

Università di Modena e Reggio Emilia

Facoltà di Ingegneria

Lezioni
di
STATISTICA MATEMATICA A

Corso di Laurea in Ingegneria Meccanica

Corso di Laurea in Ingegneria dei Materiali

- Anno Accademico 2010/11 -

1 LA PROBABILITÀ MATEMATICA

1.1 Definizioni e Proprietà

La teoria della probabilità iniziò a svilupparsi intorno al XVII secolo quando i giocatori d'azzardo iniziarono a finanziare i maggiori matematici dell'epoca per calcolare la casualità di alcuni giochi d'azzardo. Più tardi si realizzò che riguardava anche i processi scientifici e da allora queste metodologie sono state utilizzate per studiare anche il mondo fisico.

Per studiare sistematicamente la probabilità bisogna introdurre la terminologia corretta.

Definizione. Si definisce **esperimento** un processo che termina con un risultato imprevedibile con certezza in anticipo.

Lanciare una moneta o un dado, misurare il diametro di una rondella, pesare una pietra, misurare la resistenza di una lenza di una canna da pesca sono tutti esempi di esperimenti. Per poter parlare di un esperimento in termini probabilistici, bisogna prima specificare tutti i suoi possibili risultati.

Definizione. Si chiama **spazio campionario** l'insieme \mathbf{S} di tutti i possibili esiti di un dato esperimento. Un **evento** è un insieme di esiti, cioè un sottoinsieme dello spazio campionario \mathbf{S} .

Per il lancio di una moneta, lo spazio campionario è l'insieme {Testa, Croce}; per il lancio del dado a sei facce, l'insieme è {1, 2, 3, 4, 5, 6}. In questi esempi lo spazio \mathbf{S} è finito, ma non è sempre così. Alcuni esperimenti hanno lo spazio campionario formato da un numero infinito di risultati. Ad esempio, si pensi ad una perforatrice di diametro 10 mm che fa dei buchi in un foglio di metallo. A causa della variazione dell'angolo per fare i buchi e dei movimenti impercettibili del foglio di metallo, il diametro dei buchi varia tra 10.0 e 10.2 mm. In questo caso allora lo spazio campionario è l'intervallo (10.0, 10.2) o, usando la notazione insiemistica, $\{x | 10.0 < x < 10.2\}$. Tale insieme è formato da un numero infinito di elementi.

Spesso gli eventi sono costruiti combinando gli eventi elementari. Dato che gli eventi sono sottoinsiemi di spazi campionari si usa la notazione degli insiemi per descrivere gli eventi costituiti in quel modo.

Definizione. Si dice poi **classe di eventi**, e la denoteremo con $\mathbf{\Omega}$, ogni insieme non vuoto di eventi che risulti essere chiuso rispetto alle operazioni insiemistiche elementari; ossia:

- i) dati due eventi $A, B \in \mathbf{\Omega}$, allora anche $A \cup B \in \mathbf{\Omega}$ ($A \cup B$ è l'evento che si verifica se si verifica almeno uno fra gli eventi A e B);

- ii) data una successione numerabile di eventi $A_i \in \Omega$, allora anche la loro unione è un evento, cioè $\bigcup_{i=1}^{\infty} A_i \in \Omega$;
- iii) dato un evento $A \in \Omega$, allora anche il suo complementare $A^C \equiv \mathbf{S} - A \in \Omega$ (A^C è l'evento che si verifica quando A non si verifica).

Dai tre assiomi che caratterizzano una classe di eventi Ω seguono queste altre proprietà:

- Dati due eventi A e B , anche $A \cap B$ è un evento; infatti:

$$A \cap B = (A^C \cup B^C)^C \implies A \cap B \in \Omega;$$

- L'insieme vuoto \emptyset e lo spazio \mathbf{S} sono eventi; infatti, preso $A \in \Omega$, si ha

$$A \cap A^C = \emptyset \implies \emptyset \in \Omega, \quad A \cup A^C = \mathbf{S} \implies \mathbf{S} \in \Omega.$$

L'evento $A \cap B$ si realizza quando sia A che B verificano. L'evento \emptyset è detto **evento impossibile** e \mathbf{S} è detto **evento certo**.

Definizione. Due eventi A e B sono detti **incompatibili** se sono disgiunti, cioè se $A \cap B = \emptyset$.

In altre parole: due eventi sono incompatibili se non si possono mai verificare simultaneamente.

Esempio: Si consideri il seguente esperimento: si getta un dado e si guarda il risultato della prova, cioè il numero che si presenta. Lo spazio campionario consiste nei sei numeri possibili:

$$\mathbf{S} = \{1, 2, 3, 4, 5, 6\}.$$

Consideriamo i seguenti eventi: A : “il risultato è un numero pari”; B : “il risultato è un numero dispari”; C : “il risultato è un numero primo”. In termini di sottinsiemi di \mathbf{S} :

$$A = \{2, 4, 6\}, \quad B = \{1, 3, 5\}, \quad C = \{2, 3, 5\}.$$

Si ha quindi, ad esempio:

$$A^C = \{1, 3, 5\} = B;$$

$$C^C = \{1, 4, 6\}: \text{ “il risultato non è un numero primo”};$$

$$B \cap C = \{3, 5\}: \text{ “il risultato è un numero dispari e primo”};$$

$$A \cup C = \{2, 3, 4, 5, 6\}: \text{ “il risultato è un numero pari o primo”}.$$

Si noti che gli eventi A e B , essendo $A \cap B = \emptyset$, sono incompatibili.

Ogni evento in uno spazio campionario ha una probabilità di realizzarsi. Intuitivamente, la probabilità è una misura quantitativa di quanto sia ragionevole pensare che l'evento possa realizzarsi. Formalmente si può dire che la probabilità di un evento è la proporzione di volte in cui l'evento potrebbe realizzarsi se l'esperimento

fosse ripetuto un numero molto grande di volte indipendentemente e sotto le stesse condizioni. In tante situazioni il solo modo di stimare la probabilità di un evento è di ripetere l'esperimento molte volte e determinare la proporzione di volte in cui l'evento si verifica. Per esempio, se si vuole stimare la probabilità che un circuito stampato prodotto da un certo processo manifatturiero sia difettoso, è necessario di solito produrre un certo numero di circuiti e testarli per determinare la proporzione di quelli difettosi. In altri casi, le probabilità possono essere determinate dalla conoscenza della natura fisica dell'esperimento. Per esempio, se è noto che la forma di un dado è quasi un cubo perfetto e che la sua massa è distribuita quasi in maniera omogenea, si può allora assumere che ognuna delle sei facce sia ugualmente probabile in ogni lancio di dado. In pratica, scienziati e ingegneri stimano le probabilità di alcuni eventi sulla base della comprensione scientifica e dell'esperienza e quindi utilizzano le regole matematiche per calcolare le stime delle probabilità degli altri eventi. Vediamo alcune di queste regole.

Definizione. Sia \mathbf{S} uno spazio campionario ed Ω una classe di eventi in \mathbf{S} . Sia poi P una funzione definita su Ω a valori in $[0, 1]$:

$$P : \Omega \longrightarrow [0, 1].$$

Allora (\mathbf{S}, Ω, P) è detto **spazio di probabilità** e $P(A)$ è detta **probabilità** dell'evento $A \in \Omega$ se valgono i seguenti tre assiomi:

- 1) $P(\mathbf{S}) = 1$;
- 2) se A e B sono due eventi incompatibili, allora

$$P(A \cup B) = P(A) + P(B); \tag{1}$$

- 3) se $\{A_n, n \in \mathbb{N}\}$ è una successione numerabile di eventi incompatibili, si ha:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Gli assiomi 2) e 3) esprimono il fatto che le probabilità di eventi incompatibili si sommano. In particolare l'assioma 3), che ovviamente ha significato solo nel caso in cui Ω è un insieme infinito, si esprime sinteticamente dicendo che P è **numerabilmente additiva**.

Teorema. *La probabilità dell'evento impossibile è nulla, cioè*

$$P(\emptyset) = 0.$$

Dimostrazione: Sia A un qualunque evento di Ω . Poiché anche $\emptyset \in \Omega$, segue che $A \cup \emptyset \in \Omega$. Inoltre, A ed \emptyset sono eventi incompatibili essendo $A \cap \emptyset = \emptyset$. In virtù dell'assioma 2) si ha quindi

$$P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset) \implies P(\emptyset) = 0.$$

□

Teorema (regola di addizione per eventi arbitrari). *Se A, B sono eventi arbitrari in uno spazio di probabilità Ω , allora*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Dimostrazione: Scriviamo A come unione disgiunta di $A - B$ e $A \cap B$, e analogamente facciamo per B :

$$A = (A - B) \cup (A \cap B), \quad B = (B - A) \cup (A \cap B).$$

Allora applicando due volte l'additività (1):

$$\begin{aligned} P(A) + P(B) &= P(A - B) + P(A \cap B) + P(B - A) + P(A \cap B) = \\ &= P[(A - B) \cup (A \cap B) \cup (B - A)] + P(A \cap B) = P(A \cup B) + P(A \cap B) \end{aligned}$$

da cui la tesi sottraendo $P(A \cap B)$ al primo e all'ultimo membro. □

Teorema (regola di complementazione). *Se $E \subset \Omega$ è un evento ed $E^C \equiv \mathbf{S} - E$ è l'evento complementare, si ha*

$$P(E) = 1 - P(E^C).$$

Dimostrazione: Si ottiene banalmente applicando la (1) (perché E e E^C sono due eventi incompatibili). □

Teorema. *Se A e B sono due eventi tali che $A \subseteq B$, allora*

$$P(A) \leq P(B).$$

Dimostrazione: Essendo $A \subseteq B$ si può decomporre B negli eventi incompatibili A e $B - A = B \cap A^C$. Si può quindi scrivere

$$P(B) = P(A \cup (B - A)) = P(A) + P(B - A) \geq P(A),$$

essendo $P(B - A) \geq 0$. □

Teorema. *Se A e B sono due eventi qualunque, allora*

$$P(A - B) = P(A) - P(A \cap B).$$

Dimostrazione: L'evento A può essere decomposto negli eventi incompatibili $A - B$ e $A \cap B$, per cui, in virtù della (1), si ha

$$P(A) = P((A - B) \cup (A \cap B)) = P(A - B) + P(A \cap B).$$

da cui segue la tesi. □

1.2 Spazi di probabilità finiti

Sia \mathbf{S} uno spazio campionario finito:

$$\mathbf{S} = \{a_1, a_2, \dots, a_N\}$$

ed Ω l'insieme di tutti i sottinsiemi di \mathbf{S} (inclusi \mathbf{S} e \emptyset). Si ottiene uno spazio di probabilità finito assegnando a ciascun elemento a_i di \mathbf{S} un numero reale p_i , detto **probabilità** di a_i e indicato come $P\{a_i\}$, tale che

i) $p_i \geq 0$ per ogni $i = 1, 2, \dots, N$;

ii) la somma delle singole probabilità è uguale a 1, ossia $\sum_{i=1}^N p_i = 1$.

La probabilità $P(A)$ di un qualsiasi evento $A \subseteq \mathbf{S}$ viene quindi definita come la somma delle probabilità degli **eventi elementari** $\{a_i\}$ contenuti in A :

$$P(A) = P\left(\bigcup_{i:a_i \in A} \{a_i\}\right) = \sum_{i:a_i \in A} P(\{a_i\}) = \sum_{i:a_i \in A} p_i.$$

Dimostriamo che la funzione $P : \Omega \rightarrow [0, 1]$ è una funzione di probabilità facendo vedere che valgono gli assiomi 1) e 2). Per quanto riguarda la validità dell'assioma 1), si ha

$$P(\mathbf{S}) = P\left(\bigcup_{i:a_i \in \mathbf{S}} \{a_i\}\right) = P\left(\bigcup_{i=1}^N \{a_i\}\right) = \sum_{i=1}^N P(\{a_i\}) = \sum_{i=1}^N p_i = 1.$$

Inoltre, se A e B sono eventi incompatibili, si ha

$$\begin{aligned} P(A \cup B) &= P\left(\bigcup_{i:a_i \in A \cup B} \{a_i\}\right) = \sum_{i:a_i \in A \cup B} P(\{a_i\}) = \\ &= \sum_{i:a_i \in A} p_i + \sum_{i:a_i \in B} p_i = P(A) + P(B). \end{aligned}$$

per cui si ritrova l'assioma 2). Valgono dunque tutti gli assiomi richiesti (essendo lo spazio finito, l'assioma 3) non ha significato) affinché P sia una funzione di probabilità. \square

Dal punto di vista pratico ci sono diversi modi di assegnare le probabilità p_i agli eventi elementari $\{a_i\}$. Uno dei possibili modi è il seguente: se ripetiamo lo stesso esperimento n volte e chiamiamo s_i il numero di volte che si verifica $\{a_i\}$, si osserva che il rapporto

$$\frac{s_i}{n}$$

detto **frequenza relativa**, a lungo andare tende a stabilizzarsi, cioè tende ad un limite p_i (compreso, ovviamente, tra 0 ed 1). Questo valore limite p_i , così calcolato empiricamente, viene assunto come la **probabilità** dell'evento elementare $\{a_i\}$.

1.3 Spazi finiti equiprobabili

Definizione. Si dice **spazio equiprobabile** (o **uniforme**) uno spazio di probabilità finito dove tutti gli elementi dell'insieme campionario \mathbf{S} hanno la stessa probabilità.

Dalla definizione e dagli assiomi della probabilità segue immediatamente che, se lo spazio campionario \mathbf{S} è composto da N elementi, la probabilità di ciascun elemento di \mathbf{S} vale $p = \frac{1}{N}$. Avremo inoltre che, dato un qualunque evento A , la sua probabilità sarà data da

$$P(A) = \frac{\text{numero degli elementi di } A}{N} = \frac{|A|}{N}.$$

È in questo caso che vale la definizione classica di probabilità come **“numero dei casi favorevoli diviso il numero di casi possibili”**.

Osservazione: $|A|$ denota la **cardinalità di A** , cioè il numero degli eventi elementari che costituiscono A . Questa notazione sarà utilizzata anche in seguito.

Esempio: Esempio tipico è il lancio del dado non truccato, dove si definisce:

$$\mathbf{S} = \{1, 2, 3, 4, 5, 6\}, \quad P(1) = \frac{1}{6}, \dots, P(6) = \frac{1}{6}.$$

Così potremo calcolare, ad es., la probabilità degli eventi

A : esce un numero pari, B : esce un numero minore di 3

$$P(A) = P(2) + P(4) + P(6) = \frac{1}{2}, \quad P(B) = P(1) + P(2) = \frac{1}{3}$$

O in altri termini:

$$P(A) = \frac{|\{2, 4, 6\}|}{6} = \frac{1}{2}, \quad P(B) = \frac{|\{1, 2\}|}{6} = \frac{1}{3}.$$

Altro esempio tipico è il lancio, singolo o multiplo, di una moneta non truccata.

Esempio: La probabilità che in cinque lanci di una moneta esca “testa” almeno una volta si trova introducendo l'appropriato spazio di probabilità

$$\mathbf{S} = \{5\text{-uple ordinate di lettere "T" o "C"}\}$$

Siccome il numero delle possibili cinque che costituiscono \mathbf{S} è 2^5 , abbiamo $N = 32$, e quindi $p = \frac{1}{32}$. In questo spazio l'evento “non esce alcuna testa” è costituito dall'unica 5-upla (C, C, C, C, C) , per cui l'evento $A =$ “esce almeno una testa” ha probabilità

$$P(A) = 1 - P(A^C) = \frac{31}{32}.$$

1.4 Calcolo combinatorio

Da quanto detto nel paragrafo precedente, si capisce quindi che per calcolare le probabilità è necessario talvolta determinare il numero di elementi presenti nello spazio campionario. In questo paragrafo verranno descritti diversi metodi per determinare tale numerosità. Presentiamo la regola di base, detta **regola fondamentale del conteggio**, attraverso l'esempio che segue.

Esempio: Un certo modello di automobile è disponibile nei tre colori rosso, blu e verde e con un motore grande o piccolo. In quanti modi un acquirente può scegliere un'automobile?

Esistono tre scelte per il colore e due scelte per il motore. La lista completa delle scelte è presentata nella seguente tabella:

| | Rosso | Blu | Verde |
|---------|----------------|--------------|----------------|
| Grande | Rosso, Grande | Blu, Grande | Verde, Grande |
| Piccolo | Rosso, Piccolo | Blu, Piccolo | Verde, Piccolo |

Esistono pertanto $(3)(2) = 6$ possibili scelte.

Per generalizzare l'esempio precedente, se ci sono n_1 scelte per il colore ed n_2 scelte per il motore, il numero totale di scelte sarà $n_1 n_2$.

La regola fondamentale del conteggio può essere estesa ad un numero qualsiasi di attività.

Regola fondamentale del conteggio. *Se si devono svolgere k attività e ci sono n_1 modi per svolgere la prima e se per ognuno di questi modi ce ne sono n_2 per svolgere la seconda e se per ogni scelta di svolgimento delle prime due attività esistono n_3 modi per svolgere la terza attività e così via, allora il numero totale di modi per svolgere la sequenza di k attività è $n_1 n_2 \cdots n_k$.*

Esempio: Supponiamo di voler ordinare un certo tipo di computer. Esistono 3 scelte per il tipo di disco fisso, 4 scelte per l'ammontare della memoria, 2 scelte per la scheda video e 3 scelte per il monitor. In quanti modi può essere ordinato il computer.

Il computer può essere ordinato in $3 \cdot 4 \cdot 2 \cdot 3 = 72$ modi diversi.

Disposizioni semplici e Permutazioni

Definizione. Una **disposizione semplice** di n oggetti dati presi k alla volta è una k -upla ordinata di k oggetti distinti scelti tra gli n dati (ovviamente: $k \leq n$).

Esempio: Le disposizioni di 3 oggetti a, b, c presi a coppie ($k = 2, n = 3$), sono:

$$(a, b), (b, c), (c, a), (b, a), (c, b), (a, c).$$

L'aggettivo "semplice" vuol dire "senza ripetizioni".

Proposizione. *Il numero di disposizioni semplici di n oggetti dati presi k alla volta, che indichiamo con $D(n; k)$, è il prodotto dei k numeri naturali decrescenti a partire da n :*

$$D(n; k) = n(n - 1) \dots (n - k + 1)$$

Infatti se si riempiono k caselle in ordine, nella prima ho n possibilità di scelta, nella seconda $(n - 1)$, ..., nella k -esima $(n - k + 1)$.

Definizione. Una **permutazione** di n oggetti è una n -upla ordinata i cui elementi sono tutti gli n oggetti.

Detto altrimenti: è una disposizione semplice degli n oggetti (si tratta del caso $k = n$). Per quanto osservato, si ha la seguente proposizione:

Proposizione. *Il numero di permutazioni di n oggetti è il prodotto dei primi n numeri naturali:*

$$P(n) = n(n - 1) \dots 4 \cdot 3 \cdot 2 \cdot 1 \equiv n!$$

Il simbolo $n!$ si legge " n fattoriale" e indica il prodotto dei primi n numeri naturali. Per convenzione si pone $0! = 1$.

Esempio: Le permutazioni di 5 clienti di banca (cioè i possibili modi di metterli in ordine di attesa ad uno sportello) sono $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$.

Combinazioni semplici

In alcuni casi, quando si sceglie un insieme di oggetti da un insieme più grande, non interessa l'ordine con cui sono stati scelti gli oggetti, ma soltanto quali oggetti vengono scelti. Ogni gruppo distinto di oggetti distinti che possono essere selezionati senza considerare l'ordine, è detto **combinazione semplice**.

Definizione. Una **combinazione semplice** di n oggetti dati presi k alla volta è un sottoinsieme (senza struttura d'ordine) di k oggetti distinti scelti tra gli n .

Esempio: Le combinazioni di 3 oggetti a, b, c , presi due alla volta sono:

$$\{a, b\}, \{b, c\}, \{a, c\}.$$

Si noti: $\{a, b\} = \{b, a\}$. Per gli insiemi astratti (per i quali si usa la parentesi graffa) non vige alcuna struttura d'ordine.

Proposizione. Il numero di combinazioni semplici di n oggetti dati presi k alla volta, che indichiamo con $C(n; k)$, è:

$$C(n; k) = \frac{n(n-1)\dots(n-k+1)}{k!}$$

che si indica anche col simbolo $\binom{n}{k}$ e si può scrivere in forma più compatta:

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

Infatti per ciascuna combinazione semplice in cui si prendono k oggetti alla volta, esistono $P(k)$ modi di metterli in ordine. Quindi il numero di disposizioni $D(n; k)$ è più grande del numero di combinazioni $C(n; k)$ e precisamente:

$$D(n; k) = C(n; k) \cdot P(k) \quad \implies \quad C(n; k) = \frac{D(n; k)}{P(k)}$$

da cui segue la tesi della proposizione.

Esempio: Il numero di gruppi di studio di 4 persone che si possono formare da un insieme di 9 studenti è:

$$C(9; 4) = \binom{9}{4} = 9!/[4!(9-4)!] = (9 \cdot 8 \cdot 7 \cdot 6)/(4 \cdot 3 \cdot 2 \cdot 1) = 126.$$

Scegliere una combinazione di k oggetti da un insieme di n divide gli oggetti in due sottoinsiemi: i k che sono stati scelti e gli $n - k$ che non sono stati scelti. Talvolta un insieme può essere diviso in più di due sottoinsiemi.

Proposizione. Il numero di modi per dividere un gruppo di n oggetti in gruppi da k_1, \dots, k_r oggetti, con $k_1 + \dots + k_r = n$ è

$$\frac{n!}{k_1! \cdots k_r!}$$

Esempio: Un dado è stato lanciato 20 volte. Dato che per 3 volte è uscito “uno”, per 5 volte è uscito “due”, per 4 volte è uscito “tre”, per 2 volte è uscito “cinque” e per 3 volte è uscito “sei”, in quanti modi differenti possono essersi verificati i risultati?

Ci sono 20 risultati divisi in 6 gruppi (il gruppo formato dai 3 risultati in cui è uscito “uno”, il gruppo dei risultati in cui è uscito “due”...). Il numero di modi per dividere i 20 risultati in sei gruppi cosiffatti è

$$\frac{20!}{3!5!4!3!2!3!} = 1.955 \cdot 10^{12}$$

A volte può accadere di essere interessati a k -uple (ordinate o senza struttura d'ordine) di oggetti non necessariamente distinti tra loro scelti tra n oggetti dati. Se le k -uple sono ordinate, si parla di **disposizioni con ripetizione**; se l'ordine è irrilevante si parla di **combinazioni con ripetizione**.

Disposizioni con ripetizione

Definizione. Una **disposizione con ripetizione** di n oggetti a k a k è una k -upla ordinata i cui elementi, non necessariamente distinti, sono scelti fra gli n oggetti dati. Si noti che k , differentemente dal caso delle disposizioni semplici, può anche essere maggiore di n .

Esempio: Le disposizioni con ripetizione dei tre oggetti a, b, c a due a due (quindi: $n = 3, k = 2$) sono:

$$(a, a), (a, b), (b, a), (b, b), (b, c), (c, b), (a, c), (c, a), (c, c)$$

Proposizione. Il numero di disposizioni con ripetizione di n oggetti presi k alla volta è:

$$D^R(n; k) = n^k.$$

Infatti se si riempiono k caselle in ordine, nella prima casella ho n possibilità di scelta, nella seconda ho ancora n possibilità, e così per tutte le altre caselle. Quindi: numero di oggetti elevato al numero di caselle.

Esempio: Il numero delle possibili schedine del totocalcio è 3^{13} ; infatti è il numero di disposizioni con ripetizione dei 3 simboli 1, 2, x, in 13 caselle ordinate.

Combinazioni con ripetizione

Definizione. Una **combinazione con ripetizione** di n oggetti a k a k è un insieme non ordinato di k oggetti, non necessariamente distinti, scelti fra gli n oggetti dati.

Si noti che k , come per le disposizioni con ripetizione e differentemente dal caso delle combinazioni semplici, può anche essere maggiore di n .

Esempio: Le combinazioni con ripetizione dei tre oggetti a, b, c a due a due (quindi: $n = 3, k = 2$) sono:

$$\{a, a\}, \{a, b\}, \{a, c\}, \{b, b\}, \{b, c\}, \{c, c\}.$$

Analogamente, le combinazioni con ripetizione dei due oggetti a, b a tre a tre (quindi: $n = 2, k = 3$) sono:

$$\{a, a, a\}, \{a, a, b\}, \{a, b, b\}, \{b, b, b\}.$$

Proposizione. *Il numero di combinazioni con ripetizione di n oggetti presi k alla volta è:*

$$C^R(n; k) = \binom{n+k-1}{k}.$$

Dimostrazione: Si tratta di contare il numero di soluzioni (a_1, a_2, \dots, a_k) , con gli a_i numeri interi, soddisfacenti la relazione

$$1 \leq a_1 \leq a_2 \leq \dots \leq a_k \leq n.$$

Questa relazione equivale alla seguente

$$0 < a_1 < a_2 + 1 < a_3 + 2 < \dots < a_k + k - 1 < n + k,$$

che, a sua volta, equivale a

$$0 < b_1 < b_2 < b_3 < \dots < b_k < n + k,$$

con b_i interi. Si ha quindi che il numero cercato è uguale al numero di possibili scelte di k oggetti distinti presi dall'insieme $\{1, 2, \dots, n+k-1\}$ e cioè $C(n+k-1, k)$. \square

Esempio: Calcoliamo il numero di combinazioni con ripetizione di 3 oggetti presi a coppie e di 2 oggetti presi a terne utilizzando la formula della proposizione per verificare che si ottiene effettivamente 6 e 4 (come visto nell'esempio precedente).

$$\begin{aligned} C^R(3, 2) &= \binom{3+2-1}{2} = \binom{4}{2} = 6; \\ C^R(2, 3) &= \binom{2+3-1}{3} = \binom{4}{3} = 4. \end{aligned}$$

Esempio: Si consideri un gruppo costituito da 20 persone. La probabilità che almeno due fra queste 20 persone abbiano compleanno nello stesso giorno dell'anno è superiore o inferiore a $\frac{1}{2}$?

Basta calcolare la probabilità dell'evento complementare. Per l'evento complementare ("i 20 compleanni sono tutti distinti") il numero di casi favorevoli è il numero di disposizioni semplici di 365 oggetti a 20 a 20; il numero di casi possibili è il numero di disposizioni con ripetizione di 365 oggetti a 20 a 20:

$$\frac{365 \cdot 364 \cdot \dots \cdot 347 \cdot 346}{(365)^{20}} = \left(\frac{365}{365}\right) \left(\frac{364}{365}\right) \dots \left(\frac{347}{365}\right) \left(\frac{346}{365}\right) \approx 59\%$$

Pertanto il complemento a 1 di tale numero è inferiore a $1/2$.

Binomio di Newton

Proposizione. *Facendo la convenzione $0! = 1$ e chiamando anche in questi casi $\binom{n}{k}$ la quantità $n!/[k!(n-k)!]$, vale la seguente **formula binomiale di Newton**:*

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \dots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n$$

ovvero, in notazione compatta,

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k.$$

Dimostrazione:

$$(a+b)^n = (a+b)(a+b)\dots(a+b) \quad [n \text{ volte}]$$

è una lunga somma che contiene più volte l'addendo generico $a^{n-k}b^k$. Fissiamo k , con $k \leq n$. Quante volte appare tale addendo? Tante quante le scelte di k parentesi tra le n date, in cui pescare $b \cdot b \cdot \dots \cdot b$ k volte (è automatico allora che si pesca $a \cdot \dots \cdot a$ nelle rimanenti $n-k$ parentesi). In altre parole: tante volte quante sono le combinazioni semplici di k oggetti tra gli n dati, cioè $\binom{n}{k}$ volte. Quindi tale addendo va moltiplicato per $\binom{n}{k}$ e la somma va fatta rispetto a k come enunciato. \square

Esercizio: Provare la proprietà dei coefficienti binomiali

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k}.$$

(grazie alla quale si costruisce il famoso "triangolo di Tartaglia").

Per ispezione diretta:

$$\begin{aligned} \frac{(n-1)!}{(k-1)!(n-k)!} + \frac{(n-1)!}{k!(n-1-k)!} &= \frac{k(n-1)! + (n-k)(n-1)!}{k!(n-k)!} = \\ &= \frac{(n-1)!(k+n-k)}{k!(n-k)!} = \binom{n}{n-k}. \end{aligned}$$

1.5 Probabilità condizionata. Eventi indipendenti.

Spesso si vuole la probabilità di un evento B sotto la condizione che avvenga un altro evento A . Si consideri uno spazio di probabilità (\mathbf{S}, Ω, P) e due eventi A e B di Ω .

Definizione. Si dice **probabilità condizionata di B dato A**

$$P(B|A) := \frac{P(A \cap B)}{P(A)}, \quad (\text{con } P(A) > 0)$$

e analogamente la **probabilità condizionata di A dato B** :

$$P(A|B) := \frac{P(A \cap B)}{P(B)}, \quad (\text{con } P(B) > 0).$$

Essa esprime la probabilità che si verifichi il primo evento una volta che sia avvenuto il secondo.

