo insieme ad altri nel prossimi esempi.

ESEMPIO Sia $\vec{X} = (X_1, X_2) \sim f_{\vec{X}}$ e $Y = X_1 + X_2$. Mostriamo che Y è a.c e determiniamone la densità. Calcoliamo la f.d.r:

$$F_Y(t) := P(Y \le t) = P(X_1 + X_2 \le t) = P((X_1, X_2) \in B_t) \qquad B_t = \{x_1, x_2 : x_1 + x_2 < t\}$$
$$= \int_{B_t} f_{\vec{X}}(x_1, x_2)\, dx_1 dx_2 = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{t-x_1} f_{\vec{X}}(x_1, x_2)\, dx_2 \right) dx_1 =$$
$$= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{t} f_{\vec{X}}(x_1, u - x_1)\, du \right) dx_1 = \int_{-\infty}^{t} \left(\int_{-\infty}^{+\infty} f_{\vec{X}}(x_1, u - x_1)\, dx_1 \right) du =$$
$$= \int_{-\infty}^{t} g(u)\, du$$



Per definizione di v.a.a.c, possiamo concludere che Y è a.c. con densità

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{\vec{X}}(x, y-x)\, dx$$

Se poi X_1, X_2 indipendenti possiamo scrivere che:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(x_1)\, f_{X_2}(y-x)\, dx \leftarrow \text{convoluzione continua}$$

ESEMPIO Sia $\vec{X}$ un vettore aleatorio continuo (o discreto) e siano $V = \max(X_1, \dots, X_n)$ e $W = \min(X_1, \dots, X_n)$. Calcoliamo le f.d.r di V e W.

1) $$F_V(t) := P(V \leq t) = P(\max(X_1, \dots, X_n) \leq t) = P(X_1 \leq t, \dots, X_n \leq t) =$$
$$= F_{\vec{X}}(t, \dots, t)$$

Se consideriamo $X_1, \dots, X_n$ indipendenti e identicamente distribuite (i.i.d.) allora avremo:

$$F_V(t) = F_{X_1}(t)^n$$

2) $$F_W(t) := P(W \leq t) = P(\min(X_1, \dots, X_n) \leq t) =$$
$$= 1 - P(\min(X_1, \dots, X_n) \leq t) = 1 - \underbrace{P(X_1 > t, \dots, X_n > t)}_{\neq 1 - F_{\vec{X}}(\cdot)}$$

Se le v.a. sono i.i.d possiamo ridurre il tutto a:

$$F_W(t) = 1 - (1 - F_{X_1}(x))^n$$

ESERCIZIO Sia $\vec{X} = \binom{X_1}{X_2}$ un vettore aleatorio con componenti $X_i \sim \mathcal{N}(0,1)$ per $i=1,2$ indipendenti. Mostrare che $\vec{Y} = \binom{X_1+X_2}{X_1-X_2}$ è un vettore aleatorio le cui componenti $Y_1 = X_1+X_2$ e $Y_2 = X_1 - X_2$ sono ancora indipendenti e $Y_i \sim \mathcal{N}(0,2)$ $i=1,2$

L'indipendenza delle componenti di $\vec{x}$ implica che

$$f_{\vec{X}}(x_1,x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)}$$

Usando la formula che avevamo calcolato troviamo:

$$f_{\vec{Y}}(y_1,y_2) = \frac{1}{2} f_{\vec{X}}\left(\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}\right) = \frac{1}{4\pi} e^{-\frac{1}{2}\left[\left(\frac{y_1+y_2}{2}\right)^2+\left(\frac{y_1-y_2}{2}\right)^2\right]} =$$

$$\text{exp}: \quad \left(\frac{y_1+y_2}{2}\right)^2 + \left(\frac{y_1-y_2}{2}\right) = \dots = \frac{1}{2}(y_1^2+y_2^2)$$

$$= \frac{1}{4\pi} e^{-\frac{1}{4}(y_1^2+y_2^2)} = \frac{1}{\sqrt{2\cdot 2\pi}} e^{-\frac{1}{2\cdot 2}y_1^2} \cdot \frac{1}{\sqrt{2\cdot 2\pi}} e^{-\frac{1}{2\cdot 2}y_2^2}$$

Quindi $Y_i \sim \mathcal{N}(0,2)$ indipendenti.

ESERCIZIO L'attesa per parlare con un operatore di un centralino è modellabile con un tempo esponenziale di media 2 minuti. Se per accedere ad un servizio devo parlare con 2 operatori (in sequenza) qual è la distribuzione del tempo complessivo attesa supponendo le attese indipendenti?

Dobbiamo calcolare $Y = X_1 + X_2$. Per l'esempio precedente sappiamo che

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(x) f_{X_2}(y-x)\,dx = \int_{-\infty}^{+\infty} \frac{1}{2} e^{-\frac{1}{2}x} I_{(0,+\infty)}(x)\, \frac{1}{2} e^{-\frac{1}{2}(y-x)} I_{(0,+\infty)}(y-x)\,dx$$
$$= \int_{-\infty}^{+\infty} \frac{1}{2} e^{-\frac{1}{2}x} \cdot \frac{1}{2} e^{-\frac{1}{2}(y-x)} I_{(0,+\infty)}(y-x)$$

se $y \leq 0$ $\quad f_Y(y) = 0$

se $y > 0$ $\quad f_Y(y) = \int_0^y \frac{1}{4} e^{-\frac{1}{2}y}\,dx = \underbrace{\frac{1}{4} y e^{-\frac{1}{2}y} I_{(0,+\infty)}(y)}_{\Gamma(2,\frac{1}{2})}$

3.5 Media e varianza di vettori aleatori

PROPOSIZIONE - MEDIA DI FUNZ. DI VETT. A. Siano $\vec{X}$ un vettore aleatorio n-dimensionale, una funzione $g(\vec{x}): \mathbb{R}^n \to \mathbb{R}$ e $Y := g(\vec{X})$ una v.a.

1) Sia $\vec{X}$ un vettore aleatorio discreto a valori in S e densità $p_{\vec{X}}$. Se $\sum_{\vec{x} \in S} |g(\vec{x})| \, p_{\vec{X}}(x) < +\infty$, allora $Y := g(\vec{X})$ ammette media pari a

$$E(Y) = \sum_{\vec{x} \in S} g(\vec{x}) \, p_{\vec{X}}(\vec{x})$$

2) Sia $\vec{X}$ un vettore aleatorio continuo con densità $f_{\vec{X}}$. Se $\int_{\mathbb{R}^n} |g(\vec{x})| \, f_{\vec{X}}(\vec{x}) \, dx_1 \cdots dx_n < +\infty$ allora $Y := g(\vec{X})$ ammette media pari a:

$$E(Y) = \int_{\mathbb{R}^n} g(\vec{x}) \, f_{\vec{X}}(\vec{x}) \, dx_1 \dots dx_n$$

PROPOSIZIONE - PROPRIETÀ DELLA MEDIA Siano X_1 e X_2 due v.a definite sul medesimo spazio di probabilità e che ammettono media. Allora:

1) $X_1 + X_2$ ammette media e $E(X_1+X_2) = E(X_1) + E(X_2)$
2) Se X_1 e X_2 sono indipendenti, allora $X_1 \cdot X_2$ ammette media e $E(X_1 \cdot X_2) = E(X_1) E(X_2)$.

Osservazione Iterando l'applicazione di 1 e 2 otteniamo: " Se $X_1, \dots, X_n$ sono v.a. sul medesimo spazio di probabilità:

1) $X_1 + \cdots + X_n$ ammette media e $E(X_1 + \cdots + X_2) = E(X_1) + \cdots + E(X_2)$
2) $X_1 \cdots X_n$ ammette media e $E(X_1 \cdots X_2) = E(X_1) \cdots E(X_n)$ "

DIMOSTRAZIONE (Proprietà della media)

1. Dimostreremo il primo punto nel caso $\vec{X} = (X_1, X_2)$ discreto a valori in S e densità $p_{\vec{X}}$.

$$\begin{aligned}
\sum_{\vec{x} \in S} |g(\vec{x})| \, p_{\vec{X}}(\vec{x}) &= \sum_{x_1 x_2} |x_1 + x_2| \, p_{\vec{X}}(x_1, x_2) \leq \\
&\leq \sum_{x_1 x_2} (|x_1| + |x_2|) \, p_{\vec{X}}(x_1, x_2) = \\
&= \sum_{x_1} |x_1| \sum_{x_2} p_{\vec{X}}(x_1, x_2) + \sum_{x_2} |x_2| \sum_{x_1} p_{\vec{X}}(x_1, x_2) = \\
&= \sum_{x_1} |x_1| \, p_{X_1}(x_1) + \sum_{x_2} |x_2| \, p_{X_2}(x_2) < +\infty
\end{aligned}$$

Poichè per ipotesi X_1 e X_2 ammettono media.

$$E(X_1+X_2) = \sum_{x_1,x_2}(x_1+x_2)\,p_{\vec{X}}(x_1,x_2) = \sum$$
$$= \sum_{x_1} x_1 p_{X_1}(x_1) + \sum_{x_2} x_2 p_{X_2}(x_2) = E(X_1)+E(X_2)$$

2. Dimostreremo anche il secondo punto nel caso discreto.

$$\sum_{x_1,x_2} |x_1x_2|\,p_{\vec{X}}(x_1,x_2) = \sum_{x_1,x_2} |x_1||x_2|\,p_{X_1}(x_1)p_{X_2}(x_2) =$$
$$= \sum_{x_1}|x_1|p_{X_1}(x_1)\sum_{x_2}|x_2|p_{X_2}(x_2) < +\infty$$

Per l'indipendenza l'esistenza della media di X_1 e X_2

$$E(X_1X_2) = \sum_{x_1,x_2} x_1x_2\,p_{X_1}(x_1)p_{X_2}(x_2) = \sum_{x_1}x_1p_{X_1}(x_1)\sum_{x_2}x_2p_{X_2}(x_2) =$$
$$= E(X_1)E(X_2) \qquad \blacksquare$$

ESEMPIO Proviamo ora a calcolare la media di una v.a. $X \sim \text{Ipergeom}(r+b, r, n)$. Per fare questo calcolo si può procedere analiticamente oppure sfruttando un approccio più probabilistico basato sulla somma di v.a.

Consideriamo la solita estrazione di n biglie, dove r sono rosse e b bianche (senza reimmissione). Supponiamo che le biglie siano estratte in sequenza e definiamo le variabili $X_1, \dots, X_n$ come

$$X_i = \begin{cases} 1 & i\text{-esima biglia è bianca} \\ 0 & i\text{-esima biglia è rossa} \end{cases}$$

Ovviamente $X = X_1 + \dots + X_n \sim \text{Ipergeom}(r+b, r, n)$. Per calcolare $E(X)$ osserviamo che $E(X) = E(X_1) + \dots + E(X_n)$ per le proprietà della media. Quindi basterà calcolare $E(X_1) \dots E(X_n)$. Poichè X_k è per definizione bernoulliana, per calcolare $E(X_k)$ ci basta solo calcolare $P(X_k = 1)$. A tal fine pensiamo di enumerare biglie in modo che le prime $1..b$ siano bianche e le successive $b+1 .. b+r$ siano rosse. Chiaramente ogni sequenza ha la stessa probabilità di essere estratta e possiamo calcolare questa probabilità come casi favorevoli / casi possibili. Per i casi possibili abbiamo $(b+r)_n$, ossia le disposizioni di $b+r$ elementi di classe n. Per i casi favorevoli di $X_k = 1$ fissiamo la k-esima in b modi e le rimanenti in $(b+r-1)_{n-1}$ modi. Otteniamo quindi:

$$P(X_k = 1) = \frac{b\,\cancel{(b+r-1)\cdots(b+r-1-n+1+1)}}{(b+r)\,\cancel{(b+r-1)\cdots(b+r-n+1)}} = \frac{b}{b+r}$$

Segue che $E(X_k) = b/_{b+r}$ $\forall k = 1 \dots n$ e che $E(X) = n\, b/_{b+r}$.

PROPOSIZIONE - VARIANZA DELLA SOMMA Siano X_1 e X_2 due v.a sul medesimo spazio di probabilità che ammettono varianza. Allora $X_1 + X_2$ ammette varianza pari a:

$$\operatorname{var}(X_1 + X_2) = \operatorname{var}(X_1) + \operatorname{var}(X_2) + 2E[(X_1 - E(X))(X_2 - E(X))]$$

Inoltre se X_1 e X_2 sono indipendenti, allora:

$$\operatorname{var}(X_1 + X_2) = \operatorname{var}(X_1) + \operatorname{var}(X_2)$$

DIMOSTRAZIONE Dimostriamo prima che $X_1 + X_2$ ammette varianza:

$$(X_1 + X_2 - E(X_1 + X_2))^2 = (X_1 - E(X_1) + X_2 - E(X_2)) \le$$
$$\le 2(X_1 - E(X_1))^2 + 2(X_2 - E(X_2))^2$$

$$\Longrightarrow E((X_1 + X_2 - E(X_1 + X_2))^2) \le 2E((X_1 - E(X_1))^2) + 2E((X_2 - E(X_2))^2) \le +\infty$$

Calcoliamo ora la varianza:

$$\operatorname{var}(X_1 + X_2) = E((X_1 + X_2 - E(X_1 + X_2))^2) = E((X_1 - E_1(X_1) + X_2 - E(X_2))^2) =$$
$$= E[(X_1 - E(X_1))^2 + (X_2 - E(X_2))^2 + 2(X_1 - E(X_1))(X_2 - E(X_2))] =$$
$$= \operatorname{var}(X_1) + \operatorname{var}(X_2) + 2E((X_1 - E(X_1))(X_2 - E(X_2)))$$

Se inoltre X_1 e X_2 indipendenti possiamo porre $X_1' := X_1 - E(X_1)$ e $X_2' := X_2 - E(X_2)$ con X_1', X_2' ancora indipendenti. Possiamo allora sviluppare il termine misto come:

$$2E(X_1' X_2') = 2E(X_1')E(X_2') = 2E(X_1 - E(X_1))E(X_2 - E(X_2)) =$$
$$= 2[E(X_1) - E(X_1)][E(X_2) - E(X_2)] = 0$$

Dimostriamo che X_1', X_2' indipendenti:

$$F_{X_1', X_2'}(t_1, t_2) = P(X_1' \le t_1, X_2' \le t_2) = P(X_1 - E(X_1) \le t_1, X_2 - E(X_2) \le t_2) =$$
$$= P(X_1 \le t_1 + E(X_1))\, P(X_2 \le t_2 + E(X_2)) =$$
$$= P(X_1')P(X_2')$$

■

Come prima la proposizione può essere estesa ad una somma di n termini:

$$var\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} var(X_j) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E\big((X_i - E(X_i))(X_j - E(X_j))\big)$$
$$var\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} var(X_j) \quad \text{se } X_1 \dots X_n \text{ indipendenti}$$

Dove il termine misto è detto covarianza

3.6 Covarianza e coefficiente di correlazione

DEFINIZIONE - COVARIANZA Siano X_1 e X_2 due v.a. definite sul medesimo spazio di probabilità e che ammettano varianza. Si dice covarianza di X_1 e X_2:

$$cov(X_1, X_2) := E\big((X_1 - E(X_1))(X_2 - E(X_2))\big)$$

Se $0 < var(X_1)$ e $0 < var(X_2)$ si definisce coefficiente di correlazione di X_1, X_2 il numero

$$\rho_{X_1, X_2} := \frac{cov(X_1, X_2)}{\sqrt{var(X_1)\, var(X_2)}}$$

PROPOSIZIONE - PROPRIETÀ COV. Siano X_1, X_2, X_3 su $(\Omega, \mathcal{F}, P)$ che ammettono varianza σ e $a, b \in \mathbb{R}$

1. $cov(X_1, X_2) = cov(X_2, X_1)$
2. $cov(X_1, a) = 0$
3. $cov(aX_1, X_2) = a\, cov(X_1, X_2)$
4. $cov(X_1 + X_2, X_3) = cov(X_1, X_3) + cov(X_2, X_3)$
5. $cov(X_1, X_2 + X_3) = cov(X_1, X_2) + cov(X_1, X_3)$
6. $cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$
7. Se X_1 e X_2 indipendenti $cov(X_1, X_2) = 0$ ← non vale il viceversa!
8. $|\rho_{X_1X_2}| \leq 1$ e $|\rho_{X_1X_2}| = 1 \iff \exists a, b \in \mathbb{R}\ P(X_2 = aX_1 + b) = 1$. Inoltre $a = cov(X_1, X_2)/var(X_1)$

DIMOSTRAZIONE Dimostriamo solo alcune delle proprietà:

3) $$cov(aX_1, X_2) = E((aX_1 - E(aX_1))(X_2 - E(X_2))) = a E((X_1 - E(X_1))(X_2 - E(X_2))) = a\, cov(X_1, X_2)$$

4) $$\begin{aligned} cov(X_1 + X_2, X_3) &= E((X_1 + X_2 - E(X_1 + X_2))(X_3 - E(X_3))) = \\ &= E((X_1 + X_2 - E(X_1) + E(X_2))(X_3 - E(X_3))) = \\ &= E((X_1 - E(X_1))(X_3 - E(X_3)) + (X_2 - E(X_2))(X_3 - E(X_3))) = \\ &= E((X_1 - E(X_1))(X_3 - E(X_3))) + \\ &\quad + E((X_2 - E(X_2))(X_3 - E(X_3))) = \\ &= cov(X_1, X_3) + cov(X_2, X_3) \end{aligned}$$

6) $$\begin{aligned} cov(X_1, X_2) &= E((X_1 - E(X_1))(X_2 - E(X_2))) = \\ &= E(X_1X_2 - X_1E(X_2) - X_2E(X_1) + E(X_1)E(X_2)) = \\ &= E(X_1X_2) - E(X_2)E(X_1) - \cancel{E(X_1)E(X_2)} + \cancel{E(X_1)E(X_2)} = \\ &= E(X_1X_2) - E(X_2)E(X_1) \end{aligned}$$

7) Se X_1 e X_2 indipendenti allora $E(X_1X_2) = E(X_1)E(X_2)$ e per la proprietà 6 $cov(X_1, X_2) = 0$ ■

3.7 Legge debole e forte dei grandi numeri

Abbiamo detto che se lanciamo una moneta un numero molto elevato di volte otterremo che:

$$n° \text{ di volte che esce testa} / n \to 1/2$$

Se consideriamo una successione $X_1, X_2, \dots$ di v.a di Bernoulli

$$X_i = \begin{cases} 1 & \text{se all'i-esimo lancio esce testa} \\ 0 & \text{se all'i-esimo lancio esce croce} \end{cases} \sim Be(p = \tfrac{1}{2})$$

è chiaro che formano una successione indipendente e identicamente distribuita (i.i.d.). Se poniamo $S_n = \sum_{i=1}^{n} X_i$ allora $S_n/n \approx 1/2$. Come possiamo formalizzare il "$\approx 1/2$"?