Nel caso di uno spazio \mathbf{S} finito ed equiprobabile, indicato con $|E|$ il numero degli elementi di un evento $E \in \mathbf{S}$, si ha

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

e quindi

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

Esercizio. Si lanci una coppia di dadi. Se la loro somma è 6, si determini la probabilità che almeno uno dei dadi abbia dato come risultato 2.

Lo spazio campionario è

$$\mathbf{S} = \{(h, k), \quad h, k = 1, 2, 3, 4, 5, 6\},$$

per cui, indicati con A e B i due eventi

$$\begin{aligned} B &= \text{“la somma è 6”} = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}, \\ A &= \text{“almeno un 2”} = \{(2, 2), (2, k), (h, 2), \quad h, k = 1, 3, 4, 5, 6\}, \end{aligned}$$

si ha $A \cap B = \{(2, 4), (4, 2)\}$. Essendo lo spazio equiprobabile, ne consegue

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{2}{5}.$$

Esempio: In una popolazione i genotipi AA, Aa, aa abbiano probabilità rispettivamente

$$P(AA) = \frac{49}{100}, \quad P(Aa) = \frac{42}{100}, \quad P(aa) = \frac{9}{100}.$$

Supponiamo che dopo un certo tempo muoiano sistematicamente gli individui di tipo aa, sicchè gli adulti sono o AA o Aa. Qual è la probabilità di AA fra gli adulti?

Bisogna calcolare la probabilità condizionata di AA dato l'evento $C = AA \cup Aa$:

$$\begin{aligned} P(AA | AA \cup Aa) &= \frac{P(AA \cap [AA \cup Aa])}{P(AA \cup Aa)} = \frac{P(AA)}{P(AA \cup Aa)} = \\ &= \frac{0.49}{0.49 + 0.42} = \frac{0.49}{0.91} \approx 54\%. \end{aligned}$$

Teorema della probabilità composta. *Se gli eventi A, B hanno entrambi probabilità non nulla*

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B).$$

Dimostrazione: La dimostrazione segue banalmente dalla definizione di probabilità condizionata. \square

La legge appena formulata, che permette di calcolare la probabilità dell'intersezione di due eventi note la probabilità di uno e la probabilità condizionata dell'altro dato il primo, si può facilmente estendere a più eventi. Nel caso di tre eventi A_1, A_2 e A_3 si ha

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P((A_1 \cap A_2) \cap A_3) = \\ &= P(A_1 \cap A_2) \cdot P(A_3 | A_1 \cap A_2) = \\ &= P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2). \end{aligned}$$

Generalizzando al caso di n eventi A_1, A_2, \dots, A_n si ottiene

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdot \dots \cdot P(A_n | \bigcap_{i=1}^{n-1} A_i).$$

Esempio: Un'urna contiene 9 palline rosse e 6 gialle. Una dopo l'altra vengono estratte a caso, senza reimmissione, tre palline. Calcolare la probabilità che siano tutte rosse.

Denotiamo con A_k , con $k = 1, 2, 3$, l'evento "la k -esima pallina è rossa". L'evento di cui ci interessa la probabilità è $A_1 \cap A_2 \cap A_3$. Dal teorema della probabilità composta segue che

$$P(A \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) = \frac{9}{15} \cdot \frac{8}{14} \cdot \frac{7}{13} = \frac{12}{65}.$$

Definizione. Due eventi A,B sono **indipendenti** se

$$P(A \cap B) = P(A)P(B).$$

Dalla definizione di probabilità condizionata, in questo caso si ha

$$P(A|B) = P(A), \quad P(B|A) = P(B)$$

il che significa che la probabilità di A non dipende dal verificarsi o meno di B, e viceversa: ciò giustifica la terminologia.

Esempio: Un test diagnostico di una malattia è corretto nel 98% dei casi. Ripetendo due volte il test sullo stesso soggetto, qual è la probabilità di un doppio errore?

Sia A=“errore nel primo uso del test”, B=“errore nel secondo uso del test”. Essendo i due eventi indipendenti, si ha

$$P(A \cap B) = P(A)P(B) = (2/100)(2/100) = 4/10000 = 0.04\%.$$

Teorema. Se A e B sono indipendenti, lo sono anche A e B^C, A^C e B, A^C e B^C.

Dimostrazione: Dimostriamo dapprima l'indipendenza di A e B^C. Essendo

$$P(B^C) = 1 - P(B), \quad P(A) = P(A \cap B) + P(A \cap B^C),$$

si ha

$$\begin{aligned} P(A \cap B^C) &= P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) = \\ &= P(A) \cdot [(1 - P(B))] = P(A) \cdot P(B^C). \end{aligned}$$

Quindi, se A e B sono indipendenti, lo sono anche A e B^C. Scambiando l'ordine, si può dedurre che lo sono anche A^C e B, e quindi anche A^C e B^C. □

Esempio: La probabilità che il giocatore Aldo colpisca il bersaglio è $\frac{1}{4}$ e la probabilità che lo colpisca Bruno è $\frac{2}{5}$. Supposto che Aldo e Bruno sparino contemporaneamente contro il bersaglio (supponendo quindi gli eventi indipendenti), qual è la probabilità che 1) almeno uno dei due centri il bersaglio? 2) uno solo dei due centri il bersaglio?

Indicati con A l'evento “Aldo fa centro” e con B l'evento “Bruno fa centro”, l'evento “almeno uno fa centro” è A ∪ B. Siccome A e B sono indipendenti, avremo

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) =$$

$$= \frac{1}{4} + \frac{2}{5} - \frac{1}{4} \cdot \frac{2}{5} = \frac{11}{20}.$$

Per quanto riguarda invece l'evento "uno solo fa centro", esso è dato da $(A \cap B^C) \cup (A^C \cap B)$. Tenendo conto che A ed B^C sono indipendenti, così come A^C e B , e che gli eventi $A \cap B^C$ e $(A^C \cap B)$ sono incompatibili, si ha

$$\begin{aligned} P((A \cap B^C) \cup (A^C \cap B)) &= P(A \cap B^C) + P(A^C \cap B) = \\ &= P(A) \cdot P(B^C) + P(A^C) \cdot P(B) = \\ &= \frac{1}{4} \cdot \frac{3}{5} + \frac{3}{4} \cdot \frac{2}{5} = \frac{9}{20}. \end{aligned}$$

Teorema della probabilità totale. Dato uno spazio di probabilità (\mathbf{S}, Ω, P) e data una partizione finita ed esaustiva A_1, A_2, \dots, A_n di \mathbf{S} (cioè $A_1, A_2, \dots, A_n \in \Omega$ tali che $A_1 \cup A_2 \dots \cup A_n = \mathbf{S}$ con $A_i \cap A_j = \emptyset$, $i \neq j$), con $P(A_i) > 0$ per ogni i , si ha

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

Dimostrazione: Sia \mathbf{S} ripartito in eventi disgiunti ed esaustivi A_i , $i = 1, \dots, n$. Allora, per definizione di probabilità condizionata,

$$P(A_i \cap B) = P(A_i)P(B|A_i)$$

e, sommando per i che va da 1 ad n :

$$\sum_{i=1}^n P(A_i \cap B) \equiv P(B) = \sum_{i=1}^n P(A_i)P(B|A_i).$$

□

Esempio: Com'è noto, le trasfusioni di sangue sono possibili: dal gruppo O a tutti i gruppi; da A ai gruppi A, AB; da B ai gruppi B, AB; da AB al solo gruppo AB. Supponiamo anche che le frequenze dei gruppi sanguigni siano:

$$P(O) = 52\%, \quad P(A) = 32\%, \quad P(B) = 10\%, \quad P(AB) = 6\%.$$

Qual è la probabilità che un individuo, scelto a caso, possa donare sangue a un individuo pure scelto a caso ?

Si usa il teorema della probabilità totale: la probabilità di poter donare da parte di un "A" è una probabilità condizionata appunto al fatto di essere un "A", ... :

$$P(\text{don}) = P(O)P(\text{don}|\text{essere "O"}) + P(A)P(\text{don}|\text{essere "A"})$$

$$\begin{aligned}
& +P(B)P(\text{don}|\text{essere}“B”) + P(AB)P(\text{don}|\text{essere}“AB”) \\
& = (52/100) + (32/100)(32/100 + 6/100) + \\
& + (10/100)(10/100 + 6/100) + (6/100)(6/100) \simeq 66\%.
\end{aligned}$$

Formula di Bayes.

In forma semplice: Dato un evento B con $P(B) > 0$,

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

In forma generale: Dato un evento B con $P(B) > 0$ e data una partizione finita ed esaustiva A_1, A_2, \dots, A_n di \mathbf{S} , con $P(A_i) > 0$ per ogni i , si ha:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_k P(B|A_k)P(A_k)}.$$

Dimostrazione: In forma semplice segue direttamente dal teorema della probabilità composta. Sia ora lo spazio ripartito in eventi disgiunti ed esaustivi A_i , $i = 1, \dots, n$. La formula di Bayes nella forma semplice applicata ad A_i e B per i fissato dà:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)}$$

dove ora basta sostituire al denominatore la formula di probabilità totale. \square

Nota bene: dato un evento A , con $0 < P(A) < 1$, gli eventi A e A^C costituiscono la più semplice partizione di \mathbf{S} utilizzabile nell'applicazione del teorema di Bayes.

Gli eventi A_i possono essere considerati come possibili cause dell'evento B , o ipotesi che lo spiegano. Il fatto che costituiscano una partizione di \mathbf{S} , per cui certamente $B \subset \cup_i A_i$, comporta che se si verifica B , necessariamente si verifica anche uno (ed uno solo in virtù della incompatibilità) degli eventi A_i . In altre parole, l'insieme delle “cause” A_i è esaustivo: se si verifica B , una di esse deve aver agito. Una volta osservato l'evento B , ci si può chiedere quale sia la causa che ha effettivamente agito, e il teorema di Bayes risponde, naturalmente in senso probabilistico, a questa domanda.

La probabilità $P(A_i)$ è la probabilità che si verifichi A_i indipendentemente dal verificarsi o meno dell'evento B ; viene detta **probabilità a priori**. La probabilità condizionata $P(A_i|B)$ è la probabilità di A_i valutata sapendo che si è verificato B , e viene chiamata **probabilità a posteriori**.

Gli esercizi che seguono sono utili ad illustrare il significato di probabilità a priori e a posteriori, e come si applica il teorema di Bayes.

Con la formula di Bayes (in forma semplice) ottengo la probabilità di A dato B conoscendo la probabilità di B dato A. Ciò aiuta, ad esempio nelle diagnosi delle malattie.

Esempio: Se la probabilità teorica del sintomo B data la malattia A è il 30%, posso calcolare la probabilità che un paziente affetto dal sintomo B abbia la malattia A:

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

Supponendo che la percentuale della malattia e del sintomo in Emilia sia, rispettivamente, $P(A) = 0.15$ e $P(B) = 0.05$, la probabilità di malattia A dato il sintomo B è:

$$P(A|B) = \frac{(30/100)(15/100)}{5/100} = \frac{90}{100}.$$

Dunque la presenza del sintomo segnala la presenza della malattia nel 90% dei casi.

Esempio: Lampadine escono per il 60% da una linea di produzione A e per il 40% dalla linea B. Dalla prima linea esce un 2% di difettose, dall'altra esce un 3.8% di difettose. Con quale probabilità una lampadina difettosa è uscita dalla linea A?

Se D è l'evento "difettosa" i dati del problema sono:

$$P(D|A) = 0.02, \quad P(D|B) = 0.038, \quad P(A) = 0.6, \quad P(B) = 0.4$$

Il numero che cerchiamo è la probabilità condizionata di A dato D:

$$\begin{aligned} P(A|D) &= \frac{P(D|A) \cdot P(A)}{P(D|A) \cdot P(A) + P(D|A^C) \cdot P(A^C)} = \\ &= \frac{(0.02)(0.6)}{(0.02)(0.6) + (0.038)(0.4)} = \frac{0.012}{0.012 + 0.0152} = 0.441 = 44.1\% \end{aligned}$$

Esempio: Si sa che lo 0.5% dei soggetti di una città è AIDS. Si sa che i tests diagnostici danno diagnosi corretta nell'80% dei sani e nel 98% dei malati. Qual è la probabilità di esser sano posto che ti abbiano diagnosticato malato?

Consideriamo gli eventi: A= sano, A^C = malato, B=diagnosi di sanità, B^C =diagnosi di malattia. Sappiamo che

$$P(A^C) = 0.005, \quad P(B|A) = 0.80, \quad P(B^C|A^C) = 0.98.$$

Vogliamo $P(A|B^C)$, che calcoleremo con la formula di Bayes:

$$P(A|B^C) = \frac{P(B^C|A) \cdot P(A)}{P(B^C|A) \cdot P(A) + P(B^C|A^C) \cdot P(A^C)} =$$

$$= \frac{(0.995)(0.20)}{(0.20)(0.995) + (0.98)(0.005)} = 0.976$$

(incredibilmente alta: ma se stiamo dentro una categoria a rischio, avremmo una incidenza di malattia $P(A^C)$ più elevata, e dunque questa probabilità di errore più contenuta).

Ultima osservazione: notiamo che ci sono due modi di “scegliere a caso” da una popolazione:

- 1) campionamento con reimmissione
 - 2) campionamento senza reimmissione
- come si vede da questo esempio.

Esempio: Una scatola contiene 10 viti, di cui tre difettose. Si estraggono due viti a caso. Con quale probabilità nessuna delle due è difettosa?

Considero gli eventi A = “prima vite estratta non difettosa”, B = “seconda estratta non difettosa”. Estruendo con reimmissione, prima di estrarre la seconda volta abbiamo nella scatola l’identica situazione di 10 viti di cui tre difettose; pertanto:

$$P(A) = 7/10, \quad P(B) = 7/10, \quad P(A \cap B) = P(A)P(B) = 49\%.$$

Estruendo senza reimmissione l’evento B non è più indipendente da A :

$$P(A) = 7/10, \quad P(B|A) = 6/9,$$

$$P(A \cap B) = P(A)P(B|A) = \left(\frac{7}{10}\right)\left(\frac{6}{9}\right) \simeq 47\%$$

E’ chiaro che: se il numero di individui della popolazione è infinito o molto grande, non c’è differenza apprezzabile tra estrarre con reimmissione ed estrarre senza reimmissione. Allora conviene per semplicità calcolare ogni cosa “come se” si estraesse con reimmissione.

2 VARIABILI ALEATORIE

2.1 Definizioni e Proprietà

In molte situazioni si vorrebbe assegnare un valore numerico ad ogni possibile risultato di un esperimento. Tale assegnazione è chiamata **variabile casuale** o **variabile aleatoria**. Una variabile aleatoria assegna un valore numerico ad ogni elemento presente nello spazio campionario.

Definizione. Dato uno spazio di probabilità (\mathbf{S}, Ω, P) , si dice **variabile aleatoria** (o **variabile casuale**) una funzione X che ad ogni evento elementare $s \in \mathbf{S}$ fa corrispondere un numero $X(s) \in \mathbb{R}$, in modo che ogni insieme $\{s : a < X(s) \leq b\}$ sia un evento contenuto in Ω . Tale evento sarà anche indicato in modo più conciso “ $a < X \leq b$ ”.

Esempio: Se $\mathbf{S} = \{1, 2, \dots, 6\}$ è lo spazio campionario relativo all’esperimento del lancio del dado non truccato, definiamo $X :=$ “numero uscente da un lancio”, cioè

$$X(1) := 1, \dots, X(6) := 6.$$

Potremo allora introdurre e calcolare la probabilità che $X = 5$, che $1 < X \leq 4$, ecc.:

$$P(1 < X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = \frac{1}{2}$$

$$P(X \geq 3) = P(X = 3) + \dots + P(X = 6) = \frac{2}{3}, \quad P(X \leq 1, 5) = \frac{1}{6}.$$

Su uno stesso spazio di probabilità possono essere definite più variabili casuali. Ad esempio, una seconda variabile casuale Y sullo stesso \mathbf{S} può essere definita nel modo seguente:

$$\begin{array}{ll} Y := 0 & \text{se l'esito del lancio è pari;} \\ Y := 1 & \text{se l'esito del lancio è dispari.} \end{array}$$

Così:

$$P(Y = 0) = 1/2, \quad P(Y = 1) = 1/2, \quad P(1 < Y \leq 4) = 0, \dots \text{ecc.}$$

Definizione. Data una variabile aleatoria X definita sullo spazio di probabilità (\mathbf{S}, Ω, P) , si chiama **funzione di distribuzione** o **di ripartizione** di X la funzione $F : \mathbb{R} \rightarrow [0, 1]$ così definita:

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

Esempio: Consideriamo la variabile casuale Y definita nell'esempio precedente. Indicata con $F_Y(x)$ la funzione di distribuzione ad essa associata, avremo

$$F_Y(x) = P(Y \leq x) = \begin{cases} P(\emptyset) = 0 & \text{per } x < 0 \\ P(Y = 0) = \frac{1}{2} & \text{per } 0 \leq x < 1 \\ P(Y = 0) + P(Y = 1) = 1 & \text{per } x \geq 1. \end{cases}$$

La funzione di distribuzione gode di alcune proprietà che sono formalizzate nelle tre proposizioni che seguono. Di queste dimostriamo solo la prima, in quanto le altre sono facilmente deducibili a partire da questa.

Proposizione. *Sussiste la seguente relazione:*

$$P(a < X \leq b) = F(b) - F(a).$$

Dimostrazione: L'evento $X \leq b$ è l'unione dei due eventi $X \leq a$ e $a < X \leq b$, cioè degli eventi

$$\{s \in \mathbf{S} : X(s) \leq a\} \quad \text{e} \quad \{s \in \mathbf{S} : a < X(s) \leq b\},$$

che chiaramente sono incompatibili. Di conseguenza si ha

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b),$$

da cui segue banalmente la relazione che si voleva dimostrare. □

Proposizione. *La funzione di distribuzione è monotona non decrescente, cioè*

$$F(a) \leq F(b) \quad \text{se} \quad a \leq b.$$

Proposizione. *Sussistono i due limiti seguenti:*

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

Esistono due tipi di variabili aleatorie, le **variabili discrete** e le **variabili continue**. Una variabile aleatoria discreta è una variabile aleatoria i cui possibili valori formano un insieme discreto; in altre parole, i valori possono essere ordinati ed esistono dei salti tra valori adiacenti. Al contrario, i possibili valori di una variabile casuale continua contengono sempre un intervallo, ossia tutti i punti compresi fra due numeri. Nei prossimi paragrafi verrà data una definizione più precisa di questi due tipi di variabili aleatorie.

2.2 Variabili aleatorie discrete

Definizione. Una variabile aleatoria X è **discreta** se:

1) esiste un insieme finito o numerabile di valori x_j , ($j = 1, \dots, n$ oppure $j \in \mathbb{N}$) tali che

$$P(X = x_j) > 0;$$

2) la somma delle probabilità è uno:

$$\sum_j P(X = x_j) = 1.$$

Quindi una variabile aleatoria discreta è individuata dai valori x_j e dalle rispettive probabilità $p_j \equiv P(X = x_j)$ e può essere così rappresentata:

$$X : \begin{pmatrix} x_1, & x_2, & \dots \\ p_1, & p_2, & \dots \end{pmatrix}$$

In modo equivalente, essa è rappresentata dalla **funzione di probabilità** $f(x)$ definita come:

$$f(x) = \begin{cases} p_j & \text{se } x = x_j \text{ (} j = 1, 2, \dots \text{)} \\ 0 & \text{altrove.} \end{cases}$$

Infine una variabile aleatoria discreta è rappresentabile in modo equivalente con la funzione distribuzione $F(x)$ già definita per una qualunque variabile aleatoria, che nel caso discreto diventa:

$$F(x) := \sum_{x_j \leq x} f(x_j)$$

che è una funzione a gradini:

$$F(x) = \begin{cases} 0, & x < x_1 \\ p_1, & x_1 \leq x < x_2 \\ p_1 + p_2, & x_2 \leq x < x_3 \\ \dots & \\ p_1 + \dots + p_{k-1}, & x_{k-1} \leq x < x_k \\ \dots & \\ 1 & x \geq x_n \end{cases}$$

Poiché $f(x_j) = P(X = x_j)$, si ritrova per la funzione distribuzione:

$$F(x) = P(X \leq x), \quad P(a < X \leq b) = F(b) - F(a).$$

Esempio: Lancio di un dado: la funzione di probabilità è

$$f(x) = \begin{cases} \frac{1}{6} & \text{se } x=k \\ 0 & \text{altrove} \end{cases}$$

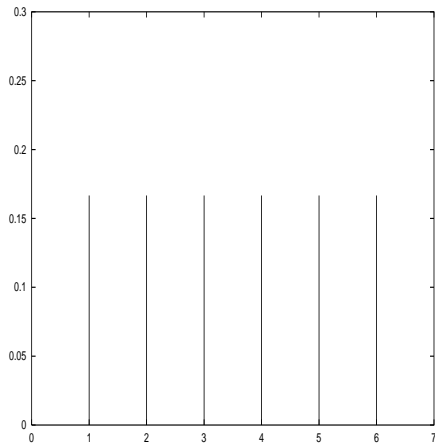


Figura 1: Funzione di probabilità $f(x)$ per il lancio del dado.

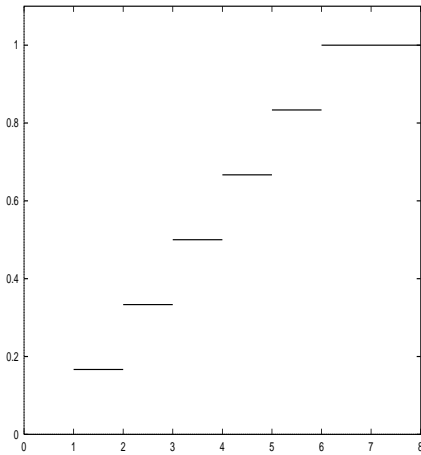


Figura 2: Funzione di distribuzione $F(x)$ per il lancio del dado.

ed $F(x) = 0$ per $x < 1$, $F(x) = 1/6$ per $1 \leq x < 2$, ..., $F(x) = 5/6$ per $5 \leq x < 6$, $F(x) = 1, \forall x \geq 6$.

Esempio: Sullo spazio di probabilità del lancio di due dadi (i cui eventi elementari sono le 36 coppie $(1, 1), (1, 2), \dots, (6, 5), (6, 6)$) sia $Z :=$ somma dei due numeri uscenti. Quindi:

$$Z : \left(\begin{array}{cccccccccccc} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{array} \right)$$

Per esercizio: descrivere il grafico di $f(x)$ e di $F(x)$.

2.3 Variabili aleatorie continue

Definizione. Una variabile aleatoria X è **assolutamente continua** se la funzione distribuzione $x \rightarrow F(x) \equiv P(X \leq x)$ è rappresentabile come funzione integrale di una funzione $f(\cdot) : \mathbb{R} \rightarrow \mathbb{R}^+$, cioè:

$$\forall x \in \mathbb{R}, \quad F(x) = \int_{-\infty}^x f(t) dt$$

La funzione $f(\cdot)$ è supposta almeno integrabile, ed è detta **densità di probabilità** della variabile aleatoria X .

Osservazione: Qui e altrove si usano integrali “impropri”, cioè integrali definiti in cui un estremo di integrazione (o entrambi gli estremi) è ∞ . Il significato è:

$$\int_{-\infty}^x f(t) dt := \lim_{a \rightarrow -\infty} \int_a^x f(t) dt, \quad \int_{-\infty}^{+\infty} f(t) dt := \lim_{a \rightarrow -\infty} \lim_{b \rightarrow +\infty} \int_a^b f(t) dt, \text{ ecc.}$$

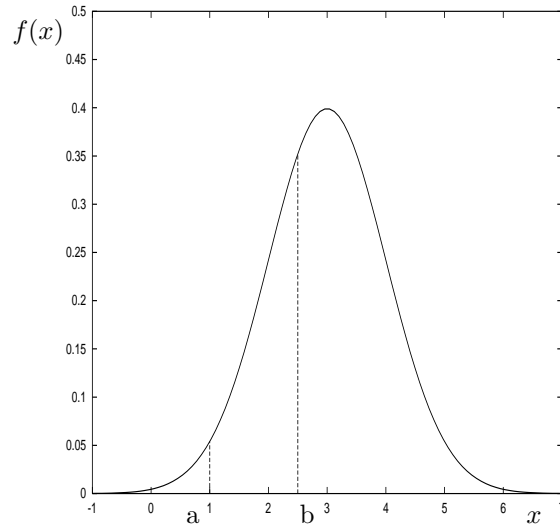


Figura 3: La probabilità $P(a < X \leq b)$ è l'area sotto la curva $f(x)$ tra gli estremi $x = a$ ed $x = b$.

Osservazione: La funzione f , che appare sotto il segno di integrale, si chiama **densità di probabilità** o semplicemente **densità** della variabile casuale. Derivando ambo i membri, avremo

$$F'(x) = f(x), \text{ in ogni } x \text{ dove } f \text{ sia continua.}$$

Osservazione: Nel seguito, salvo diversa indicazione, considereremo solo variabili aleatorie o discrete o assolutamente continue. Per brevità chiameremo queste ultime *variabili aleatorie continue*. Considereremo qualche volta anche *variabili aleatorie miste*, ma solo quando la rappresentazione sopra descritta di $F(\cdot)$ valga a tratti, a causa solo di qualche discontinuità di prima specie.

Osservazione importante: Poiché $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$, il modo standard di calcolare le probabilità come integrali della funzione densità è il seguente:

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(t)dt.$$

Perciò questa probabilità è uguale all'area sottesa dalla curva $f(x)$ tra gli estremi $x = a$ ed $x = b$ (Fig. 3).

Ciò comporta, ad es., che per una variabile aleatoria continua si ha sempre $P(X = a) = \int_a^a f(x)dx = 0$, (mentre può essere $P(X = a) > 0$ nel caso di variabile X discreta). Analogamente, se X è continua si ha

$$P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b),$$

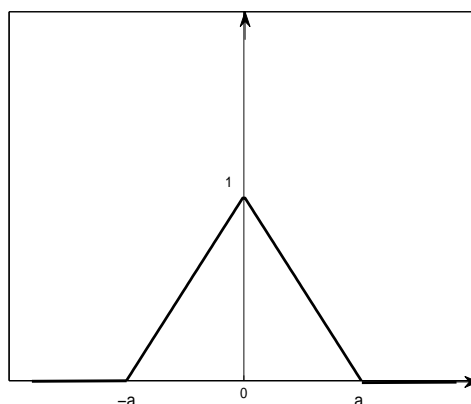


Figura 4: grafico della funzione triangolare.

(queste stesse probabilità possono differire tra loro nel caso di X discreta).

Esercizio: In un processo automatico si riempiono bottigliette di sciroppo. Il contenuto risulta $Y = 100 + X$ ml, dove X è una variabile casuale di densità

$$f(x) = \begin{cases} 1 - |x|, & \text{se } |x| \leq 1 \\ 0, & \text{altrove.} \end{cases}$$

Fare il grafico di $f(x)$ e di $F(x)$. In una partita di 1000 confezioni, quante approssimativamente conterranno meno di 99.5 unità di misura?

E' facile verificare che $f(x)$ è una densità, perché $\int_{-1}^1 f(x)dx = 1$, essendo questa l'area di un triangolo di base 2 e altezza 1 (vedi Fig. 4 con $a = 1$).

Il numero di bottigliette cercato è 1000 moltiplicato per $P(Y \leq 99.5) = P(100 + X \leq 99.5)$, cioè per

$$P(X \leq -0.5) = \int_{-1}^{-0.5} (1 - |x|)dx = \frac{1}{8}.$$

essendo questa l'area di un triangolo di base $\frac{1}{2}$ e altezza $\frac{1}{2}$. Il numero approssimativo di bottiglie aventi contenuto inferiore a quanto richiesto è dunque

$$\frac{1}{8} \cdot 1000 = 125.$$

2.4 Media e varianza

Definizione: Si dice **media** (o **valor medio** o **valore atteso** o **aspettazione matematica** o **speranza matematica**) di una variabile aleatoria X discreta:

$$\mu_X \equiv E(X) := \sum x_j f(x_j) \quad (2)$$

sotto l'assunzione che sia assolutamente convergente la corrispondente serie numerica: $\sum |x_j|f(x_j) < +\infty$.

Definizione: Si dice **media** (o **valor medio** o **valore atteso** o **aspettazione matematica** o **speranza matematica**) di una variabile aleatoria X continua:

$$\mu_X \equiv E(X) := \int_{-\infty}^{+\infty} xf(x)dx. \quad (3)$$

sotto l'assunzione che sia assolutamente convergente il corrispondente integrale: $\int_{-\infty}^{+\infty} |x|f(x)dx < +\infty$.

Osservazione: La media appena definita dipende dalla variabile aleatoria esaminata; essa, nel caso discreto, è la somma dei valori x_j moltiplicati per le rispettive probabilità $f(x_j) \equiv P(X = x_j)$. Essa rappresenta dunque la *media ponderata* dei possibili valori di X , ciascuno pesato con la sua probabilità.