PROPOSIZIONE - LEGGE DEBOLE DEI GRANDI NUMERI Sia $X_1, X_2, \dots$ una successione di variabili aleatorie indipendenti e identicamente distribuite con media μ e varianza σ^2 finite. Sia $S_n = \sum_{i=1}^{n} X_i$, allora per ogni $\varepsilon > 0$

$$\lim_{n \to +\infty} P\left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) = 0$$

Indichiamo con $\bar{X}_n = \frac{S_n}{n}$ la media campionaria di $X_1, \dots, X_n$.

DIMOSTRAZIONE Essendo le v.a. i.i.d., per le proprietà della varianza otteniamo:

$$var(\bar{X}_n) = \frac{\sum var(X_i)}{n^2} = \frac{\cancel{n} var(X_1)}{n^{\cancel{2}}} = \sigma^2/n$$

e per le proprietà della media:

$$E(\bar{X}_n) = E\left(\frac{S_n}{n}\right) = \frac{\sum E(X_i)}{n} = \frac{\cancel{n} E(X_1)}{\cancel{n}} = \mu$$

La disuguaglianza di Chebyshev ci dice che:

$$P\left(|\bar{X}_n - E(\bar{X}_n)| > \varepsilon \right) \le \frac{var(\bar{X}_n)}{\varepsilon^2} \Rightarrow 0 \le P\left(|\bar{X} - \mu| > \varepsilon \right) \le \frac{\sigma^2}{n\varepsilon^2}$$

che per $n \to +\infty$ tende a 0. ∎

PROPOSIZIONE - LEGGE FORTE DEI GRANDI NUMERI Sia $X_1, X_2, \dots$ una successione di variabili aleatorie i.i.d. che ammettono media μ. Allora:

$$P\left(\left\{\omega : \lim_{n\to+\infty} \bar{X}_n(\omega) = \mu\right\}\right) = 1$$

3.8 Teorema centrale del limite

TEOREMA - TEOREMA CENTRALE DEL LIMITE Sia $X_1, X_2, \dots$ di v.a. i.i.d con media μ e varianza σ^2 finite e $\sigma^2 > 0$. Allora posto $S_n = \sum_{i=1}^{n} X_i$ vale:

$$\lim_{n\to+\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \leq x\right) = \Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du \qquad \forall x \in \mathbb{R}$$

Osservazioni

1) $E(S_n) = E(X_1) + \dots + E(X_n) = n\mu$ e $var(S_n) = var(X_1) + \dots + var(X_n) = n\sigma^2$ quindi la v.a $\frac{S_n - n\mu}{\sqrt{n}\,\sigma}$ non è altro che la standardizzata di S_n e quindi il TCL ci dice che per n grande la standardizzata di S_n ha f.d.r approssimabile con quella della gaussiana standard. Ciò significa anche che per n grande S_n è approssimabile con $\mathcal{N}(n\mu, n\sigma^2)$.

2. Se sviluppiamo la forma della standardizzata otteniamo:

$$\frac{S_n - n\mu}{\sqrt{n}\,\sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - E(\bar{X}_n)}{var(\bar{X}_n)}$$

Quindi la tesi del TCL può essere data in forma equivalente

$$\lim_{n\to+\infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) = \Phi(x)$$

cioè per n grande $\bar{X}_n \approx \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$

3) Sia $S_n :=$ n° successi in n prove di Bernoulli. Allora $S_n = X_1 + \dots + X_n \sim Bi(n, p)$ con $E(S_i) = nE(X_i) = np$ e $var(S_n) = n\, var(X_i) = np(1-p)$. Se consideriamo la standardizzata

di S_n per il TCL abbiamo:

$$\lim_{n\to+\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x) \rightarrow \text{Teorema di De Moivre - Laplace}$$

Se $n \geq 30$, $np > 5 \wedge n(1-p) > 5$ o $np(1-p) > 10$ usiamo l'approssimazione della binomiale fornita dal TCL. Altrimenti, per $np \leq 5$, approssimiamo con la distribuzione di Poisson.

ESEMPIO 1 Il tempo di vita di un certo componente elettrico è una v.a. con media 100 ore e deviazione standard 20 ore. Se si provano 100 componenti di questo tipo, quanto vale la probabilità che la somma delle loro durate sia compresa tra 9800 ore e 10400 ore?

Siano $X_1, \dots, X_{100}$ i tempi di vita. Sono v.a. i.i.d. con media $\mu = 100\,h$ e $\sigma^2 = 400\,h$ quindi soddisfano le Hp del TCL.
Dobbiamo calcolare $P(9800 \leq S_{100} \leq 10400)$:

assumiamo a.c.

$$P(9800 \leq S_{100} \leq 10400) = P(9800 < S_{100} \leq 10400) =$$
$$= P(S_{100} \leq 10400) - P(S_{100} \leq 9800) =$$
$$= P\left(\frac{S_{100} - 100\mu}{\sqrt{100}\cdot\sigma} \leq \frac{10400 - 100^2}{\sqrt{100}\cdot 20}\right) - P\left(\frac{S_{100} - 100\mu}{\sqrt{100}\cdot\sigma} \leq \frac{9800 - 100^2}{\sqrt{100}\cdot 20}\right) \approx$$

TCL

$$\approx \Phi(2) - \Phi(-1) = \Phi(2) - 1 + \Phi(1) = 0,8185$$

ESEMPIO 2 Un astronomo vuole misurare la distanza di una stella lontana. Tuttavia a causa dei disturbi dovuti all'atmosfera, le misure effettuate dal suo osservatorio non restituiscono la distanza esatta d. Per questo motivo egli ha deciso di fare una serie di misurazioni e di usare la media campionaria come "stimatore" di d. È infatti convinto che misurazioni successive siano v.a. i.i.d. di media d e deviazione standard 2 (l'unità di misura è l'anno luce). Quante misure deve effettuare per avere il 95% di probabilità che la sua stima sia accurata entro ± 0.5 anni luce?

Sia $X_1, \dots, X_n$ la serie di misurazioni. Le singole X_j sono v.a i.i.d. con $E(X_j) = d$ e $\sqrt{var(X_j)} = 2$. Per il TCL sappiamo che $\bar{X}_n \approx \mathcal{N}(d, 4/n)$. Il nostro obiettivo è determinare n affinché $P(|\bar{X}_n - d| \leq 0,5) \geq 0,95$:

$$0.95 \leq P(-0,5 \leq \bar{X}_n - d \leq 0,5) = P\left(-\frac{0.5}{2/\sqrt{n}} \leq \frac{\bar{X}_n - d}{2/\sqrt{n}} \leq \frac{0.5}{2/\sqrt{n}}\right) \approx$$
$$\approx \Phi(\sqrt{n}/4) - 1 + \Phi(\sqrt{n}/4) =$$
$$= 2\Phi(\sqrt{n}/4) - 1$$
$$\Rightarrow \Phi(\sqrt{n}/4) \geq 0.975 \iff z_{0.025} \leq \sqrt{n}/4$$
$$1.96 \leq \sqrt{n}/4 \iff n \geq 62$$

3.9 Matrice di covarianza

DEFINIZIONE - MATRICE DI COVARIANZA Sia $\vec{X} = (X_1, \dots, X_n)$ un vettore aleatorio n-dimensionale con componenti che ammettono varianza. Si chiama matrice di covarianza di $\vec{X}$ la matrice $n \times n$ $C_{\vec{X}} = (c_{ij})_{i=1,n}^{j=1,n}$ il cui elemento c_{ij} è $c_{ij} := cov(X_i, X_j)$

PROPOSIZIONE - PROPRIETÀ MATRICE DI COV.

1. $C_{\vec{X}}$ è una matrice simmetrica e semidefinita positiva
2. Se le componenti di $\vec{X}$ sono scorrelate ($cov(X_i, X_j) = 0$ $i \neq j$) allora $C_{\vec{X}}$ è diagonale
3. Se $A = (a)_{i=1,m}^{j=1,n}$ matrice $m \times n$ e $\vec{b} \in \mathbb{R}^m$, allora la matrice di covarianza di $\vec{Y} = A\vec{X} + b$ è $C_{\vec{Y}} = A C_{\vec{X}} \cdot A^T$

DIMOSTRAZIONE

1) $C_{\vec{X}}$ è simmetrica poiché $cov(X_i, X_j) = cov(X_j, X_i)$ e di conseguenza $a_{ij} = a_{ji}$ e quindi la matrice è simmetrica. Inoltre poiché

$$\vec{x}^T C_{\vec{X}} \vec{x} = \sum_{i,j=1}^n x_i x_j \, cov(X_i, X_j) = \sum_{i,j=1}^n x_i x_j E((X_i - E(X_i))(X_j - E(X_j))) =$$
$$= E\left(\sum_{i,j=1}^n x_i x_j (X_i - E(X_i))(X_j - E(X_j))\right) =$$
$$= E\left(\sum_{i=1}^n x_i (X_i - E(X_i))^2\right) \geq 0$$

possiamo dire che $C_{\vec{X}}$ è semidefinita positiva

2) Se $cov(X_i, X_j) = 0$ $\forall i, j$, seguirà che gli unici elementi non nulli saranno $cov(X_i, X_i)$, rendendo $C_{\vec{X}}$ diagonale

3) A è un matrice $m \times n$, $\vec{b} \in \mathbb{R}^m$ e $Y = A\vec{X} + b$. Calcoliamo la matrice di covarianza di $\vec{Y}$ $C_{\vec{Y}}$:

$$\hat{c}_{ij} := \text{cov}(Y_i, Y_j) = \text{cov}\left(\sum_{k=1}^n a_{ik} X_k + b_i, \sum_{l=1}^n a_{jl} X_l + b_j\right)$$
$$= \sum_{k,l=1}^n a_{jk} a_{jl} \,\text{cov}(X_k, X_l) = \sum_{k,l=1}^n a_{jk} a_{jl} c_{kl} =$$
$$= (A C_{\vec{X}} A^T)_{ij}$$

■

3.10 Vettori gaussiani

DEFINIZIONE - VETTORI GAUSSIANI STD Un vettore aleatorio $\vec{Z} = (Z_1, \dots, Z_m)^T$ è detto vettore gaussiano standard m-dimensionale (m-variato) se $Z_1, \dots, Z_m$ sono i.i.d gaussiani standard.

Conseguenze della definizione sono:

1. $\vec{Z}$ è un vettore aleatorio a.c. poiché le componenti sono a.c e indipendenti:

$$f_{\vec{Z}}(z_1, \dots, z_m) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z_i^2} = \frac{1}{(2\pi)^{m/2}} e^{-\frac{1}{2}\sum_{i=1}^m z_i^2} =$$
$$= \frac{1}{(2\pi)^{m/2}} e^{-\frac{1}{2}\vec{z}^T \vec{z}}$$

   e scriveremo $\vec{Z} \sim \mathcal{N}(0, I)$

2. $E(\vec{Z}) = 0$ e $C_{\vec{Z}} = I$

DEFINIZIONE - VETTORI GAUSSIANI Un vettore aleatorio $\vec{X} = (X_1, \dots, X_n)^T$ è detto gaussiano n dimensionale se esistono una matrice $A = (a_{ij})$ $n \times m$, $\vec{\mu} \in \mathbb{R}^n$ e un vettore gaussiano standard m-dimensionale Z tali che $\vec{X} = A\vec{Z} + \vec{\mu}$

Si può notare che $E(\vec{X}) = \mu$ per la linearità della media e che $C_{\vec{X}} = A C_{\vec{Z}} A^T = AA^T$. Scriveremo $X \sim \mathcal{N}(\vec{\mu}, AA^T)$.
Se per esempio gli elementi della i-esima riga di A sono nulli, allora $X_i = a_{i1} Z_1 + a_{i2} Z_2 + \dots + a_{im} Z_m + \mu_i = \mu_i$, cioè X_i è una v.a. costante. Quindi è chiaro che $\vec{X}$ non è a.c.. Esistono, tuttavia, delle condizioni per verificare l'assoluta continuità di $\vec{X}$.
Se consideriamo $m = n$, $\vec{X}$ è trasf. affine di $\vec{Z}$ e quindi

assolutamente continuo di densità

$$f_{\vec{X}}(\vec{x}) = \frac{1}{|\det A|} f_{\vec{Z}}\left(A^{-1}(\vec{x}-\vec{\mu})\right) =$$

$$= \frac{1}{|\det A| \cdot (2\pi)^{n/2}} e^{-\frac{1}{2}[A^{-1}(\vec{x}-\vec{\mu})]^T[A^{-1}(\vec{x}-\vec{\mu})]}$$

$$= \frac{1}{\sqrt{(\det A)^2 (2\pi)^n}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T A^{-1\,T} A^{-1}(\vec{x}-\vec{\mu})}$$

$$= \frac{1}{\sqrt{\det AA^T (2\pi)^n}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T (AA^T)^{-1}(\vec{x}-\vec{\mu})}$$

$$= \frac{1}{\sqrt{\det C_{\vec{X}} (2\pi)^n}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T C_{\vec{X}}^{-1}(\vec{x}-\vec{\mu})} \qquad (*)$$

Questo non è l'unico caso in cui $\vec{X}$ è a.c. Vale, infatti:

PROPOSIZIONE - CNS DI A.C. DI VETTORI GAUSSIANI

1) Un vettore gaussiano $\vec{X} = A\vec{Z} + \vec{\mu}$ ha densità (a.c.) su $\mathbb{R}^n$ se e solo se $C = AA^T$ è non singolare. In questo caso la densità di $\vec{X}$ è (*)
2) Se $\vec{X}$ è un vettore aleatorio a.c. con densità data da (*) e C simmetrica definita positiva, allora, indicata con A (la radice di C) invertibile tale che $C = AA^T$, il vettore $\vec{Z}$ sarà $\vec{Z} = A^{-1}(\vec{X}-\vec{\mu}) \sim \mathcal{N}(\underline{0}, \underline{I})$

PROPOSIZIONE - PROPRIETÀ VETT. G. Siano $\vec{X} = A\vec{Z} + \vec{\mu}$ un vettore gaussiano n-dimensionale, $C = AA^T = (c_{ij})_{i,j=1,n}$ la matrice di covarianza di $\vec{X}$. Allora valgono:

1. Se $c_{ii} > 0$ allora la i-esima componente X_i è una v.a. gaussiana, $X_i \sim \mathcal{N}(\mu_i, c_{ii})$. Se invece $c_{ii} = 0$, allora $P(X_i = \mu_i) = 1$
2. Se G è una matrice $K \times n$ e $\vec{h} \in \mathbb{R}^n$, allora $\vec{Y} = G\vec{X} + \vec{h}$ è un vettore gaussiano con vettore delle medie $G\vec{\mu} + \vec{h}$ e matrice di covarianza GCG^T
3. Se $X_1, \dots, X_n$ sono scorrelate allora sono anche indipendenti

DIMOSTRAZIONE

1) $X_i = a_{i1} Z_1 + \dots + a_{im} Z_m + \mu_i$

$$m_{X_i} = E(e^{tX_i}) = E(e^{t(a_{i1}Z_1 + \dots + a_{im}Z_m + \mu_i)}) = e^{t\mu_i} E(\Pi_{j=1}^m e^{ta_{ij}Z_j})$$
$$= e^{t\mu_i} \Pi_{j=1}^m E(e^{ta_{ij}Z_j}) = e^{t\mu_i} \Pi_{j=1}^m m_{Z_j}(ta_{ij}) =$$
$$= e^{t\mu_i} \Pi_{j=1}^m e^{\frac{1}{2}a_{ij}^2 t^2} = e^{t\mu_i + \frac{1}{2}\sum_{j=1}^m a_{ij}^2 t^2} = e^{t\mu_i + \frac{1}{2}c_{ii}t^2} = m_X \text{ con}$$

$\hookrightarrow c_{ii} = \sum_{j=1}^m a_{ij}^2$

$X \sim \mathcal{N}(\mu_i, c_{ii})$. Poiché $m_{X_i} = m_X$, allora necessariamente X_i è gaussiana di media μ_i. Se invece $c_{ii} = 0 = var(X_i) \Rightarrow X_i$ è una costante.