Invece, per evitare confusioni, si rammenti che la somma di tutte le probabilità $f(x_j)$ è uno, qualunque sia la variabile aleatoria X : $\sum f(x_j) = \sum P(X = x_j) = 1$. Nel caso continuo, l'integrale su tutto \mathbb{R} della densità è 1:

$$\int_{-\infty}^{+\infty} f(x)dx = P(-\infty < X < +\infty) = 1.$$

Definizione: Si dice **varianza** (o **variazione standard**) di una variabile aleatoria X il numero, che indicheremo con σ_X^2 o $Var(X)$, così definito:

$$\sigma_X^2 \equiv Var(X) := E[(X - \mu_X)^2].$$

Quindi, tenendo conto della definizione di μ_X , la varianza di una variabile aleatoria X discreta sarà:

$$\sigma_X^2 \equiv Var(X) := \sum_j (x_j - \mu_X)^2 f(x_j), \quad (4)$$

mentre varianza di una variabile aleatoria X continua:

$$\sigma_X^2 \equiv Var(X) := \int_{-\infty}^{+\infty} (x - \mu_X)^2 f(x)dx. \quad (5)$$

Definizione: La radice quadrata della varianza si dice **deviazione standard** (o **scarto quadratico medio**) e si indica σ_X .

Osservazione: La varianza σ_X^2 è sempre non negativa; essa è nulla solo quando X è una variabile aleatoria discreta con funzione di probabilità tale che $f(x_1) = 1$ in un certo punto x_1 , ed $f(x) = 0$ altrove. Tranne questo unico caso, di nessun interesse probabilistico, si ha sempre $\sigma_X^2 > 0$.

Osservazione: La varianza (e quindi anche la deviazione standard) in un certo senso misura quanto è dispersa la variabile aleatoria rispetto alla media, ossia misura la “dispersione” dei valori assunti da X rispetto al suo valor medio μ_X : tanto più grande è σ_X^2 , tanto più i valori di X saranno lontani dal valor medio; viceversa, tanto più σ_X^2 è piccola, tanto più i valori di X saranno raccolti attorno a μ_X . Si può anche dire che *la media di una variabile casuale è tanto più attendibile quanto più piccola è la sua deviazione standard (o varianza)*.

Esempio: Se un’epidemia colpisce il 30% della popolazione, la probabilità di contagio per un singolo è di $p = 0.30$. La variabile aleatoria

$$X : \begin{pmatrix} 0 & 1 \\ 0.7 & 0.3 \end{pmatrix}$$

ha media e varianza rispettivamente

$$\mu_X = 0 \cdot 0.7 + 1 \cdot 0.3 = 0.3$$

$$\sigma_X^2 = (0 - 0.3)^2 \cdot 0.7 + (1 - 0.3)^2 \cdot 0.3 = 63/1000 + 147/1000 = 0.21$$

Sommando n variabili casuali identiche ad X si ottiene la variabile aleatoria $Z =$ numero di individui contagiati in un gruppo di n persone. Ad es. se $n = 2$ avremo la variabile aleatoria che può assumere i valori 0 o 1 o 2 :

$$Z := \begin{pmatrix} 0 & 1 & 2 \\ (7/10)^2 & 2 \cdot (7/10)(3/10) & (3/10)^2 \end{pmatrix}$$

Quindi in un gruppo di due persone il numero atteso di persone contagiate è:

$$\mu_Z = E(Z) = 0 \cdot 0.49 + 1 \cdot 0.42 + 2 \cdot 0.09 = 0.6$$

(non è detto che la media sia uno dei valori assunti dalla variabile aleatoria), con deviazione standard dalla media:

$$\sigma_Z = \sqrt{(0 - 0.6)^2 \cdot 0.49 + (1 - 0.6)^2 \cdot 0.42 + (2 - 0.6)^2 \cdot 0.09}.$$

Proposizione. *Si ha:*

$$E(X - \mu_X) = 0, \quad \sigma_X^2 = E(X^2) - \mu_X^2.$$

Dimostrazione: Nel caso discreto:

$$E(X - \mu_X) \equiv \sum_j (x_j - \mu_X) f(x_j) = E(X) - \mu_X = 0,$$

$$\begin{aligned}\sigma_X^2 &\equiv \sum_j (x_j - \mu_X)^2 f(x_j) = \sum_j (x_j^2 - 2\mu_X x_j + \mu_X^2) f(x_j) \\ &= \sum_j x_j^2 f(x_j) - 2\mu_X \sum_j x_j f(x_j) + \mu_X^2 \sum_j f(x_j) = E(X^2) - 2\mu_X^2 + \mu_X^2.\end{aligned}$$

Nel caso continuo è analogo il calcolo, ricordando che $\int_{-\infty}^{+\infty} f(x)dx = 1$, $\int_{-\infty}^{+\infty} xf(x)dx = \mu_X$. \square

Teorema (trasformazione lineare di variabile aleatoria). *Se una variabile aleatoria X ha media μ_X e varianza σ_X^2 , allora la variabile aleatoria $X^* = c_1X + c_2$, $c_1 \neq 0$, ha media e varianza:*

$$\mu_{X^*} = c_1\mu_X + c_2, \quad \sigma_{X^*}^2 = c_1^2\sigma_X^2.$$

Dimostrazione: Lo proviamo nel caso discreto. La variabile aleatoria $X^* = c_1X + c_2$ è la seguente:

$$\begin{pmatrix} c_1x_1 + c_2 & c_1x_2 + c_2 & \dots & c_1x_n + c_2 & \dots \\ f(x_1) & f(x_2) & \dots & f(x_n) & \dots \end{pmatrix}$$

Pertanto

$$\begin{aligned}\mu_{X^*} &= \sum_j (c_1x_j + c_2)f(x_j) = c_1 \sum_j x_j f(x_j) + c_2 \sum_j f(x_j) = c_1\mu_X + c_2 \\ \sigma_{X^*}^2 &= \sum_j (c_1x_j + c_2 - c_1\mu_X - c_2)^2 f(x_j) = \\ &= (c_1)^2 \sum_j (x_j - \mu_X)^2 f(x_j) = (c_1)^2 \sigma_X^2\end{aligned}$$

\square

Corollario (variabile standardizzata). *Se X ha media μ_X e varianza σ_X^2 , allora la corrispondente variabile aleatoria*

$$Z = \frac{X - \mu_X}{\sigma_X}$$

*ha media 0 e varianza 1 e viene detta **variabile standardizzata**.*

Dimostrazione: Basta prendere $c_1 = \frac{1}{\sigma_X}$ e $c_2 = -\frac{\mu_X}{\sigma_X}$. Pertanto, la variabile aleatoria $c_1X + c_2 = \frac{1}{\sigma_X} X - \frac{\mu_X}{\sigma_X}$ ha media e varianza rispettivamente:

$$\frac{\mu_X}{\sigma_X} - \frac{\mu_X}{\sigma_X} = 0, \quad c_1^2\sigma_X^2 = (\sigma_X^{-1})^2\sigma_X^2 = 1.$$

\square

3 DISTRIBUZIONI PIÙ COMUNI

L'inferenza statistica, come vedremo nel Cap.6, è quella parte della statistica che consente di estendere all'intera popolazione le informazioni fornite da un campione. In molte situazioni si ha a disposizione una conoscenza approssimativa della funzione di probabilità della popolazione. In questi casi, la funzione di probabilità può essere ben approssimata da una delle diverse funzioni note. In questo capitolo esamineremo tali funzioni e per ognuna di esse saranno descritte le condizioni di applicazione.

3.1 La distribuzione uniforme

Definizione. Una variabile casuale X continua si dice **uniformemente distribuita** o **equidistribuita** se la sua funzione densità $f(x)$ o, equivalentemente, la sua funzione di distribuzione $F(x)$ sono così definite:

$$f(x) = \begin{cases} 0 & \text{per } x \leq a \\ \frac{1}{b-a} & \text{per } a < x < b \\ 0 & \text{per } x \geq b \end{cases}; \quad F(x) = \begin{cases} 0 & \text{per } x \leq a \\ \frac{x-a}{b-a} & \text{per } a < x < b \\ 1 & \text{per } x \geq b \end{cases}.$$

Proposizione. La variabile aleatoria uniforme nell'intervallo $[a, b]$, $X \sim U([a, b])$, ha media e varianza:

$$E(X) = \frac{a+b}{2}, \quad \text{Var}(X) = \sigma_X^2 = \frac{(b-a)^2}{12} \quad (6)$$

Dimostrazione:

$$\begin{aligned} \mu_X &= \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}; \\ \sigma_X^2 &= E(X^2) - \mu_X^2 = \int_{-\infty}^{+\infty} x^2 f(x) dx - \mu_X^2 = \\ &= \int_a^b \frac{x^2}{b-a} dx - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}. \end{aligned}$$

□

3.2 Legge di probabilità di Bernoulli

Consideriamo esperimenti casuali con due soli risultati possibili: ad esempio, il lancio di una moneta (che produce come risultato testa o croce); controllare se un individuo

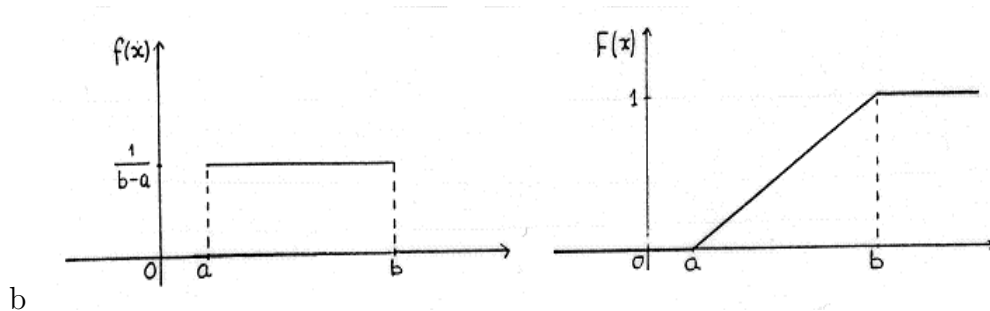


Figura 5: grafico della densità $f(x)$ e della funzione di distribuzione $F(x)$ della variabile aleatoria equidistribuita.

preso a caso da un gruppo è malato; controllare se un individuo scelto a caso da una popolazione batterica possiede una determinata caratteristica; se un dato prodotto, uscito da una linea di produzione, è difettoso oppure no. Un risultato è chiamato “successo”, l’altro “insuccesso”. Supponiamo che la probabilità di successo sia p . Di conseguenza, la probabilità di insuccesso sarà $1 - p$. Questo esperimento è detto **bernoulliano** o di **Bernoulli**.

Per ogni esperimento di Bernoulli la variabile aleatoria X è così definita:

Definizione. Sia $0 < p < 1$. Si dice **variabile aleatoria di Bernoulli** di parametro p (e la si denota con $X \simeq \text{Bern}(p)$) la variabile aleatoria X discreta che vale 1 se l’esperimento si conclude con un successo, 0 in caso contrario, ossia

$$X : \begin{pmatrix} 0 & 1 \\ 1 - p & p \end{pmatrix}.$$

La funzione di probabilità di Bernoulli sarà pertanto:

$$P(X = x) \equiv f(x) = \begin{cases} q = 1 - p & \text{se } x = 0 \\ p & \text{se } x = 1 \\ 0 & \text{altrove} \end{cases}$$

Media e Varianza della variabile aleatoria di Bernoulli

Teorema. La variabile aleatoria di Bernoulli $X \simeq \text{Bern}(p)$ ha media e varianza date da:

$$\mu_X = p, \quad \sigma_X^2 = p(1 - p).$$

Dimostrazione: Utilizzando le definizioni (2) e (4), si ottiene

$$\begin{aligned} \mu_X &= (0)(1 - p) + (1)(p) = p \\ \sigma_X^2 &= (0 - p)^2(1 - p) + (1 - p)^2p = p(1 - p) \end{aligned}$$

□

3.3 Legge di probabilità binomiale

Come abbiamo visto nel paragrafo precedente, si ha un esperimento di Bernoulli quando si considera un esperimento casuale con due soli esiti possibili. Se si suppone di fare più prove indipendenti di un esperimento casuale di Bernoulli, in cui la probabilità di successo è p , la variabile aleatoria che conta il numero di successi in queste prove si chiama **variabile aleatoria binomiale**.

Definizione. Sia $0 < p < 1$, $n \in \mathbb{N}$. Si dice **variabile aleatoria binomiale** di parametri p ed n (e la si denota con $X \simeq \text{Bin}(n, p)$) la variabile aleatoria X discreta con funzione di probabilità:

$$P(X = k) \equiv f(k) = \binom{n}{k} p^k q^{n-k}, \quad q = 1 - p \quad k = 0, 1, \dots, n$$

ossia:

$$X : \begin{pmatrix} 0 & 1 & 2 & \dots & n-2 & n-1 & n \\ q^n & npq^{n-1} & \frac{n!}{2!(n-2)!} p^2 q^{n-2} & \dots & \frac{n!}{(n-2)!2!} p^{n-2} q^2 & np^{n-1} q & p^n \end{pmatrix}$$

Ebbene, sapendo che in una singola prova si dà un evento A con probabilità p , (oppure l'evento complementare con probabilità $q = 1 - p$), la variabile aleatoria binomiale descrive la situazione nel caso in cui si facciano n prove indipendenti dell'esperimento casuale. Generalizzando, siano A e $B \equiv A^C$ i due possibili risultati del nostro esperimento, e siano $p = P(A)$ e $q = P(B) = 1 - p$ le loro probabilità. Supponiamo poi che dell'esperimento in questione siano fatte n prove e sia X la variabile aleatoria che ne descrive il risultato: $X = k$ se l'evento A si verifica k volte. Ebbene, il teorema che segue ci dice che X è una variabile aleatoria Binomiale.

Teorema di Bernoulli. *La probabilità che in n prove indipendenti l'evento A avvenga esattamente k volte è*

$$\binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n$$

dove p è la probabilità di A in una singola prova e $q = 1 - p$.

Dimostrazione: Definiamo:

$$X = \text{“numero di volte in cui } A \text{ avviene nell'ambito di } n \text{ prove”}$$

Vogliamo provare che X è la variabile aleatoria binomiale quale è stata sopra definita. Ovviamente X può assumere i valori $k = 0, 1, \dots, n$. Inoltre dobbiamo provare che X ha per funzione di probabilità f . \square

Ora, nello spazio di probabilità Ω gli eventi elementari sono n -uple ordinate di oggetti uguali ad “A” o a “B $\equiv A^C$ ”. Un particolare evento elementare è:

$(AA\dots ABB\dots B)$, dove A compare k volte, B compare $n - k$ volte

e significa che A avviene le prime k volte e B avviene le rimanenti $n - k$ volte. Poiché le prove sono indipendenti la probabilità di questa particolare n -upla è il prodotto delle probabilità:

$$pp\dots p qq\dots q = p^k q^{n-k}.$$

Ma questa n -upla è solo *un particolare* modo di disporre in ordine k volte A ed $n - k$ volte B. Ora, posso etichettare le n prove con $1, \dots, n$, e ci sono $C(n; k)$ modi di scegliere k di queste etichette tra le n date: proprio il numero di combinazioni di n oggetti a k a k . Cioè ci sono $C(n; k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$ modi di mettere k volte l'evento A nelle n prove. Quindi moltiplicando per questo numero otteniamo:

$$P(X = k) \equiv f(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n$$

che è proprio la funzione di probabilità di Bernoulli. □

Si noti che effettivamente la somma di tutte le probabilità è 1:

$$\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1^n = 1,$$

per la formula binomiale di Newton. Inoltre si può dimostrare il seguente teorema:

Teorema. *La variabile aleatoria binomiale $X \simeq \text{Bin}(n, p)$ ha media e varianza date da:*

$$\mu_X = np, \quad \sigma_X^2 = npq.$$

Esempio: Se la probabilità di avere un figlio maschio è $\frac{1}{2}$, per una famiglia con 5 figli, qual è la probabilità di avere: (i) due maschi? (ii) almeno un maschio? (iii) almeno 3 femmine?

Sia $X =$ “numero di maschi fra $n = 5$ figli”:

$$P(X = 2) = \binom{5}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = 10 \cdot (1/2)^5 = 10/32$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{5}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = 1 - (1/32) = 31/32$$

$$P(X \leq 2) = \sum_{k=0}^2 \binom{5}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{5-k} = (1/32) + 5 \cdot (1/32) + 10 \cdot (1/32) = 1/2.$$

Esempio: Sia $p = 98\%$ la probabilità che un test diagnostico dia risposta vera su un individuo. In un gruppo di 7 persone qual è la probabilità che il test dia risposta vera: (i) su tutti e 7? (ii) su almeno 6? (iii) su meno della metà? Qual è il valore atteso di diagnosi veritiere in un gruppo di 75 persone? con quale deviazione standard?

Se $X =$ “numero di diagnosi veritiere in un gruppo di $n=7$ individui”

$$P(X = 7) = \binom{7}{7} (98/100)^7 (2/100)^0 = (98/100)^7$$

$$P(X \geq 6) = \binom{7}{6} (98/100)^6 (2/100) + \binom{7}{7} (98/100)^7.$$

$$P(X \leq 3) = \sum_{k=0}^3 \binom{7}{k} (98/100)^k (2/100)^{7-k}.$$

Se $Y =$ “numero di diagnosi veritiere in un gruppo di $n=75$ individui”

$$E(Y) = np = 75 \cdot 98/100, \quad \sigma_Y = \sqrt{npq} = \sqrt{75 \cdot (98/100)(2/100)}.$$

Questo esempio mostra che conoscere il parametro p di una popolazione distribuita secondo Bernoulli permette di fare previsioni circa la composizione di un campione di lunghezza n : valore atteso, probabilità di estrarre un campione di composizione diversa da quella attesa,...

Quando si estrae un campione da una popolazione finita, se la numerosità della popolazione è elevata rispetto alla numerosità del campione, allora le unità del campione possono essere considerate come indipendenti. In caso contrario le unità non sono indipendenti. In alcuni casi lo scopo per cui si estrae un campione è quello di classificare ogni elemento estratto in una di due categorie (difettoso/non difettoso). In questo caso ogni estrazione può essere considerata una prova di Bernoulli in cui una categoria è denotata come successo e l'altra come insuccesso. Quando il numero di oggetti nella popolazione è grande rispetto al numero degli oggetti estratti nel campione, le prove di Bernoulli associate ad ogni estrazione possono essere considerate come indipendenti e il numero di successi in queste prove ha, per tutti gli scopi pratici una distribuzione binomiale. Quando invece l'ampiezza della popolazione non è grande, se paragonata a quella del campione, allora le prove di Bernoulli non sono indipendenti ed il numero di successi in queste prove non ha distribuzione binomiale.

Una regola pratica è quella di considerare indipendenti le prove se l'ampiezza del campione è più piccola del 5% di quella della popolazione e allora il numero di successi nel campione può essere considerato una variabile casuale binomiale.

3.4 Legge di probabilità di Poisson

Definizione. X è una **variabile aleatoria di Poisson** di parametro μ ($\mu > 0$) e si indica con la notazione $X \simeq Poisson(\mu)$ se può assumere gli infiniti valori $k = 0, 1, 2, \dots$ con probabilità

$$P(X = k) = f(k) = \frac{\mu^k}{k!} e^{-\mu}$$

ossia:

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & \dots & k & \dots \\ e^{-\mu} & \mu e^{-\mu} & \frac{\mu^2}{2!} e^{-\mu} & \frac{\mu^3}{3!} e^{-\mu} & \dots & \frac{\mu^k}{k!} e^{-\mu} & \dots \end{pmatrix}$$

Si osservi che effettivamente la somma di tutte le probabilità è 1:

$$\sum_{k=0}^{+\infty} \frac{\mu^k}{k!} e^{-\mu} = e^{\mu} \cdot e^{-\mu} = 1,$$

essendo $\sum_{k=0}^{+\infty} \frac{x^k}{k!} = e^x$ (è la ben nota serie esponenziale). Inoltre si dimostra che:

Teorema. *La variabile aleatoria di Poisson di parametro μ ha media μ e varianza anch'essa uguale a μ .*

Ciò significa che, aumentando la media, aumenta di pari passo anche la dispersione dei valori rispetto alla media.

La distribuzione di Poisson è tra le più importanti del calcolo delle probabilità. Per farne capire l'importanza nelle applicazioni e il gran numero di situazioni in cui essa si applica, elenchiamo alcuni casi in cui la distribuzione di Poisson descrive sufficientemente bene i dati osservati:

- il numero di particelle in una piccola porzione di sospensione;
- il numero casuale delle particelle α emesse da un corpo radioattivo e rilevate in un intervallo di tempo fissato; lo stesso vale per molte altre variabili aleatorie osservate in connessione con la radioattività;
- il numero di stelle osservate in una piccola area scelta a caso in un settore omogeneo del cielo;

- il numero delle bombe cadute su Londra nella seconda guerra mondiale, dividendo la pianta della città in quadrati della stessa area;
- il numero delle chiamate in arrivo ad un centralino telefonico, così come il numero di collegamenti ad un numero sbagliato;
- il numero di soldati prussiani morti in una settimana in seguito a calcio di cavallo.

La variabile aleatoria di Poisson è adatta quindi a descrivere il numero di fenomeni casuali distribuiti con una data densità media μ nell'unità di tempo o nell'unità di volume o nell'unità di superficie.... Essa può essere pensata come un'approssimazione di una variabile aleatoria binomiale con n grande e p piccolo. Gli esempi che seguono illustrano quanto affermato.

Esempio: Nel 1910 Rutherford e Geiger provarono che il numero di particelle α emesse al secondo da una sostanza radioattiva era una variabile aleatoria di Poisson con $\mu = 0.5$. Qual è la probabilità di osservare almeno due particelle durante un secondo ?

$$P(X \geq 2) = \sum_{k=2}^{+\infty} \frac{(0.5)^k}{k!} e^{-\mu} =$$

$$= 1 - P(X = 0) - P(X = 1) = 1 - e^{-0.5} - 0.5 \cdot e^{-0.5} = 1 - 0.91 = 9\%$$

Esempio: Una certa sospensione batterica contiene 5 batteri per cm^3 (valor medio). Qual è la probabilità che un campione causale di 1 cm^3 contenga (i) nessun batterio (ii) al più due batteri (iii) almeno 5 batteri?

$$P(X = 0) = e^{-5} \approx 0.007;$$

$$P(X \leq 2) = \left(1 + 5 + \frac{5^2}{2!}\right)e^{-5} \approx .125;$$

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \left(1 + 5 + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!}\right)e^{-5} \approx 0.561.$$

Esempio: Si desidera determinare la carica batterica di un campione di acqua.

Per valutare il numero di batteri in una sospensione se ne cerca la diluizione limite alla quale si trova ancora almeno un batterio capace di riprodursi. Ad esempio, se diluendo $1cm^3$ di acqua di canale prima con fattore $\frac{1}{10}$, poi $\frac{1}{100}$, quindi $\frac{1}{10^3}$, infine $\frac{1}{10^4}$ troviamo, dopo incubazione, sviluppo dei batteri mentre troviamo sterile la diluizione con fattore $\frac{1}{10^5}$, allora, grossolanamente, diremo che quel canale conteneva circa 10.000 germi per cm^3 . Per raffinare l'approssimazione della carica batterica presente nell'acqua di canale, usiamo la distribuzione di Poisson e inoculiamo in

20 provette la sospensione diluita con fattore $\frac{1}{10^4}$, mettendone 1cm^3 in ognuna. La distribuzione di Poisson permette di dire che se vi sono in media μ germi per cm^3 di diluito, vi sarà una proporzione $P(X = 0) = e^{-\mu}$ di tubi che non riceveranno alcun germe e perciò saranno sterili. Poniamo di trovare sterili 12 tubi su 20. Avremo dunque $e^{-\mu} = \frac{12}{20} = 0.6$ cioè $\mu = -\log(0.6) = -(\log_e 10) \cdot \log_{10}(0.6) = -2.3026 \cdot (-0.222) = 0.51$. Allora la concentrazione di germi nel canale è $0.51 \cdot 10^4 = 5.1 \cdot 10^3$ germi per cm^3 .

3.5 Altre leggi di probabilità discrete

Legge di probabilità ipergeometrica

Quando in una popolazione finita esistono due tipi di unità classificabili in successo e insuccesso e si estrae un campione casuale semplice, ogni elemento estratto costituisce un esperimento di Bernoulli. Ma dopo che ciascun elemento estratto viene incluso nel campione, la proporzione di successi o di insuccessi cambia. Per questa ragione le prove non sono tra loro indipendenti e quindi il numero di successi non segue una distribuzione binomiale. La distribuzione che esprime il numero di successi in questo caso prende il nome di **distribuzione ipergeometrica**.

Problema. Da un'urna contenente b palline bianche ed r rosse, se ne estraggono n ($n \leq b + r$) senza reimmissione. Qual è la probabilità che esattamente k di esse siano rosse?

Risposta. Supponiamo che le palline siano numerate da 1 a $b+r$ e che le palline rosse siano quelle con i numeri $\leq r$. Lo spazio Ω , degli eventi elementari è l'insieme di tutti i sottoinsiemi $\omega = \{\omega_1, \dots, \omega_n\}$ di $\{1, 2, \dots, b+r\}$. Quindi lo spazio di probabilità Ω è lo spazio delle combinazioni semplici di $b+r$ oggetti ad n ad n :

$$\#\Omega = |\Omega| = C(b+r; n)$$

Se poniamo

$$A_k = \{\omega : \omega \text{ ha esattamente } k \text{ elementi con indice } \leq r\}$$

la probabilità richiesta è il quoziente

$$P(A_k) = \frac{\# A_k}{\# \Omega} = \frac{C(r; k) \cdot C(b; n-k)}{C(b+r; n)} = \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{b+r}{n}}$$

La variabile aleatoria $X =$ “numero di palline rosse estratte nell'ambito di n estratte senza restituzione”, sapendo che “il numero di rosse è r su un totale iniziale di

$N = r + b$ è una **variabile aleatoria ipergeometrica** con parametri r, N, n e si denota con $X \sim H(N, r, n)$. Essa può assumere i valori $k = 0, 1, 2, \dots, r$, dove $n \leq N = b + r$. La sua funzione densità di probabilità è

$$P(X = k) = f(k) = \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{b+r}{n}}, \quad k = 0, 1, \dots, r$$

Inoltre, posto $N = b + r$, si ha:

$$E(X) = n \frac{r}{b+r} = n \frac{r}{N} \quad \sigma_X^2 = n \frac{r}{N} \frac{b}{N} \frac{N-n}{N-1}.$$

Esempio: Qual è la probabilità della terna $\{5, 51, 63\}$ nel gioco del lotto? Supponiamo di dividere gli $N = 90$ numeri in due classi: da una parte gli $r = 3$ numeri indicati, dall'altra parte gli altri $b = 87$. Inoltre ci sono $n = 5$ estrazioni senza reimmissione (quindi non indipendenti, qui non serve la variabile aleatoria binomiale). La variabile aleatoria

$X =$ numero di estratti dal primo gruppo nell'ambito di 5 estrazioni

$$P(X = k) = \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{b+r}{n}} = \frac{\binom{3}{3} \binom{87}{5-3}}{\binom{90}{5}} = \frac{1}{11748} = 0.0085\%.$$

Esempio: Una partita di 150 libri ne contiene 30 che presentano un difetto nella rilegatura. Se 10 libri vengono scelti a caso per un controllo, qual è la probabilità che 3 libri tra i 10 estratti siano difettosi? Effettuare il calcolo sia nell'ipotesi di estrazione senza reimmissione che in quella di estrazione con reimmissione.

Applicando la formula della distribuzione ipergeometrica con parametri $r = 30$, $N = 150$, $n = 10$ ($b = 120$ e $k = 3$) si ha

$$P(X = 3) = f(3) = \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{b+r}{n}} = \frac{\binom{30}{3} \binom{120}{7}}{\binom{150}{10}} \approx 0.2065$$

Se invece applichiamo la distribuzione binomiale $Bin(n, p)$ con $n = 10$ e $p = \frac{30}{150} = 0.2$, otteniamo

$$f(3) = \binom{10}{3} (0.2)^3 (0.8)^7 \approx 0.2013.$$

Osservazione: L'esempio appena presentato mostra che in certi casi la distribuzione ipergeometrica e quella binomiale producono risultati pressochè uguali (nel caso specifico la differenza tra i due è minore dell'1%). La spiegazione sta nel fatto che se r , b ed N sono grandi rispetto ad n , allora non è molto rilevante se il campionamento viene effettuato con o senza reimmissione, in quanto la distribuzione ipergeometrica può essere ben approssimata dalla distribuzione binomiale (con $p = \frac{r}{N}$). In una "popolazione infinita" si usa sempre la distribuzione binomiale indipendentemente dal tipo di campionamento.

Legge di probabilità geometrica

Si supponga di effettuare una sequenza di prove indipendenti di Bernoulli in cui ciascuna prova ha la stessa probabilità p di successo. Sia X il numero di prove che bisogna effettuare prima di ottenere un successo. Allora X è una variabile aleatoria discreta chiamata **geometrica**

La legge di probabilità geometrica interviene in risposta alla seguente domanda: *quante prove si devono aspettare per raggiungere il primo successo in una sequenza di prove di Bernoulli ripetute (indipendenti), in cui la probabilità di successo in ogni prova è p ?*

Definizione. Si dice che una variabile aleatoria X si distribuisce secondo la legge **geometrica** di parametro p ($0 < p \leq 1$) e si indica con $X \sim \text{Geom}(p)$ se la sua funzione di probabilità è data da

$$f(k) = P(X = k) = \begin{cases} p(1-p)^{k-1} & \text{per } k = 1, 2, \dots ; \\ 0 & \text{altrove.} \end{cases}$$

La variabile aleatoria geometrica indica, pertanto, il numero di prove necessarie per ottenere il primo successo (o, analogamente, il tempo di attesa del primo successo). Si ha, inoltre, che la media e la varianza della variabile aleatoria geometrica sono date, rispettivamente, da:

$$E(X) = \frac{1}{p}, \quad \sigma_X^2 = \frac{1-p}{p^2}.$$

Problema. Un dado viene lanciato più volte finché non si ottiene 6. Qual è la probabilità che occorran esattamente k lanci?