2) $$\vec{Y} = G\vec{X} + \vec{h} = G(A\vec{Z} + \vec{\mu}) + \vec{h} = GA\vec{Z} + G\vec{\mu} + \vec{h} \sim$$
$$\sim \mathcal{N}(G\vec{\mu} + \vec{h},\ GAA^T G^T = GCG^T)$$

3) Ci limitiamo dimostrare la proprietà nel caso in cui vi sia densità su $\mathbb{R}^n$.
Sappiamo che $c_{ii} = \sigma^2 = var(X_i)$ $i = 1, n$. Quindi $\det C = \Pi_{i=1}^n c_{ii}$ sarà diverso da 0 se e solo $c_{ii} = \sigma^2 > 0$ $\forall i$. Esplicitando la densità su $\mathbb{R}^n$ di $\vec{X}$ otteniamo che:

$$f_{\vec{X}}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n \det C}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T C^{-1}(\vec{x}-\vec{\mu})} =$$
$$= \frac{1}{\sqrt{(2\pi)^n \det C}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \begin{pmatrix} 1/\sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_n^2 \end{pmatrix} (\vec{x}-\vec{\mu})} =$$
$$= \frac{1}{\sqrt{(2\pi)^n \sigma_1^2 \cdots \sigma_n^2}} e^{-\frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu_i)^2}{\sigma_i^2}} =$$
$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\frac{(x_i-\mu_i)^2}{\sigma_i^2}} = \prod_{i=1}^n f_{X_i}(x_i)$$

$X_1, \dots, X_n$ sono, quindi, v.a. indipendenti. ■

ESERCIZIO Sia $\vec{X} = (X_1, \dots, X_n)^T$ un vettore gaussiano con vettore delle medie $\vec{\mu}$ e matrice di covarianza C. Mostrate che per ogni scelta di $a_1, \dots, a_n$ numeri reali di almeno 1 diverso da 0, $a_1X_1 + \dots + a_nX_n$ è una v.a. gaussiana e determinarne i parametri
Prendiamo $G = (a_1, \dots, a_n)$, allora $G\vec{X} = a_1X_1 + \dots + a_nX_n$. Per punto 1 della precedente proposizione $G\vec{X}$ è una v.a. gaussiano con $E(G\vec{X}) = \sum_{i=1}^n a_i\mu_i$ e $var(G\vec{X}) = GCG^T = \sum_{i,j=1}^n a_{ij}\, cov(X_i, X_j)$

$= \sum_{i=1}^{n} a_i^2 \, var(X_i) + \sum_{i \neq j} a_{ij} \, cov(X_i, X_j)$. Possiamo quindi dire che $G\vec{X} \sim \mathcal{N}(\sum_{i=1}^{n} a_i a_j, \sigma^2)$.

ESERCIZIO Sia $\vec{X} = (X_1, \dots, X_n)^T$ un vettore gaussiano con vettore delle medie $\vec{\mu}$ e matrice di covarianza C.

1) Usando la 2ª proprietà dei vettori gaussiani, mostrare che ogni vettore $(X_i, X_j)^T$ $i \neq j$ estratto da $\vec{X}$ è un vettore gaussiano bidimensionale e determinarne i parametri
2) Dedurre che se $cov(X_i, X_j) = 0$ per una fissata coppia $i \neq j$, allora X_i e X_j sono indipendenti

1.

$$G = \begin{array}{c} \begin{array}{ccccc} & & i & & \\ & & \downarrow & & \end{array} \\ \begin{pmatrix} 0 & \cdots & 1 & \cdots\cdots & 0 \\ 0 & \cdots & \cdots & 1 \;\cdots & 0 \end{pmatrix} \\ \begin{array}{ccccc} & & & \uparrow & \\ & & & j & \end{array} \end{array} \Rightarrow G\vec{X} = \begin{pmatrix} X_i \\ X_j \end{pmatrix} = \vec{Y}$$

Per la seconda proprietà dei vettori gaussiani $\vec{Y} = G\vec{X}$ è un vettore gaussiano di parametri:

$$\vec{\mu} = \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix} \qquad C_{\vec{Y}} = \begin{pmatrix} var(X_i) & cov(X_i, X_j) \\ cov(X_i, X_j) & var(X_j) \end{pmatrix}$$

2. Conseguenza del fatto che $\vec{Y}$ è gaussiano è il fatto che se se $cov(X_i, X_j) = 0$ allora X_i e X_j sono indipendenti.

ESERCIZIO Sia $(Z_1, Z_2)^T$ un vettore gaussiano standard e siano

$$U = Z_1 + aZ_2 + b$$
$$V = cZ_1 + Z_2 + d$$

con $a, b, c, d \in \mathbb{R}$.

1) Determinare in funzione dei parametri la distribuzione del vettore $(U, V)^T$
2) Determinare per quali valori dei parametri U e V sono indipendenti.

1. $\begin{pmatrix} U \\ V \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & a \\ c & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}}_{\vec{Z}} + \underbrace{\begin{pmatrix} b \\ a \end{pmatrix}}_{\rightarrow}$

$\Rightarrow \begin{pmatrix} U \\ V \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} b \\ a \end{pmatrix}, \begin{pmatrix} 1+a^2 & c+a \\ c+a & 1+c^2 \end{pmatrix} \right)$

2. Essendo $(U,V)^T$ vettore gaussiano U e V sono indipendenti se e solo se $c+a=0$.

4 INFERENZA STATISTICA E STATISTICHE CAMPIONARIE

La statistica è la scienza che si occupa di trarre conclusioni dai dati sperimentali. La situazione tipica è quella in cui si studia un insieme molto grande, detto popolazione, di oggetti a cui sono associate quantità misurabili. L'approccio statistico consiste nel selezionare un sottoinsieme ridotto di oggetti detto campione e "analizzarlo" per trarre conclusioni, ossia fare inferenza.

Per basare i dati del campione sulle inferenze che riguardino l'intero popolazione è necessario assumere qualche relazione tra il campione e la popolazione. Un'ipotesi fondamentale è che vi sia una distribuzione di probabilità tipica della popolazione, nel senso che da essa si estraggono in modo casuale degli oggetti e le quantità numeriche loro associate possono essere pensate come valori assunti da v.a. indipendenti e tutte con la stessa distribuzione (i.i.d)

DEFINIZIONE - CAMPIONE Le v.a. $X_1, \dots, X_n$ sono dette campione casuale (aleatorio) di dimensione n estratto da popolazione con f.d.r F se sono i.i.d.

DEFINIZIONE - REALIZZAZIONE CAMPIONE Se osserviamo i valori $X_1 = x_1, \dots, X_n = x_n$ allora $(x_1, \dots, x_n)$ è detta realizzazione del campione

DEFINIZIONE - PROBL. DI INFERENZA PARAM. La f.d.r F è nota a meno di un parametro incognito

DEFINIZIONE - PROBL. DI INFERENZA NON PARAM. Non conosciamo la forma analitica della f.d.r F.

4.1 Statistiche

DEFINIZIONE - STAT. BASATA SUL CAMPIONE Sia $X_1, \dots, X_n$ un campione estratto da F, una statistica basata sul campione è un funzione nota del campione:

$$D_n = d_n(X_1, \dots, X_n)$$

Dove d_n è una funzione nota delle n v.a. $X_1, \dots, X_n$. La statistica D_n sarà anch'essa una v.a.

MEDIA CAMPIONARIA Sia $X_1, \dots, X_n$ un campione aleatorio estratto da F, la media μ e la varianza σ^2 sono dette media/varianza della popolazione. Nel capitolo sul TCL abbiamo definito la media campionaria come

$$\bar{X}_n := \frac{X_1 + \dots + X_n}{n} \Rightarrow \bar{X}_n = d_n(X_1, \dots, X_n)$$

La media campionaria è, quindi, una statistica. Abbiamo già dimostrato le sue principali proprietà:

$$E(\bar{X}_n) = \mu \qquad var(\bar{X}_n) = \sigma^2/n \qquad \bar{X}_n \simeq \mathcal{N}(\mu, \sigma^2/n)$$

VARIANZA CAMPIONARIA Diciamo varianza campionaria:

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \qquad \text{per } n = 2, 3, \dots$$

e deviazione standard campionaria $\sqrt{S_n^2}$. Come anche la media campionaria, la varianza campionaria possiede importanti proprietà legate a media e varianza:

PROPOSIZIONE - PROPRIETÀ

1) $E(S_n^2) = \sigma^2$

2) $var(S_n^2) = \frac{1}{n}\left[\mu_4 - \frac{n-3}{n-1}\sigma^4\right]$ dove $\mu_4 = E((X_1-\mu)^4)$ e $\sigma^4 = (\sigma^2)^2$

DIMOSTRAZIONE

1) $(n-1)S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i^2 + \bar{X}_n^2 - 2X_i\bar{X}_n) =$
$= \sum_{i=1}^n X_i^2 + n\bar{X}_n^2 - 2n\bar{X}_n^2 = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2$

$E(S_n^2) = \frac{1}{n-1}\left[E(\sum_{i=1}^n X_i^2) - nE(\bar{X}_n^2)\right] = \frac{1}{n-1}\left[nE(X_1^2) - nE(\bar{X}_n^2)\right] =$
$= \frac{n}{n-1}\left[var(X_1) + E(X_1)^2 - var(\bar{X}_n) - E(\bar{X}_n)^2\right] =$
$= \frac{n}{n-1}\left[\sigma^2 + \cancel{\mu^2} - \sigma^2/n - \cancel{\mu^2}\right] = \sigma^2$

2) Analogo a 1. ■

Si può notare che $var(S_n^2) \to 0 \quad n \to +\infty$

4.1.1 Distribuzione congiunta di $\bar{X}_n$ e S_n^2 nel caso di popolazioni Gaussiane

Sia $X_1, \dots, X_n$ un campione estratto da una popolazione gaussiana, come calcoliamo la distribuzione congiunta di $\bar{X}_n$ e S_n^2?

Sappiamo che, poiché $\bar{X}_n$ c.l. di v.a. gaussiane, è anch'essa gaussiana, quindi:

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \iff \frac{(\bar{X}_n - \mu)\sqrt{n}}{\sigma} \sim \mathcal{N}(0,1)$$

PROPOSIZIONE Se $X_1, \dots, X_n$ è un campione di dimensione n estratto da popolazione gaussiana di media μ e varianza σ^2 e $\bar{X}_n$ e S_n^2 sono indipendenti, allora

$$(\bar{X}_n - \mu)\sqrt{n}/\sigma \sim \mathcal{N}(0,1) \quad , \quad (n-1)\, S_n^2/\sigma^2 \sim \chi^{2*}(n-1)$$

COROLLARIO Se $X_1, \dots, X_n$ è un campione di dimensione n estratto da popolazione gaussiana di media μ e varianza σ^2 e $\bar{X}_n$ e S_n^2 sono indipendenti, allora:

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t^*(n-1)$$

* vedi es del 26/04/21

DIMOSTRAZIONE Sappiamo che se $Z \sim \mathcal{N}(0,1)$ e $\chi^2_k \sim \chi^2(K)$ sono indipendenti, allora $Z/\sqrt{\chi^2_k/K} \sim t(K)$. Poiché dalla precedente proposizione sappiamo che:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \qquad (n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

E sono indipendenti, segue che:

$$\frac{\bar{X}_n - \mu}{\cancel{\sigma}/\sqrt{n}} \cdot \sqrt{\frac{\cancel{(n-1)}\,\cancel{\sigma^2}}{\cancel{(n-1)}\,S_n^2}} = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$$

■

4.2 Inferenza statistica parametrica

Sia un $X_1, \dots, X_n$ campione aleatorio di dimensione n estratto da una popolazione con distribuzione F_θ nota a meno di un parametro incognito θ. Indicheremo con P_θ, E_θ, var_θ la probabilità / media / varianza calcolate per ogni θ ammissibile, cioè in funzione di θ.

DEFINIZIONE - CARATTERISTICA DELLA POPOLAZIONE Sia $X_1, \dots, X_n$ un campione aleatorio estratto F_θ. Si definisce caratteristica della popolazione una funzione (non costante) di θ.

Il problema è, quindi, fare inferenza sui parametri incogniti o sulla caratteristica

4.2.1 Stima puntuale

Sia $X_1, \dots, X_n$ campione aleatorio estratto da F_θ e sia $K(\theta)$ una caratteristica su cui vogliamo fare inferenza:

DEFINIZIONE - STIMATORE PUNTUALE Chiamiamo stimatore puntuale di $K(\theta)$ una statistica $\hat{D} = d(X_1, \dots, X_n)$ usata per fare inferenza su $K(\theta)$.

DEFINIZIONE - STIMA Data la realizzazione campionaria $x_1, \dots, x_n$ e lo stimatore $\hat{D} = d(X_1, \dots, X_n)$ di $K(\theta)$, si definisce stima di $K(\theta)$ il valore $\hat{d} = d(x_1, \dots, x_n)$ assunto dalla statistica $\hat{D}$ in corrispondenza delle osservazioni $x_1, \dots, x_n$.

4.2.2 Proprietà di uno stimatore

Siano $X_1, \dots X_n$ un campione aleatorio estratto F_θ, f_θ la sua densità e $\hat{D}$ lo stimatore di $K(\theta)$, caratteristica della popolazione. Ci servono dei criteri per valutare la bontà di $\hat{D}$ come stimatore. Possiamo considerare la "distanza" di $\hat{D}$ da $K(\theta)$:

$$E_\theta((\hat{D} - K(\theta))^2) = E_\theta((d(X_1, \dots, X_n) - K(\theta))^2)$$

Lo stimatore sarà tanto migliore quanto è piccola questa distanza. Formalizzando, introduciamo:

DEFINIZIONE - ERRORE QUADRATICO MEDIO Sia $\hat{D} = d(X_1, \dots, X_n)$ uno stimatore di $K(\theta)$ che ammette momento secondo finito. Si definisce errore quadratico medio di $\hat{D}$ come stimatore di $K(\theta)$ la funzione di θ definita da:

$$r_\theta(d, K(\theta)) := E_\theta((d(X_1, \dots, X_n) - K(\theta))^2) \leftarrow \text{anche } MSE_\theta(\hat{D})$$

Come possiamo calcolarlo?

caso continuo

$$r_\theta(d, K(\theta)) = \int_{\mathbb{R}^n} (d(x_1, \dots, x_n) - K(\theta))^2 \prod_{i=1}^n f_\theta(x_i)\, dx_1 \cdots dx_n$$

Possiamo rendere minimo lo MSE? Purtroppo no.

ESEMPIO Sia $X_1, \dots, X_n$ il campione da f_θ, $\theta > 0$ incognito.

Imponiamo $K(\theta) = \theta$. sarà, quindi, $MSE_\theta(\hat{D}) = E_\theta((\hat{D} - \theta)^2)$.
Non si può trovare $\hat{D}$ che renda minimo $\forall \theta$ $MSE_\theta(\hat{D})$ rispetto a tutti gli altri stimatori perché per esempio se il "vero" valore di θ è $\theta = 5$ allora lo stimatore $\hat{D}^* = 5$ rende lo MSE nullo e quindi nessun altro stimatore in $\theta = 5$ può fare meglio.

Restringiamo, quindi, il campo e introduciamo:

DEFINIZIONE - DISTORSIONE (BIAS) Sia $\hat{D} = d(X_1, \dots, X_n)$ uno stimatore di $K(\theta)$ che ammette media. Si definisce distorsione (bias) di $\hat{D}$ come stimatore di $K(\theta)$ la funzione:

$$b_\theta(d) := E_\theta(d(X_1, \dots, X_n)) - K(\theta)$$

DEFINIZIONE - STIMATORE NON DISTORTO Lo stimatore $\hat{D} = d(X_1, \dots, X_n)$ è detto non distorto (o corretto) se $b_\theta(d) = 0$.

La media campionaria è uno stimatore corretto per la media, mentre la varianza campionaria è corretta per la varianza.

PROPOSIZIONE - DECOMPOSIZIONE MSE Sia $\hat{D} = d(X_1, \dots, X_n)$ uno stimatore di $K(\theta)$ basato sul campione $X_1, \dots, X_n$ che ammette momento secondo finito, allora:

$$r_\theta(d, K(\theta)) = var_\theta(d(X_1, \dots, X_n)) + (b_\theta(d))^2$$

DIMOSTRAZIONE

$$r_\theta(d, K(\theta)) = E_\theta[(\hat{D} - K(\theta))^2] = var(\hat{D} - K(\theta)) + (E_\theta(\hat{D} - K(\theta)))^2 =$$
$$= var_\theta(\hat{D}) + (E_\theta(\hat{D}) - K(\theta))^2 = var_\theta(d(X_1, \dots, X_n)) + (b_\theta(d))^2$$

■

Conseguenza della proposizione è che se $\hat{D}$ non è distorto, allora $r_\theta(d, K(\theta)) = var_\theta(d(X_1, ..., X_n))$.

4.2.3 Proprietà asintotiche di uno stimatore

Sia $X_1, X_2, ...$ una successione di v.a. i.i.d $X_i \sim f_\theta$ con θ parametro/i incognito/i. Sia $(\hat{D}_n)_n = (d(X_1, ..., X_n))_n$ una successione di stimatori che ammettono media e momento secondo finiti.

DEFINIZIONE - NON DISTORSIONE AS. La successione di stimatori $(\hat{D}_n)_n$ è detta asintoticamente non distorta (o corretta) per $K(\theta)$ se:

$$b_\theta(d_n) \to 0 \quad \text{per } n \to +\infty \quad \forall \theta \text{ ammissibile}$$

OSSERVAZIONE Se $\hat{D}_n$ è non distorto $\forall n$, allora $(\hat{D}_n)_n$ è asintoticamente non distorta.

DEFINIZIONE - DEBOLEZZA CONSISTENTE Una successione di stimatori $(\hat{D}_n)_n$ è detta debolmente consistente per $K(\theta)$ se

$$\forall \varepsilon > 0 \quad P_\theta(|\hat{D}_n - K(\theta)| > \varepsilon) \to 0 \quad n \to +\infty \quad \forall \theta \text{ ammissibile}$$

DEFINIZIONE - CONSISTENZA IN MEDIA QUADRATICA Una successione di stimatori $(\hat{D}_n)_n = (d_n(X_1, ..., X_n))_n$ è detta consistente in media quadratica per $K(\theta)$ se:

$$r_\theta(d_n, K(\theta)) = E_\theta((\hat{D}_n - K(\theta))^2) \to 0 \quad \text{per } n \to +\infty \quad \forall \theta \text{ amm.}$$

OSSERVAZIONI Per la disuguaglianza di Markov abbiamo:

$$0 \leq P_\theta(|\hat{D}_n - K(\theta)| > \varepsilon) \leq E_\theta((\hat{D}_n - K(\theta))^2)/\varepsilon^2 = r_\theta(d_n, K(\theta))/\varepsilon^2 \to 0$$

Possiamo, quindi, affermare che la consistenza in media quadratica implica la consistenza debole. Inoltre, per la decomposizione, $r_\theta \to 0$ sse entrambi gli addendi tendono a zero in quanto tutte quantità positive. La successione di stimatori $(\hat{D}_n)_n$ è consistente in media quadratica per $K(\theta)$ sse $(\hat{D}_n)$ è asintoticamente non distorta e la varianza di $\hat{D}_n \to 0$ $\forall\theta$ ammissibile.