Risposta. È la probabilità che per $k - 1$ lanci esca "insuccesso" ed esca "successo" la k -esima volta; se X è il numero di lanci necessari ad avere successo,

$$P(X = k) = p(1-p)^{k-1} = \frac{1}{6} \left(\frac{5}{6}\right)^{k-1}, \quad k = 1, 2, \dots$$

La variabile aleatoria $T = X - 1$, cioè

$T =$ num. di prove bernoulliane precedenti il primo successo

ha legge: $P(T = k) = P(X = k + 1) = p(1 - p)^k$, $k = 0, 1, 2, \dots$

Sommando questi termini si ottiene la serie geometrica:

$$\sum_{k=0}^{\infty} p(1 - p)^k = p \frac{1}{1 - (1 - p)} = \frac{p}{p} = 1$$

□

Esempio: Un arciere ha probabilità $\frac{1}{3}$ di far centro in un bersaglio. Trovare la probabilità che gli occorra un numero di prove maggiore di 3.

Sia $X =$ “numero di prove necessarie per ottenere il primo centro nel bersaglio”, sapendo che la probabilità di far centro è $1/3$. Allora X è geometrica.

$$\begin{aligned} P(X \geq 4) &= 1 - P(X = 1) - P(X = 2) - P(X = 3) = \\ &= 1 - \frac{1}{3} - \frac{1}{3} \cdot \frac{2}{3} - \frac{1}{3} \cdot \left(\frac{2}{3}\right)^2 = \\ &= 1 - \frac{1}{3}[1 + 0.666 + 0.444] = 1 - (2.111)/3 \simeq 29.6\% \end{aligned}$$

Legge di probabilità binomiale negativa

La legge di probabilità binomiale negativa è una generalizzazione della geometrica. Essa interviene in risposta alla seguente domanda: *quante prove si devono aspettare per raggiungere l' r -esimo successo in una sequenza di prove di Bernoulli ripetute (indipendenti), in cui la probabilità di successo in ogni prova è p ?*

Definizione. Sia r un intero positivo. Si assuma che vengano effettuate delle prove di Bernoulli indipendenti, ognuna con probabilità di successo p e sia X il numero di prove necessarie per ottenere r successi. Si dice che X si distribuisce secondo la legge **binomiale negativa** di parametri r e p ($0 < p \leq 1$) e si indica con $X \sim NB(r, p)$ se la sua funzione di probabilità è data da

$$f(k) = P(X = k) = \begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r} & \text{per } k = r, r+1, \dots ; \\ 0 & \text{altrove.} \end{cases}$$

Il valore più piccolo che X può assumere è r , dal momento che saranno necessarie almeno r prove per avere r successi. Si osservi che quando $r = 1$, la distribuzione binomiale negativa coincide con la distribuzione geometrica, cioè $NB(1, p) = \text{Geom}(p)$.

La media e la varianza della variabile aleatoria binomiale negativa sono date, rispettivamente, da:

$$E(X) = \frac{r}{p}, \quad \sigma_X^2 = \frac{r(1-p)}{p^2}.$$

Esempio: Viene testata la resistenza di alcune saldature. Esse vengono sottoposte a sforzo finché non si rompono. Per un certo tipo di saldatura, l'80% delle rotture avviene al centro della saldatura mentre il 20% avviene ad un'estremità della saldatura. Vengono estratte a caso delle saldature per testarne la resistenza. Calcolare la probabilità che occorran 8 prove prima che 3 saldature si rompano ad un'estremità.

Sia X la variabile aleatoria che indica il numero di prove che vengono effettuate prima che tre saldature si rompano ad un'estremità. Pertanto $X \sim NB(3, 0.2)$. Allora si cerca

$$P(X = 8) = \binom{7}{2} (0.2)^3 (0.8)^5 \simeq 0.05505$$

3.6 Legge di probabilità normale o di Gauss

La legge di probabilità normale o gaussiana è tra le più utilizzate in ambito statistico. Essa fornisce un buon modello per molte popolazioni (anche se non per tutte). Il motivo è legato al teorema del Limite Centrale che verrà discusso nel paragrafo 4.6.

Definizione. Siano $\mu \in \mathbb{R}$, $\sigma > 0$. La variabile aleatoria continua X è detta **normale** o **di Gauss** con parametri μ e σ , e si scrive $X \sim N(\mu, \sigma^2)$, se la funzione densità è:

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

La funzione $f(x)$ è detta **funzione di Gauss**. È la funzione "a campana" simmetrica rispetto ad $x_o = \mu$, che ha un massimo per $x_o = \mu$, dove assume il valore massimo $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$. Quest'ultimo ha il significato di fattore di normalizzazione, cioè è quel numero tale che

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

Questa uguaglianza, la cui dimostrazione viene omessa, dice che f è una densità di probabilità, cioè $P(-\infty < X < +\infty) = 1$.

Si dimostra che la variabile aleatoria X ha media μ e varianza σ^2 . Si nota che più è piccolo σ , più è alto il picco $f(\mu)$, e dunque è più concentrata la campana intorno alla media μ : ciò concorda perfettamente con il significato di varianza che possiede σ^2 . Ecco il calcolo di media e varianza:

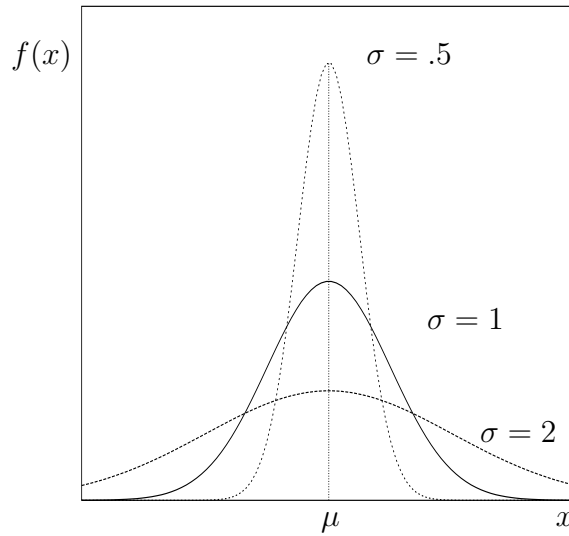


Figura 6: grafico della funzione di Gauss.

Lemma.

$$E(X) = \mu, \quad Var(X) = \sigma^2.$$

Dimostrazione: La media di X è ovviamente μ a causa della simmetria del grafico della densità attorno ad $x_0 = \mu$.

Per la varianza basta moltiplicare e dividere per $-\sigma^2$ e integrare per parti riconoscendo $-\frac{(x-\mu)}{\sigma^2} e^{-(x-\mu)^2/2\sigma^2}$ come fattore differenziale :

$$\int_{-\infty}^{+\infty} (x - \mu)^2 \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} dx = \lim_{R \rightarrow \infty} (-\sigma^2) \left[(x - \mu) \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \right]_{-R}^R +$$

$$-(-\sigma^2) \int_{-\infty}^{+\infty} f(x) dx = \sigma^2 \cdot P(-\infty < X < +\infty) = \sigma^2.$$

□

Dalla espressione della densità otteniamo la funzione distribuzione:

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/2\sigma^2} dt,$$

tale che

$$P(a < X < b) = F(b) - F(a) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/2\sigma^2} dt.$$

Naturalmente, trattandosi di una variabile aleatoria continua, si ha $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$.

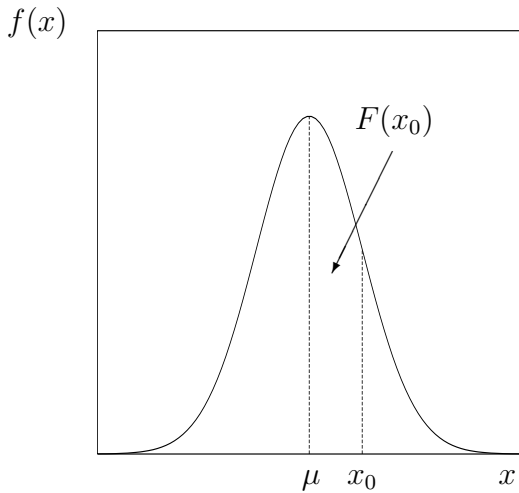


Figura 7: Funzione di probabilità $f(x)$ (funzione di Gauss) della variabile aleatoria normale.

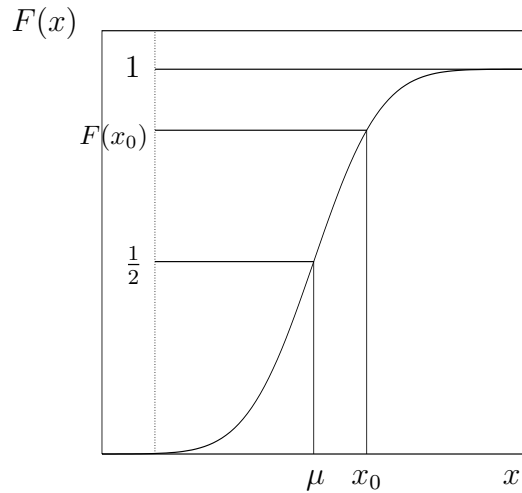


Figura 8: Funzione di distribuzione $F(x)$ della variabile aleatoria normale.

La funzione integrale F non si può calcolare coi metodi di integrazione elementari. Tuttavia, detta

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

la funzione distribuzione della variabile aleatoria **normale standardizzata**, cioè la variabile aleatoria normale con media 0 e varianza 1, si ha:

Proposizione. *La funzione distribuzione normale $F(x)$ di media μ e varianza σ^2 si può rappresentare in termini della funzione distribuzione normale $\Phi(x)$ di media 0 e varianza 1 nel seguente modo:*

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

In particolare:

$$P(a < X \leq b) = F(b) - F(a) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Dimostrazione: Ponendo $\frac{t - \mu}{\sigma} = u$, si ottiene $\frac{du}{dt} = \frac{1}{\sigma}$, $dt = \sigma du$, e quindi

$$\begin{aligned} F(x) &\equiv \lim_{R \rightarrow -\infty} \int_R^x \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/2\sigma^2} dt \\ &= \lim_{R \rightarrow -\infty} \int_{\frac{R-\mu}{\sigma}}^{\frac{x-\mu}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} e^{-u^2/2} \sigma du = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \equiv \Phi\left(\frac{x - \mu}{\sigma}\right). \end{aligned}$$

□

Osservazione: Quindi si usano le tavole di Φ per avere i valori di F . In particolare:

$$P(\mu - \sigma < X < \mu + \sigma) = \Phi(1) - \Phi(-1) \simeq 68\%$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = \Phi(2) - \Phi(-2) \simeq 95.5\%$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = \Phi(3) - \Phi(-3) \simeq 99.7\%.$$

Queste probabilità sono indicative del comportamento di una variabile aleatoria normale. La prima, ad esempio, ci dice che è ragionevole aspettarsi che più dei due terzi dei valori osservati di $X \simeq N(\mu, \sigma^2)$ cadano nell'intervallo $(\mu - \sigma, \mu + \sigma)$. La terza ci dice invece che fra mille osservazioni di X , mediamente solo tre cadono fuori dall'intervallo $(\mu - 3\sigma, \mu + 3\sigma)$.

Sarà utile ricordare:

$$P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 95\%$$

$$P(\mu - 2.58\sigma < X < \mu + 2.58\sigma) = 99\%.$$

Osservazione: I libri riportano una tavola che riguarda la variabile aleatoria normale standard Z e che si può usare in due modi:

- 1) dato un valore $z \in \mathbb{R}$, si cerca la probabilità $P(Z \leq z) = \Phi(z)$;
- 2) data una probabilità α (a volte assegnata come percentuale) si cerca il valore $z \in \mathbb{R}$ tale che $\alpha = P(Z \leq z)$. Tale z è denotato ϕ_α , e chiamato **quantile** relativo ad α , ovvero **percentile n -esimo** se $\alpha = n/100$.

Ricerche di quantità simili sono riconducibili alla tavola di $N(0, 1)$ tramite la proposizione precedente e considerazioni geometriche sulle aree sottese al grafico della densità: ad es. $\Phi(-z) = 1 - \Phi(z)$; $\phi_{0.95} = -\phi_{0.05}$.

Esempio: Sia $X \simeq N(0.8; 4)$ ossia X è normale con media 0.8 e varianza 4. Calcoliamo ad esempio:

$$P(X \leq -1.16) = \Phi[(-1.16 - 0.8)/2] = \Phi(-0.98) = 1 - \Phi(0.98) = 16.35\%$$

$$P(X \geq 1) = 1 - \Phi[(1 - 0.8)/2] = 1 - \Phi(0.1) = 46.02\%$$

$$P(2 \leq X \leq 3) = \Phi[(3 - 0.8)/2] - \Phi[(2 - 0.8)/2] = \Phi(1.1) - \Phi(0.6) = 13.86\%$$

Esercizio: Sia $X \simeq N(-2; 0.25)$. Determinare $c \in \mathbb{R}$ tale che:

- (a) $P(X \geq c) = 0.2$;
- (b) $P(-2 - c \leq X \leq -2 + c) = 90\%$.

$$0.2 = 1 - \Phi[(c + 2)/0.5]; \implies \Phi[(c + 2)/0.5] = 0.8; \implies (c + 2)/0.5 = 0.84;$$

$$\implies c = -1.58.$$

$$\begin{aligned} 0.9 &= \Phi[(-2 + c + 2)/0.5] - \Phi[(-2 - c + 2)/0.5] = \Phi(2c) - \Phi(-2c) = \\ &= \Phi(2c) - (1 - \Phi(2c)) = 2\Phi(2c) - 1, \end{aligned}$$

da cui

$$\Phi(2c) = 0.95; \implies 2c = 1.64 \implies c = 0.82$$

Enunciamo senza dimostrazione il seguente teorema.

Teorema. *Se X è normale con media μ e varianza σ^2 , allora $X^* = c_1X + c_2$ ($c_1 > 0$) è normale con media $\mu^* = c_1\mu + c_2$ e varianza $(\sigma^*)^2 = c_1^2\sigma^2$.*

3.7 Teorema di approssimazione di De Moivre e Laplace

Teorema di approssimazione di De Moivre e Laplace. *Siano a e b interi qualunque non negativi. Sia X la variabile aleatoria binomiale di parametri n e p . Sia Y la variabile aleatoria normale avente media np e varianza npq e sia Z la variabile aleatoria normale standardizzata, cioè $Z = (Y - np)/\sqrt{npq}$. Vale l'approssimazione:*

$$P(a \leq X \leq b) \simeq P(a - 0.5 \leq Y \leq b + 0.5) \quad \text{per } n \rightarrow \infty$$

cioè

$$P(a \leq X \leq b) \simeq P\left(\frac{a - 0.5 - np}{\sqrt{npq}} \leq Z \leq \frac{b + 0.5 - np}{\sqrt{npq}}\right) \quad \text{per } n \rightarrow \infty$$

dove “ \simeq ” significa che il quoziente tra le due espressioni tende ad 1 per $n \rightarrow \infty$.

Osservazione: Il termine 0.5 è una correzione dovuta al passaggio da una variabile aleatoria discreta a una continua. Il caso più evidente: se $a = b$, avremmo “probabilità = 0” per il semplice fatto di usare una variabile aleatoria continua; si rimedia prendendo $[a - 0.5, a + 0.5]$, intervallo lungo esattamente 1, che funge da base al rettangolo avente per altezza il valore della densità, che è circa la probabilità binomiale di a ; così l’area del rettangolo approssima la probabilità binomiale $P(X = a)$.

Esempio (approssimazione della Binomiale alla Normale): Determinare la probabilità di ottenere più di 25 “sette” in 100 lanci di una coppia di dadi equi.

La variabile aleatoria X = “numero di ‘sette’ nell’ambito di cento lanci” è binomiale con parametri $n = 100$ e $p = \frac{6}{36} = \frac{1}{6}$. Se Y è normale con la stessa media np e la stessa varianza npq abbiamo:

$$P(X \geq 26) \simeq P\left(N(0, 1) \geq \frac{25.5 - 100/6}{\sqrt{500/36}}\right) \simeq 1\%$$

Esempio (*approssimazione della Binomiale alla Normale*): Il 10% di bulloni prodotti da una certa macchina è difettoso. Trovare la probabilità che, in un campione casuale di 400, al massimo 30 siano difettosi.

$X \sim \text{Bin}(400, \frac{1}{10})$ ha media $\mu = np = 40$, e varianza $\sigma^2 = npq = 36$. Essendo $np > 5$ ed $n > 50$ è lecita l'approssimazione normale:

$$\begin{aligned} P(X \leq 30) &\simeq P\left(N(0, 1) \leq \frac{30.5 - np}{\sqrt{npq}}\right) = \\ &= P[N(0, 1) \leq -1.58] = 1 - P[N(0, 1) \leq 1.58] = 1 - 0.9429 = 0.0571 \end{aligned}$$

Osservazione: Sussiste anche un'approssimazione della variabile aleatoria binomiale alla variabile aleatoria di Poisson, sia pure in un diverso regime: la Poissoniana di media μ si può ottenere come caso limite della Binomiale se

$$p \rightarrow 0, \quad \text{ed } n \rightarrow \infty \quad \text{in modo tale che la media } np \rightarrow \mu.$$

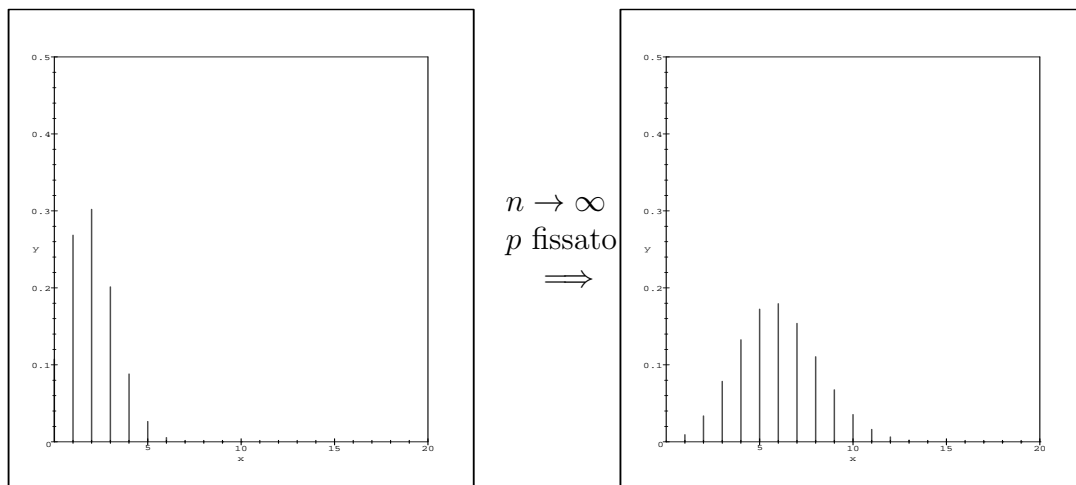
In questo regime, in sostanza, n è grande e p è piccolo in modo che $np \simeq npq$ (media \simeq varianza). L'idea che giustifica questa approssimazione è molto semplice: consideriamo una variabile aleatoria binomiale $X \sim \text{Bin}(n, \frac{\lambda}{n})$ e studiamo il suo comportamento per $n \rightarrow \infty$:

$$\begin{aligned} P(X = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

dove abbiamo i limiti:

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda} \\ \lim_{n \rightarrow \infty} \frac{n(n-1)\dots(n-k+1)}{n^k} &= 1 \end{aligned}$$

(il limite per $n \rightarrow \infty$ del rapporto di due polinomi aventi lo stesso grado è uguale al quoziente dei coefficienti del termine di grado massimo, qui entrambi uguali ad 1). Per questo la Poissoniana è detta a volte “la variabile aleatoria degli eventi rari”: infatti vive in un regime poissoniano ogni binomiale con p molto piccolo ed n grande. Tuttavia, se n cresce ulteriormente in rapporto a p , si entra allora nell'altro regime, la funzione di probabilità binomiale diventa sempre più simmetrica e sempre più simile alla funzione densità di Gauss. Ai fini pratici siamo già in regime “gaussiano” se $n \geq 50$ ed $np \geq 5$.



Esempio (*approssimazione della Binomiale alla Poissoniana*):

Un'azienda vende un preparato in partite di 200 confezioni con la garanzia che tutte siano non difettose; se la probabilità che una confezione sia difettosa è 0.5%, con quale probabilità almeno una partita viola la garanzia ?

La variabile aleatoria X = “numero di confezioni difettose nell'ambito di 200” è binomiale con parametri $n = 200$, $p = \frac{5}{1000}$. Essa è bene approssimata dalla Poissoniana Y di media $np = 200 \frac{5}{1000}$. Quindi:

$$P(X \geq 1) \simeq P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-1} \simeq 63\%$$

3.8 Legge di probabilità esponenziale e gamma

La legge di probabilità geometrica e binomiale negativa intervengono in risposta alla domanda: quante prove si devono aspettare per raggiungere l' r -esimo successo in una sequenza di prove di Bernoulli ripetute, indipendenti, in cui la probabilità di successo in ogni prova è p ? Allo stesso modo, la legge di probabilità esponenziale e la legge Γ intervengono in risposta alla seguente domanda: *quanto tempo deve aspettare un osservatore di una sequenza di eventi che si verificano nel tempo conformemente ad una legge di probabilità di Poisson di media λ , per osservare l' r -esimo verificarsi dell'evento?*

La densità esponenziale risponde allora alla stessa domanda a cui risponde la funzione di probabilità geometrica nei riguardi di quella binomiale: la legge esponenziale, cioè, regola il tempo di attesa di un evento A nell'ipotesi che il numero di volte che si verifica A nell'intervallo di lunghezza t sia retto da una legge poissoniana.

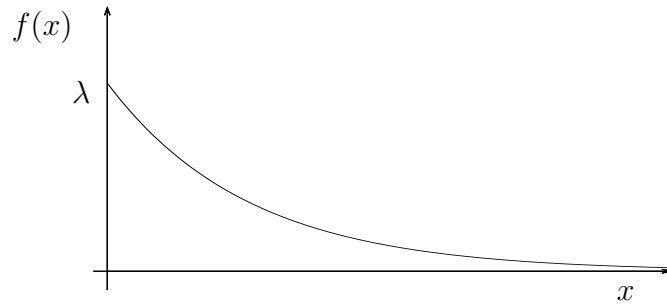


Figura 9: grafico della densità $f(x)$ della variabile aleatoria esponenziale.

Teorema: *Se il numero di occorrenze di un evento nell'unità di tempo è una variabile aleatoria di Poisson con media λ , allora il tempo T di attesa del primo verificarsi dell'evento è una variabile aleatoria **esponenziale** con parametro λ :*

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{per } t \geq 0 \\ 0 & \text{per } t < 0 \end{cases}$$

La variabile aleatoria esponenziale ha media $\frac{1}{\lambda}$ (“tempo medio di attesa”) e varianza $\frac{1}{\lambda^2}$.

Più in generale, il tempo di attesa dell' r -esimo verificarsi dell'evento ($r = 1, 2, \dots$) segue una legge **Gamma** con parametri r e λ , denotata $\Gamma(r, \lambda)$:

$$f_{T_r}(t) = \begin{cases} \lambda^r \frac{t^{r-1}}{(r-1)!} e^{-\lambda t} & \text{per } t \geq 0 \\ 0 & \text{per } t < 0 \end{cases}$$

la cui media e varianza sono, rispettivamente, r/λ ed $r/(\lambda)^2$.

Dimostrazione: Indichiamo con $F_T(t)$ la probabilità $P(T \leq t)$. Allora $1 - F_T(t)$ è la probabilità che il tempo di attesa della prima occorrenza sia maggiore di t . Ovvero, $1 - F_T(t)$ è la probabilità che il numero di eventi occorrenti da 0 a t sia zero. Poiché il numero di occorrenze in $[0, t]$ è di Poisson con media λt ,

$$1 - F_T(t) = e^{-\lambda t}, \quad \forall t > 0$$

Derivando si ottiene la densità di T : $F'_T(t) = \lambda e^{-\lambda t}, \forall t > 0$. Media e varianza si ottengono integrando per parti:

$$E(T) = \int_0^\infty t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}, \quad \text{Var}(T) = \int_0^\infty (t - \frac{1}{\lambda})^2 \lambda e^{-\lambda t} dt = \frac{1}{\lambda^2}$$

Stesso argomento per trovare la densità di $T_2 =$ “tempo di attesa per la seconda occorrenza dell'evento”:

$$1 - F_{T_2}(t) = e^{-\lambda t} + \frac{\lambda t}{1!} e^{-\lambda t}$$

$$\implies F'_{T_2}(t) = \frac{d}{dt}[-(e^{-\lambda t} + \lambda t e^{-\lambda t})] = \lambda e^{-\lambda t} - \lambda e^{-\lambda t} + \lambda^2 t e^{-\lambda t}$$

che è proprio una densità $\Gamma(2, \lambda)$. Analogamente si dimostra:

$$T_r \sim \Gamma(r, \lambda), \quad f_{T_r}(t) = \lambda^r \frac{t^{r-1}}{(r-1)!} e^{-\lambda t}, \quad t > 0$$

con media r/λ e varianza $r/(\lambda)^2$. □

Le variabili aleatorie con distribuzione esponenziale hanno notevole interesse applicativo in quanto utilizzabili per rappresentare diversi fenomeni che si incontrano nelle osservazioni scientifiche o nelle applicazioni tecnologiche. Di solito esse rappresentano i *tempi d'attesa* affinché un dato evento si verifichi. Ad esempio, se X indica il tempo misurato a partire dall'inizio del funzionamento di un dato pezzo di una macchina, ci si può chiedere *qual è la probabilità che il pezzo non si rompa prima che sia decorso un dato tempo x* . Ebbene, la risposta è data da $P(X \geq x)$, ossia

$$P(X \geq x) = 1 - F(x) = e^{-\lambda x}.$$

Una proprietà caratteristica delle *variabili casuali esponenziali* è che *non hanno memoria*. Questo fatto, che non dimostreremo, matematicamente è espresso dalla seguente relazione fra probabilità:

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

Ciò significa che se X è il tempo d'attesa fino al primo guasto di una data apparecchiatura, questo tempo non dipende dal fatto che l'apparecchiatura abbia già funzionato per un dato tempo s . In altre parole, la distribuzione di probabilità di X *non dipende dall'istante iniziale*.

Presentiamo qui di seguito due esercizi (che fanno parte dell'elenco finale degli esercizi proposti) i quali utilizzano nella risoluzione la distribuzione esponenziale.

Esercizio 3.3 *Il numero di chilometri (misurato in migliaia) che un dato pneumatico può percorrere prima di deteriorarsi è rappresentabile con una variabile aleatoria X avente distribuzione esponenziale con parametro $\lambda = 0.05$. Determinare la probabilità che un pneumatico di questo tipo duri (i) almeno 30 Km; (ii) tra i 35 e i 40 km.*

Si ha

$$P(X \geq 30) = 1 - F(30) = e^{-30\lambda} = e^{-0.05 \cdot 30} = e^{-1.5} \approx 0.223;$$

$$\begin{aligned}
P(35 \leq X \leq 40) &= F(40) - F(35) = (1 - e^{-40\lambda}) - (1 - e^{-35\lambda}) = \\
&= e^{-1.75} - e^{-2} \approx 0.174 - 0.135 = 0.039.
\end{aligned}$$

Esercizio 4.5 *Un apparecchio elettronico è composto da due elementi in parallelo, l'uno indipendente dall'altro e ciascuno con un tempo di vita esponenziale di media 8 giorni. Con quale probabilità l'apparecchio durerà un tempo non superiore a 12 giorni, supposto che esso funzioni se una almeno delle due componenti funziona?*

Poiché una variabile aleatoria esponenziale ha media uguale all'inverso del parametro λ , nel nostro caso si ha $\lambda = \frac{1}{8}$. Di conseguenza ciascuna componente ha un tempo di vita X_i , $i = 1, 2$, avente densità

$$f_{X_i} = \begin{cases} 0 & \text{per } t < 0 \\ \frac{1}{8}e^{-\frac{1}{8}x} & \text{per } t \geq 0 \end{cases}$$

Indicato quindi con Y il tempo di vita dell'apparecchio, si ha $Y = \max\{X_1, X_2\}$. Sapendo poi che X_1 e X_2 sono indipendenti, si ha

$$P(Y \leq t) = P(X_1 \leq t, X_2 \leq t) = P(X_1 \leq t) \cdot P(X_2 \leq t) = (1 - e^{-\frac{t}{8}})^2,$$

e quindi

$$P(Y \leq 12) = (1 - e^{-\frac{12}{8}})^2 = (1 - e^{-\frac{3}{2}})^2 \approx (1 - 0.223)^2 \approx 0.6035.$$

4 TRASFORMAZIONI DI VARIABILI ALEATORIE

4.1 Leggi congiunte di due variabili aleatorie

In un esperimento, invece che ad un unico risultato numerico, possiamo essere interessati a più valori (ad esempio, per una persona, a peso, altezza, età, ecc.). Ciascuno di tali valori è una variabile aleatoria, ma anche la n -upla di valori ottenuti può essere considerata come una variabile aleatoria multipla o n -dimensionale. In questo primo paragrafo per semplicità, ci limiteremo a trattare, seppure brevemente, le variabili aleatorie bidimensionali.