ESERCIZIO

1. Mostrare che la successione delle medie campionarie $(\bar{X}_n)_{n\geq 1}$ è sempre una successione di stimatori della media della popolazione non distorti e consistenti in media quadratica.
2. Mostrare che la successione delle varianze campionarie $(S_n^2)_{n\geq 2}$ è sempre una successione di stimatori non distorti della varianza della popolazione e, assumendo l'esistenza del momento quarto della popolazione è anche consistente in media quadratica.
3. Mostrare che, se la media della popolazione è nota e uguale a μ_0, si può prendere come successione di stimatori della varianza, non distorti e consistente in media quadratica la successione $(S^2_{0n})_{n\geq 1}$ con

$$S^2_{0n} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu_0)^2$$

Sia $X_1, \dots, X_n$ un campione estratto da una popolazione con densità f_θ che ammette media e varianza; consideriamo le caratteristiche:

$$K_1(\theta) := E_\theta(X_1)$$
$$K_2(\theta) := var_\theta(X_1)$$

1) $(\bar{X}_n)_{n\geq 1}$ successione di stimatori di $K_1(\theta)$, $\bar{X}_1$ è non distorto per $K_1(\theta)$ $\forall n$ e a maggior ragione $(\bar{X}_n)_{n\geq 1}$ è asintoticamente non distorta.

$MSE_\theta(\bar{X}_n) = var_\theta(\bar{X}_n)$ perché non distorto.

$var_\theta(\bar{X}_n) = K_2(\theta)/n \to 0$ $\forall\theta \Rightarrow (\bar{X}_n)_{n\geq 1}$ è consistente in media quadratica per $K_1(\theta)$ e quindi anche debolmente consistente.

2) $(S_n^2)_{n\geq 2} = \left(\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2\right)_{n\geq 2}$

$E_\theta(S_n^2) = var_\theta(X_1) = K_2(\theta) \Rightarrow S_n^2$ è uno stimatore non distorto per la varianza.

$MSE_\theta(S_n^2) = var_\theta(S_n^2) = \frac{1}{n}\left[\mu_4 - \frac{n-3}{n-1}\theta^2\right] \to 0 \quad n\to+\infty \quad \forall\theta$

$\Rightarrow (S_n^2)_{n\geq 2}$ è consistente in media quadratica e quindi anche debolmente per $K_2(\theta)$.

3) Sia $S_{0n}^2 = \frac{1}{n}\sum_{j=1}^n (X_j - \mu_0)^2$ ATTENZIONE: poichè μ_0 è nota, S_{0n}^2 è una statistica e quindi posso prenderla come stimatore.

$E_\theta(S_{0n}^2) = E_\theta\left(\frac{1}{n}\sum_{j=1}^n (X_j-\mu_0)^2\right) = \frac{1}{n}\sum_{j=1}^n E_\theta((X_j-\mu_0)^2)$

$= E((X_1-\mu_0)^2) = K_2(\theta) \Rightarrow S_{0n}^2$ non distorto per var

$MSE_\theta(S_{0n}^2) = var_\theta(S_{0n}^2) = var_\theta\left(\frac{1}{n}\sum_{j=1}^n (X_j-\mu_0)^2\right) =$

$\frac{1}{n^2}\sum_{j=1}^n var_\theta((X_j-\mu_0)^2) = var_\theta((X_j-\mu_0)^2) \leq E_\theta((X_j-\mu_0)^2)/n$

$\to 0 \quad n\to+\infty \quad \forall\theta \Rightarrow S_{0n}^2$ consistente in m.q.

4.2.4 Come si ottiene uno stimatore puntuale?

ESEMPIO Sia $X_1, \dots, X_n$ un campione aleatorio da una densità $\Gamma(\alpha,\lambda)$, $\theta = (\alpha,\lambda)$ incogniti. Allora:

$E(X_1) = \alpha/\lambda = \mu_1 \qquad \mu_2 = \mu_2(\alpha,\lambda) = E_\theta(X_1^2) = \frac{\alpha(\alpha+1)}{\lambda}$

Per ricavare uno stimatore usiamo quello che viene chiamato METODO DEI MOMENTI che consiste nel ricavare gli stimatori di α e λ uguagliando i momenti μ_1 e μ_2 al momento primo e secondo campionario. Si imposta, cioè, il sistema:

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n X_i = \frac{\alpha}{\lambda} \\ \frac{1}{n}\sum_{i=1}^n X_i^2 = \frac{\alpha(\alpha+1)}{\lambda} \end{cases} \Rightarrow \begin{cases} \hat{A}_n = \bar{X}_n^2 / \sum_{i=1}^n (X_i - \bar{X}_n)^2/n \\ \hat{\Lambda}_n = \bar{X}_n / \sum_{i=1}^n (X_i - \bar{X}_n)^2/n \end{cases}$$

METODO DEI MOMENTI Sia $X_1, \dots, X_n$ un campione aleatorio estratto da f_θ nota a meno di $\theta = (\theta_1, \dots, \theta_K)$. Supponiamo siano definiti i primi K momenti della popolazione:

$$\mu_j = E_\theta(X_1^j) = \begin{cases} \int_{-\infty}^{+\infty} x^j f_\theta(x)\,dx & v.a.a.c. \\ \sum_h x_h^j f_\theta(x_h) & v.a.d. \end{cases}$$

Eguagliamo i K momenti della popolazione ai primi K momenti campionari. Otteniamo:

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu_1 \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu_2 \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_i^K = \mu_K \end{cases}$$

Supponiamo che il sistema abbia soluzione $\hat{\Theta}_1, \dots, \hat{\Theta}_K$, esse sono in funzione del campione $X_1, \dots, X_n$:

$$\begin{cases} \hat{\Theta}_1 = d_1(X_1, \dots, X_n) \\ \hat{\Theta}_2 = d_2(X_1, \dots, X_n) \\ \vdots \\ \hat{\Theta}_K = d_K(X_1, \dots, X_n) \end{cases} \Rightarrow (\hat{\Theta}_1, \dots, \hat{\Theta}_K) \text{ stimatore del metodo dei momenti}$$

Se osserviamo $X_1 = x_1, \dots, X_n = x_n$ allora $\hat{\theta}_1 = d_1(x_1, \dots, x_n), \dots, \theta_K = d_K(x_1, \dots, x_n)$ è detta stima dei momenti di $\theta = (\theta_1, \dots, \theta_K)$ in corrispondenza della realizzazione campionaria $(x_1, \dots, x_n)$

CRITERIO DI MASSIMA VEROSOMIGLIANZA Sia $X_1, \dots, X_n$ un campione estratto da f_θ nota a meno di θ, $X_j \sim f_\theta$. La densità del campione è $L(\theta, \vec{x}) = \prod_{j=1}^{n} f_\theta(x_j)$. Osserviamo i dati $X_1 = \tilde{x}_1, \dots, X_n = \tilde{x}_n$ e consideriamo la funzione di θ:

$$L(\theta) = L(\theta, \vec{\tilde{x}}) = \prod_{j=1}^{n} f_\theta(\tilde{x}_j)$$

È detta verosomiglianza del campione in corrispondenza dell'osservazione $\vec{\tilde{x}}$. Se $X_1, \dots, X_n$ sono v.a.d, allora:

$$L(\theta) = \prod_{j=1}^{n} f_\theta(\tilde{x}_j) = \prod_{j=1}^{n} P_\theta(X_j = \tilde{x}_j) = P_\theta(X_1 = \tilde{x}_1, \dots, X_n = \tilde{x}_n)$$

ossia la probabilità di osservare i valori $\tilde{x}_1, \dots, \tilde{x}_n$ del campione. Se $X_1, \dots, X_n$ sono v.a.a.c. allora $L(\theta)$ non è una probabilità e parliamo di "verosomiglianza". Il criterio sarà: prendere come stima di θ quel valore che massimizza la "verosomiglianza" dei dati osservati.

DEFINIZIONE - STIMATORE M.L.E. La stima di massima verosomiglianza o stima ML in corrispondenza dell'osservazione $\vec{x} = (x_1, \dots, x_n)$ è:

$$\hat{\theta}_n = \arg_\theta \max L(\theta, \vec{x}) = \arg_\theta \max \prod_{j=1}^{n} f_\theta(x_j)$$

Ovviamente $\hat{\theta}_n$ è una funzione di $(x_1, \dots, x_n)$, diciamo $\hat{\theta}_n = t_n(x_1, \dots, x_n)$ per una opportuna t_n. Il corrispondente stimatore $\hat{\Theta}_n = t_n(X_1, \dots, X_n)$ è detto stimatore di massima verosomiglianza o M.L.E. di θ.

Supponiamo di voler stimare una caratteristica $k(\theta)$:

PROPOSIZIONE - PRINCIPIO D'INVARIANZA DEGLI M.L.E. Se $\hat{\Theta}_n = t_n(X_1, \dots, X_n)$ è lo M.L.E. di θ basato sul campione $X_1, \dots, X_n$ estratto da f_θ, allora per ogni funzione k, lo M.L.E. di $k(\theta)$ è:

$$\hat{K}_n := k(\hat{\Theta}_n) = k(t_n(X_1, \dots, X_n))$$

ESERCIZIO

1) Determinare lo stimatore M.L. del parametro (i.e. il tasso) di una popolazione esponenziale, basato su un campione di dimensione n.
2) Se osservo i valori $x_1 = 1.2$, $x_2 = 1.8$, $x_3 = 1.5$ e $x_4 = 0.7$, determinare la stima corrispondente.
3) Dedurre lo stimatore M.L. e la relativa stima della media e della varianza della popolazione.

Sia $X_1, \dots, X_n$ campione aleatorio da $\mathcal{E}(\theta)$, $\theta > 0$ incognito.

$$f_\theta(x) = \theta e^{-\theta x} I_{(0;+\infty)}(x)$$

$L(\theta) = \prod_{j=1}^{n} f_\theta(x_j) = \prod_{j=1}^{n} \theta e^{-\theta x_j} \cancel{I_{(0;+\infty)}(x)} = \theta^n e^{-\theta \sum_{j=1}^n x_j}$

1) $L(\theta) = \theta^n e^{-\theta \sum_{j=1}^n x_j}$

log-verosomiglianza: $l(\theta) = \ln L(\theta) = n \ln\theta - \theta \sum_{j=1}^n x_j$

$\hookrightarrow l'(\theta) = n/\theta - \sum_{j=1}^n x_j \geq 0 \iff \theta < n/\sum x_j \Rightarrow \hat{\theta}_n = 1/\bar{x}_j$ p.d.m.

$\Rightarrow$ $\hat{\theta}_n$ è la stima M.L. di θ in corrispondenza dell'osservazione $(x_1, \dots, x_n)$

$\Rightarrow$ MLE di θ è $\hat{\Theta}_n = 1/\bar{X}_n$

2) $x_1 = 1.2 \quad x_2 = 1.8 \quad x_3 = 1.5 \quad x_4 = 0.7$

$\Rightarrow \bar{x}_n = \dots = 1.3 \Rightarrow$ stima M.L. $\hat{\theta}_n = 1/1.3$

3) $k_1(\theta) = E_\theta(X_1) = 1/\theta \qquad k_2(\theta) = var_\theta(X_1) = 1/\theta^2$

$\hat{\Theta}_n = 1/\bar{X}_n$

$\Rightarrow$ per principio di invarianza degli MLE:

$k_1(\hat{\Theta}_n) = 1/\hat{\Theta}_n = \bar{X}_n \qquad k_2(\hat{\Theta}_n) = 1/\hat{\Theta}_n^2 = \bar{X}_n^2$

ESERCIZIO

1) Determinare lo MLE del parametro di densità di un Poisson e determinarne le proprietà.
2) Determinare la sua distribuzione asintotica

1. $L(\lambda) = \prod_{j=1}^{n} f_\lambda(x_j) = \prod_{j=1}^{n} e^{-\lambda} \lambda^{x_j}/x_j! \cancel{I_{\{0,1,2,\dots\}}(x)} =$

$= e^{-n\lambda} \dfrac{\lambda^{\sum_{j=1}^n x_j}}{\prod_{j=1}^n x_j!} \Rightarrow l(\lambda) = \ln L(\lambda) = -n\lambda + (\sum_{j=1}^n x_j) \ln\lambda - \ln(\prod_{j=1}^n x_j!)$

Cerco il massimo di $l(\lambda)$ ... $\Rightarrow \hat{\lambda}_n = \sum_{j=1}^n x_j/n$

$\Rightarrow$ MLE di λ è $\hat{\Lambda}_n = \bar{X}_n$.

1. $\bar{X}_n$ è non distorto per λ: $E_\lambda(\bar{X}_n) = E_\lambda(X_1) = \lambda$
2. $(\bar{X}_n)_n$ è debolmente consistente per λ. Considerando la legge debole di grandi numeri vale: $P(|\bar{X}_n - \lambda| > \varepsilon) \to 0$
3. $(\bar{X}_n)$ è consistente in media quadratica:
$MSE_\lambda(\bar{X}_n) = var_\lambda(\bar{X}_n) = var_\lambda(X_1)/n = \lambda/n \to 0$

2. Per il TCL: $P\left(\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}/\sqrt{n}} < t\right) \to \Phi(t) \Rightarrow \bar{X}_n \approx \mathcal{N}\left(\lambda, \frac{n}{\lambda}\right)$

PERCHÉ GLI MLE SONO MOLTO USATI Gli M.L.E sono molto usati perché con delle ipotesi abbastanza leggere possiamo garantire un po' di importanti proprietà:

"TEOREMA" Se la densità f_θ soddisfa opportune condizioni di regolarità, se $\hat{\Theta}_n = t_n(X_1, \dots, X_n)$ $n = 1, 2 \dots$ è la successione degli MLE di θ e $K(\theta)$ è una funzione differenziabile di θ, allora la successione $(K(\hat{\Theta}_n))_n$ degli MLE di $K(\theta)$ è:

1) asintoticamente non distorta
2) consistente in media quadratica
3) asintoticamente gaussiana con media $K(\theta)$ e varianza:

$$\frac{\sigma^2}{n} = \frac{(K'(\theta))^2}{n E_\theta\left(\left(\frac{\partial}{\partial \theta} \ln(f_\theta(X_1))\right)^2\right)} = \frac{(K'(\theta))^2}{n E_\theta\left(\frac{\partial^2}{\partial \theta^2} \ln f_\theta(X_1)\right)}$$

ESEMPIO Sia $X_1, X_2, \dots$ una successione di campioni estratti da una popolazione $\mathcal{E}(\theta)$, $\theta > 0$ incognito. Abbiamo visto che lo MLE di θ è $\hat{\Theta}_n = 1/\bar{X}_n$ in base al teorema precedente essendo $f_\theta(x) = \theta e^{-\theta x} I_{(0,+\infty)}(x)$. Siccome $K(\theta) > 0$ è differenziabile:

1. $E(1/\bar{X}_n) \to \theta$
2. $E((1/\bar{X}_n - \theta)^2) \to 0$
3. $1/\bar{X}_n \approx \mathcal{N}(\theta, \sigma^2(\theta)/n)$ dove $\sigma^2(\theta) = 1/E\left(\left(\frac{\partial}{\partial \theta} \ln(f_\theta(X_1))\right)^2\right)$
   $\Rightarrow \sigma^2(\theta) = \dots = \theta^2$.

ESERCIZIO Sia $X_1, \dots, X_n$ un campione aleatorio estratto da una popolazione uniforme sull'intervallo $[0, \theta]$, $\theta > 0$ parametro incognito.

1) Determinare lo MLE $\hat{\Theta}_n$ di θ basato sul campione
2) Dedurre lo MLE della media della popolazione
3) Calcolare la densità dello MLE $\hat{\Theta}_n$
4) Calcolare quindi, per ogni $\theta > 0$, la distorsione e lo MSE di $\hat{\Theta}_n$ come stimatore di θ.

5) Confrontare $\hat{\Theta}_n$ con lo stimatore di θ ottenuto con il metodo dei momenti.

$X_1, \dots, X_n$ campione da una densità $\mathcal{U}([0,\theta])$, $\theta > 0$ incognito.