Definizione. Dato uno spazio di probabilità (\mathbf{S}, Ω, P) , si dice **variabile aleatoria bidimensionale** una coppia di funzioni (X, Y) che ad ogni $s \in S$ associa una coppia di numeri reali $(X(s), Y(s))$, tali che ogni insieme $\{s : a < X(s) \leq b, c < Y(s) \leq d\}$ sia un evento contenuto in Ω .

Pertanto, si dice che due variabili aleatorie X, Y sono distribuite congiuntamente (o che sono un sistema di variabili aleatorie) se sono definite sullo stesso spazio di probabilità. Allora si possono fare affermazioni di probabilità contemporaneamente su X e su Y , per le quali non sarebbe sufficiente conoscere le due leggi singole di X e di Y . Ad esempio se in un territorio X ed Y sono misure del CO e della densità di autoveicoli, c'è da aspettarsi una certa dipendenza, ossia informazioni sulla concomitanza di variabilità che non sono contenute nelle due singole leggi di probabilità.

La **legge di probabilità congiunta** di due variabili aleatorie distribuite congiuntamente è definita su insiemi misurabili B contenuti in \mathbb{R}^2 come segue:

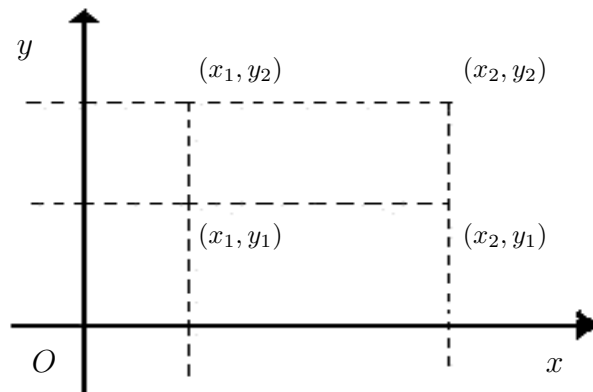
$$P_{X,Y}(B) = P(\{s \in \mathbf{S} : (X(s), Y(s)) \in B\})$$

Tale è il significato della scrittura più concisa da tutti usata:

$$P_{X,Y}(B) = P[(X, Y) \in B].$$

In particolare se l'insieme B è un prodotto cartesiano $B_1 \times B_2$, con $B_1, B_2 \subset \mathbb{R}$, allora $P_{X,Y}(B) = P(X \in B_1, Y \in B_2)$.

Definizione. Data una variabile aleatoria bidimensionale definita sulla spazio di probabilità (\mathbf{S}, Ω, P) , si dice **funzione di distribuzione congiunta** $F_{X,Y}(\cdot, \cdot)$ del sistema di due v.a X, Y la funzione di due variabili definita nel piano \mathbb{R}^2 a valori in



$[0, 1]$:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = P_{X,Y}((-\infty, x) \times (-\infty, y))$$

Essa contiene tutte le informazioni sulla legge congiunta $P_{X,Y}$, perché tutti gli insiemi misurabili di \mathbb{R}^2 si ottengono con unioni numerabili e complementazioni a partire dai rettangoli. Il lettore può verificare la seguente formula mediante considerazioni geometriche in \mathbb{R}^2 (vedi Fig.4.1):

$$\begin{aligned} P(x_1 < X \leq x_2, y_1 < Y \leq y_2) &= \\ &= F_{X,Y}(x_2, y_2) + F_{X,Y}(x_1, y_1) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2). \end{aligned}$$

È possibile dimostrare che:

$$\lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F_{X,Y}(x, y) = 1$$

e che

$$\lim_{x \rightarrow -\infty} F_{X,Y}(x, y) = \lim_{y \rightarrow -\infty} F_{X,Y}(x, y) = 0.$$

Dalla funzione di distribuzione congiunta $F_{X,Y}$ si possono ricavare le singole F_X, F_Y , dette **funzioni di distribuzione marginali**:

$$F_X(x) = P(X \leq x, Y < +\infty) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$$

$$F_Y(y) = P(X < +\infty, Y \leq y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y)$$

Ora specializziamo il discorso, fin qui valido in generale, alle variabili aleatorie discrete e poi a quelle continue.

Una variabile bidimensionale (X, Y) è **discreta** se esiste un insieme finito o numerabile di coppie di numeri reali (x_r, y_s) , $r = 1, 2, \dots$, $s = 1, 2, \dots$, tali che

$$P(X = x_r, Y = y_s) \geq 0, \quad \text{con} \quad \sum_{r,s} P(X = x_r, Y = y_s) = 1.$$

Definiamo **funzione di probabilità congiunta** di due variabili aleatorie discrete X, Y la funzione di due variabili

$$f_{X,Y}(x, y) = \begin{cases} P(X = x_r, Y = y_s) & \text{se } (x, y) = (x_r, y_s), r, s = 1, 2, \dots \\ 0 & \text{altrove.} \end{cases}$$

La **funzione di distribuzione congiunta** sarà, nel caso discreto:

$$F_{X,Y}(x_o, y_o) = \sum_{(x,y): x \leq x_o, y \leq y_o} f_{X,Y}(x, y)$$

La **funzione di probabilità marginale** f_X sarà allora:

$$f_X(x) = \sum_{y: f_{X,Y}(x,y) > 0} f_{X,Y}(x, y)$$

e analoga espressione avrà f_Y :

$$f_Y(y) = \sum_{x: f_{X,Y}(x,y) > 0} f_{X,Y}(x, y).$$

Le **distribuzioni marginali** $F_X(x_o)$ e $F_Y(y_o)$ saranno date da:

$$F_X(x_o) = \sum_{(x,y): x \leq x_o, y < +\infty} f_{X,Y}(x, y)$$

$$F_Y(y_o) = \sum_{(x,y): x < +\infty, y \leq y_o} f_{X,Y}(x, y)$$

Nel caso di una variabile aleatoria (X, Y) discreta finita, supposto $r = 1, 2, \dots, N$, $s = 1, 2, \dots, M$ e $p_{rs} = P(X = x_r, Y = y_s)$, le funzioni di probabilità congiunta e marginali vengono rappresentate attraverso la seguente tabella:

| | | Y | | | | |
|---|-------|-----------------|-----------------|-------|-----------------|----------------|
| | | y_1 | y_2 | | y_M | |
| X | x_1 | p_{11} | p_{12} | | p_{1M} | $p_{1\bullet}$ |
| | x_2 | p_{21} | p_{22} | | p_{2M} | $p_{2\bullet}$ |
| | ... | ... | ... | | ... | ... |
| | ... | ... | ... | | ... | ... |
| | x_N | p_{N1} | p_{N2} | | p_{NM} | $p_{N\bullet}$ |
| | | $p_{\bullet 1}$ | $p_{\bullet 2}$ | | $p_{\bullet M}$ | |

dove $p_{r\bullet} = \sum_s p_{rs}$ e $p_{\bullet s} = \sum_r p_{rs}$.

Una legge congiunta (o sistema) di due variabili aleatorie X, Y si dice **continua** se è determinata da una **funzione densità di probabilità congiunta**, cioè se esiste una funzione $f_{X,Y}(x, y)$ non negativa tale che

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = 1.$$

In altri termini, $P_{X,Y}$ è determinata da una densità congiunta $f_{X,Y}(\cdot, \cdot)$ se avviene

$$P_{X,Y}(B) = P((X, Y) \in B) = \int_B f_{X,Y}(x, y) dx dy$$

per ogni insieme misurabile B di coppie del piano. Dalla densità si ottiene la **funzione distribuzione congiunta**:

$$F_{X,Y}(x_0, y_0) = \int_{-\infty}^{x_0} dx \int_{-\infty}^{y_0} dy f_{X,Y}(x, y)$$

e viceversa

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

Le **densità marginali** e le **distribuzioni marginali** si ottengono in modo naturale:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx;$$

$$F_X(x_0) = \int_{-\infty}^{x_0} \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy \right) dx,$$

$$F_Y(y_0) = \int_{-\infty}^{y_0} \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx \right) dy.$$

Tutte le nozioni descritte si estendono ai sistemi (o leggi congiunte) di n variabili aleatorie X_1, X_2, \dots, X_n .

Esercizio: In due punti di un lago si misura l'intensità del suono causato da rumore di fondo generale (detto "rumore di ambiente"). Siano X, Y le due variabili aleatorie intensità del suono. Supponiamo che la loro legge congiunta sia continua con densità

$$f_{X,Y}(x, y) = \begin{cases} xy e^{-\frac{1}{2}(x^2+y^2)} & \text{se } x, y \geq 0 \\ 0 & \text{altrove} \end{cases}$$

Trovare: (a) le densità marginali; (b) $P(X + Y \leq 1)$; (c) la distribuzione della intensità massima di rumore, $Z = \max(X, Y)$; (d) la distribuzione dell'intensità minima di rumore $U = \min(X, Y)$.

$$f_X(x) = \int_0^{\infty} xy e^{[-\frac{1}{2}(x^2+y^2)]} dy = x e^{(-\frac{1}{2}x^2)},$$

$$f_Y(y) = \int_0^{\infty} xy e^{[-\frac{1}{2}(x^2+y^2)]} dx = y e^{(-\frac{1}{2}y^2)}.$$

La probabilità che la somma delle intensità sia ≤ 1 :

$$\begin{aligned} P(X + Y \leq 1) &= \int_{\{(x,y):x+y \leq 1\}} f_{X,Y}(x,y) dx dy = \\ &= \int_0^1 dx x e^{-\frac{1}{2}x^2} \int_0^{1-x} dy y e^{-\frac{1}{2}y^2} = \\ &= \int_0^1 (1 - e^{[-\frac{1}{2}(1-x)^2]}) x e^{-\frac{1}{2}x^2} dx \approx 0.2433 \end{aligned}$$

Per un numero positivo z , la probabilità che Z sia minore o uguale a z è data da

$$\begin{aligned} P[\max(X, Y) \leq z] &= P[X \leq z, Y \leq z] = \int_0^z dx \int_0^z dy xy e^{[-\frac{1}{2}(x^2+y^2)]} = \\ &= \left(\int_0^z x e^{-\frac{1}{2}x^2} dx \right)^2 = (1 - e^{-\frac{1}{2}z^2})^2 = F_Z(z) \end{aligned}$$

Infine per la variabile aleatoria “minimo” conviene considerare la probabilità che essa sia $\geq u$, e poi trovarne il complemento ad 1 :

$$\begin{aligned} P(U \geq u) &= P(\min(X, Y) \geq u) = P(X \geq u, Y \geq u) = \\ &= \int_u^{+\infty} dx \int_u^{+\infty} dy xy e^{[-\frac{1}{2}(x^2+y^2)]} = \\ &= \left(\int_u^{+\infty} dx x e^{-x^2/2} \right)^2 = e^{-u^2} = 1 - F_U(u) \end{aligned}$$

Quindi, per $u \geq 0$, la funzione distribuzione è $F_U(u) = 1 - e^{(-u^2)}$, mentre la densità è $f_U(u) = 2u e^{-u^2}$.

4.2 Indipendenza

Una nozione che generalizza quella dell'indipendenza tra due o più eventi riguarda le variabili aleatorie.

Definizione. Due variabili aleatorie distribuite congiuntamente X, Y sono **indipendenti** se la loro funzione di distribuzione congiunta $F_{X,Y}(\cdot, \cdot)$ può scriversi come prodotto delle singole funzioni di distribuzione marginali:

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

In altri termini, due variabili aleatorie X, Y sono indipendenti se, per ogni coppia di insiemi di Borel $A, B \subset \mathbb{R}$, si ha:

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

I due teoremi seguenti, che non dimostriamo, forniscono la condizione necessaria e sufficiente affinché due variabili aleatorie, discrete o continue, siano indipendenti.

Teorema. *Due variabili aleatorie congiuntamente continue X, Y , sono indipendenti se e solo se la loro funzione densità di probabilità è prodotto delle singole densità:*

$$f_{X,Y}(x, y) = f_X(x)f_Y(y). \quad (7)$$

Teorema. *Due variabili aleatorie congiuntamente discrete X, Y , sono indipendenti se e solo se la loro funzione di probabilità è prodotto delle singole funzioni di probabilità.*

$$P(X = x_r, Y = y_s) = P(X = x_r)P(Y = y_s). \quad (8)$$

La nozione di variabili aleatorie indipendenti si estende al caso di n variabili aleatorie: esse si dicono indipendenti se

$$\begin{aligned} P[X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n] = \\ = P[X_1 \in B_1] \cdot P[X_2 \in B_2] \dots \cdot P[X_n \in B_n] \end{aligned}$$

e analogamente ciò si trasmette alla espressione della densità congiunta, ecc.

Esempi. Le due variabili aleatorie X e Y i.i.d. (=indipendenti identicamente distribuite) siano uniformi in $[0, 1]$. Trovare:

- (a) $P(X + Y < \frac{1}{2})$;
- (b) $P(X^2 + Y^2 < \frac{1}{2})$;
- (c) $P(\cos(\pi Y) < \frac{1}{2})$.

(a) Per indipendenza, la densità congiunta è $f_X(x)f_Y(y)$:

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{se } (x, y) \in [0, 1]^2 \subset \mathbb{R}^2 \\ 0 & \text{altrove} \end{cases}$$

Dunque

$$P(Y < \frac{1}{2} - X) = \int_{y < 0.5-x} dx dy = \int_0^{\frac{1}{2}} dx \int_0^{\frac{1}{2}-x} dy = \frac{1}{8} = 12.5\%$$

Si può anche tracciare, nel supporto $[0, 1]^2$ della densità congiunta, il segmento congiungente $(0, \frac{1}{2})$ con $(\frac{1}{2}, 0)$ e vedere che l'area del triangolino sotteso è $\frac{1}{8}$.

(b) Traccio in $[0, 1]^2$ l'arco di cerchio $x^2 + y^2 = (\frac{1}{\sqrt{2}})^2$, e noto che l'area sottesa in $[0, 1]^2$ è $(\frac{1}{\sqrt{2}})^2 \cdot \pi \cdot \frac{1}{4} = 39.27\%$.

(c) Infine

$$P(\cos(\pi Y) < \frac{1}{2}) = P(\pi Y > \frac{\pi}{3}) = P(Y > \frac{1}{3}) = \frac{2}{3} = 66.67\%$$

4.3 Covarianza

Per una coppia di variabili aleatorie (X, Y) , ecco il parametro più utile a dare indicazioni sul comportamento congiunto di X e di Y :

Definizione. Si dice **covarianza** di una legge congiunta per le variabili aleatorie X ed Y con medie μ_X e μ_Y e si indica con $\sigma_{X,Y}$ o $Cov(X, Y)$, la quantità:

$$\sigma_{X,Y} = Cov(X, Y) := E[(X - \mu_X)(Y - \mu_Y)].$$

Dunque, se la legge congiunta è discreta, si ha:

$$Cov(X, Y) = \sum_{r,s} (x_r - \mu_X)(y_s - \mu_Y) f_{X,Y}(x_r, y_s).$$

Se la legge congiunta è continua, si ha:

$$Cov(X, Y) = \int_{R^2} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) dx dy.$$

Proposizione. La covarianza di due variabili aleatorie è uguale alla differenza “media del prodotto meno prodotto delle medie”:

$$Cov(X, Y) = E(X \cdot Y) - \mu_X \mu_Y. \quad (9)$$

Inoltre, mentre $E(X \pm Y) = E(X) \pm E(Y)$, la varianza di una somma soddisfa:

$$\sigma_{X \pm Y}^2 = \sigma_X^2 \pm 2 \cdot Cov(X, Y) + \sigma_Y^2. \quad (10)$$

Dimostrazione: Siano, per fissare le idee, X ed Y due variabili aleatorie discrete che assumono i valori $\{x_i\}$ ed $\{y_j\}$, rispettivamente, in modo che $P(X = x_i, Y = y_j) = f_{X,Y}(x_i, y_j)$. Poiché le due funzioni di probabilità marginali sono queste somme:

$$\sum_j f_{X,Y}(x_i, y_j) = f_X(x_i), \quad \sum_i f_{X,Y}(x_i, y_j) = f_Y(y_j),$$

allora:

$$\begin{aligned} E(X \pm Y) &= \sum_{i,j} (x_i \pm y_j) f_{X,Y}(x_i, y_j) = \\ &= \sum_i x_i \sum_j f_{X,Y}(x_i, y_j) \pm \sum_j y_j \sum_i f_{X,Y}(x_i, y_j) \\ &= \sum_i x_i f_X(x_i) \pm \sum_j y_j f_Y(y_j) = \mu_X \pm \mu_Y. \end{aligned}$$

Inoltre, usando questa additività della media,

$$\begin{aligned} Var(X \pm Y) &= E[((X \pm Y) - (\mu_X \pm \mu_Y))^2] = E[((X - \mu_X) \pm (Y - \mu_Y))^2] = \\ &= E[(X - \mu_X)^2] \pm 2 \cdot E[(X - \mu_X)(Y - \mu_Y)] + E[(Y - \mu_Y)^2] \end{aligned}$$

□

Teorema. *Date due variabili aleatorie indipendenti X ed Y , valgono le relazioni seguenti:*

- i) $Cov(X, Y) = 0$;
- ii) $\mu_{XY} = \mu_X \cdot \mu_Y$;
- iii) $\sigma_{X \pm Y}^2 = \sigma_X^2 + \sigma_Y^2$.

N.B. Covarianza sensibilmente positiva significa che le due variabili tendono a disporsi nella stesso verso rispetto alla loro media, come ad esempio nel caso della densità di autoveicoli e del CO nell'aria. Covarianza accentuatamente negativa rivela che $X - \mu_X$ ed $Y - \mu_Y$ tendono ad avere segno opposto, cioè una variabile è decrescente al crescere dell'altra, come nel caso di longevità e di stress. Quando la covarianza è nulla, le due variabili aleatorie si dicono *incorrelate*.

Definizione. Si dice **coefficiente di correlazione lineare** fra due variabili aleatorie X ed Y

$$\rho_{X,Y} := \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \equiv E\left[\frac{X - \mu_X}{\sigma_X} \frac{Y - \mu_Y}{\sigma_Y} \right].$$

N.B. Si ha per definizione $-1 \leq \rho_{X,Y} \leq 1$. Due situazioni antitetiche sono il caso $\rho_{X,Y} = 0$ (variabili aleatorie incorrelate) e il caso $\rho_{X,Y} = \pm 1$ (variabili aleatorie totalmente correlate). Quando $\rho_{X,Y} = \pm 1$, le variabili aleatorie X, Y sono *linearmente dipendenti*:

$$\exists a, b : Y = aX + b$$

In particolare Y cresce ($a > 0$) o decresce ($a < 0$) al crescere di X a seconda che $\rho_{X,Y} = 1$ oppure $\rho_{X,Y} = -1$.

Osservazione:

$$X, Y \text{ indipendenti} \implies X, Y \text{ incorrelate}$$

Infatti, se X, Y sono indipendenti, la media del loro prodotto è uguale al prodotto delle medie:

$$E(X \cdot Y) = \int_{R^2} xy f_{X,Y}(x, y) dx dy = \int_R x f_X(x) dx \int_R y f_Y(y) dy = \mu_X \cdot \mu_Y.$$

Tuttavia le due nozioni non sono equivalenti:

$$X, Y \text{ incorrelate} \not\Rightarrow X, Y \text{ indipendenti}$$

Ciò non toglie che, nel caso di due variabili aleatorie congiuntamente normali, esse sono indipendenti se e solo se $\rho_{X,Y} = 0$. Analoga situazione se si considerano n variabili aleatorie congiuntamente normali.

Esercizio: legge normale bivariata. La densità congiunta di due variabili aleatorie normali indipendenti è

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-(x-\mu_1)^2/2\sigma_1^2} \cdot e^{-(y-\mu_2)^2/2\sigma_2^2}$$

dove $X \simeq N(\mu_1, \sigma_1^2)$ e $Y \simeq N(\mu_2, \sigma_2^2)$. Se X ed Y sono dipendenti, la densità congiunta non è uguale al prodotto delle densità, ma dipende da un ulteriore parametro ρ :

$$f(x, y) = \frac{\exp\{-Q(x, y)\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}, \text{ dove}$$

$$Q(x, y) = \frac{1}{2(1-\rho^2)} \cdot \left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) \right]$$

Questa viene anche chiamata densità **normale bivariata**. Il parametro $\rho = \rho_{X,Y}$ varia tra -1 ed 1 , ed è il **coefficiente di correlazione** tra X ed Y .

La curva $Q(x, y) = \text{costante}$ è un'ellisse. Se si pone $Z = Q(X, Y)$, allora $P(Z > z) = e^{-z}$ per $z > 0$.

4.4 Combinazioni lineari di variabili aleatorie

Definizione. Se X_1, X_2, \dots, X_n sono n variabili casuali e c_1, c_2, \dots, c_n sono costanti, allora la variabile casuale $c_1X_1 + \dots + c_nX_n$ è detta **combinazione lineare** di X_1, \dots, X_n .

Vediamo nei seguenti teoremi i risultati relativi alla media ed alla varianza di una combinazione lineare di variabili aleatorie.

Proposizione. Se X_1, X_2, \dots, X_n sono n variabili casuali e c_1, c_2, \dots, c_n sono costanti, allora la media e la varianza della combinazione lineare sono date da

$$\mu_{c_1X_1+\dots+c_nX_n} = c_1\mu_{X_1} + \dots + c_n\mu_{X_n}; \quad (11)$$

$$\sigma_{c_1X_1+\dots+c_nX_n}^2 = \sum_{i=1}^n c_i^2 \sigma_{X_i}^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_i c_j \text{Cov}(X_i, X_j). \quad (12)$$

La dimostrazione di questo teorema è stata presentata per il caso $n = 2$ nel paragrafo 4.3.

Proposizione. Se X_1, X_2, \dots, X_n sono n variabili casuali indipendenti e c_1, c_2, \dots, c_n sono costanti, allora la varianza della combinazione lineare è data da

$$\sigma_{c_1X_1+\dots+c_nX_n}^2 = c_1^2 \sigma_{X_1}^2 + \dots + c_n^2 \sigma_{X_n}^2. \quad (13)$$

In particolare, la varianza della loro somma è uguale alla somma delle loro varianze

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n).$$

e, se le X_i hanno tutte la stessa varianza σ^2 , si ha:

$$\text{Var}(X_1 + X_2 + \dots + X_n) = n\sigma^2.$$

Per la somma di variabili aleatorie normali e indipendenti si ha il fatto seguente:

Teorema. Se X ed Y sono variabili aleatorie indipendenti distribuite normalmente:

$$X \sim N(\mu_X, \sigma_X^2), \quad Y \sim N(\mu_Y, \sigma_Y^2)$$

allora la somma $X \pm Y$ è normale con parametri $\mu = \mu_X \pm \mu_Y$ e $\sigma^2 = \sigma_X^2 + \sigma_Y^2$.

Esercizio: Siano X ed Y variabili aleatorie indipendenti ed ambedue normali $\sim N(0, 1)$. Trovare $P(X > Y + \frac{1}{2})$.

Basta notare che la variabile aleatoria $X - Y$ è normale con media 0 e varianza 2. Standardizzando:

$$P\left(X - Y > \frac{1}{2}\right) = P\left(\frac{X - Y}{\sqrt{2}} > \frac{1}{2\sqrt{2}}\right) = P[N(0, 1) > 0.35] = 36\%$$

4.5 Applicazione alla gestione del portafoglio

L'equazione (12) e le sue varianti giocano un ruolo importante nel campo finanziario. si assuma che un investitore abbia un numero prefissato di dollari da investire. Sceglie da una varietà di investimenti, per esempio, stock, bond e beni immobili. Dopo un anno venderà i suoi investimenti; sia X il suo profitto (o la sua perdita). Il valore di X non può essere determinato con certezza, così gli economisti trattano X come una variabile aleatoria. La media μ_X indica l'ammontare di guadagno atteso in media. La deviazione standard σ_X riflette la *volatilità*, o *rischio*, dell'investimento. Se σ_X è molto piccolo, allora è quasi certo che l'investimento avrà un guadagno prossimo al suo profitto medio μ_X e quindi il rischio è basso. Se σ_X è grande, il guadagno può variare su un range molto ampio, così che il rischio è alto. In generale, se due investimenti hanno lo stesso guadagno medio, quello con la deviazione standard più bassa è preferibile, dato che avrà lo stesso guadagno medio ma con un rischio più basso.

Esercizio: Un investitore ha 200\$ da investire. Investirà 100\$ in ognuno dei due investimenti. Siano X e Y i profitti per i due investimenti. Si assuma che $\mu_X = \mu_Y = 5\%$, $\sigma_X = \sigma_Y = 2\%$ e $\rho_{X,Y} = 0.5$. Si determinino la media e la deviazione standard del profitto totale sui due investimenti.

Il profitto totale è $X + Y$. La media è

$$\mu_{X+Y} = \mu_X + \mu_Y = 5\% + 5\% = 10\%.$$

Usando l'equazione (10) (che è la (12) con $n = 2$) si ottiene che la deviazione standard è

$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + 2Cov(X, Y) + \sigma_Y^2}$$

Ora $Cov(X, Y) = \rho_{X,Y}\sigma_X\sigma_Y = (0.5)(2)(2) = 2$. Quindi

$$\sigma_{X+Y} = \sqrt{2^2 + 2(2) + 2^2} = 3.46\%.$$

Può essere utile confrontare il risultato di questo esercizio con il risultato che si potrebbe ottenere se l'intero ammontare di 200\$ fosse riposto in un unico investimento. L'esempio che segue analizza questa possibilità.

Esercizio: Supponiamo che l'investitore dell'esempio precedente investa l'intera somma di 200\$ in uno dei due investimenti. Si calcolino media e deviazione standard del profitto.

Si assuma che l'investitore investa sull'investimento il cui profitto per 100\$ è X (il risultato non cambierebbe se si scegliesse Y). Dato che vengono investiti 200\$,

anziché 100\$, il guadagno sarà $2X$. Il guadagno medio è

$$\mu_{2X} = 2\mu_X = 2(5) = 10\$.$$

La deviazione standard è

$$\sigma_{2X} = 2\sigma_X = 2(2) = 4\$.$$

Confrontando i risultati dei due esercizi si vede che il profitto medio delle due strategie di investimento è lo stesso, ma la deviazione standard (cioè il rischio) è più basso quando il capitale dell'investimento è diviso in due investimenti. Questo è il principio di diversificazione. Quando sono disponibili due investimenti i guadagni dei quali presentano stessa media e stesso rischio, è sempre più vantaggioso dividere il capitale tra di essi, piuttosto che investirlo in uno solo di essi.

4.6 Approssimazione

La nozione di indipendenza, e poi la nozione di convergenza in legge che ora richiamiamo, ci permettono di enunciare l'importante teorema di limite centrale.

Definizione. La successione di variabili aleatorie reali $\{X_n\}_n$ **converge in legge** (o **in distribuzione**) alla v.a X se e solo se, dette F_n ed F le rispettive funzioni distribuzione, si ha

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

per ogni punto $x \in \mathbb{R}$ di continuità per F .

Teorema di limite centrale. Sia $\{X_n\}_n$ una successione di variabili aleatorie indipendenti e identicamente distribuite, di media μ e varianza $\sigma^2 > 0$. Allora la loro somma n -esima standardizzata

$$S_n^* = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

converge in legge ad una variabile aleatoria $N(0, 1)$.

Osservazione: La legge di S_n^* , che in generale è complicata da esprimere, si approssima, **per n grande**, con una legge $N(0, 1)$, e questo **qualunque** sia la legge delle variabili X_n . Il teorema di limite centrale essenzialmente dice che se un fenomeno aleatorio può essere riguardato come sovrapposizione di un gran numero di fenomeni aleatori indipendenti, aventi ciascuno una qualsiasi legge dello stesso tipo, allora tale fenomeno ha una distribuzione che, all'aumentare del numero dei fenomeni, converge alla normale. Ad esempio, si assume spesso che un errore di misurazione segua una

legge normale. Infatti, in assenza di errore sistematico, è ragionevole pensare che la discrepanza tra il valore vero e quello misurato sia la risultante di numerosi piccoli errori che si sono sovrapposti. Spesso l'esperienza conferma la validità di questa approssimazione.

N.B. Il teorema di limite centrale si applica senza difficoltà al caso delle prove di Bernoulli. Infatti se Y è il numero di successi in n prove indipendenti, allora $Y = X_1 + \dots + X_n$ dove ciascun X_i è la variabile aleatoria “numero di successi nella singola i -esima prova”:

$$Y = X_1 + X_2 + \dots + X_n, \quad \text{dove } X_i \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

Poiché le prove di Bernoulli sono indipendenti, ne consegue direttamente il tipo di approssimazione di De Moivre e Laplace nel senso di convergenza delle funzioni distribuzione.

Pur essendo originato dal problema di approssimare la binomiale, l'utilità del teorema di limite centrale va molto oltre: ogni volta che si studia una somma di variabili aleatorie, ma è sconosciuta la loro densità o ne è proibitivo il calcolo.

Esempio: Il tempo di sopravvivenza di una lampada è variabile aleatoria esponenziale di media $\mu = 10$ giorni. Appena si brucia, essa è sostituita. Trovare la probabilità che 40 lampade siano sufficienti per un anno.

Detta X_i la “durata della i -esima lampada”, per $i = 1, \dots, 40$ le X_i sono indipendenti ed esponenziali con parametro $\lambda = \frac{1}{\mu} = 1/10$. Sappiamo che $E(X_i) = \lambda^{-1} = 10$, $Var(X_i) = \lambda^{-2} = 100$. Allora la loro somma ha media $40 \cdot 10$ e varianza $40 \cdot 100$:

$$P(X_1 + X_2 + \dots + X_{40} \geq 365) \simeq P[N(0, 1) \geq \frac{365 - 400}{\sqrt{4000}}] = 71\%$$

4.7 Condizionamenti, Leggi condizionali

Consideriamo due variabili aleatorie X ed Y con una nota legge di probabilità congiunta. Ad es. sia nota la densità di probabilità congiunta $f_{X,Y}(x, y)$, e per suo mezzo ciascuna densità marginale. Il conoscere che la variabile aleatoria X ha assunto un particolare valore x induce *un condizionamento* della variabile aleatoria Y ; cioè può apportare una modifica della legge di probabilità marginale di Y .

Definizione: La **funzione distribuzione condizionata** della variabile aleatoria Y dato l'evento $\{X = x\}$ è per definizione

$$F_{Y|X}(y | x) := \left(\int_{-\infty}^y f_{X,Y}(x, \eta) d\eta \right) / f_X(x)$$

La **funzione densità di probabilità condizionata** $f_{Y|X}(y|x)$ della variabile aleatoria Y dato l'evento $\{X = x\}$ si ricava derivando la distribuzione condizionata

$$f_{Y|X} := \frac{d F_{Y|X}(y|x)}{dy} = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

Osservazione: Si ha dunque in generale che la densità congiunta è uguale alla densità condizionata di Y dato $X = x$ moltiplicata per la densità marginale di X : :

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) \cdot f_X(x).$$

In particolare, se la densità di probabilità *marginale* $f_Y(y)$ della variabile aleatoria Y coincide con la densità di probabilità *condizionata* $f_{Y|X}(y|x)$, allora le due variabile aleatoria sono *independenti*; infatti la densità congiunta risulta essere il prodotto delle due marginali:

$$f_Y(y) = f_{Y|X}(y|x) \implies f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y).$$

Esercizio: Il tempo di decadimento X di una particella in una camera a nebbia sia una variabile aleatoria esponenziale con parametro y . Tuttavia il valore y non è uguale per tutte le particelle bensì è distribuito come la variabile aleatoria Y . Si richiede la densità di X assumendo che Y sia uniforme in $[0, 1]$.

L'ipotesi appena descritta sul tempo X non è altro che un'ipotesi sulla legge condizionata di X dato Y :

$$f_{X|Y}(x|y) = y e^{-xy}, \quad \text{per } x > 0.$$

Cerchiamo la densità singola del tempo X , impiegato da una particella a caso per decadere: per $x > 0$

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy = \\ &= \int_{-\infty}^{+\infty} f_{X|Y}(x|y) \cdot f_Y(y) dy = \int_0^1 y e^{-xy} dy = \\ &= \left[y \frac{e^{-xy}}{-x} \right]_0^1 - \int_0^1 \left(\frac{e^{-xy}}{-x} \right) dy = -\frac{e^{-x}}{x} + \frac{(1 - e^{-x})}{x^2}. \end{aligned}$$

Esercizio: Ogni anno un tipo di macchina deve essere sottoposto ad alcuni arresti per manutenzione. Questo numero di arresti X è variabile aleatoria di Poisson con parametro y . Ma anche y è aleatorio (ad es. può dipendere dalla macchina) e assumiamo che esso segua una legge

$$f_Y(y) = y e^{-y}, \quad y > 0; \quad f_Y(y) = 0, \quad y < 0$$

(a) Qual è la probabilità che una singola macchina sia sottoposta a k arresti in un anno? (b) Sapendo che una macchina ha subito $k = 5$ arresti l'anno scorso, qual è la migliore stima della sua "propensione alla difettosità" Y ?

Risposta. Pur non dicendolo esplicitamente, il problema fornisce la funzione di probabilità condizionata di X dato $Y = y$:

$$P(X = k|Y = y) \equiv f_{X|Y}(k|y) = \frac{y^k}{k!} e^{-y}.$$

La densità congiunta è il prodotto fra questa e la densità della variabile aleatoria condizionante:

$$f_{X,Y}(k, y) = f_{X|Y}(k|y) \cdot f_Y(y) = \frac{y^k}{k!} e^{-y} \cdot y e^{-y}$$

per cui

$$\begin{aligned} P(X = k) &= \int_0^\infty f_{X,Y}(k, y) dy = \int_0^\infty \frac{1}{k!} y^{k+1} e^{-2y} dy = \\ (\text{posto } \tau = 2y) &= \int_0^\infty \left(\frac{1}{2}\right)^{k+2} \tau^{k+1} k! e^{-\tau} d\tau = \frac{k+1}{2^{k+2}} \end{aligned}$$

(b) Si tratta di calcolare la media condizionata di Y dato $X = 5$, quindi occorre la densità condizionata:

$$\begin{aligned} f_{Y|X}(y|k) &= \frac{f_{X,Y}(k, y)}{f_X(k)} = \\ &= \left[\frac{1}{k!} y^{k+1} e^{-2y} \right] \cdot \frac{2^{k+2}}{k+1} = \frac{(2y)^{k+1} \cdot 2}{(k+1)!} e^{-2y} \end{aligned}$$

Allora

$$\begin{aligned} E(Y|X = k) &= \int_0^\infty y f_{Y|X}(y|k) dy = \\ &= \int_0^\infty \frac{2^{k+1} \cdot 2}{(k+1)!} y^{k+2} e^{-2y} dy = \frac{2^{k+1} \cdot 2}{(k+1)!} \int_0^\infty e^{-\tau} \left(\frac{\tau}{2}\right)^{k+2} \frac{1}{2} d\tau = \\ &= \frac{(k+2)! \cdot 2}{2^2 (k+1)!} = \frac{k+2}{2} \quad \text{cioè } E(Y|X = 5) = \frac{7}{2} = 3.5 \end{aligned}$$

Essa è la migliore stima di Y in base al dato sperimentale $k = 5$.

È possibile anche condizionare eventi tramite la conoscenza del valore assunto da una variabile aleatoria. Consideriamo due variabili aleatorie X, Y distribuite congiuntamente. Sia $g(\cdot, \cdot)$ una funzione di due variabili, che determina l'evento $A = \{g(X, Y) \leq z\}$, per un numero reale z . Allora si usa spesso la riduzione:

$$P[g(X, Y) \leq z | X = x] = P[g(x, Y) \leq z].$$

Esercizio: Due giovani decidono di incontrarsi tra le 17 e le 18 con l'accordo che nessuno deve aspettare l'altro per più di 10 minuti. Supponiamo che gli orari X ed Y in cui arrivano siano indipendenti e casuali, variabili fra le 17 e le 18. Trovare la probabilità condizionata che i due giovani si incontrino, dato che lei arriva alle 17 : 30.

Risoluzione. Se lei arriva a un orario x e lui a un orario Y , si incontrano solo se $|Y - x| \leq 10$, cioè se $-10 + x \leq Y \leq x + 10$. Indichiamo con A l'evento che i due si incontrino. Allora per un $x \in [0, 60]$,

$$\begin{aligned} P(A|X = x) &= P(-10 \leq Y - X \leq 10|X = x) = \\ &= P(-10 + x \leq Y \leq x + 10|X = x) = \\ &= P(-10 + x \leq Y \leq x + 10) = F_Y(x + 10) - F_Y(-10 + x). \end{aligned}$$

Poiché Y è uniforme in $[0, 60]$, distinguiamo i casi: $-10 + x \leq 0$, e $x + 10 \leq 60$; $0 \leq -10 + x$ e $x + 10 \leq 60$; $-10 + x \leq 60$ ma $x + 10 \geq 60$; e infine $-10 + x \geq 60$.

$$P(A|X = x) = \begin{cases} \frac{10+x}{60} & \text{se } 0 \leq x \leq 10 \\ \frac{1}{3} & \text{se } 10 \leq x \leq 50 \\ \frac{70-x}{60} & \text{se } 50 \leq x \leq 60 \\ \text{non definita} & \text{se } x < 0 \text{ o } x > 60. \end{cases}$$

Quindi $P(A|X = 30) = 1/3$.

Infine menzioniamo anche un condizionamento di variabile aleatoria concettualmente più semplice: cerchiamo la legge condizionata di una variabile aleatoria X dato un evento B . In altre parole, data la funzione distribuzione di X , $F_X(x) = P(X \leq x)$, cerchiamo la funzione distribuzione di X condizionata al verificarsi di B avente probabilità positiva. A tal fine basta la definizione di probabilità condizionata come rapporto fra la probabilità dell'intersezione e la probabilità condizionante:

$$F_{X|B}(x|B) = \frac{P[(X \leq x) \cap B]}{P(B)}.$$

Si definisce, poi, la *funzione densità di probabilità condizionata* $f_{X|B}(x|B)$ mediante derivazione della F :

$$f_{X|B}(x|B) = \frac{dF_{X|B}(x|B)}{dx}.$$

Tali due funzioni godono appieno le proprietà di ogni funzione distribuzione e di ogni densità di probabilità.

Esempio: *Tempo di guasto dopo rodaggio.* Collaudiamo lampadine lasciandole accese fino al guasto (bruciatura del filamento). Un modello per il *tempo di guasto*

T di una lampadina è la variabile aleatoria esponenziale, avente densità non nulla per $t > 0$:

$$\forall t > 0, \quad f_T(t) = \lambda e^{-\lambda t}; \quad \frac{1}{\lambda} = \text{tempo medio di guasto.}$$

Modifichiamo ora il modo di collaudare: attendiamo un tempo fisso t_o dall'accensione, scartiamo le lampadine che all'istante t_o sono già guaste, e collaudiamo solo quelle ancora funzionanti. L'evento condizionante è insomma $B = \{T \geq t_o\}$: cerchiamo la distribuzione condizionata $F_{T|B}(t|B)$.

$$P(B) = P(T \geq t_o) = 1 - P(T \leq t_o) = 1 - F_T(t_o).$$

La probabilità dell'intersezione è

$$P[(T \leq t) \cap (T \geq t_o)] = \begin{cases} F_T(t) - F_T(t_o) & t \geq t_o \\ 0 & t < t_o. \end{cases}$$

Infatti è vuota l'intersezione fra i due eventi proprio quando $t < t_o$. Otteniamo allora:

$$F_{T|B}(t|B) = \frac{P[(T \leq t) \cap (T \geq t_o)]}{P(T \geq t_o)} = \begin{cases} \frac{F_T(t) - F_T(t_o)}{1 - F_T(t_o)} & t \geq t_o \\ 0 & t < t_o. \end{cases}$$

Quindi la densità della variabile aleatoria “tempo di guasto dopo rodaggio” è ottenuta derivando:

$$f_{T|B}(t|B) = \frac{d F_{T|B}(t|B)}{dt} = \begin{cases} \frac{f_T(t)}{1 - F_T(t_o)} & t \geq t_o \\ 0 & t < t_o. \end{cases}$$

Allora, in che modo il rodaggio modifica la densità esponenziale $f_T(t)$ originaria? Primo: la densità del tempo di guasto dopo rodaggio è ovviamente nulla per $t < t_o$: il rodaggio è stato superato con certezza, la lampadina si guasta solo dopo t_o . Inoltre per $t > t_o$ la densità modificata ha l'espressione di $f_T(t)$, ma divisa per $1 - F_T(t_o)$, al fine di garantire probabilità totale uguale ad 1.

4.8 Esempi di variabili aleatorie congiunte

Esercizio 4.a *Da un'urna contenente due palline bianche, una nera e due rosse, si estraggono una dopo l'altra con reimmissione due palline. Sia X_1 la variabile aleatoria che descrive l'esito della prima estrazione e X_2 quella che descrive l'esito della seconda estrazione. Ciascuna delle due variabili assume valore 1 se la pallina estratta è bianca, valore 0 se è nera o rossa. Si chiede di descrivere le leggi di probabilità congiunta e marginali, calcolare la covarianza e il coefficiente di correlazione.*

Essendo l'estrazione con reimmissione, gli esiti delle due estrazioni sono eventi indipendenti e quindi si ha

$$P(X_1 = x_{1r}, X_2 = x_{2s}) = P(X_1 = x_{1r}) \cdot P(X_2 = x_{2s}),$$

per $r, s = 1, 2$ e $x_{11} = x_{21} = 0$, $x_{12} = x_{21} = 1$. Questa relazione tra le probabilità assicura l'indipendenza delle due variabili casuali X e Y e permette immediatamente di rappresentarne la funzione di probabilità congiunta mediante la seguente tabella:

| | | X_2 | | |
|-------|---|---------------------------------|---------------------------------|---------------|
| | | 0 | 1 | |
| X_1 | 0 | $\frac{3}{5} \cdot \frac{3}{5}$ | $\frac{3}{5} \cdot \frac{2}{5}$ | $\frac{3}{5}$ |
| | 1 | $\frac{2}{5} \cdot \frac{3}{5}$ | $\frac{2}{5} \cdot \frac{2}{5}$ | $\frac{2}{5}$ |
| | | $\frac{3}{5}$ | $\frac{2}{5}$ | |

Calcoliamo le medie μ_{X_1} , μ_{X_2} e $\mu_{X_1 X_2}$ e le varianze $\sigma_{X_1}^2$ e $\sigma_{X_2}^2$. Una volta ottenuti questi valori, attraverso i teoremi visti calcoleremo immediatamente la covarianza σ_{X_1, X_2} e il coefficiente di correlazione ρ_{X_1, X_2} .

$$\begin{aligned} \mu_{X_1} &= \mu_{X_2} = 0 \cdot \frac{3}{5} + 1 \cdot \frac{2}{5} = \frac{2}{5}; \\ \sigma_{X_1}^2 &= \sigma_{X_2}^2 = \left(0 - \frac{2}{5}\right)^2 \cdot \frac{3}{5} + \left(1 - \frac{2}{5}\right)^2 \cdot \frac{2}{5} = \frac{6}{25}; \\ \mu_{X_1 X_2} &= \sum_{r,s} x_{1r} x_{2s} p_{rs} = 0 \cdot 0 \cdot \frac{9}{25} + 0 \cdot 1 \cdot \frac{6}{25} + 1 \cdot 0 \cdot \frac{6}{25} + 1 \cdot 1 \cdot \frac{4}{25} = \frac{4}{25}; \\ \text{Cov}(X_1, X_2) &= \mu_{X_1 X_2} - \mu_{X_1} \cdot \mu_{X_2} = \frac{4}{25} - \frac{2}{5} \cdot \frac{2}{5} = 0; \\ \rho_{X_1, X_2} &= \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}} = 0. \end{aligned}$$

Osserviamo che in virtù dell'indipendenza di X e Y sapevamo già, grazie ad un teorema che abbiamo visto, che $\text{Cov}(X_1, X_2)$ era nulla. Il calcolo è stato fatto ugualmente per fare pratica.

Esercizio 4.b *Si risolva il problema dell'esercizio precedente senza reimmissione.*

Non reimmettendo la pallina estratta per prima, l'esito della seconda estrazione dipenderà dall'esito della prima. Di conseguenza le probabilità congiunte cambiano e si ha

$$P(X_1 = x_{1r}, X_2 = x_{2s}) = P(X_1 = x_{1r}) \cdot P(X_2 = x_{2s} | X_1 = x_{1r}).$$

Le variabili casuali X e Y non sono indipendenti e la tabella che rappresenta la funzione di probabilità congiunta risulta così modificata:

| | | X_2 | | |
|-------|---|---------------------------------|---------------------------------|---------------|
| | | 0 | 1 | |
| X_1 | 0 | $\frac{3}{5} \cdot \frac{2}{4}$ | $\frac{3}{5} \cdot \frac{2}{4}$ | $\frac{3}{5}$ |
| | 1 | $\frac{2}{5} \cdot \frac{3}{4}$ | $\frac{2}{5} \cdot \frac{1}{4}$ | $\frac{2}{5}$ |
| | | $\frac{3}{5}$ | $\frac{2}{5}$ | |

Ripetendo i calcoli fatti per l'esercizio precedente otteniamo

$$\begin{aligned} \mu_{X_1} &= \mu_{X_2} = 0 \cdot \frac{3}{5} + 1 \cdot \frac{2}{5} = \frac{2}{5}; \\ \sigma_{X_1}^2 &= \sigma_{X_2}^2 = \left(0 - \frac{2}{5}\right)^2 \cdot \frac{3}{5} + \left(1 - \frac{2}{5}\right)^2 \cdot \frac{2}{5} = \frac{6}{25}; \\ \mu_{X_1 X_2} &= \sum_{r,s} x_{1r} x_{2s} p_{rs} = 0 \cdot 0 \cdot \frac{6}{20} + 0 \cdot 1 \cdot \frac{6}{20} + 1 \cdot 0 \cdot \frac{6}{20} + 1 \cdot 1 \cdot \frac{2}{20} = \frac{1}{10}; \\ \text{Cov}(X_1, X_2) &= \mu_{X_1 X_2} - \mu_{X_1} \cdot \mu_{X_2} = \frac{1}{10} - \frac{2}{5} \cdot \frac{2}{5} = -\frac{3}{50}; \\ \rho_{X_1, X_2} &= \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}} = -\frac{1}{4}. \end{aligned}$$

Il fatto che la media e la varianza delle variabili singole siano le stesse dell'esercizio precedente non deve sorprendere: le distribuzioni marginali non sono cambiate. Giova piuttosto osservare che ora le variabili X e Y non sono indipendenti e, essendo $\rho_{X_1, X_2} \neq 0$, neppure incorrelate.

Esercizio 4.c Un'urna contiene 112 dadi di cui 56 (cioè la metà) sono equi, mentre gli altri sono stati manipolati in modo che, per ciascuno di essi, la probabilità di ottenere 1 sia $\frac{1}{2}$, mentre ogni altro risultato si verifica con probabilità $\frac{1}{10}$. Si chiede:

- Un dado viene estratto a caso e lanciato; indichiamo con X la variabile aleatoria che rappresenta il risultato del lancio. Qual è la probabilità di ottenere 3? Quanto vale $E(X)$?
- Un dado viene estratto a caso e lanciato due volte. Indicato con X il risultato del primo lancio e con Y quello del secondo, qual è la probabilità di ottenere $X = 2$ e $Y = 3$?
- Sapendo che i due lanci hanno dato come risultato $X = 2$ e $Y = 3$, qual è la probabilità che si tratti di uno dei dadi truccati?

d) Le variabili casuali X e Y sono indipendenti?

a) Le probabilità con cui la variabile aleatoria X assume i valori $1, 2, \dots, 6$ dipendono dal fatto che il dado estratto sia oppure no equo. Indicato con A l'evento "il dado estratto è equo" e quindi con A^C l'evento "il dado estratto è alterato", si ha

$$X|A : \left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array} \right), \quad X|A^C : \left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \end{array} \right).$$

Applicando la formula della probabilità totale, per cui

$$P(X = k) = P(X = k|A) \cdot P(A) + P(X = k|A^C) \cdot P(A^C),$$

la variabile non condizionata X risulta così definita:

$$X : \left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{3} & \frac{2}{15} & \frac{2}{15} & \frac{2}{15} & \frac{2}{15} & \frac{2}{15} \end{array} \right).$$

Si ha dunque $P(X = 3) = \frac{2}{15}$ e, facendo i calcoli, $E(X) = \sum_{k=1}^6 k \cdot P(X = k) = 3$.

b) Consideriamo la variabile bidimensionale (X, Y) , con Y variabile identica alla X . La sua funzione di probabilità congiunta sarà definita dalle relazioni

$$\begin{aligned} P[(X=j, Y=k)] &= P[(X=j, Y=k)|A]P(A) + \\ &\quad + P[(X=j, Y=k)|A^C]P(A^C) = \\ &= P(X=j|A)P(Y=k|A)\frac{1}{2} + P(X=j|A^C)P(Y=k|A^C)\frac{1}{2} \end{aligned}$$

In particolare

$$\begin{aligned} P[(X=2, Y=3)] &= P(X=2|A)P(Y=3|A)\frac{1}{2} + P(X=2|A^C)P(Y=3|A^C)\frac{1}{2} = \\ &= \frac{1}{6} \frac{1}{6} \frac{1}{2} + \frac{1}{10} \frac{1}{10} \frac{1}{2} = \frac{17}{900}. \end{aligned}$$

c) Indicato con B l'evento $\{X = 2, Y = 3\}$, ci si chiede ora $P(A^C|B)$. Utilizzando la formula di Bayes, otteniamo

$$P(A^C|B) = \frac{P(B|A^C) \cdot P(A^C)}{P(B)} = \frac{\frac{1}{10} \cdot \frac{1}{10} \cdot \frac{1}{2}}{\frac{17}{900}} = \frac{9}{34}.$$

d) Perché due variabili aleatorie X e Y siano indipendenti (vedi (8)) deve essere

$$P(X = x_r, Y = y_s) = P(X = x_r) \cdot P(Y = y_s),$$

per ogni coppia (x_r, y_s) . Nel nostro caso, con $x_r = 2$ e $y_s = 3$, si ha

$$\frac{17}{900} = P(X = 2, Y = 3) \neq P(X = 2) \cdot P(Y = 3) = \frac{2}{15} \cdot \frac{2}{15} = \frac{4}{225}.$$

Di conseguenza X e Y non sono indipendenti.

Volendo, anche se non richiesto dall'esercizio, calcolare e mostrare la tabella completa che rappresenta la funzione di probabilità congiunta della variabile aleatoria (X, Y) , abbiamo:

| | | Y | | | | | | |
|---|---|-----------------|------------------|------------------|------------------|------------------|------------------|----------------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | |
| X | 1 | $\frac{5}{36}$ | $\frac{7}{180}$ | $\frac{7}{180}$ | $\frac{7}{180}$ | $\frac{7}{180}$ | $\frac{7}{180}$ | $\frac{1}{3}$ |
| | 2 | $\frac{7}{180}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{2}{15}$ |
| | 3 | $\frac{7}{180}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{2}{15}$ |
| | 4 | $\frac{7}{180}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{2}{15}$ |
| | 5 | $\frac{7}{180}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{2}{15}$ |
| | 6 | $\frac{7}{180}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{17}{900}$ | $\frac{2}{15}$ |
| | | $\frac{1}{3}$ | $\frac{2}{15}$ | $\frac{2}{15}$ | $\frac{2}{15}$ | $\frac{2}{15}$ | $\frac{2}{15}$ | |

Esercizio 4.d Due variabili aleatorie X e Y sono indipendenti ed uniformi su $[0, 1]$. Calcolare:

$$a) P\left(XY > \frac{1}{2}\right); \quad b) P\left(XY < \frac{1}{4} \mid X > \frac{1}{2}\right).$$

Essendo X e Y uniformi ed indipendenti sull'intervallo $[0, 1]$, la variabile congiunta (X, Y) ha una funzione densità $f(x, y)$ che è il prodotto delle funzioni densità di X e Y . Pertanto, detto \mathcal{Q} il "quadrato" $\equiv (0, 1) \times (0, 1)$, si ha:

$$f(x, y) = \begin{cases} 0 & \text{se } (x, y) \notin \mathcal{Q} \\ 1 & \text{se } (x, y) \in \mathcal{Q} \end{cases}$$

Inoltre ricordando che

$$P(\Phi(X, Y) \in B) = \iint_{\mathcal{A}} f(x, y) dx dy \quad \text{con } \mathcal{A} = \{(x, y) : \Phi(x, y) \in B\},$$

per $\Phi(X, Y) = XY$, si ha

$$P(XY \in B) = \iint_{\mathcal{A}} f(x, y) dx dy = \iint_{\mathcal{A} \cap \mathcal{Q}} dx dy, \quad \text{con } \mathcal{A} = \{(x, y) : xy \in B\}.$$

a) Dovendo essere $XY > \frac{1}{2}$, si ha $B = \{z > \frac{1}{2}\}$ e quindi \mathcal{A} è la regione, tutta contenuta in \mathcal{Q} costituita dai punti (x, y) per i quali $xy > \frac{1}{2}$ (vedi figura 10). Pertanto

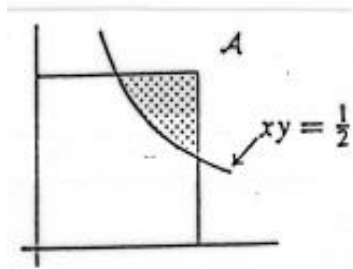


Figura 10: Caso a).

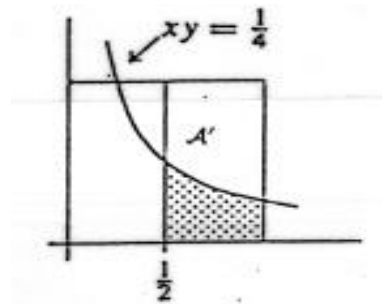


Figura 11: Caso b).

la probabilità richiesta è:

$$\begin{aligned}
 P\left(XY > \frac{1}{2}\right) &= \iint_{\mathcal{A} \cap \mathcal{Q}} dx dy = \int_{\frac{1}{2}}^1 dx \int_{\frac{1}{2x}}^1 dy = \\
 &= \int_{\frac{1}{2}}^1 \left(1 - \frac{1}{2x}\right) dx = \frac{1}{2}(1 - \log_e 2).
 \end{aligned}$$

b) Per calcolare la probabilità richiesta si procede analogamente a quanto fatto per il punto a) tenendo però conto che si tratta di una probabilità condizionata:

$$P\left(XY < \frac{1}{4} \mid X > \frac{1}{2}\right) = \frac{P[XY < \frac{1}{4}, X > \frac{1}{2}]}{P[X > \frac{1}{2}]}.$$

Posto $\mathcal{A}' = \{(x, y) : xy < \frac{1}{4}, x > \frac{1}{2}\}$, si ha

$$\begin{aligned}
 P\left(XY < \frac{1}{4} \mid X > \frac{1}{2}\right) &= \frac{\iint_{\mathcal{A}' \cap \mathcal{Q}} dx dy}{\frac{1}{2}} = 2 \int_{\frac{1}{2}}^1 dx \int_0^{\frac{1}{4x}} dy = 2 \int_{\frac{1}{2}}^1 \frac{1}{4x} dx = \\
 &= \frac{1}{2} \log_e 2.
 \end{aligned}$$

5 STATISTICA DESCRITTIVA

Per **statistica descrittiva** o **metodologica** si intende il complesso di quelle norme utilizzate dallo sperimentatore per raccogliere, rappresentare ed elaborare insiemi di dati osservati.

I dati raccolti riguardano solo un **campione** e non l'intera **popolazione**. L'elaborazione statistica ha l'obiettivo di ricavare informazioni sulla popolazione estraendole dai (pochi) dati che sono stati osservati sul campione. Naturalmente le informazioni a cui siamo interessati riguardano una o più **caratteristiche** della popolazione in esame.

Volendo dare una veste matematica a quanto appena detto, sia X una variabile aleatoria, di tipo discreto o continuo, definita su un insieme S (la popolazione). Sono noti i valori che X assume in corrispondenza degli elementi di un sottoinsieme C di S (il campione). Sia $N = |S|$ e $n = |C|$. Il campione è dunque una n -pla (x_1, x_2, \dots, x_n) , dove ciascun x_i rappresenta il valore noto che $X(s)$ assume per $s = s_i \in C$. Essendo, in generale, $n \ll N$, la variabile aleatoria X è incognita in molti (moltissimi) elementi su cui è definita. Il compito della statistica è quello di desumere dai dati del campione il maggior numero di informazioni circa la distribuzione di X , avendo anche un'idea, il più possibile precisa, del grado di affidabilità di queste informazioni. A questa variabile aleatoria ci riferiremo d'ora in poi come alla **variabile aleatoria sottostante** al nostro esperimento.

Un'indagine statistica di tipo descrittivo può essere articolata nei seguenti quattro passi:

- 1) **rilevazione** dei dati;
- 2) **organizzazione** dei dati;
- 3) **presentazione** dei dati organizzati;
- 4) **interpretazione** e **conclusioni**.

5.1 Organizzazione e Descrizione dei Dati

Rilevazione dei dati

La rilevazione, che è l'inizio del procedimento statistico, è l'insieme dei meccanismi che permettono di ottenere le informazioni necessarie da elaborare. Strumenti basilari di questo momento sono i questionari, i modelli di rilevazione, le inchieste telefoniche, l'accesso e la consultazione di banche dati, ecc.

Le modalità di rilevazione dei dati x_i sono particolarmente importanti. Occorre infatti aver chiaramente fissati gli obiettivi, valutata la fattibilità, definita l'estensione in termini geografici, temporali, economici. Infine, è fondamentale aver scelto in modo appropriato la tecnica di campionamento (di cui però non ci occuperemo).