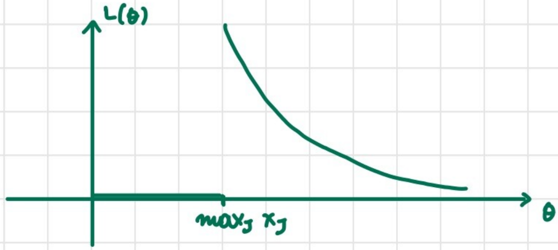
1) $x_1, \dots, x_n$ osservazione del campione

$L(\theta) = \prod_{j=1}^n f_\theta(x_j) = \prod_{j=1}^n 1/\theta^n \, I_{[0,\theta]}(x_j) = 1/\theta^n \prod_{j=1}^n I_{[0;\theta]}(x_j) =$

$\to \prod I_{[0;\theta]}(x_j) = 1 \iff I_{[0,\theta]}(x_j) = 1 \ \forall j = 1,n \iff x_j \leq \theta \ \forall j = 1$

$\to \iff \max_j x_j \leq \theta$

$= 1/\theta^n \, I_{[\max_j x_j; +\infty]}(\theta)$



$\Rightarrow$ MLE di θ è $\max\{x_1, \dots, x_n\}$

2) $E_\theta(X_1) = \int_0^\theta 1/\theta \, x \, dx = \theta/2 \Rightarrow E_\theta(X_1) =: K_1(\theta) = \theta/2$

$\Rightarrow$ per l'invarianza degli MLE lo stimator di $K_1(\theta)$ è $K_1(\hat{\Theta}_n)$

$\Rightarrow \hat{K}_n = \max\{x_1, \dots, x_n\} \cdot \frac{1}{2}$

⋮

4) $\bar{X}_n = \theta/2 \Rightarrow \hat{\Theta}_n = 2\bar{X}_n$ stimatore del metodo dei momenti

5 INTERVALLI DI CONFIDENZA

ESEMPIO Sia $X_1, \dots, X_n$ un campione aleatorio da una popolazione gaussiana $\mathcal{N}(\mu, 4)$, μ incognita. Lo MLE di μ è $\bar{X}_n$, quindi $\bar{X}_n \sim \mathcal{N}(\mu, 4/n)$. Se standardizzo, otteniamo che $\bar{X}_n - \mu / 2/\sqrt{n} \sim \mathcal{N}(0,1)$, che è una distribuzione nota. Per esempio, posso dire:

$$P_\mu\left(-1.96 < \frac{\bar{X}_n - \mu}{2/\sqrt{n}} < 1.96\right) = \Phi(1.96) - \Phi(-1.96) = 2\Phi(1.96) - 1 = 0.95 \quad \forall \mu$$

$$\Rightarrow 0.95 = P_\mu(-1.96 < \% < 1.96) = P_\mu\left(\bar{X}_n - \frac{2}{\sqrt{n}} 1.96 < \mu < \frac{2}{\sqrt{n}} 1.96 + \bar{X}_n\right) \ \forall \mu$$

Possiamo concludere che:

1. Qualunque sia il "vero" valore di μ, la $\bar{X}_n$ è ad una distanza da μ inferiore a $2/\sqrt{n}\, 1.96$ con probabilità minore di

0.95

2. $T_1 = t_1(X_1, \dots, X_n) = \bar{X}_n - 1.96 \frac{2}{\sqrt{n}}$ e $T_2 = t_2(X_1, \dots, X_n) = \bar{X}_n + \frac{2}{\sqrt{n}} 1.96$ sono due statistiche tali che $T_1 < T_2$ e considerato l'intervallo "aleatorio" (T_1, T_2) è tale che $P_\mu(\mu \in (T_1, T_2)) = 0.95$ $\forall \mu$
3. Se osserviamo $X_1 = x_1, \dots, X_n = x_n$ e $\bar{X}_n = \bar{x}_n$, allora $\bar{x}_n - \frac{2}{\sqrt{n}} 1.96 < \mu < \bar{x}_n + \frac{2}{\sqrt{n}} 1.96$ è un intervallo reale

L'intervallo aleatorio $(\bar{X}_n - \frac{2}{\sqrt{n}} 1.96, \bar{X}_n + \frac{2}{\sqrt{n}} 1.96)$ è detto intervallo di confidenza per μ al 95% (a livello di confidenza 0.95). Se osservo $\bar{x}_n$, allora l'intervallo reale $(\bar{x}_n - \frac{2}{\sqrt{n}} 1.96, \bar{x}_n + \frac{2}{\sqrt{n}} 1.96)$ è chiamata stima intervallare di confidenza 95% e diremo con CONFIDENZA (non probabilità) 95% μ appartiene a questo intervallo. La notazione alternativa è: $\mu \in \bar{x}_n \mp \frac{2}{\sqrt{n}} 1.96$.

DEFINIZIONE - INTERVALLO DI CONFIDENZA BILATERO Sia $X_1, \dots, X_n$ un campione aleatorio estratto da una popolazione con densità f_θ dipendente da un parametro incognito (o vettore di incognite) θ. Sia $K(\theta)$ una caratteristica della popolazione e sia $\alpha \in (0,1)$ fissato. Siano $T_1 = t_1(X_1, \dots, X_n)$ e $T_2 = t_2(X_1, \dots, X_n)$ due statistiche tali che $T_1 < T_2$ e per le quali:

$$P_\theta(T_1 < K(\theta) < T_2) = 1 - \alpha \quad \forall \theta$$

Allora (T_1, T_2) è detto intervallo all'$(1-\alpha)100\%$ per $K(\theta)$ e $1-\alpha$ il livello di confidenza.

Questi intervalli sono detti bilaterali perché delimitati da due statistiche. Successivamente incontreremo anche intervalli unilaterali.

Se osservo $X_1 = x_1, \dots, X_n = x_n$ e se $\bar{t}_1 = t_1(x_1, \dots, x_n)$ e $\bar{t}_2 = t_2(x_1, \dots, x_n)$ sono i valori corrispondenti all'osservazione campionaria delle statistiche T_1 e T_2, allora l'intervallo reale $(\bar{t}_1, \bar{t}_2)$ è detto stima intervallare (o intervallo di confidenza) di $K(\theta)$ con livello di confidenza $1-\alpha$ in corrispondenza dell'osservazione $(x_1, \dots, x_n)$. Diremo che con confidenza $1-\alpha$

$K(\theta)$ è $(\bar{t}_1, \bar{t}_2)$.

5.1 Metodo della quantità pivotale

DEFINIZIONE - QUANTITÀ PIVOTALE Sia $X_1, \dots, X_n$ un campione aleatorio estratto da una densità f_θ, θ incognito. Si definisce quantità pivotale una v.a. $Q = q(X_1, \dots, X_n, \theta)$ funzione del campione e del parametro incognito la cui distribuzione non dipende dal parametro incognito θ.

Poiché la distribuzione di Q non dipende da θ, allora per ogni $\alpha \in (0,1)$ posso determinare q_1 e q_2 che dipendono da α ma non dipendono da θ:

$$1-\alpha = P_\theta(q_1 < Q < q_2)$$

Supponiamo che per ogni realizzazione campionaria $X_1 = x_1, \dots, X_n = x_n$:

$$q_1 < q(x_1, \dots, x_n, \theta) < q_2 \iff t_1(x_1, \dots, x_n) < K(\theta) < t_2(x_1, \dots, x_n)$$

per opportune t_1 e t_2. Segue che:

$$\begin{aligned} 1-\alpha &= P_\theta(q_1 < q(X_1, \dots, X_n, \theta) < q_2) = \\ &= P_\theta(\underbrace{t_1(X_1, \dots, X_n)}_{T_1} < K(\theta) < \underbrace{t_2(X_1, \dots, X_n)}_{T_2}) \qquad \forall \theta \end{aligned}$$

Quindi (T_1, T_2) è un i.c. all' $(1-\alpha)100\%$ per $K(\theta)$.

5.2 Intervalli di confidenza unilateri

DEFINIZIONE - INTERVALLO DI CONFIDENZA ILLIMITATO SUPERIORMENTE

Sia $T_1 = t_1(X_1, \dots, X_n)$ tale che:

$$P_\theta(T_1 < K(\theta)) = 1-\alpha \qquad \forall \theta, \ \alpha \in (0,1)$$

Allora $(T_1, +\infty)$ è detto intervallo di confidenza unilatero (non limitato superiormente) all' $(1-\alpha)100\%$ per $K(\theta)$.

Se osservo $X_1 = x_1, \dots, X_n = x_n$ e sia $\bar{t}_1 = t_1(x_1, \dots, x_n)$, allora l'intervallo dell'asse reale $(\bar{t}_1, +\infty)$ è detto stima intervallare unilatera (o intervallo di confidenza unilatero) di $K(\theta)$ con livello di confidenza $1-\alpha$ in corrispondenza dell'osservazione campionaria $(x_1, \dots, x_n)$. Diremo che con confidenza $1-\alpha$ vale $K(\theta) \in (\bar{t}_1, +\infty)$.

Vale il simmetrico per gli intervalli di confidenza illimitati superiormente.

5.3 Intervalli di confidenza per media di una popolazione $\mathcal{N}(\mu, \sigma^2)$

$\sigma^2 = \sigma_0^2$ NOTA Consideriamo $X_1, \dots, X_n$ i.i.d con $X_j \sim \mathcal{N}(\mu, \sigma_0^2)$. La media campionaria sarà $\bar{X}_n \sim \mathcal{N}(\mu, \sigma_0^2/n)$. Otteniamo che la quantità pivotale sarà:

$$\frac{\bar{X}_n - \mu}{\sigma_0/\sqrt{n}} \sim \mathcal{N}(0,1)$$

Se $\alpha \in (0,1)$, per ogni μ possiamo scrivere:

$$1-\alpha = P_\mu\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu}{\sigma_0/\sqrt{n}} < z_{\alpha/2}\right) = $$ per proprietà dei quantili della gaussiana

$$= P_\mu\left(\bar{X}_n - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} < \mu < \bar{X}_n + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right)$$

Otteniamo quindi che $\left(\bar{X}_n - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} ; \bar{X}_n + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right)$ è l'i.c. bilatero per la media all'$(1-\alpha)100\%$. Per gli intervalli unilateri con passaggi analoghi ai precedenti otteniamo:

$$1-\alpha = P_\mu\left(\frac{\bar{X}_n - \mu}{\sigma_0/\sqrt{n}} > -z_\alpha\right) = P_\mu\left(\mu < \bar{X}_n + z_\alpha \frac{\sigma_0}{\sqrt{n}}\right)$$

$\Rightarrow \left(-\infty ; \bar{X}_n + z_\alpha \frac{\sigma_0}{\sqrt{n}}\right)$ i.c. unilatero ill. inf. per μ all'$(1-\alpha)100\%$

$$1-\alpha = P_\mu\left(\frac{\bar{X}_n - \mu}{\sigma_0/\sqrt{n}} < z_\alpha\right) = P_\mu\left(\mu > \bar{X}_n - z_\alpha \frac{\sigma_0}{\sqrt{n}}\right)$$

$\Rightarrow \left(\bar{X}_n - z_\alpha \frac{\sigma_0}{\sqrt{n}} ; +\infty\right)$ i.c. unilatero ill. sup. per μ all'$(1-\alpha)100\%$.

σ^2 NON NOTA La quantità pivotale trovata precedentemente non va bene perché dipende da σ^2, che sarebbe una incognita. Le nostre incognite saranno $\theta = (\mu, \sigma^2)$ con $K(\theta) = \mu$. Sappiamo che

$\bar{X}_n - \mu / S_n/\sqrt{n} \sim t(n-1) \Rightarrow$ quantità pivotale

$\hookrightarrow S_n = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X}_n)^2}$

Per quanto riguarda gli intervalli bilateri troviamo:

$$1-\alpha = P_{(\mu,\sigma^2)}\left(-t_{\alpha/2,n-1} < \bar{X}_n - \mu / S_n/\sqrt{n} < t_{\alpha/2,n-1}\right) =$$

proprietà dei quantili della T di student

$$= P_{(\mu,\sigma^2)}\left(\bar{X}_n - t_{\alpha/2,n-1} S_n/\sqrt{n} < \mu < \bar{X}_n + t_{\alpha/2,n-1} S_n/\sqrt{n}\right)$$

$\Rightarrow \bar{X}_n \mp t_{\alpha/2,n-1} S_n/\sqrt{n}$ i.c. bilatero per μ all'$(1-\alpha)100\%$.

Mentre per gli unilateri:

$$1-\alpha = P_{(\mu,\sigma^2)}\left(\bar{X}_n - \mu / S_n/\sqrt{n} < t_{\alpha,n-1}\right) = P_{(\mu,\sigma^2)}\left(\bar{X}_n - \mu / S_n/\sqrt{n} > -t_{\alpha,n-1}\right) =$$
$$= P_{(\mu,\sigma^2)}\left(\mu > \bar{X}_n - t_{\alpha,n-1} S_n/\sqrt{n}\right)$$

$\Rightarrow (\bar{X}_n - t_{\alpha,n-1} S_n/\sqrt{n}, +\infty)$ i.c. ill. sup. per μ all'$(1-\alpha)100\%$.

$$1-\alpha = \dots = P_{(\mu,\sigma^2)}\left(\mu < \bar{X}_n + t_{\alpha,n-1} S_n/\sqrt{n}\right)$$

$\Rightarrow (-\infty, \bar{X}_n + t_{\alpha,n-1} S_n/\sqrt{n})$ i.c. ill. inf. per μ all'$(1-\alpha)100\%$

5.4 Intervalli di confidenza per la varianza di una popolazione $\mathcal{N}(\mu,\sigma^2)$

μ INCOGNITA Consideriamo sempre $X_1, \dots, X_n$ i.i.d $X_j \sim \mathcal{N}(\mu,\sigma^2)$ con $\theta = (\mu, \sigma^2)$ incognite. Sappiamo che:

$(n-1) S_n^2/\sigma^2 \sim \chi^2(n-1) \Rightarrow$ quantità pivotale

Per intervalli bilateri:

$$1-\alpha = P_{(\mu,\sigma^2)}\left(\chi^2_{1-\alpha/2,n-1} < (n-1)S_n^2/\sigma^2 < \chi^2_{\alpha/2,n-1}\right) =$$

per proprietà dei quantili di $\chi^2(n-1)$

$$= P_{(\mu,\sigma^2)}\left((n-1)S_n^2/\chi^2_{\alpha/2,n-1} < \sigma^2 < (n-1)S_n^2/\chi^2_{1-\alpha/2,n-1}\right)$$

$\Rightarrow \left(\frac{(n-1)S_n^2}{\chi^2_{\alpha/2,n-1}} ; \frac{(n-1)S_n^2}{\chi^2_{1-\alpha/2,n-1}}\right)$ i.c. per σ^2 all'$(1-\alpha)100\%$.

Per intervalli unilateri:

$$1-\alpha = P_{(\mu,\sigma^2)}\left((n-1)S_n^2/\sigma^2 > \chi^2_{1-\alpha,n-1}\right) =$$
$$= P_{(\mu,\sigma^2)}\left(\sigma^2 < (n-1)S_n^2/\chi^2_{1-\alpha,n-1}\right)$$

$\Rightarrow \left(0; \, ^{(n-1)S_n^2}/\chi^2_{1-\alpha, n-1}\right)$ i.c. unilatero per σ^2 all'$(1-\alpha)100\%$

$$1-\alpha = P_{(\mu,\sigma^2)}\left(^{(n-1)S_n^2}/\sigma^2 < \chi^2_{\alpha, n-1}\right) =$$
$$= P_{(\mu,\sigma^2)}\left(\sigma^2 > \, ^{(n-1)S_n^2}/\chi^2_{\alpha, n-1}\right)$$

$\Rightarrow \left(^{(n-1)S_n^2}/\chi^2_{\alpha, n-1} ; +\infty\right)$ i.c. ill. sup. per σ^2 all'$(1-\alpha)100\%$

$\mu = \mu_0$ NOTA Supponiamo $\mu = \mu_0$ fissata reale. $X_1, \dots, X_n$ i.i.d $X_j \sim \mathcal{N}(\mu_0, \sigma^2)$ con $\theta = \sigma^2$ incognita.

$$^{X_i - \mu_0}/\sigma \sim \mathcal{N}(0,1) \Rightarrow \sum_{i=1}^n \, ^{(X_i-\mu_0)^2}/\sigma^2 \sim \chi^2(n)$$

Ricordandoci che abbiamo indicato con $S_{0n}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2$ lo stimatore della varianza con media nota pari a μ_0, possiamo scrivere la seguente quantità pivotale:

$$\sum_{i=1}^n \, ^{(X_i-\mu_0)^2}/\sigma^2 = \, ^{nS_{0n}^2}/\sigma^2 \sim \chi^2(n)$$

Quindi per intervalli bilateri abbiamo:

$$1-\alpha = P_{\sigma^2}\left(\chi^2_{1-\alpha/2, n} < \, ^{nS_{0n}^2}/\sigma^2 < \chi^2_{\alpha/2, n}\right)$$
$$= P_{\sigma^2}\left(^{nS_{0n}^2}/\chi^2_{\alpha/2, n} < \sigma^2 < \, ^{nS_{0n}^2}/\chi^2_{1-\alpha/2, n}\right)$$

$$\Rightarrow \left(\frac{nS_{0n}^2}{\chi^2_{\alpha/2, n}} ; \frac{nS_{0n}^2}{\chi^2_{1-\alpha/2, n}}\right) \quad \text{i.c. per } \sigma^2 \text{ all'}(1-\alpha)100\%$$

Per gli intervalli unilateri:

$$1-\alpha = P_{\sigma^2}\left(^{nS_{0n}^2}/\sigma^2 > \chi^2_{1-\alpha, n}\right) = P_{\sigma^2}\left(\sigma^2 < \, ^{nS_{0n}^2}/\chi^2_{1-\alpha, n}\right)$$

$\Rightarrow \left(0; \, ^{nS_{0n}^2}/\chi^2_{1-\alpha, n}\right)$ i.c. ill. inf. per σ^2 all'$(1-\alpha)100\%$

$$1-\alpha = P_{\sigma^2}\left(^{nS_{0n}^2}/\sigma^2 < \chi^2_{\alpha, n}\right) = P_{\sigma^2}\left(\sigma^2 > \, ^{nS_{0n}^2}/\chi^2_{\alpha, n}\right)$$

$\Rightarrow \left(^{nS_{0n}^2}/\chi^2_{\alpha, n} ; +\infty\right)$ i.c. ill. sup. per σ^2 all'$(1-\alpha)100\%$

5.5 Intervallo di confidenza per la media di popolazione esponenziale

ESERCIZIO Sia $X_1, \dots, X_n$ un campione aleatorio estratto da una popolazione esponenziale di media θ incognita.

1) Determinare lo MLE $\hat{\Theta}_n$ di θ.
2) Determinare la f.g.m di $2n\hat{\Theta}_n/\theta$ e dedurre che è una quantità pivotale
3) Usare il risultato del punto 2 per ottenere un intervallo di confidenza bilatero di livello $1-\alpha$ per θ.
4) Supponiamo ora che gli oggetti prodotti da una azienda abbiano tempi di vita (espressi in ore) indipendenti ed esponenziali di uguale media incognita. Se la somma dei tempi di 10 esemplari è 1740 ore che stima intervallare bilatera al 95% ne risulta per la media della popolazione.