Organizzazione dei dati

In genere i dati grezzi ottenuti dalla rilevazione sono difficilmente interpretabili: occorre organizzarli opportunamente. Quando i dati sono di tipo numerico, e lo sono nella grande maggioranza dei casi, il modo più semplice di farlo consiste nell'ordinarli in modo crescente o decrescente. Ciò permette immediatamente di stabilire il **campo di variazione** degli x_i (o **rango**), cioè il minimo intervallo che li contiene tutti. Questo **indice** ci dice già qualcosa (ad esempio i valori minimo e massimo della variabile campionata); tuttavia esso può essere poco indicativo, soprattutto se n è grande. Può dunque essere conveniente organizzare i dati in **classi**.

Come si formano le classi? Si tratta di un punto importante in quanto una cattiva scelta delle classi può portare ad una cattiva interpretazione della distribuzione dei dati (cioè della variabile casuale X ad essi associata). Proponiamo dunque alcuni criteri di formazione delle classi ritenuti ottimali.

Il **numero delle classi** è importante. Se le classi sono troppe, in ogni classe ci sarebbero pochissimi elementi (o addirittura nessuno); se sono poche, essendovi concentrati molti elementi, potrebbe sfuggirci la globalità della distribuzione. In genere il numero delle classi è compreso fra 6 e 20. Secondo Sturges il numero ottimale di classi è

$$n_c = [1 + 1.443 \lg n],$$

con $\lg n$ che indica il logaritmo naturale di n e $[a]$ che rappresenta l'intero più vicino ad a .

È conveniente che le classi abbiano la stessa ampiezza. In questo caso, se r è il campo di variazione dei dati ed n_c il numero delle classi in cui abbiamo deciso di organizzare i dati, se ne deduce per ciascuna classe un'ampiezza ℓ data da

$$\ell = \frac{r}{n_c}.$$

Tale ampiezza, tuttavia, in genere non è quella più conveniente; torna utile "aggiustarla" in modo tale che i punti di mezzo di ciascun intervallo siano della stessa grandezza, come ordine di approssimazione, dei dati x_i e che nessun x_i cada su un estremo dell'intervallo. Ad esempio, se gli x_i sono interi qualunque (nel senso che non sono dei multipli di un intero k), allora conviene prendere ℓ intero e dispari, e ciascun intervallo del tipo $(h - \frac{1}{2}, h + \ell - \frac{1}{2})$, dove h è un intero. Scelte analoghe possono essere fatte se gli x_i sono numeri decimali (tutti con lo stesso numero di

decimali). I due esempi che verranno proposti fra poco saranno utili a chiarire il senso di quanto appena detto.

Funzioni di frequenza

Per avere altri tipi di informazione sempre più precisi ed esaurienti, si possono definire altri indici statistici, quali quelli che seguono:

- la **funzione di frequenza**, che associa ad ogni classe il numero degli elementi che la compongono; la indicheremo con $\varphi(x)$;
- la **funzione di frequenza relativa**, che esprime il rapporto fra il numero degli elementi della classe ed il numero totale n di elementi del campione; indicata con $\varphi_r(x)$, si ha dunque $\varphi_r(x) \equiv \frac{\varphi(x)}{n}$;
- la **funzione di frequenza cumulativa**, cioè il numero degli elementi della classe e delle classi precedenti; sarà rappresentata da $\varphi_c(x)$;
- la **funzione di frequenza cumulativa relativa**, ovvero il rapporto tra il numero degli elementi dato dalla frequenza cumulativa e il numero totale n di elementi del campione; denotata con $\varphi_{cr}(x)$, si ha perciò $\varphi_{cr}(x) \equiv \frac{\varphi_c(x)}{n}$.

Rappresentazioni grafiche

Nella statistica descrittiva la rappresentazione grafica dei dati riveste un ruolo molto importante, in quanto serve a fornire in modo immediato una descrizione del fenomeno oggetto di studio. Gli strumenti disponibili sono diversi, a seconda degli obiettivi che si intendono conseguire mostrando i dati. Tra le rappresentazioni più utili a visualizzare una serie di numeri troviamo gli **istogrammi**, i **grafici a bastoni** e i **poligoni di frequenza**, i **boxplot** e gli **scatter plot** o grafici a dispersione.

Vediamo cos'è un **istogramma**, che probabilmente costituisce lo strumento più comune di rappresentazione di dati statistici. È un grafico che dà un'idea della "forma" di un campione, indica gli intervalli in cui sono molto concentrate le osservazioni campionarie e gli intervalli in cui lo sono poco. Si ottiene nel modo seguente. Sull'asse delle ascisse si riportano le classi indicandone i loro punti di mezzo; sull'asse delle ordinate si possono riportare i valori della funzione $\varphi(x)$ oppure quelli di $\varphi_r(x)$. Il grafico si ottiene disegnando poi per ciascuna classe un rettangolo avente come base l'ampiezza ℓ e come altezza $\varphi(x)$ o, equivalentemente, $\varphi_r(x)$. Se poi si riportano nel grafico sia $\varphi(x)$ che $\varphi_r(x)$ (in opportuna scala), rispettivamente a sinistra e a destra del grafico, si ottiene il duplice obiettivo di poter **leggere** entrambi i valori.

Un **grafico a bastoni** è del tutto equivalente ad un istogramma, e si costruisce in maniera del tutto analoga. Per quanto riguarda poi i **poligoni di frequenza**, l'esempio che segue permetterà facilmente di capire come si costruiscono e qual è il loro significato.

Esempio 5.1.1 La tabella che segue riporta i pesi (in chilogrammi) di 50 studentesse, che per brevità sono già stati ordinati (in ordine crescente). Naturalmente, ogni numero è ripetuto tante volte quante sono le studentesse aventi quel peso.

53 55 56 57 57 58 58 59 59 60
 60 60 61 61 61 61 62 62 62 62
 63 63 63 63 63 64 64 64 64 64
 64 65 65 65 65 65 66 66 66 66
 67 67 67 68 68 69 70 71 71 73

Dalla tabella si deduce immediatamente che il campo di variazione è [53,73]. Applicando poi la formula di Sturges per determinare il numero ottimale di classi, si ha

$$n_c = [1 + 1.443 \lg 50] = [1 + 5.64] = 7, \quad \text{equindi} \quad \ell = \frac{20}{7} \approx 2.86$$

In base a quanto detto in precedenza, essendo l'unità di misura adottata un numero intero (i chili), è conveniente che ℓ sia un intero dispari e che gli intervalli abbiano come punto medio un intero. Scegliamo dunque $\ell = 3$ e prendiamo gli intervalli di ampiezza 3 a partire da 52.5. La tabella che segue riporta gli intervalli relativi a ciascuna classe, il loro punto di mezzo x , il numero di elementi di ogni classe e le quattro funzioni di frequenza precedentemente definite $\varphi(x)$, $\varphi_r(x)$, $\varphi_c(x)$ e $\varphi_{cr}(x)$.

| Classi di pesi | Punto x di mezzo | $\varphi(x)$ | $\varphi_r(x)$ | $\varphi_c(x)$ | $\varphi_{cr}(x)$ |
|----------------|--------------------|--------------|----------------|----------------|-------------------|
| 52.5 – 55.5 | 54 | 2 | 0.04 | 2 | 0.04 |
| 55.5 – 58.5 | 57 | 5 | 0.10 | 7 | 0.14 |
| 58.5 – 61.5 | 60 | 9 | 0.18 | 16 | 0.32 |
| 61.5 – 64.5 | 63 | 15 | 0.30 | 31 | 0.62 |
| 64.5 – 67.5 | 66 | 12 | 0.24 | 43 | 0.86 |
| 67.5 – 70.5 | 69 | 4 | 0.08 | 47 | 0.94 |
| 70.5 – 73.5 | 72 | 3 | 0.06 | 50 | 1.00 |

In figura 12 è rappresentato l'istogramma e il grafico a bastoni, in figura 13 il poligono di frequenza ed il poligono di frequenza relativa cumulativa.

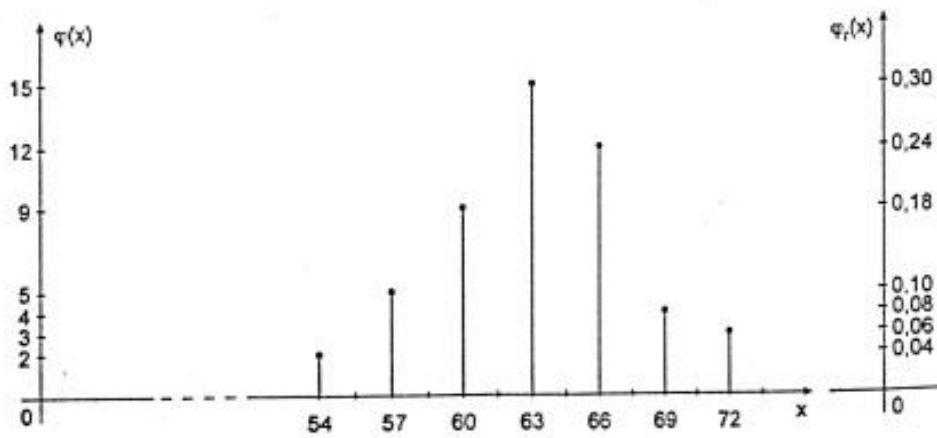
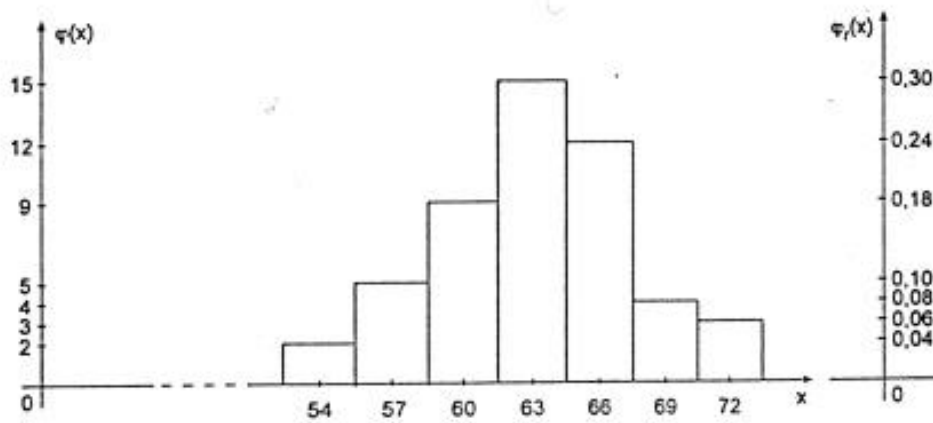


Figura 12: Istogramma e grafico a bastoni dell'esempio 5.1.1

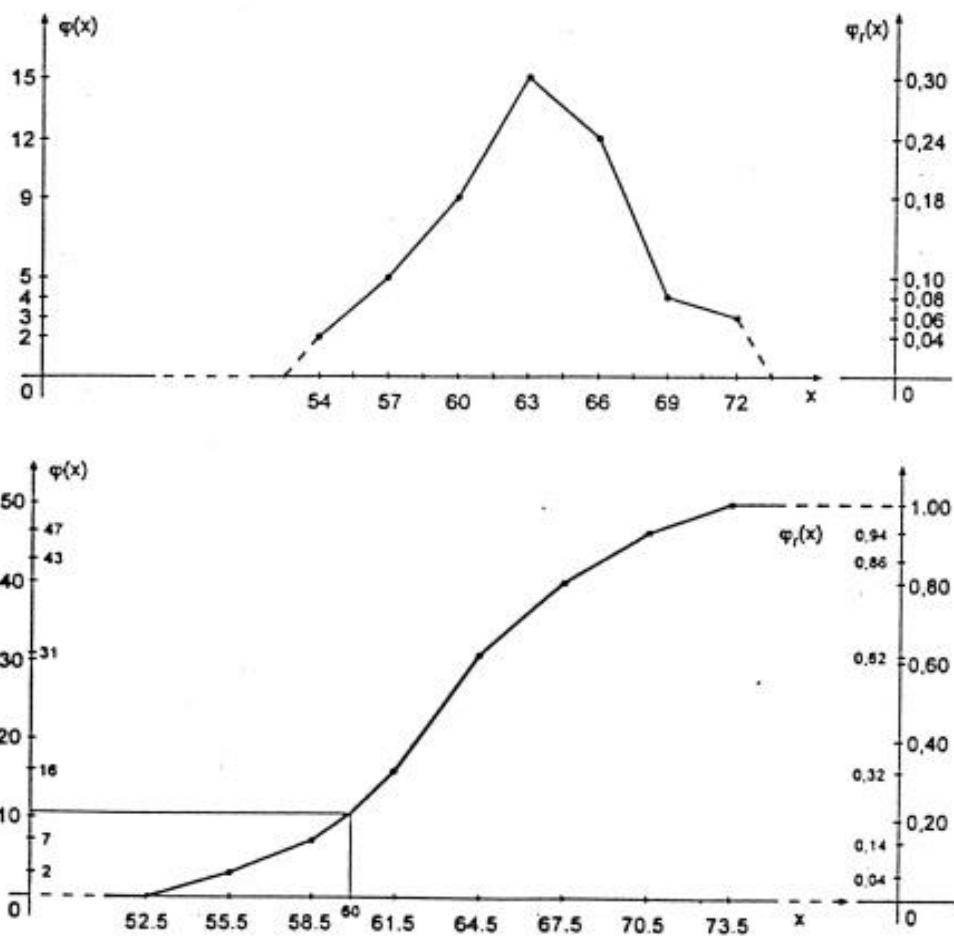


Figura 13: Poligono di frequenza e poligono di frequenza relativa cumulativa dell'esempio 5.1.1

Osserviamo che, in un certo senso, il poligono di frequenza (primo grafico di fig. 13) “rappresenta” la funzione densità della variabile aleatoria X sottostante al fenomeno studiato; analogamente il poligono di frequenza cumulativa (secondo grafico di fig.13) “rappresenta” la funzione di distribuzione di X .

Anche i **boxplot** sono rappresentazioni grafiche molto usate in statistica. Poiché per la costruzione di un boxplot occorrono le definizioni di quartili, rimandiamo la descrizione di questo oggetto alla fine di questo capitolo.

I dati le cui unità possiedono più di un valore sono chiamati **dati multivariati**. Quando ogni unità ha una coppia di valori, allora i dati sono chiamati **bivariati**. Mediante lo **scatter plot** o *grafico a dispersione* (scatter) possiamo visualizzare dati *bivariati*, rappresentati, ad esempio, come coordinate x e y di un piano. Questi diagrammi sono utili per studiare relazioni tra due variabili.

Consideriamo il seguente esempio: abbiamo seguito la carriera scolastica di 15 studenti diplomati lo stesso anno nello stesso liceo e iscritti successivamente ad Ingegneria, avendo rilevato sia il voto di maturità (espresso in 60-esimi) che il voto di Analisi I (espresso in 30-esimi) rappresentiamo graficamente la relazione tra questi dati. I voti sono raccolti in una tabella come in figura (dove Mat. rappresenta il voto di maturità e An.I il voto dell’esame di analisi I):

| Mat. | An.I | Mat. | An.I | Mat. | An.I |
|------|------|------|------|------|------|
| 47 | 27 | 48 | 24 | 55 | 25 |
| 53 | 26 | 45 | 24 | 47 | 21 |
| 45 | 22 | 56 | 28 | 45 | 21 |
| 48 | 24 | 55 | 27 | 47 | 23 |
| 57 | 28 | 50 | 25 | 49 | 23 |

Se le due variabili hanno una qualche relazione di tipo lineare (ad esempio ad un alto voto nell’esame di maturità corrisponde un alto voto nell’esame di analisi I) allora i punti nel grafico sparso sono distribuiti lungo una retta.

Esempio 5.1.2 La tabella che segue riporta le altezze (in centimetri) di 80 atleti, anche in questo caso già ordinati (in modo crescente). Si chiede di organizzare questi dati in classi e di calcolarne le quattro funzioni di frequenza, rappresentando poi il tutto con una tabella del tipo di quella dell’esercizio 5.1.1.

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 160 | 162 | 164 | 165 | 167 | 168 | 168 | 169 | 169 | 170 |
| 170 | 171 | 171 | 172 | 172 | 172 | 172 | 173 | 173 | 174 |
| 174 | 174 | 175 | 175 | 175 | 176 | 176 | 176 | 177 | 177 |
| 177 | 177 | 178 | 178 | 178 | 178 | 178 | 178 | 179 | 179 |
| 179 | 179 | 179 | 179 | 179 | 180 | 180 | 180 | 180 | 181 |
| 181 | 181 | 181 | 182 | 182 | 182 | 182 | 182 | 183 | 183 |
| 184 | 184 | 185 | 185 | 186 | 186 | 187 | 187 | 188 | 189 |
| 190 | 190 | 191 | 192 | 192 | 193 | 194 | 197 | 199 | 201 |

Dalla tabella si legge subito che il campo di variazione è [160,201]. Applicando poi la formula di Sturges per determinare il numero ottimale di classi, si ha

$$n_c = [1 + 1.443 \lg 80] = [1 + 6.32] = 7, \quad \text{e quindi} \quad \ell = \frac{41}{7} \approx 5.86$$

Volendo scegliere come ℓ un intero dispari, o si sceglie 5, che comporta poi di prendere $n_c=9$, oppure si sceglie 7, che comporta $n_c=6$. Per non avere un numero di classi troppo piccolo, scegliamo $\ell = 5$ e quindi $n_c = 9$. I dati organizzati in classi portano dunque a questa tabella:

| Classi di pesi | Punto x di mezzo | $\varphi(x)$ | $\varphi_r(x)$ | $\varphi_c(x)$ | $\varphi_{cr}(x)$ |
|----------------|--------------------|--------------|----------------|----------------|-------------------|
| 158.5 – 163.5 | 161 | 2 | 0.025 | 2 | 0.025 |
| 163.5 – 168.5 | 166 | 5 | 0.063 | 7 | 0.088 |
| 168.5 – 173.5 | 171 | 12 | 0.150 | 21 | 0.263 |
| 173.5 – 178.5 | 176 | 19 | 0.237 | 38 | 0.475 |
| 178.5 – 183.5 | 181 | 22 | 0.275 | 60 | 0.750 |
| 183.5 – 188.5 | 186 | 9 | 0.113 | 69 | 0.863 |
| 188.5 – 193.5 | 191 | 7 | 0.087 | 76 | 0.950 |
| 193.5 – 198.5 | 196 | 2 | 0.025 | 78 | 0.975 |
| 198.5 – 205.5 | 201 | 2 | 0.025 | 80 | 1.000 |

5.2 Grandezze che sintetizzano i dati

Ci proponiamo ora di caratterizzare una distribuzione statistica, cioè un insieme di dati x_i , $i = 1, 2, \dots, n$, del tipo di quelli visti finora, attraverso misure che ne riassumano le principali proprietà. In tal modo si parla anche di **misure di tendenza centrale**: si chiamano così alcune caratterizzazioni sintetiche della distribuzione che servono a dare un'idea di **dove la distribuzione sia collocata** e **quanto sia concentrata**.

Media

Definizione: Date n osservazioni numeriche x_i , $i = 1, 2, \dots, n$, si chiama **media aritmetica**, o più semplicemente **media**, delle osservazioni il numero

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Ai fini di collegare questa definizione a quella di media di una variabile casuale, osserviamo che in generale tra i dati x_i ce ne sono di quelli che sono ripetuti più volte. Ebbene, supposto che gli x_i distinti siano m (ovviamente $m \leq n$), indichiamo questi numeri con z_1, z_2, \dots, z_m . Denotata poi con α_k la molteplicità (cioè il numero di presenze) di z_k , ovviamente con $\alpha_1 + \alpha_2 + \dots + \alpha_m = n$, potremo scrivere

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{k=1}^m \alpha_k z_k = \sum_{k=1}^m \frac{\alpha_k}{n} z_k = \sum_{k=1}^m p_k z_k.$$

Il numero $p_k = \frac{\alpha_k}{n}$ rappresenta la frequenza relativa del dato z_k . Confrontando quest'ultima espressione di \bar{x} con la definizione di media di una variabile aleatoria finita, ne deduciamo che la media aritmetica appena definita altro non è che la media di una variabile aleatoria che assume gli m valori z_k con probabilità p_k . La media \bar{x} dei dati x_i può dunque essere vista come la media di una variabile aleatoria \tilde{X} finita, che assume i valori x_i con probabilità uguali alla loro frequenza relativa nel campione, ossia

$$P(\tilde{X} = x_i) = p_i, \quad p_i = \frac{\alpha_i}{n},$$

essendo α_i il numero di volte in cui ciascun x_i è presente nel campione.

La variabile aleatoria \tilde{X} costituisce una rozza approssimazione della vera variabile aleatoria sottostante al problema. La media, che abbiamo appena definito, così come la mediana e la varianza che definiremo in seguito, sono indici coerenti con questa approssimazione.

Ricordando le proprietà della media di una variabile aleatoria, si può affermare che

- se ogni osservazione di un campione è letta in una scala diversa, ovvero se ogni dato è moltiplicato per una costante a , allora

$$\overline{ax} = a\bar{x};$$

- se (x_1, x_2, \dots, x_n) e (y_1, y_2, \dots, y_n) sono due serie di osservazioni di uno stesso fenomeno, allora la **media della somma è uguale alla somma delle medie**, cioè

$$\overline{x + y} = \bar{x} + \bar{y};$$

– se due osservazioni sono legate da una relazione funzionale del tipo $y = a + bx$, con a e b costanti, allora

$$\bar{y} = a + b\bar{x}.$$

Quando i dati sono forniti già organizzati in classi, la media può essere ugualmente calcolata con la formula seguente:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n_c} \bar{x}_k \varphi(\bar{x}_k),$$

dove \bar{x}_k è il punto medio dell'intervallo k -esimo e $\varphi(\bar{x}_k)$ fornisce, come abbiamo già visto, il numero degli x_i appartenenti alla classe k -esima.

Osserviamo che questa formula può essere utilizzata anche quando ci sono assegnate tutte le n osservazioni x_i e la loro organizzazione in classi viene fatta da noi solo successivamente al fine di una rappresentazione più sintetica dei dati. In tal caso la media così calcolata è una approssimazione, in generale molto buona, di quella vera (cioè di quella che si ottiene dalla definizione). Il vantaggio di quest'ultima formula sta nel fatto che, utilizzando le classi, è richiesto un numero molto minore di calcoli.

Mediana

Definizione: Date n osservazioni numeriche x_i , $i = 1, 2, \dots, n$, si chiama **mediana** delle osservazioni il valore “centrale” dell'insieme ordinato.

Quindi, a seconda che n sia pari o dispari, si ha

$$x_{med} = \begin{cases} x_{\frac{n+1}{2}} & \text{se } n \text{ dispari} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{se } n \text{ pari} \end{cases}$$

Anche per la mediana, così come abbiamo fatto per la media, ci si può porre il problema di come determinarla quando i dati x_i non sono noti individualmente in quanto forniti già organizzati in classi. In questo caso, per poter definire operativamente la mediana, occorre introdurre alcune ulteriori notazioni.

Supposto che le classi si susseguano in ordine crescente, indichiamo con $(\lambda_{i-1}, \lambda_i)$ l'intervallo associato alla classe i -esima e con \bar{x}_i il suo punto medio. Allora $\varphi_c(\bar{x}_i)$ denota il valore della funzione di frequenza cumulativa della classe i -esima, cioè il numero complessivo di elementi contenuti nelle prime i classi. Chiamiamo **classe mediana**, indicando con m il suo numero d'ordine, quella classe per cui

$$\varphi_c(\bar{x}_m) \geq \frac{n}{2}, \quad \text{con} \quad \varphi_c(\bar{x}_{m-1}) < \frac{n}{2}.$$

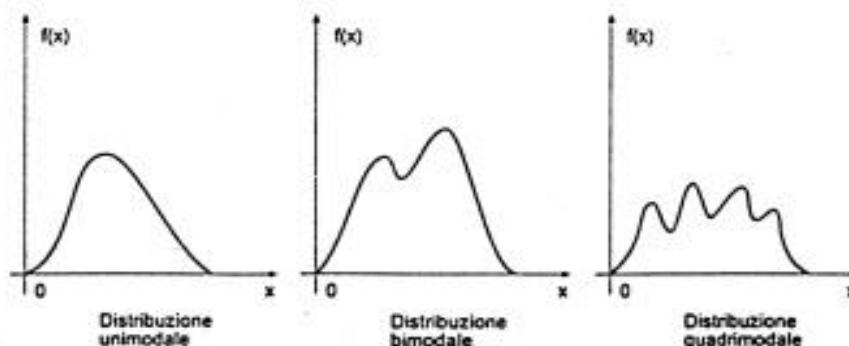


Figura 14:

Ciò posto, la mediana x_{med} può essere così definita:

$$x_{med} = \lambda_{m-1} + \frac{\frac{n}{2} - \varphi_c(\bar{x}_{m-1})}{\varphi_c(\bar{x}_m) - \varphi_c(\bar{x}_{m-1})} \ell = \lambda_{m-1} + \frac{\frac{n}{2} - \varphi_c(\bar{x}_{m-1})}{\varphi(\bar{x}_m)} \ell.$$

Osserviamo che x_{med} appartiene certamente alla classe mediana (cioè all'intervallo $(\lambda_{m-1}, \lambda_m)$) se $\varphi_c(\bar{x}_m) > \frac{n}{2}$, mentre si ha $x_{med} = \lambda_m$ se $\varphi_c(\bar{x}_m) = \frac{n}{2}$ (il che può accadere solo se n è pari).

Moda

Molto spesso i dati sono divisi in classi che non sono di tipo numerico (ad esempio sesso, gruppo sanguigno, professione, provincia di appartenenza, etc...). In questo caso non ha alcun senso parlare di media o mediana, per cui può tornare utile un'altra misura di tendenza centrale, valida per qualunque tipologia di dati. Questa misura, però, non esiste per tutte le distribuzioni, ma solo per quelle unimodali. La figura 14 mostra una distribuzione unimodale assieme a due multimodali.

Definizione: Si definisce **moda** di una distribuzione unimodale di dati il valore fra questi più ripetuto. La moda, che può anche non essere unica, sarà indicata con x_{mod} .

Per definire la moda quando i dati sono forniti già divisi in classi, occorre determinare innanzitutto la **classe modale**, cioè la classe nella quale si trova la moda. Di solito la classe modale è quella in cui $\varphi(x)$ è massima. Supposto che tale classe sia unica, che $(\lambda_{j-1}, \lambda_j)$ sia l'intervallo associato e \bar{x}_j il suo punto medio, la moda è così definita:

$$x_{mod} = \lambda_{j-1} + \frac{|\varphi(\bar{x}_j) - \varphi(\bar{x}_{j-1})|}{|\varphi(\bar{x}_j) - \varphi(\bar{x}_{j-1})| + |\varphi(\bar{x}_{j+1}) - \varphi(\bar{x}_j)|} \ell.$$

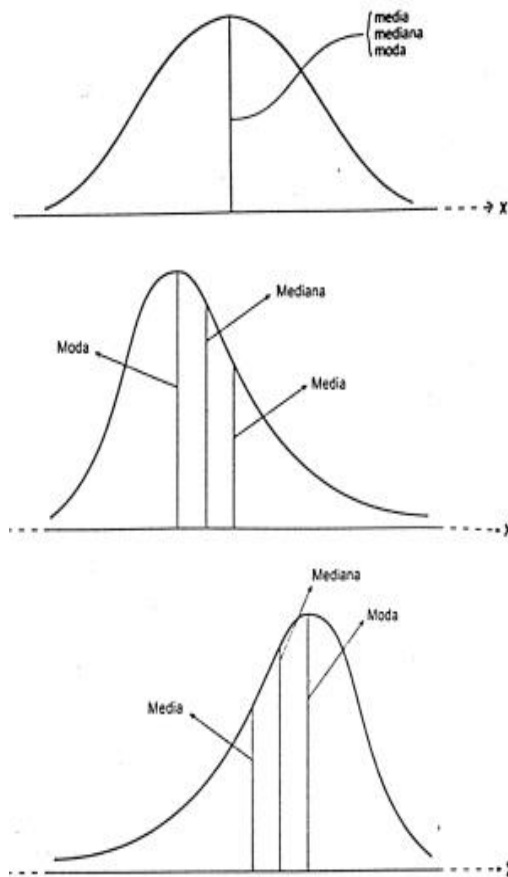


Figura 15:

Se la classe non è unica, si hanno più mode.

Ci si può chiedere come sono disposte l'una rispetto all'altra le tre misure di tendenza centrale che abbiamo definito (quando esistono tutte tre). Disegnata la distribuzione dei dati, la loro reciproca disposizione dipende dalla simmetria o asimmetria di questo grafico. Supposto che la distribuzione sia unimodale (vedi figura 14), se il grafico è perfettamente simmetrico, allora media, mediana e moda coincidono. Se invece il grafico è asimmetrico, allora la moda corrisponde ovviamente al massimo del grafico, mentre media e mediana sono sempre disposte con la mediana più vicina della media alla moda (vedi Fig.15).

Esempio 5.2.1: Calcoliamo media, mediana e moda dei dati dell'esempio 5.1.1.