Sia $X_1, \dots, X_n$ campione aleatorio $\mathcal{E}(1/\theta)$

1. Lo MLE di θ è $\bar{X}_n$

2. $2n\hat{\Theta}/\theta = 2n\bar{X}_n/\theta = 2/\theta \sum_{i=1}^n X_i$

$$m_{2/\theta \sum X_i}(t) = E_\theta\left(e^{t2/\theta \sum X_i}\right) = E_\theta\left(\prod_{i=1}^n e^{2tX_i/\theta}\right) =$$
$$= \prod_{i=1}^n E_\theta\left(e^{2tX_i/\theta}\right) = \prod_{i=1}^n m_{X_i}(2t/\theta) = \left(m_{X_1}(2t/\theta)\right)^n =$$
$$= \left(\frac{1/\theta}{1/\theta - 2t/\theta}\right)^n = \left(\frac{1}{1-2t}\right)^{\frac{2n}{2}} \Rightarrow \text{f.g.m di } \chi^2(2n)$$
$\hookrightarrow 2t/\theta < 1/\theta \Rightarrow t < \frac{1}{2}$

$\Rightarrow 2n\bar{X}_n/\theta \sim \chi_i(2n)$ quantità pivotale

3. $$1-\alpha = P_\theta\left(\chi^2_{1-\alpha/2,2n} < 2n\bar{X}_n/\theta < \chi^2_{\alpha/2,2n}\right)$$
$$= P_\theta\left(2n\bar{X}_n/\chi^2_{\alpha/2,2n} < \theta < 2n\bar{X}_n/\chi^2_{1-\alpha/2,n}\right)$$
$\Rightarrow \left(\frac{2n\bar{X}_n}{\chi^2_{\alpha/2,2n}} ; \frac{2n\bar{X}_n}{\chi^2_{1-\alpha/2,2n}}\right)$ i.c. per θ all'$(1-\alpha)100\%$.

4. $\sum_{i=1}^n X_i = 1740$ per $n=10$, $1-\alpha = 0{,}95 \Rightarrow \alpha/2 = 0.025$

$\Rightarrow \chi^2_{0.025,20} \cong 34.170 \quad \chi^2_{0.975,20} \cong 9.591$

$\Rightarrow$ con confidenza 95% $\theta \in (101.84 ; 362.84)$

5.6 Intervallo di confidenza approssimato per la media di una popolazione bernoulliana

Consideriamo una popolazione di oggetti che possiedono un certo requisito con probabilità p incognita indipendenti l'uno dall'altro. Si testano n oggetti della popolazione:

$$X := n° \text{ di oggetti tra gli } n \text{ testati che soddisfano il requisito}$$

$$X_1, \dots, X_n, \quad X_j \sim Be(p) \Rightarrow X_1 + \dots + X_n \sim Bi(n, p)$$

Se prendiamo un n molto grande, per il TCL:

$$X - np / \sqrt{np(1-p)} \approx \mathcal{N}(0,1)$$

Allora per ogni $p \in (0,1)$

$$1-\alpha \approx P_p\left(-z_{\alpha/2} < X - np / \sqrt{np(1-p)} < z_{\alpha/2}\right)$$

Se osservo $X = \tilde{x}$, ottengo la regione di confidenza:

$$\left\{ p : -z_{\alpha/2} \leq \tilde{x} - np / \sqrt{np(1-p)} < z_{\alpha/2} \right\}$$

Se voglio ottenere un vero e proprio intervallo, considero la statistica $\hat{p} = X/n = \bar{X}_n$:

$$\frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} = \frac{X/n - p}{\sqrt{\hat{p}(1-\hat{p})/n}} = \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} = \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)/n}}$$

Si può dimostrare che per n grande vale che:

$$\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \approx \mathcal{N}(0,1) \Rightarrow \text{asintoticamente è quantità pivotale}$$

Possiamo quindi scrivere un intervallo di confidenza bilatero approssimato:

$$1-\alpha = P_p\left(-z_{\alpha/2} < \hat{p}-p/\sqrt{\hat{p}(1-\hat{p})/n} < z_{\alpha/2}\right) =$$
$$= P_p\left(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}\right)$$

$$\Rightarrow \left(\max\left\{\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}, 0\right\}; \max\left\{\hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}, 1\right\}\right)$$

poiché $p \in (0,1)$

6 VERIFICA D'IPOTESI

Supponiamo di disporre di un campione aleatorio proveniente da una distribuzione nota a meno di uno o più parametri incogniti. Come possiamo utilizzare il campione per verificare qualche ipotesi sul campione?

DEFINIZIONE - IPOTESI STATISTICA Una ipotesi statistica è una affermazione su uno o più parametri della distribuzione della popolazione.

DEFINIZIONE - TEST PER LA VERIFICA D'IPOTESI Data un'ipotesi statistica, un test per la verifica di questa ipotesi è una procedura per determinare se i valori di un campione aleatorio e l'ipotesi sono compatibili oppure no.

Se secondo un test il campione sarà giudicato compatibile con l'ipotesi considerata diremo che quest'ultima è accettata, altrimenti rifiutata.

Consideriamo una popolazione avente distribuzione F_θ che dipende da un parametro incognito θ e supponiamo di voler verificare una qualche ipotesi su θ che chiameremo ipotesi nulla e indicheremo con H_0.

DEFINIZIONE - IPOTESI SEMPLICE E COMPOSTA Una ipotesi statistica si dice semplice se caratterizza completamente la distribuzione della popolazione, altrimenti è detta composta.

6.1 Regione critica e livello di significatività di un test

Sia $X_1, \dots, X_n$ un campione aleatorio estratto da F_θ. Supponiamo di voler usare il campione per eseguire una verifica o test su certa ipotesi H_0. Poiché dobbiamo decidere se accettare o meno H_0 basandoci esclusivamente sui valori assunti dal campione, il test sarà definito da quello che viene detta regione critica:

DEFINIZIONE - REGIONE CRITICA Un test per la verifica di una ipotesi H_0 ha regione critica $C \subseteq \mathbb{R}^n$ se, avendo osservato $X_1 = x_1, \dots, X_n = x_n$, accetto H_0 se $(x_1, \dots, x_n) \notin C$ e rifiuto H_0 altrimenti.

ESEMPIO Sia $X_1, \dots, X_n$ un campione aleatorio estratto da $\mathcal{N}(\theta, 4)$. Vogliamo testare $H_0: \theta = 1$. Una regione critica è

$$C = \left\{ (x_1, \dots, x_n) : \left| \frac{1}{n} \sum x_i - 1 \right| \geq K \right\}$$

per K opportuno.

È importante notare che in un qualunque test per verificare un'ipotesi nulla H_0 si può sbagliare in due modi:

- errore di 1° specie: i dati ci portano a rifiutare H_0 e H_0 è corretta
- errore di 2° specie: i dati ci portano ad accettare H_0 e H_0 è falsa.

Non vi è simmetria tra i due tipi di errore. L'obiettivo della verifica d'ipotesi non è quello di dire se l'ipotesi H_0 sia vera o falsa, ma di stabilire se l'ipotesi sia compatibile con i dati raccolti. Quindi vi è un'ampia tolleranza nell'accettare H_0, mentre per rifiutarla occorre che i dati campionari siano molto improbabili quando H_0 è soddisfatta.

Questo bilanciamento si ottiene specificando una soglia α per l'errore di prima specie:

DEFINIZIONE - LIVELLO DI SPECIFICITÀ DI UN TEST Si dice che un test ha specificità (almeno) α se:

$$P(\text{errore di } 1° \text{ specie}) = P_{\theta \in H_0}(\text{rifiutare } H_0) \leq \alpha$$

Valori tipici di α sono 0.1, 0.05 e 0.01.

DEFINIZIONE - REGIONE CRITICA DI LIVELLO α La regione critica C di un test è detta livello di significatività α se

$$\sup_{\theta \in H_0}((X_1, \dots, X_n) \in C) \leq \alpha$$

Quindi il livello di significatività di un test è la soglia per la probabilità di rifiutare H_0, calcolata supponendo H_0 vera.

6.2 Ipotesi nulla e ipotesi alternativa

In generale un problema di verifica d'ipotesi si traduce nel decidere tra due ipotesi in competizione: $H_0: \theta \in \Theta_0$ contro $H_1: \theta \in \Theta_1$ ($\Theta_0 \cap \Theta_1 \neq \emptyset$) sulla base dei dati $x_1, \dots, x_n$.

DEFINIZIONE - IPOTESI ALTERNATIVA Si dice che H_1 è l'ipotesi alternativa se è quella che si accetta quando si rifiuta H_0.

6.3 Curva OC, probabilità dell'errore di 2° specie e potenza

DEFINIZIONE - CURVA OC Dato il test di regione critica C, si definisce curva OC la funzione del parametro incognito θ:

$$\beta(\theta) := P_\theta(\text{accettare } H_0) = P(X_1, \dots, X_n \notin C)$$

OC: Operating Characteristic

Se $\theta \in H_1$, $\beta(\theta)$ rappresenta una probabilità di un errore di 2° specie.

DEFINIZIONE - FUNZIONE DI POTENZA Si definisce potenza del test la funzione del parametro incognito θ, per θ che soddisfa H_1

$\pi(\theta) := P_\theta(\text{rifiutare } H_0) = 1 - \beta(\theta)$

6.4 Come costruiamo la regione critica di un test?

Supponiamo di voler verificare l'ipotesi nulla $H_0 : \theta \in \Theta_0$ con Θ_0 è un insieme di valori ammissibili per il parametro. Un approccio naturale è individuare uno stimatore puntuale $d(X_1, \dots, X_n)$ di θ e quindi rifiutare l'ipotesi quando lo stimatore è "lontano" da Θ_0.

Quanto deve essere "lontano" lo stimatore da Θ_0 per giustificare un rifiuto di H_0 ad un livello di significatività α? Occorre conoscere la distribuzione di $d(X_1, \dots, X_n)$ quando H_0 è vera. Questa conoscenza permette di determinare la regione critica del test, cioè la regione di lontananza in modo tale che la probabilità dell'errore di 1° specie sia minore o uguale ad α.

6.5 Verifica d'ipotesi per la media di una popolazione gaussiana quando la varianza è nota (test z)

Sia $X_1, \dots, X_n$ un campione estratto da una popolazione $\mathcal{N}(\mu, \sigma^2)$ con varianza σ^2 nota.

TEST BILATERO Fissata una costante μ_0 verificare l'ipotesi nulla $H_0 : \mu = \mu_0$ contro l'ipotesi alternativa $H_1 : \mu \neq \mu_0$. Essendo $\bar{X}_n$ lo MLE di μ appare ragionevole rifiutare H_0 quando $\bar{X}_n$ è lontana da μ_0, cioè considerare una regione critica del test tipo:

$$C := \{(x_1, \dots, x_n) : |\bar{x}_n - \mu_0| \geq K\} \quad \text{con} \quad \bar{x}_n = x_1 + \dots + x_n / n$$

Vogliamo che il test abbia livello di significatività α, quindi fissiamo K in modo tale che la probabilità dell'errore di 1° specie sia α:

$$\alpha \geq P(\text{errore di } 1° \text{ specie}) = P_{\mu \in H_0}((X_1, \dots, X_n) \in C) =$$
$$P_{\mu_0}(|\bar{X}_n - \mu_0| \geq K)$$

Se $\mu = \mu_0$, allora $\bar{X}_n \sim \mathcal{N}(\mu_0, \sigma_0^2/n)$ e quindi:

$$\alpha = P_{\mu_0}\left(\frac{|\bar{X}_n - \mu_0|}{\sigma_0/\sqrt{n}} \geq \frac{K\sqrt{n}}{\sigma_0} \right) =$$
$$= P_{\mu_0}\left(|Z| \geq \frac{K\sqrt{n}}{\sigma_0} \right) = 2P\left(Z \geq \frac{K\sqrt{n}}{\sigma_0} \right) \Rightarrow P\left(Z \geq \frac{K\sqrt{n}}{\sigma_0} \right) = \frac{\alpha}{2}$$

$$\iff \frac{K\sqrt{n}}{\sigma_0} = z_{\alpha/2} \Rightarrow C = \left\{ (x_1, \dots, x_n) : \frac{|\bar{x}_n - \mu_0|\sqrt{n}}{\sigma_0} \geq z_{\alpha/2} \right\}$$

Possiamo concludere che un test di livello di significatività α è quello che, avendo osservato il valore della media campionaria $\bar{X}_n = \bar{x}_n$:

- rifiuta H_0 se $|\bar{x}_n - \mu_0|/\sigma_0 \geq z_{\alpha/2}$
- accetta H_0 se $|\bar{x}_n - \mu_0|/\sigma_0 \not\geq z_{\alpha/2}$

$(\bar{X}_n - \mu_0)\sqrt{n}/\sigma_0$ è detta statistica test. Rifiuterò H_0 se il modulo del valore osservato della statistica test è maggiore o uguale al quantile $z_{\alpha/2}$.

ESERCIZIO Un segnale di valore μ trasmesso da una sorgente A, viene raccolto dal ricevente B con un rumore gaussiano di media nulla e varianza 4, il segnale ricevuto da B è quindi una v.a. gaussiana $\mathcal{N}(\mu, 4)$. Per ridurre il rumore, viene inviato 5 volte e la media campionaria dei valori ricevuti è $\bar{X}_5 = 9.5$. Si sa inoltre che B aveva motivo di supporre che il valore inviato dove essere 8. Si verifichi questa ipotesi.

Si tratta di verificare $H_0: \mu = 8$ contro $H_1: \mu \neq 8$. La regione critica di un test di livello $\alpha = 0.05$ è:

$$C = \left\{ (x_1, \dots, x_5) : \frac{|\bar{x}_5 - 8|\sqrt{5}}{2} \geq z_{0.025} \simeq 1.96 \right\}$$

Quindi rifiuto H_0 se il valore osservato $\bar{X}_5 = 9.5$ sta in C:

$$\frac{|\bar{x}_5 - 8|\sqrt{5}}{2} \simeq 1.68 < 1.96 \Rightarrow \text{accetto}$$

+ α è grande, + rifiuto

Se $\alpha = 0.1$, $z_{\alpha/2} \simeq 1.645$ $\quad \frac{|\bar{x}_5 - 8|\sqrt{5}}{2} \simeq 1.68 > 1.645 \Rightarrow$ rifiuto

Come possiamo scegliere il livello di significatività giusto?
Dipende dal problema.

P-VALUE Come visto nel precedente esempio, esiste un valore critico del livello di significatività α, detto p-value, al di sopra del quale rifiuto H_0 con i dati a disposizione:

$$|\bar{X}_n - \mu_0|\sqrt{n}/\sigma_0 = |9.5-8|\sqrt{5}/4 \cong 1.68 \quad \begin{cases} \forall \alpha : z_{\alpha/2} \leq 1.68 \Rightarrow \text{rifiuto} \\ \forall \alpha : z_{\alpha/2} > 1.68 \Rightarrow \text{accetto} \end{cases}$$

DEFINIZIONE - P-VALUE (P DEI DATI) Data una famiglia di test al variare del livello di significatività α, si definisce p-value il più piccolo valore della significatività per cui rifiuto H_0 con i dati a disposizione.

CURVA OC Per il test considerato calcoliamo curva OC:

$$\begin{aligned}\beta(\mu) :&= P_\mu(\text{accettare } H_0) = P_\mu\left(-z_{\alpha/2} < (\bar{X}_n-\mu_0)\sqrt{n}/\sigma_0 < z_{\alpha/2}\right) = \\ &= P_\mu\left(-z_{\alpha/2} < (\bar{X}_n-\mu)\sqrt{n}/\sigma_0 + (\mu-\mu_0)\sqrt{n}/\sigma_0 < z_{\alpha/2}\right) = \\ &= P_\mu\left(-z_{\alpha/2} + (\mu_0-\mu)\sqrt{n}/\sigma_0 < (\bar{X}_n-\mu)\sqrt{n}/\sigma_0 < z_{\alpha/2} + (\mu_0-\mu)\sqrt{n}/\sigma_0\right) = \\ &= \Phi\left(z_{\alpha/2} + (\mu_0-\mu)\sqrt{n}/\sigma_0\right) - \Phi\left(-z_{\alpha/2} + (\mu_0-\mu)\sqrt{n}/\sigma_0\right)\end{aligned}$$

ESERCIZIO Dati $\mu_0 = 8$, $\sigma_0^2 = 4$, $n = 5$, $\alpha = 0.05$ e $z_{\alpha/2} = 1.96$ calcolare il valore della potenza del test $\mu = 10$

$$\begin{aligned}\pi(10) &= 1 - \beta(10) = 1 - \left[\Phi\left(z_{\alpha/2} + (\mu-\mu_0)\sqrt{n}/\sigma_0\right) - \Phi\left(-z_{\alpha/2} + (\mu_0-\mu)\sqrt{n}/\sigma_0\right)\right] = \\ &= 1 - [\Phi(1.96 - \sqrt{5}) - \Phi(-1.96 - \sqrt{5}) \simeq 0.609\end{aligned}$$

La curva OC può essere usata per dimensionare il campione in modo tale che l'errore di 2° specie soddisfi delle condizioni specifiche.

ESEMPIO Determinare n in modo tale che la probabilità di accettare $H_0: \mu = \mu_0$ quando il vero valore è $\mu = \mu_1$ sia minore o uguale a β:

$$\beta(\mu_1) = \Phi\left(\frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + z_{\alpha/2} \right) - \Phi\left(\frac{\mu_0-\mu_1}{\sigma_0/\sqrt{n}} - z_{\alpha/2} \right) \leq \beta$$

Supponiamo $\mu_1 > \mu_0$. Osserviamo che:

$$\beta(\mu_1) \leq \Phi\left(\frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + z_{\alpha/2} \right) \leq \beta$$
$$\iff \frac{\mu_0-\mu_1}{\sigma_0/\sqrt{n}} + z_{\alpha/2} < -z_\beta \iff n \geq \sigma_0^2 \left(\frac{z_{\alpha/2} + z_\beta}{\mu_0 - \mu_1} \right)^2$$

Anche nel caso $\mu_1 < \mu_0$ si ottiene la stessa formula.

TEST UNILATERALE Consideriamo $H_0 : \mu = \mu_0$ contro $H_1 : \mu > \mu_0$. Calcoliamo la statistica test, la regione critica e il p-value. Sappiamo che $\bar{X}_n$ è lo MLE di μ e, poiché $H_1 : \mu > \mu_0$, valori bassi di $\bar{X}_n$ sono improbabili quando H_0 è falsa. Rifiuto, quindi, H_0 se $\bar{X}_n > \mu_0 + K$ con K determinato in base al livello di significatività:

$$\alpha = P_{\mu_0}(\bar{X}_n > \mu_0 + K) = P_{\mu_0}\left(\frac{\bar{X}_n - \mu_0}{\sigma_0}\sqrt{n} > \frac{K\sqrt{n}}{\sigma_0} \right) = P\left(Z > \frac{K\sqrt{n}}{\sigma_0} \right)$$
$$\Rightarrow \frac{K\sqrt{n}}{\sigma_0} = z_\alpha$$

$$\Rightarrow \alpha = P_{\mu_0}\left(\frac{(\bar{X}_n - \mu_0)\sqrt{n}}{\sigma_0} \geq z_\alpha \right) \Rightarrow C_\alpha = \left\{ (x_1, \dots, x_n) : \frac{(\bar{x}_n - \mu_0)\sqrt{n}}{\sigma_0} \geq z_\alpha \right\}$$

$$\Rightarrow Z^* = \frac{(\bar{X}_n - \mu_0)\sqrt{n}}{\sigma_0}$$

ESERCIZIO

1. Mostrare che se osservo il valore z^* della statistica test $\frac{(\bar{x}_n - \mu_0)\sqrt{n}}{\sigma_0}$, allora $\text{p-value} = P(Z \geq z^*) = 1 - \Phi(z^*)$
2. Determinare la curva OC del test di livello α

1. $C_\alpha = \left\{ (x_1, \dots, x_n) : \frac{(\bar{x}_n - \mu_0)\sqrt{n}}{\sigma_0} \geq z_\alpha \right\}$

Prendiamo $\alpha < 1 - \Phi(z^*) = P(Z \geq z^*) \Rightarrow z^* < z_\alpha \Rightarrow$ accetto con i dati a disposizione.

Supponiamo ora $\alpha \geq 1 - \Phi(z^*) \Rightarrow z^* \geq z_\alpha \Rightarrow$ rifiuto con i dati a disposizione

$\Rightarrow \text{p-value} = 1 - \Phi(z^*)$

2. $\beta(\mu) = P_\mu(\text{accettare } H_0) = P_\mu((X_1, \dots, X_n) \in C^c) =$
$= P_\mu\left((\bar{x}_n - \mu_0)\sqrt{n}/\sigma_0 < z_\alpha\right) = P_\mu\left((\bar{x}_n - \mu + \mu - \mu_0)\sqrt{n}/\sigma_0 < z_\alpha\right) =$
$= P_\mu\left((\bar{x}_n - \mu)\sqrt{n}/\sigma_0 < z_\alpha + (\mu_0 - \mu)\sqrt{n}/\sigma_0\right)$
$= P_\mu\left(Z < z_\alpha + (\mu_0 - \mu)\sqrt{n}/\sigma_0\right) = \Phi\left(z_\alpha + (\mu_0 - \mu)\sqrt{n}/\sigma_0\right)$

$\mu \uparrow \Rightarrow \beta(\mu) \downarrow$

$\pi(\mu) = 1 - \beta(\mu) = 1 - \Phi(z_\alpha + (\mu_0 - \mu)\sqrt{n}/\sigma_0) \Rightarrow \mu \uparrow \Rightarrow \pi(\mu) \uparrow$

Supponiamo ora $H_0: \mu \leq \mu_0$ contro $H_1: \mu > \mu_0$. Ancora un test di livello di significatività α è quello che, avendo osservato $\bar{X}_n = \bar{x}_n$, rifiuta H_0 se $(\bar{x}_n - \mu_0)\sqrt{n}/\sigma_0 \geq z_\alpha$ e accetta se $(\bar{x}_n - \mu_0)\sqrt{n}/\sigma_0 < z_\alpha$. Tuttavia, poiché H_0 non è un'ipotesi semplice, potremmo dire che quello è un test di livello α se dimostriamo:

$$\alpha = \sup_{\mu \leq \mu_0} P_\mu(\text{rifiutare } H_0)$$

Calcoliamo questo estremo superiore:

$\sup_{\mu \leq \mu_0} P_\mu(\text{rifiutare } H_0) = \sup_{\mu \leq \mu_0} P_\mu\left((\bar{x}_n - \mu_0)\sqrt{n}/\sigma_0 > z_\alpha\right) =$
$= \sup_{\mu \leq \mu_0}\left[1 - P_\mu\left((\bar{x}_n - \mu_0)\sqrt{n}/\sigma_0 < z_\alpha\right)\right] = \sup_{\mu \leq \mu_0}\left[1 - \beta(\mu)\right]$ → vedi es.
$= 1 - \beta(\mu_0) = \alpha$
↳ $\beta(\mu_0) = 1 - \alpha$

Con analoghi ragionamenti possiamo dire che un test di livello α per $H_0: \mu = \mu_0$ (o $\mu \geq \mu_0$) contro $H_1: \mu < \mu_0$ è quello che, avendo osservato $\bar{X}_n = \bar{x}_n$, rifiuta H_0 se $(\bar{x}_n - \mu_0)\sqrt{n}/\sigma_0 \leq -z_\alpha$ e accetta H_0 se $(\bar{x}_n - \mu_0)\sqrt{n}/\sigma_0 > -z_\alpha$. La statistica test è $(\bar{X}_n - \mu_0)\sqrt{n}/\sigma_0$.

6.6 Verifica d'ipotesi per la media di una popolazione gaussiana con varianza incognita (test t)

Consideriamo un campione $X_1, \dots, X_n$ estratto da una popolazione gaussiana $\mathcal{N}(\mu, \sigma^2)$ con (μ, σ) incognite. Sia μ_0 costante fissata, vogliamo testare $H_0: \mu = \mu_0$ contro $H_1: \mu \neq \mu_0$. Stimiamo σ con la deviazione standard campionaria e sostituiamola sulla statistica test del test z, ottenendo:

$$T^* = \bar{X}_n - \mu_0 / S_n/\sqrt{n} \implies C_\alpha = \{(x_1, \dots, x_n) : |\bar{x}_n - \mu_0| / S_{0n}/\sqrt{n} \geq K\}$$

Poiché H_0 non è una ipotesi semplice, dobbiamo dimostrare:

$$\sup_{(\mu, \sigma^2) \in H_0} P_{(\mu,\sigma^2)}((X_1, \dots, X_n) \in C) =$$
$$= \sup_{\sigma^2 > 0} P_{(\mu_0,\sigma^2)}(|\bar{X}_n - \mu_0| / S_n/\sqrt{n} \geq K) \quad \text{se } \mu = \mu_0 \ \forall \sigma^2 > 0 \ \bar{X}_n - \mu_0 / S_n/\sqrt{n} \sim t(n-1)$$
$$\implies \text{sotto } H_0 \text{ e } \forall \sigma^2 > 0 \quad T^* \sim t(n-1)$$
$$\implies \sup_{\sigma^2 > 0} P_{(\mu_0,\sigma^2)}(T^* > t_{\frac{\alpha}{2}, n-1}) = P(|T_{n-1}| > t_{\frac{\alpha}{2}, n-1}) = \alpha$$

Quindi la regola di un test di livello α per il problema di verifica d'ipotesi considerato è: rifiuto H_0 se $|\bar{x}_n - \mu_0| / S_n/\sqrt{n} \geq t_{\alpha, n-1}$ e accetto H_0 se $|\bar{x}_n - \mu_0| / S_n/\sqrt{n} < t_{\alpha, n-1}$. Questa verifica è basata sulla statistica test $T^* = \bar{X}_n - \mu_0 / S_n/\sqrt{n} \sim t(n-1)$.

Sia t^* il valore assunto da T^* in corrispondenza delle osservazioni, si può dimostrare che il p. value sia:

$$\text{p. value} = P_{(\mu_0, \sigma^2)}\left(\left| \bar{X}_n - \mu_0 / S_n/\sqrt{n} \right| \geq |t^*| \right) =$$
$$= P_{(\mu_0, \sigma^2)}(|T_{n-1}| \geq |t^*|)$$

6.7 Test per la varianza di una popolazione gaussiana

MEDIA INCOGNITA Sia $X_1, \dots, X_n$ un campione estratto da una popolazione $\mathcal{N}(\mu, \sigma^2)$. Vogliamo verificare $H_0: \sigma^2 = \sigma_0^2$ contro $H_1: \sigma^2 \neq \sigma_0^2$. Ricordiamo che $(n-1)S_n^2/\sigma_0^2 \sim \chi^2(n-1)$ sotto H_0. Quindi:

$$\forall \mu \quad P_{(\mu,\sigma_0^2)}(\chi_{1-\alpha/2, n-1} < (n-1)S_n^2/\sigma_0^2 < \chi^2_{\alpha/2, n-1}) = 1 - \alpha$$
$$\implies \sup_\mu P_{(\mu,\sigma_0^2)}(\{\chi^2_{1-\alpha/2, n-1} \geq (n-1)S_n^2/\sigma_0^2\} \cup \{\chi^2_{\alpha/2, n-1} \leq (n-1)S_n/\sigma_0^2\}) = \alpha$$

Con statistica test $(n-1)S_n^2/\sigma_0^2 \sim \chi^2(n-1)$. Quindi un test di livello di significatività α è quello che avendo osservato $S_n^2 = s_n^2$ rifiuta H_0 se $(n-1)s_n^2/\sigma_0^2 \leq \chi^2_{1-\alpha/2, n-1}$ o $(n-1)s_n^2/\sigma_0^2 \geq \chi^2_{\alpha/2, n-1}$ e accetta

H_0 se $\chi^2_{1-\alpha/2,\, n-1} < {(n-1)s_n^2}/{\sigma_0^2} < \chi^2_{\alpha/2,\, n-1}$.

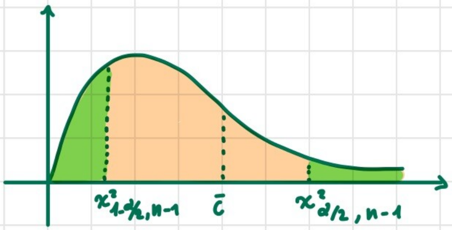
ESERCIZIO Mostrare che, se $\bar{c}$ è il valore assunto dalla statistica test $X_0 := {(n-1)S_n^2}/{\sigma_0^2}$ in corrispondenza dei dati, allora

$$p\text{-value} = 2 \min\{P(W \geq \bar{c}), P(W \leq \bar{c})\}$$

dove $W \sim \chi^2(n-1)$.

Poniamo $\tilde{p} = 2 \min\{P(W \geq \bar{c}), P(W \leq \bar{c})\}$. Si tratta di dimostrare che se $\alpha < \tilde{p} \Rightarrow$ accetto H_0 e se $\alpha \geq \tilde{p}$ rifiuto.

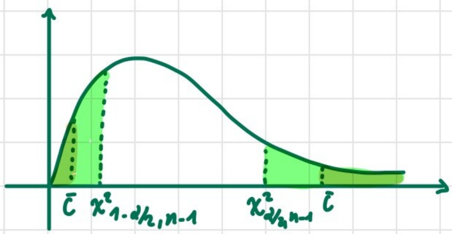
$\boxed{\alpha < \tilde{p}} \Rightarrow \alpha/2 < \min\{P(W \geq \bar{c}), P(W \leq \bar{c})\}$
$\Rightarrow \alpha/2 < P(W \geq \bar{c}) \wedge \alpha/2 < P(W \leq \bar{c})\}$



$\Rightarrow \chi^2_{1-\alpha/2,\, n-1} < \bar{c} < \chi^2_{\alpha/2,\, n-1}$

$\Rightarrow$ accetto H_0 a livello α

$\boxed{\alpha \geq \tilde{p}} \Rightarrow \alpha/2 \geq P(W > \bar{c}) \vee \alpha/2 \geq P(W < \bar{c})$



$\Rightarrow \bar{c} \leq \chi^2_{1-\alpha/2,\, n-1} \vee \bar{c} \geq \chi^2_{\alpha/2,\, n-1}$

$\Rightarrow$ rifiuto H_0 a livello α

$\tilde{p}$ è, quindi, il p-value.

MEDIA $\mu = \mu_0$ NOTA Vogliamo sempre testare $H_0: \sigma^2 = \sigma_0^2$ contro $H_1: \sigma^2 \neq \sigma_0^2$. Possiamo usare la varianza campionaria S_n come stimatore visto che conosciamo la media. La statistica test sarà:

$$T^* := n S_0^{*}/\sigma_0^2 = \sum_{i=1}^{n} (X_i - \mu_0)^2/\sigma_0^2 \overset{H_0}{\sim} \chi^2(n)$$

Un test di livello α sarà quello che avendo osservato $S^2_{0n} = \bar{s}_0^2$ rifiuta H_0 se $n\bar{s}_0^2/\sigma_0^2 \leq \chi^2_{1-\alpha/2,n}$ o $n\bar{s}_0^2/\sigma_0^2 > \chi^2_{\alpha/2,n}$. Come nel caso di media incognita si ha che il p-value sarà:

$$\text{p-value} = 2 \min \{ P(W^* \leq c^*), P(W^* \geq c^*) \}$$
$$c^* = n\bar{s}_0^2/\sigma_0^2 \quad e \quad W^* \sim \chi^2(n)$$

6.8 Verifica d'ipotesi su una popolazione bernoulliana

Sia $X_1, \dots, X_n$ un campione numeroso di varibili $X_i \sim Be(p)$ con p incognita. Sappiamo che:

$$X = X_1 + \dots + X_n \qquad \forall p \quad E(X) = np, \quad var(X) = np(1-p)$$
$$\Rightarrow TCL: \frac{X - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0,1)$$

Consideriamo il problema di verifica d'ipotesi $H_0: p \leq p_0$ contro $H_1: p > p_0$. In questo caso, si usa la statistica test

$$X_1 - np_0/\sqrt{np_0(1-p_0)} \overset{H_0}{\approx} \mathcal{N}(0,1)$$

Quindi un test di livello approssimato α è quello che, avendo osservato $\bar{y} = X - np_0/\sqrt{np_0(1-p_0)}$ rifiuta H_0 se il valore osservato $\bar{y} \geq z_\alpha$ e lo accetta se $\bar{y} < z_\alpha$. Il p-value sarà:

$$\text{p-value} = 1 - \Phi(\bar{y}) \simeq P_{p_0}\left(X - np_0/\sqrt{np_0(1-p_0)} \geq \bar{y} \right).$$

7 VERIFICA D'IPOTESI CON PIÙ POPOLAZIONI

7.1 Confronto tra medie per popolazioni gaussiane indipendenti

Consideriamo qualche esempio pratico. Per tutti i casi lavoreremo su due campioni $X_1, \dots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y_1, \dots, Y_m \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ i.i.d. indipendenti tra di loro. Siano $H_0: \mu_X = \mu_Y$ contro $H_1: \mu_Y \neq \mu_X$ le nostre ipotesi

VARIANZE NOTE Consideriamo le due medie campionarie $\bar{X}_n = 1/n \sum_{i=1}^{n} X_i$ e $\bar{Y}_m = 1/m \sum_{i=1}^{m} Y_i$ come stimatori non distorti di μ_X e μ_Y rispettivamente. Avremo che $\bar{X}_n - \bar{Y}_m$ sarà lo stimatore non distorto della differenza delle medie. Sarà ragionevole, quindi, rifiutare H_0 se $|\bar{X}_n - \bar{Y}_m| \geq K$ con K scelto in base al livello di significatività. Poichè $\bar{X}_n \sim \mathcal{N}(\mu_x, \sigma_x^2/n)$ e $\bar{Y}_m \sim \mathcal{N}(\mu_y, \sigma_y^2/m)$, la differenza sarà $\bar{X}_n - \bar{Y}_m \sim \mathcal{N}(\mu_x - \mu_y, \sigma_x^2/n + \sigma_y^2/m)$. Possiamo allora dire che:

$$\alpha = P_{H_0}(|\bar{X}_n - \bar{Y}_m| \geq K) = P_{H_0}\left(\frac{|\bar{X}_n - \bar{Y}_m|}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \geqslant \frac{K}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}\right) =$$

$$\iff \frac{K}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} = z_{\alpha/2} \implies P_{H_0}\left(\left|\frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}\right| \geqslant z_{\alpha/2}\right) = \alpha$$

$$\implies Z_0 = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \overset{H_0}{\sim} \mathcal{N}(0,1) \quad \text{STATISTICA TEST}$$

Quindi un test di livello α è quello che rifiuta H_0 se il valore osservato $\bar{z}_0$ di Z_0 è tale che $|\bar{z}_0| \geq z_{\alpha/2}$ e accetta se $|\bar{z}_0| < z_{\alpha/2}$. Il p-value sarà:

$$\text{p-value} = P_{H_0}(|Z_0| > |\bar{z}_0|) = 2(1 - \Phi(|\bar{z}_0|))$$

VARIANZE UGUALI INCOGNITE Sia $\sigma^2 = \sigma_x^2 = \sigma_y^2$. Consideriamo le varianze campionarie dei due campioni:

$$(n-1)S_x^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \quad ; \quad (m-1)S_y^2 = \sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2$$

$$\implies \begin{cases} (n-1)S_x^2/\sigma^2 \sim \chi^2(n-1) \\ (m-1)S_y^2/\sigma^2 \sim \chi^2(m-1) \end{cases} \xRightarrow{\text{IND.}} \frac{(n-1)S_x^2 + (m-1)S_y^2}{\sigma^2} \sim \chi^2(n+m-2)$$

$$\implies S_p^2 = \frac{(n-1)\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + (m-1)\sum_{i=1}^{m}(Y_i - \bar{Y}_m)}{n+m-2}$$

$$\implies (n-m+2)S_p^2/\sigma^2 \sim \chi^2(n+m-2)$$

Con S_p^2 lo stimatore "pooled" di σ^2. Inoltre per l'indipendenza:

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_x - \mu_y)}{\sigma\sqrt{1/n + 1/m}} \sim \mathcal{N}(0,1)$$

Permettendoci di scrivere:

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_x - \mu_y)}{\cancel{\sigma}\sqrt{1/n + 1/m}} \Big/ \sqrt{\frac{\cancel{(n+m-2)}\, S_p^2}{\cancel{\sigma^2}\cancel{(n+m-2)}}} = \frac{\bar{X}_n - \bar{Y}_m - (\mu_x - \mu_y)}{S_p\sqrt{1/n + 1/m}} \sim t(n+m-2)$$

$$\Rightarrow T_0 = \frac{\bar{X}_n - \bar{Y}_m}{S_p\sqrt{1/n + 1/m}} \overset{H_0}{\sim} t(n+m-2) \qquad \text{STATSTICA TEST}$$

$$\Rightarrow P_{H_0}(|T_0| \geq t_{\alpha/2,\, n+m-2}) = \alpha$$

Possiamo, quindi, considerare un test di livello di significatività α quello che, dato il valore osservato di $T_0 = t_0$, rifiuta H_0 se $|t_0| \geq t_{\alpha/2, n+m-2}$ e accetta H_0 se $|t_0| < t_{\alpha/2, n+m-2}$. Il p-value sarà:

$$\text{p-value} = P_{H_0}(|T_0| \geq |t_0|)$$

ESERCIZIO Vengono messe a confronto le durate di batterie (in ore) di due tipologie diverse 1 e 2. Si può ritenere si tratti di campioni normali indipendenti con medie incognite μ_1 e μ_2 rispettivamente e deviazioni standard note e pari a $\sigma_1 = 0.12$ e $\sigma_2 = 0.15$ rispettivam. Si osservano i seguenti dati:

TIPO 1	1.49	1.50	1.46	1.51	1.54	1.33	1.37	1.63
TIPO 2	1.49	1.47	1.40	1.17	1.1	1.31		

1) Eseguire un test per l'ipotesi H_0 che la durata media delle batterie di tipo 1 e di tipo 2 sia uguale contro l'alternativa che sia diversa, con un livello di significatività del 10%.
2) Determinare il p-value e stabilire per quali valori della significatività si rifiuta l'ipotesi nulla con i dati a disposizione
3) Fornire un intervallo di confidenza di tipo $(c; +\infty)$ al 95% per la differenza delle medie $\Delta := \mu_1 - \mu_2$ e calcolare la stima intervallare in corrispondenza dei dati a disposizione.

$X_1, \dots, X_8 \sim \mathcal{N}(\mu_1, (0.12)^2)$; $Y_1, \dots, Y_6 \sim \mathcal{N}(\mu_2, (0.15)^2)$

1) $Z_0 = \bar{x}_8 - \bar{y}_6 / \sqrt{(0.12)^2/8 + (0.15)^2/6} \Rightarrow |z_0| \simeq |2.09|$

$z_{\alpha/2} = z_{0.05} \simeq 1.645 \Rightarrow$ RIFIUTO

2) p-value $= P_{H_0}(|Z_0| \geq 2.09) = 2(1 - \Phi(2.09)) \simeq 0.04$

ESERCIZIO I dati presentati sotto costituiscono i tempi di vita (in centinaia di ore) di due tipi di valvole termoioniche. Lo studio passato di questo tipo di dati ci permette di dire che la loro distribuz. deve essere lognormale, ovvero i logaritmi dei tempi di vita hanno distribuzione normale. Assumendo che le varianze dei logaritmi siano uguali per i due campioni, verificare al 5% di significatività l'ipotesi che le distribuzioni coincidano.

TIPO 1	32	84	37	42	78	62	59	74
TIPO 2	39	111	55	106	90	87	85	

Abbiamo $X_1, \dots, X_8$ i.i.d lognormali e $Y_1, \dots, Y_7$ i.i.d lognormali indipendenti.

$\hookrightarrow X_i^* = \ln(X_i) \sim \mathcal{N}(\mu_x, \sigma^2)$

$Y_i^* = \ln(Y_i) \sim \mathcal{N}(\mu_y, \sigma^2)$

X_i^*	3.47	4.43	3.61	3.74	4.36	4.13	4.08	4.30
Y_j^*	3.66	4.71	4.01	4.66	4.50	4.47	4.44	

$H_0: \mu_x = \mu_y$ vs. $H_1: \mu_x \neq \mu_y$

$\Rightarrow T_0^* = \bar{x}_8^* - \bar{y}_7^* / s_p^* \sqrt{1/8 + 1/7} \simeq -0.0201 \Rightarrow |T_0^*| \overset{?}{\geq} t_{0.025, 13} \rightarrow$ NO

$\Rightarrow$ accetto al 5%

VARIANZE DIVERSE INCOGNITE Se n e m sono grandi, si può dimostrare che la statistica test è:

$$Z^* = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_x^2/n + S_y^2/m}} \overset{H_0}{\approx} \mathcal{N}(0,1) \Rightarrow P_{H_0}\left(\left|\frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_x^2/n + S_y^2/m}}\right| \leq z_{\alpha/2}\right) \simeq \alpha$$

Il test di livello approssimato α sarà quello che rifiuta H_0 se il valore osservato di Z^* z^* è tale che $|z^*| \geq z_{\alpha/2}$ e accetta se $|z^*| < z_{\alpha/2}$. Il p-value sarà:

$$p\text{-value} = P_{H_0}(|Z^*| \geq |z^*|) = 2\left(1 - \Phi(|z^*|)\right)$$

7.2 Confronto tra medie per popolazioni gaussiane non indip.

Consideriamo ora il caso in cui i nostri due campioni normali non siano indipendenti. Introduciamo, allora, il test t.

Consideriamo i campioni $X_1, \dots, X_n$ e $Y_1, \dots, Y_n$ non indipendenti. Siano $H_0: \mu_X \leq \mu_Y$ contro $H_1: \mu_X > \mu_Y$ le nostre ipotesi. Sia $W_1, \dots, W_n$ dove $W_i = Y_i - X_i$. Avremo:

$$E(W_i) = E(Y_i - X_i) = \mu_Y - \mu_X$$

Abbiamo, quindi, a disposizione un campione di dimensione n $W_1, \dots, W_n$ e le ipotesi possono essere scritte come $H_0: \mu_W \geq 0$ contro $H_1: \mu_W < 0$.

Supponiamo che (X_i, Y_i) sia un campione gaussiano bidimensionale. Allora per le proprietà dei vettori gaussiani anche W_i è gaussiana con media μ_W e varianza σ_W^2 entrambe incognite. Sulla base del campione gaussiano $W_1 = Y_1 - X_1, \dots, W_n = Y_n - X_n$ vogliamo verificare $H_0: \mu_W \geq \mu_0 = 0$ contro $H_1: \mu_W < \mu_0 = 0$. Ci siamo, quindi, ridotti al test sulla media di una popolazione gaussiana con varianza σ_W^2 incognita:

$$T^* = \frac{\overline{W}_n - \mu_0}{S_W/\sqrt{n}} = \frac{\overline{W}_n}{S_W/\sqrt{n}} \overset{\mu_W = 0}{\sim} t(n-1)$$

Poiché $P_{\mu_W = 0}(T^* < -t_{\alpha, n-1}) = \alpha$, un test di livello α è quello che rifiuta H_0 se il valore osservato di T^* $\bar{t}$ è $\bar{t} \leq -t_{\alpha, n-1}$ e accetta se $\bar{t} > -t_{\alpha, n-1}$.

ESERCIZIO Un gruppo di 6 pazienti vengono sottoposti ad una cura dimagrante. Si osservano i seguenti dati:

PRIMA	77	87	104	98	91	78
DOPO	75	88	97	99	83	70

1. C'è evidenza sperimentale ad un livello del 5% che la cura è stata efficace?
2. Se la cura promette un calo medio di più di 5 Kg, c'è evidenza sperimentale che è stata efficace?

$W_1, \dots, W_6 \quad \mu_W = \mu_Y - \mu_X \quad \Rightarrow$ calo di peso se $\mu_W < 0$

1. $H_0: \mu_W \geq 0$ vs. $H_1: \mu_W < 0$, $\alpha = 5\%$
   $T^* \cong -2.156$; $-t_{0.05,5} = -2.015 \Rightarrow$ RIFIUTO $H_0 \Rightarrow$ c'è stato dimagrimento

2. $H_0: \mu_W \geq -5$ vs. $H_1: \mu_W < -5$,
   $T^* = \overline{W}_6 - \mu_0 / S_6/\sqrt{6} \cong 0.656$; $-t_{0.05,5} = -2.015 \Rightarrow$ ACCETTO H_0

7.3 Confronto di varianze di popolazioni gaussiane indipendenti

Introduciamo prima un risultato intermedio che ci sarà d'utilità:

COROLLARIO Se $X_1, \dots, X_n$ è un campione da popolazione gaussiana di media μ_X e varianza σ_X^2 e $Y_1, \dots, Y_m$ è un campione indipendente dal precedente, gaussiano di media μ_Y e varianza σ_Y^2, allora

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(n-1, m-1)$$

Quindi se $\sigma_X^2 = \sigma_Y^2$ otteniamo $S_X^2/S_Y^2 \sim F(n-1, m-1)$

DIMOSTRAZIONE Sappiamo che $(n-1)S_X^2/\sigma_X^2 \sim \chi(n-1)$. Stessa cosa per l'altro campione. Poiché sono indipendenti, applichiamo la definizione di distribuzione di Fisher:

$$\frac{\cancel{(n-1)} S_x^2}{\cancel{(n-1)}\sigma_x^2} \Big/ \frac{\cancel{(m-1)} S_y^2}{\cancel{(m-1)}\sigma_y^2} = \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \sim F(n-1, m-1)$$

È evidente la formula per $\sigma_x^2 = \sigma_y^2$ ■

Possiamo ora introdurre il test. Siano i campioni indipendenti $X_1, \dots, X_n \sim \mathcal{N}(\mu_x, \sigma_x^2)$ e $Y_1, \dots, Y_m \sim \mathcal{N}(\mu_y, \sigma_y^2)$ con medie e varianze incognite. Siano $H_0: \sigma_x^2 = \sigma_y^2$ contro $H_1: \sigma_x^2 \neq \sigma_y^2$ le nostre ipotesi. Per il corollario la statistica test sarà:

$$F_0 = S_x^2/S_y^2 \overset{H_0}{\sim} F(n-1, m-1) \Rightarrow P_{H_0}(f_{1-\alpha/2, n-1, m-1} < F_0 < f_{\alpha/2, n-1, m-1}) = 1-\alpha$$

Quindi un test di livello α è quello che rifiuta H_0 se il valore osservato F_0 $\bar{f}$ è $\bar{f} \leq f_{1-\alpha/2, n-1, m-1} \vee \bar{f} \geq f_{\alpha/2, n-1, m-1}$ e accetta H_0 se $f_{1-\alpha/2, n-1, m-1} < \bar{f} < f_{\alpha/2, n-1, m-1}$. Il p-value sarà:

$$\text{p-value} = 2\min\{P(F_0 > \bar{f}), P(F_0 \leq \bar{f})\}$$

ESERCIZIO Due blogger hanno monitorato per alcuni giorni il numero di visualizzazioni giornaliere di un loro video su Youtube ottenendo i seguenti risultati:

BLOGGER X	2270	2490	2190	2030	2540	2710		
BLOGGER Y	2740	3600	3120	3280	3310	3280	2340	2460

Supponiamo si possa assumere che i dati provengano da due popolazioni normali indipendenti.

1. Usando un opportuno test di livello 10%, si può rifiutare l'ipotesi che le varianze delle due popolazioni sono uguali? Più in generale, cosa si decide agli usuali livelli di significatività del 5%, 2.5% e 1%?
2. Utilizzando il risultato ottenuto al punto precedente, c'è evidenza statistica al 5% che il blogger Y ha in media un

numero di visualizzazioni giornaliere maggiore del blogger X?
3. Determinare un intervallo di confidenza al 97.5%, illimitato superiormente, per la media del numero di visualizzazioni per il blogger Y e la stima corrispondente

1. Si tratta di verificare $H_0: \sigma_X^2 = \sigma_Y^2$ vs. $H_1: \sigma_X^2 \neq \sigma_Y^2$
$f_{0.95,5,7} = 1/f_{0.05,7,5} = 1/4.88 = 0.205 < S_X^2/S_Y^2 = 0.3317 < f_{0.05,5,7} = 3.97$
$\Rightarrow$ ACCETTO H_0

Se accetto al 10%, accetterò anche a livelli minori.

8 VERIFICA DEL MODELLO

Consideriamo il problema di riconoscere quando un modello probabilistico si adatta ad un fenomeno casuale. Definiamo test di buon adattamento i test statistici che servono a verificare se un modello probabilistico è compatibile con i dati.

8.1 Test χ^2 di buon adattamento

Sia un'ampia popolazione formata da oggetti classificabili in K categorie diverse. Per $i = 1, K$ sia p_i la probabilità che un oggetto faccia parte della classe i. Siano $p_1^0, \dots, p_K^0$ i numeri assegnati alle p_i.

Vogliamo verificare $H_0: p_i = p_i^0$ contro $H_1: p_i \neq p_i^0$. Consideriamo un campione di n osservazioni indipendenti e ciascuna ha probabilità p_i di appartenere alla classe i. Sia, per $i = 1, \dots, K$, $X_i :=$ "n° delle osservazioni nel campione che hanno classe i" chiamata frequenza assoluta nel campione. Si avrà:

$$X_i > 0 \rightarrow \sum_{i=1}^{K} X_i = n$$
$$X_i \sim Bin(n, p_i) \Rightarrow E(X) = n p_i$$

Poiché $X_i \sim Bin(n, p_i)$ e $E(X_i) = n p_i$ segue che $(X_i - n p_i^0)^2$ è un indicatore di quanto sia verosimile $p_i = p_i^0$. Quando è troppo grande è molto probabile il rifiuto di H_0. È naturale usare la statistica test:

$$T := \sum_{i=1}^{K} \frac{(X_i - np_i^0)^2}{np_i^0}$$

L'ipotesi nulla va rifiutata se $T > c$, c determinata in base alla significatività.

$$P_{H_0}(T \geq c) = \alpha$$

Se n è grande si ha:

$$T \overset{H_0}{\approx} \chi^2(\overset{\text{categorie}}{K} - 1) \Rightarrow c \simeq \chi^2_{\alpha, K-1}$$

Quindi un test di livello approssimato α è quello che osservato $T = \bar{t}$ rifiuta H_0 se $\bar{t} \geq \chi^2_{\alpha,K-1}$ e accetta H_0 se $\bar{t} < \chi^2_{\alpha,K-1}$. Il p-value è:

$$\text{p-value: } P_{H_0}(T \geq \bar{t}) \simeq P(W \geq \bar{t}) \quad \text{con} \quad W \sim \chi^2(\overset{\text{categorie}}{K} - 1)$$

Una regola empirica per stabilire quando n è "abbastanza grande" è che 80% delle $np_i^0 > 5$ e le rimanenti $np_i^0 > 1$.

Ricordando $\sum_{i=1}^{K} p_i = 1$ e $\sum_{i=1}^{K} X_i = n$, possiamo ricavare

$$T = \sum_{i=1}^{K} \frac{(X_i - np_i^0)^2}{np_i^0} = \sum_{i=1}^{K} \frac{X_i + n^2 p_i^{0^2} - 2np_i^0 X_i}{np_i^0}$$
$$= \sum_{i=1}^{K} \frac{X_i^2}{np_i} + \sum_{i=1}^{K} p_i^0 - 2\sum_{i=1}^{K} X_i = \sum_{i=1}^{K} \frac{X_i^2}{np_i} - n$$

Dimostriamo perché $T \approx \chi^2(K-1)$:

DIMOSTRAZIONE ($T \approx \chi^2(K-1)$). Sia $K = 2$: $X_1 + X_2 = n$ e $p_1^0 + p_2^0 = 1$

$$T = \frac{(X_1 - np_1^0)^2}{np_1^0} + \frac{(X_2 - np_2^0)^2}{np_2^0} =$$

$$= \frac{(X_1 - np_1^0)^2}{np_1^0} + \frac{(n - X_1 - n(1-p_1^0))^2}{n(1-p_1^0)}$$

$$= \frac{(X_1 - np_1^0)^2}{np_1^0} + \frac{(X_1 - np_1^0)^2}{n(1-p_1^0)} = \frac{(X_1 - np_1^0)^2}{np_1^0(1-p_1^0)}$$

Se n è grande e H_0 è vera avremo:

$$\frac{X_1 - np_1^0}{\sqrt{np_1^0(1-p_1^0)}} \approx \mathcal{N}(0,1) \Rightarrow T = \frac{(X_1 - np_1^0)^2}{np_1^0(1-p_1^0)} \approx \chi^2(1)$$

Per $K>2$ la dimostrazione è più complessa e non la affronteremo. ∎

Nel caso in cui T dipenda da un parametro (e.g. $H_0: Y \sim Poiss(\lambda)$ vs. $H_1: Y \nsim Poiss(\lambda)$) consideriamo $\tilde{T}$ costruita con una stima dei parametri. Si dimostra che $\tilde{T} \approx \chi^2(K-1-n)$ con n il numero dei parametri stimati (per n grande).

8.2 Test χ^2 di buon adattamento per dati continui

Sia $Y_1, \dots, Y_n$ un campione aleatorio estratto da distribuzione F continua. Sia F_0 una distribuzione continua assegnata. Vogliamo verificare $H_0: F = F_0$ contro $H_1: F \neq F_0$. Un approccio è dividere i possibili valori delle v.a. Y_j, sotto H_0, in K intervalli disgiunti $(y_0, y_1], \dots, (y_{K-1}, y_K]$. I K intervalli $(y_{i-1}, y_i]$ saranno le K classi in cui suddivideremo le osservazioni. Le probabilità p_i saranno $p_i^0 = P(Y_j \in (y_{i-1}, y_i])$. Possiamo quindi riformulare le nostre ipotesi in $H_0: p_i = p_i^0$ contro $H_1: p_i \neq p_i^0$ e procedere con il test χ^2 come nel caso discreto.