Per quanto riguarda la media, facendo uso della definizione, si ottiene

$$\bar{x} = \frac{1}{50} \sum_{i=1}^{50} x_i = \frac{1}{50} (73 + 71 + \dots + 51) = \frac{3161}{50} = 63.22.$$

Calcolando invece la media utilizzando le classi, indicato con \bar{x}_k il punto medio dell'intervallo corrispondente alla k -classe, si ha

$$\begin{aligned} \bar{x} &\approx \frac{1}{50} \sum_{k=1}^{n_c} \bar{x}_k \varphi(\bar{x}_k) = \frac{1}{50} (54 \cdot 2 + 57 \cdot 5 + 60 \cdot 9 + 63 \cdot 15 + 66 \cdot 12 + 69 \cdot 4 + 72 \cdot 3) = \\ &= \frac{3162}{50} = 63.24. \end{aligned}$$

Come si vede, per quanto approssimato, il valore della media così ottenuto è molto prossimo a quello corretto ottenuto in precedenza.

Venendo alla mediana, il suo calcolo è immediato. Infatti, essendo $n=50$, cioè pari, dalla tabella contenente i dati ordinati si legge che $x_{25}=63$ e $x_{26}=64$. Si ha quindi

$$x_{med} = \frac{x_{25} + x_{26}}{2} = \frac{63 + 64}{2} = 63.5.$$

Anche x_{med} può essere calcolato utilizzando la formula per i dati organizzati in classi; in tal caso si ottiene

$$x_{med} = \lambda_{m-1} + \frac{\frac{n}{2} - \varphi_c(\bar{x}_{m-1})}{\varphi(\bar{x}_m)} \ell = 61.5 + \frac{25 - 16}{15} 3 = 63.3.$$

Per quanto riguarda la moda, si ha

$$x_{mod} = 64,$$

in quanto è il valore più ripetuto. D'altra parte, se x_{mod} viene calcolata in base all'organizzazione dei dati in classi, essendo la classe modale quella centrata in $x_j = 63$, si ha

$$x_{mod} = 61.5 + \frac{|15 - 9|}{|15 - 9| + |12 - 15|} 3 = 63.5.$$

Esempio 5.2.2: Calcoliamo media, mediana e moda dei dati dell'esempio 5.1.2.

Facendo uso della definizione per calcolare la media, si ottiene

$$\bar{x} = \frac{1}{80} \sum_{i=1}^{80} x_i = \frac{1}{80} (160 + 162 + \dots + 201) = \frac{14332}{80} = 179.15.$$

Se calcoliamo la media utilizzando la formula per i dati organizzati in classi, abbiamo

$$\begin{aligned} \bar{x} &\approx \frac{1}{80} \sum_{k=1}^{n_c} \bar{x}_k \varphi(\bar{x}_k) = \frac{1}{80} (161 \cdot 2 + 166 \cdot 5 + 171 \cdot 12 + 176 \cdot 19 + 181 \cdot 22 + \\ &+ 186 \cdot 9 + 191 \cdot 7 + 196 \cdot 2 + 201 \cdot 2) = \frac{14335}{80} \approx 179.19. \end{aligned}$$

Anche in questo caso il valore della media ottenuto utilizzando la formula per le classi è molto prossimo a quello corretto ottenuto in precedenza.

Per quanto concerne la mediana, dalla tabella dei dati ordinati, essendo $x_{40} = x_{41} = 179$, segue ovviamente $x_{med} = 179$. Se poi si effettua il calcolo con la formula per i dati organizzati in classi, si ha

$$x_{med} = \lambda_{m-1} + \frac{\frac{n}{2} - \varphi_c(\bar{x}_{m-1})}{\varphi(\bar{x}_m)} \ell = 178.5 + \frac{40 - 38}{22} 5 \approx 178.5 + 0.45 = 178.95,$$

che costituisce un'ottima approssimazione di 179, che è il valore esatto di x_{med} .

Infine, dalla tabella dei dati, si ha $x_{mod} = 179$. Facendo il calcolo sulla base dell'organizzazione dei dati in classi, si ha

$$x_{mod} = 178.5 + \frac{|22 - 19|}{|22 - 19| + |9 - 22|} 5 = 179.44.$$

Abbiamo finora visto misure di tendenza centrale che servono ad individuare il “centro” della distribuzione. Ciò però non indica come i dati siano distribuiti intorno al centro. In certi casi i dati possono essere estremamente concentrati attorno a questo valore centrale, in altri possono essere estremamente sparsi. Torna quindi utile avere delle **misure di dispersione**. Ovviamente il caso limite di dispersione nulla si ha quando tutti i dati coincidono.

Il primo indice di dispersione è il **campo di variazione** o **rango**, che abbiamo già definito. Questo intervallo ci dà una prima, anche se spesso grossolana, idea di come stanno le cose. Ad esempio, se i dati riguardano le temperature di un giorno in una data città, conoscere le temperature minima e massima può essere una indicazione preliminare abbastanza utile. È però evidente che questo indice risente in maniera significativa di valori particolarmente alti o bassi.

Deviazione standard e varianza

La **deviazione standard** σ , o **scarto quadratico medio**, già introdotta per una variabile aleatoria come radice quadrata della **varianza**, è l'indice di dispersione probabilmente più usato. Nel caso di un campione di dati x_1, x_2, \dots, x_n , la deviazione standard è definita nel modo seguente:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Anche la varianza σ^2 , definita come

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

costituisce una misura di dispersione molto comune. Ad essa sono espressamente rivolti alcuni capitoli della statistica matematica.

Come la media \bar{x} e la mediana x_{med} , anche σ può essere associato alla variabile casuale \tilde{X} definita in precedenza come quella variabile casuale finita per la quale $P(\tilde{X} = x_i)$ è uguale alla frequenza relativa di x_i nel campione. Si ha infatti $\sigma^2 = Var(\tilde{X})$.

Le quantità $(x_i - \bar{x})$ rappresentano gli **scarti dalla media** dei dati. Di qui il nome di “scarto quadratico medio” per σ e l’affermazione che *la varianza è uguale alla media dei quadrati degli scarti dalla media*. Osserviamo che quando la deviazione standard, a differenza della varianza, ha la “dimensione” dei dati del problema in esame.

Due formule molto importanti che riguardano la varianza sono:

$$\sigma_{aX+b}^2 = a^2 \sigma_X^2, \quad \sigma_X^2 = E(X^2) - E^2(X).$$

La prima formula è utile quando ci sono dei cambiamenti di scala e/o delle traslazioni dei dati: se si moltiplicano tutti i dati per uno stesso fattore, allora anche la deviazione standard risulterà moltiplicata per lo stesso fattore; se invece si traslano tutti i dati, la deviazione standard non ne viene influenzata. Quest’ultimo fatto risulta perfettamente comprensibile se si pensa al significato di questo indicatore come misura di dispersione: solo la posizione dei dati x_i rispetto alla media è significativa, e non la dislocazione dell’insieme di questi dati sull’asse x .

La seconda formula ci permette invece di calcolare la varianza (e quindi la deviazione standard) anche in questo modo:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2.$$

Anche per il calcolo della varianza σ^2 (e quindi della deviazione standard), se i dati sono raggruppati in classi, si possono utilizzare i punti di mezzo \bar{x}_k degli intervalli associati alle classi e le loro frequenze $\varphi(\bar{x}_k)$. La formula che dà σ^2 (in modo approssimato) è la seguente:

$$\sigma^2 = \frac{1}{n} \sum_{k=1}^{n_c} (\bar{x}_k - \bar{x})^2 \varphi(\bar{x}_k).$$

Varianza campionaria

Sia x_1, x_2, \dots, x_n un campione. Si definisce **varianza campionaria** la quantità

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Una formula equivalente, più semplice da calcolare, è

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right).$$

Anche se la varianza campionaria è una quantità importante, essa presenta problemi se utilizzata come misura di dispersione. La sua unità di misura infatti non è quella dei valori del campione ma è l'unità di misura del campione al quadrato. Una misura di dispersione è quindi ottenuta facendo la radice quadrata della varianza campionaria. Questa quantità è nota con il nome di **deviazione campionaria standard** ed è data da

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

È naturale chiedersi come mai la somma delle deviazioni al quadrato sia divisa per $n-1$ anziché per n . Lo scopo del calcolo della deviazione campionaria standard è quello di stimare la dispersione nella popolazione da cui il campione è stato estratto. Idealmente bisognerebbe calcolare le distanze di tutti i valori dalla media della popolazione piuttosto che dalla media del campione. Dato che, solitamente, la media della popolazione è sconosciuta, viene usata la media del campione per poterla stimare. Si può dimostrare che le distanze dalla media del campione risultano essere tendenzialmente più piccole delle distanze dalla media della popolazione e quindi dividere per $n-1$ anziché per n fornisce l'adeguata correzione.

Deviazioni medie

Altri due indici di dispersione sono la **deviazione media dalla media** e la **deviazione media dalla mediana**, che indichiamo rispettivamente con $D_{med}(\bar{x})$ e $D_{med}(x_{med})$. Tali indici sono dati dalla media aritmetica delle differenze in valore assoluto rispettivamente dalla media \bar{x} e dalla mediana x_{med} , ossia da

$$D_{med}(\bar{x}) = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|, \quad D_{med}(x_{med}) = \frac{1}{n} \sum_{i=1}^n |x_i - x_{med}|.$$

Esempio 5.2.3 Calcoliamo la varianza, la deviazione standard e le deviazioni medie dalla media e dalla mediana dei dati dell'esempio 5.1.1.

Calcoliamo la varianza utilizzando la relazione $\sigma_X^2 = E(X^2) - E^2(X)$, sapendo che $\bar{x} = 63.22$ (vedi esempio 5.2.1):

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1}{50} \sum_{i=1}^{50} x_i^2 - (63.22)^2 \approx 17.13.$$

A questo punto per avere la deviazione standard basta calcolare la radice quadrata di σ^2 :

$$\sigma = \sqrt{17.02} \approx 4.14.$$

Il calcolo della varianza poteva essere semplificato mediante la formula che utilizza i punti di mezzo delle classi e le loro frequenze. In questo modo si ottiene:

$$\begin{aligned} \sigma^2 &\approx \frac{1}{n} \sum_{k=1}^{n_c} (\bar{x}_k - \bar{x})^2 \varphi(\bar{x}_k) = \frac{1}{50} \sum_{k=1}^7 (\bar{x}_k - 63.22)^2 \varphi(\bar{x}_k) = (54 - 63.22)^2 2 + \\ &+ (57 - 63.22)^2 5 + (60 - 63.22)^2 9 + (63 - 63.22)^2 15 + (66 - 63.22)^2 12 + \\ &+ (69 - 63.22)^2 4 + (72 - 63.22)^2 3 \approx 18.30, \end{aligned}$$

da cui $\sigma \approx 4.28$. Di qui si vede come la formula basata sulla suddivisione in classi, essendo ovviamente la distribuzione che ne deriva più grossolana rispetto a quella dei dati di partenza, fornisca (in questo caso) un valore della deviazione standard con un errore di circa il 3.6%.

Calcoliamo infine le deviazioni medie dalla media e dalla mediana (sapendo dall'esempio 5.2.1 che $x_{med} = 63.5$):

$$\begin{aligned} D_{med}(\bar{x}) &= \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| = \frac{1}{50} \sum_{i=1}^{50} |x_i - 63.22| = 3.26; \\ D_{med}(x_{med}) &= \frac{1}{n} \sum_{i=1}^n |x_i - x_{med}| = \frac{1}{50} \sum_{i=1}^{50} |x_i - 63.5| = 3.26. \end{aligned}$$

(N.B.: I calcoli sono stati eseguiti con un programma di calcolo).

Esempio 5.2.4 Calcoliamo la varianza, la deviazione standard e le deviazioni medie dalla media e dalla mediana dei dati dell'esempio 5.1.2.

Procedendo come nell'esempio precedente, essendo ora $\bar{x} = 179.15$ e $x_{med} = 179$ (vedi esempio 5.2.2), si ha

$$\sigma^2 = \frac{1}{80} \sum_{i=1}^{80} x_i^2 - (179.15)^2 \approx 67.05, \quad \text{da cui} \quad \sigma = \sqrt{67.05} \approx 8.19.$$

Se poi si effettua il calcolo (approssimato) mediante la formula che usa i punti di mezzo degli intervalli delle classi, si ha

$$\sigma^2 \approx \frac{1}{n} \sum_{k=1}^{n_c} (\bar{x}_k - \bar{x})^2 \varphi(\bar{x}_k) \approx 68.90, \quad \text{da cui} \quad \sigma \approx 8.30,$$

con un errore su σ di poco superiore all'1%. Calcoliamo infine le deviazioni medie dalla media e dalla mediana:

$$\begin{aligned} D_{med}(\bar{x}) &= \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| = \frac{1}{80} \sum_{i=1}^{80} |x_i - 179.15| \approx 6.24; \\ D_{med}(x_{med}) &= \frac{1}{n} \sum_{i=1}^n |x_i - x_{med}| = \frac{1}{80} \sum_{i=1}^{80} |x_i - 179| \approx 6.22. \end{aligned}$$

Come per l'esempio precedente, i calcoli sono stati fatti con un programma di calcolo.

Outlier

Talvolta un campione può contenere alcuni valori che sono molto più grandi o molto più piccoli rispetto agli altri. Tali valori vengono chiamati **valori anomali**, in inglese **outlier**. A volte gli outlier sono errori provocati dal processo di data entry, per esempio, una virgola per indicare il decimale digitata in una posizione errata. Gli outlier dovrebbero sempre essere analizzati e ogni outlier che risulti da un errore dovrebbe essere corretto o eliminato dall'insieme dei dati. Ma non tutti i valori anomali sono errori. Spesso una popolazione può contenere alcuni valori che differiscono molto dagli altri e gli outlier presenti nel campione riflettono tale situazione.

I valori anomali rappresentano un vero problema nell'analisi dei dati. Per questa ragione, quando nel campione si osservano dei valori anomali, si cerca di trovare una ragione o una scusa per eliminarli, spesso senza motivo. Ma un outlier non dovrebbe essere cancellato a meno che esso non sia chiaramente un errore. Se in una popolazione sono presenti degli outlier, ma questi ultimi non vengono compresi nel campione, il campione non rappresenterà correttamente la popolazione.

Percentili

Il **p -esimo percentile** di un campione, dato p un numero tra 0 e 100, divide il campione in modo tale che almeno il $p\%$ dei valori campionari siano più piccoli del p -esimo percentile ed il restante $(100 - p)\%$ siano più grandi. Ci sono diversi modi per calcolare i percentili e tutti producono gli stessi risultati. Il metodo più semplice è il seguente. Prima di tutto si ordinano i valori del campione in ordine crescente e si calcola la quantità $(p/100)(n + 1)$, dove n è l'ampiezza del campione. Se $(p/100)(n + 1)$ è un numero intero allora il valore del campione che occupa questa posizione è il p -esimo percentile. In caso contrario è la media dei due valori tra cui giace $(p/100)(n + 1)$.

Quartili

La mediana divide il campione in due parti della stessa numerosità. I **quartili** dividono il campione in quattro parti aventi la stessa numerosità. Utilizzando anche per i quartili un metodo analogo a quello visto per i percentili, si ha che per calcolare il primo quartile bisogna calcolare, dopo aver ordinato i valori del campione in ordine crescente, il valore $0.25(n + 1)$ e per il terzo quartile il valore $0.75(n + 1)$. Il secondo quartile utilizza il valore $0.50(n + 1)$ in quanto coincide con la mediana. Si

osservi infatti che il primo quartile è il 25-esimo percentile, la mediana è il 50-esimo percentile e il terzo quartile è il 75-esimo percentile.

Spesso i percentili (così come i quartili) sono utilizzati per poter interpretare i risultati sui test standardizzati. Per esempio, se il punteggio del test standardizzato per entrare all'università di uno studente è pari al 25-esimo percentile, questo significa che il 64% degli studenti ha conseguito un punteggio più basso.

Boxplot

Un **boxplot** è un grafico che riporta la mediana, il primo e il terzo quartile e gli outlier presenti nel campione. Un boxplot è semplice da interpretare, anche se bisogna introdurre alcune definizioni. La differenza interquartile (IQR) è la differenza tra il terzo e il primo quartile. Dato che almeno il 75% dei dati ha un valore inferiore al terzo quartile e che almeno il 25% di essi presenta un valore inferiore al primo quartile, segue che il 50%, o la metà dei dati presentano un valore compreso tra il primo e il terzo quartile.

Poiché gli outlier sono quelle unità che presentano valori molto grandi o molto piccoli rispetto agli altri dati, se IQR rappresenta lo scarto interquartile, allora ogni unità che è più grande di 1.5 IQR del terzo quartile o più piccola di 1.5 IQR del primo quartile, è considerato un outlier. Alcuni testi indicano un'unità più piccola o più grande di 3 IQR rispetto al primo e al terzo quartile come un **outlier estremo**. Queste definizioni di outlier sono convenzionali e servono solo per disegnare il boxplot.

La figura 16 riporta un boxplot per una distribuzione ipotetica. Il grafico consiste in un rettangolo nella cui parte in basso viene riportato il primo quartile e nella parte in alto il terzo quartile. Una linea orizzontale rappresenta la mediana. Gli outlier sono disegnati individualmente e sono indicati con delle croci. Le linee verticali che si estendono sopra e sotto la scatola sono chiamate **baffi**. I punti finali dei baffi sono i valori estremi delle unità che non sono outlier.

Tralasciando gli outlier, un boxplot può essere visto come diviso in quattro parti: le due parti del rettangolo separate dalla linea mediana e i due baffi. Ognuna di queste quattro parti rappresenta un quarto dei dati. Il boxplot indica come ogni quarta parte dei dati viene divisa su ogni intervallo, in questo modo si possono determinare gli intervalli in cui i valori campionari sono più addensati e quelli in cui sono più dispersi.

Osservazione

Le sintesi statistiche descritte vengono chiamate **statistiche descrittive** poichè forniscono una descrizione dei dati. Ogni statistica descrittiva fatta sul campione ha una formulazione analoga per la popolazione. Questo è semplice da vedere quando

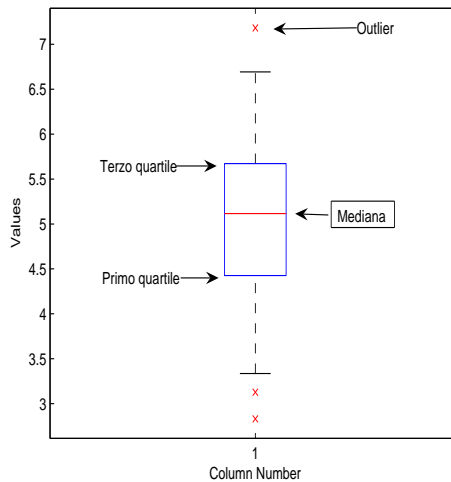


Figura 16: Anatomia di un boxplot

la popolazione è finita. Per esempio, la media della popolazione è semplicemente la media calcolata su tutti i valori della popolazione; la mediana della popolazione è il valore centrale o la media tra i due valori centrali e così via. Infatti ogni sintesi statistica calcolata sul campione può essere calcolata su una popolazione finita, utilizzando le stesse formule sui valori degli elementi della popolazione. Una piccola eccezione è fatta per la varianza della popolazione, dove, come si è visto, il numeratore viene diviso per n anziché per $n - 1$. Esistono però delle differenze per quanto riguarda la terminologia: le sintesi numeriche calcolate su un campione vengono definite **statistiche**, mentre quelle della popolazione **parametri**. Naturalmente, nella pratica, non è mai osservata l'intera popolazione, così che i parametri non possono essere calcolati direttamente. Le statistiche campionarie, allora, servono per stimare i parametri sconosciuti della popolazione.

Esercizio 5.a *Calcolare la media, la mediana, lo scarto quadratico medio e le deviazioni medie dalla media e dalla mediana dei dati*

46 31 1 33 2 44 66 8 54 99 92 98 69 50

Innanzitutto ordiniamo i 14 dati in senso crescente:

1 2 8 31 33 44 46 50 54 66 69 92 98 99

Calcoliamo la media:

$$\bar{x} = \frac{1}{14}(1 + 2 + 8 + \dots + 98 + 99) = \frac{693}{14} = 49.5.$$

Per quanto riguarda la mediana si ha

$$x_{med} = \frac{x_7 + x_8}{2} = \frac{46 + 50}{2} = 48.$$

Dovendo poi calcolare lo scarto quadratico medio, serve la varianza:

$$\sigma^2 = \frac{1}{14}(1^2 + 2^2 + 8^2 + \dots + 98^2 + 99^2) - (49.5)^2 = 1019.25,$$

da cui

$$\sigma = \sqrt{1019.25} \approx 31.93.$$

Infine

$$\begin{aligned} D_{med}(\bar{x}) &= \frac{1}{14} \sum_{i=1}^{14} |x_i - 49.5| = \frac{363}{14} \approx 25.93; \\ D_{med}(x_{med}) &= \frac{1}{14} \sum_{i=1}^{14} |x_i - 47| = \frac{363}{14} \approx 25.93. \end{aligned}$$

Il fatto che queste due ultime medie siano uguali ha una facile spiegazione geometrica: *quando i dati sono in numero pari e anche la media è compresa fra i due dati di mezzo (cioè $x_{\frac{n}{2}}$ e $x_{\frac{n}{2}+1}$), si ha sempre $D_{med}(\bar{x}) = D_{med}(x_{med})$.*

Esercizio 5.b *Calcolare la media, la mediana e le deviazioni medie dalla media e dalla mediana dei dati dell'esercizio precedente sostituendo 91 a 1*

Sostituito il numero 1 con 91 il nuovo campione ordinato è il seguente:

$$2 \ 8 \ 31 \ 33 \ 44 \ 46 \ 50 \ 54 \ 66 \ 69 \ 91 \ 92 \ 98 \ 99.$$

Calcoliamo la nuova media e la nuova mediana

$$\begin{aligned} \bar{x} &= \frac{1}{14}(2 + 8 + 31 + \dots + 98 + 99) = \frac{783}{14} \approx 55.93; \\ x_{med} &= \frac{x_7 + x_8}{2} = \frac{50 + 54}{2} = 52. \end{aligned}$$

Calcoliamo ora le due deviazioni medie:

$$\begin{aligned} D_{med}(\bar{x}) &= \frac{1}{14} \sum_{i=1}^{14} |x_i - 55.93| \approx 25.63; \\ D_{med}(x_{med}) &= \frac{1}{14} \sum_{i=1}^{14} |x_i - 52| = \frac{355}{14} \approx 25.36. \end{aligned}$$

Si può verificare che ora, essendo \bar{x} esterno all'intervallo $[x_7, x_8]$ (di cui la mediana è il punto medio), $D_{med}(\bar{x})$ e $D_{med}(x_{med})$ sono diversi.

Esercizio 5.c *Uno studente di ingegneria ha sostenuto 16 esami, ciascuno dei quali con un dato numero di crediti formativi. I voti riportati dallo studente, ciascuno con a fianco il numero dei crediti relativi a quell'esame, sono i seguenti:*

$$\begin{array}{cccccccc} 28 (6) & 21 (8) & 22 (5) & 24 (6) & 24 (8) & 25 (4) & 25 (6) & 26 (8) \\ 27 (5) & 27 (4) & 27 (6) & 19 (10) & 28 (5) & 29 (7) & 30 (8) & 30 (4) \end{array}$$

Si chiede di calcolare: a) la media, la mediana e la deviazione standard dei voti; b) la media, la mediana e la deviazione standard dei crediti; c) la media ponderata dei voti assumendo come pesi i crediti.

a) Ordiniamo innanzitutto i 16 voti. Si ha

19 21 22 24 24 25 25 26 27 27 27 28 28 29 30 30

Indicati con v_1, v_2, \dots, v_{16} i voti così ordinati e con \bar{v} , v_{med} e σ_v rispettivamente la media, la mediana e la deviazione standard, abbiamo

$$\begin{aligned}\bar{v} &= \frac{1}{16} \sum_{i=1}^{16} v_i = \frac{412}{16} = 25.75; & v_{med} &= \frac{v_8 + v_9}{2} = 26.5; \\ \sigma_v &= \left(\frac{1}{16} \sum_{i=1}^{16} (v_i - 25.75)^2 \right)^{\frac{1}{2}} \approx 3.07.\end{aligned}$$

b) Ordiniamo anche i crediti:

4 4 4 5 5 5 6 6 6 6 7 8 8 8 8 10

Indicati con c_1, c_2, \dots, c_{16} i crediti così ordinati e rispettivamente con \bar{c} , c_{med} e σ_c le relative media, mediana e deviazione standard, abbiamo

$$\begin{aligned}\bar{c} &= \frac{1}{16} \sum_{i=1}^{16} c_i = \frac{100}{16} = 6.25; & c_{med} &= \frac{c_8 + c_9}{2} = 6; \\ \sigma_c &= \left(\frac{1}{16} \sum_{i=1}^{16} (c_i - 6.25)^2 \right)^{\frac{1}{2}} \approx 1.71.\end{aligned}$$

c) Calcoliamo infine la media ponderata dei voti, v_{pond} , assumendo come pesi i relativi crediti. Riordinati i c_i in modo che ci sia corrispondenza fra voti e crediti:

10 8 5 6 8 4 6 8 4 5 6 6 5 7 4 8

abbiamo

$$v_{pond} = \frac{\sum_{i=1}^{16} v_i \cdot c_i}{\sum_{i=1}^{16} c_i} = 25.38.$$

6 STATISTICA MATEMATICA

6.1 Popolazioni e Campioni

Definizione. Si definisce **popolazione** un insieme i cui elementi hanno in comune almeno una caratteristica (od attributo).

Esempi di popolazioni: gli ingegneri che si sono laureati in Italia dal 1950 al 1980; i giorni con vento superiore ai 100 Km/h a Trieste nel mese di aprile dal 1900 al 1999; gli italiani aventi diritto al voto per il senato alle elezioni politiche del 2001; i corpi celesti dell'universo; gli alberi passati e presenti di tutte le foreste del mondo.

Esempi di caratteristiche nel caso degli ingegneri: l'età al momento della laurea; l'età al momento del primo impiego come ingegnere; l'altezza; il peso; il sesso; il primo stipendio; ecc.

Le popolazioni possono essere finite o infinite. In genere popolazioni molto numerose sono considerate infinite anche se non lo sono (ad esempio i corpi celesti dell'universo).

Ogni caratteristica della popolazione, nella maggior parte dei casi, viene misurata da un valore numerico per ciascuno degli N elementi che la compongono. Di conseguenza uno studio completo della popolazione implicherebbe un insieme di N numeri. In genere, però, N è così grande da rendere impraticabile, per ovvi motivi, la misurazione della caratteristica per l'intera popolazione. Ci si limita dunque a farlo solo per un suo sottoinsieme, spesso assai limitato, detto **campione**. Uno scopo delle ricerche statistiche è quello di *inferire* (da cui il nome di **inferenza statistica**), cioè fare delle deduzioni o delle previsioni sulla popolazione mediante l'esame di un campione.

Compito della statistica è risalire dai campioni ai parametri della popolazione.

Matematicamente la caratteristica oggetto di studio è una variabile aleatoria X la cui distribuzione è, in generale completamente incognita. In ogni caso la variabile aleatoria X sottostante alla popolazione in questione avrà una media ed una varianza, che nel seguito indicheremo semplicemente con μ e σ^2 , ossia

$$\mu = \mu_X = E(X), \quad \sigma^2 = \sigma_X^2 = Var(X).$$

Nel seguito ci riferiremo spesso a μ e σ^2 come alla media e alla varianza della popolazione oggetto di indagine, sottintendendo ovviamente con ciò μ_X e σ_X^2 .

Definizione. Si chiama **campione casuale di dimensione n** , estratto da una popolazione avente X come variabile aleatoria sottostante, una variabile n -dimensionale (X_1, X_2, \dots, X_n) , con le X_i indipendenti e aventi la stessa distribuzione di X .

Quando si misura la caratteristica della popolazione limitandosi ad un campione di dimensione n , si ottengono n misure x_1, x_2, \dots, x_n : ciò equivale ad una singola esecuzione dell'esperimento rappresentato dalla variabile n -dimensionale (X_1, X_2, \dots, X_n) con risultato (x_1, x_2, \dots, x_n) .

Sul problema della scelta del campione, che nella pratica consiste nell'estrarre n elementi da un insieme di N , con $n \ll N$, ci limitiamo ad osservare l'importanza che ciò venga realmente fatto **a caso** e che esistono diverse tecniche utili allo scopo.

Riassumendo: Una **popolazione** è, quindi, un'ampia collezione di valori di una variabile aleatoria. Essa si può descrivere con una funzione densità, nel caso continuo, o di probabilità, nel caso discreto, che dipende da alcuni **parametri** (ad es. media, varianza). Un **campione** è un sottoinsieme della popolazione, cioè un insieme di valori $\{x_1, \dots, x_n\}$ estratti dalla popolazione.

Secondo quanto si è definito, x_1, x_2, \dots, x_n sono numeri. E' molto utile considerarli anche come variabili aleatorie indipendenti, ciascuna distribuita come la popolazione da cui si effettua l'estrazione (cioè x_i è indipendente da x_j per i diverso da j ed ogni x_i ha per media e varianza la media μ e la varianza σ^2 della popolazione, ha la stessa funzione distribuzione, ecc.). Ossia assumiamo *l'ipotesi che le x_i siano variabili aleatorie indipendenti e identicamente distribuite (i.i.d.)*. Per quanto detto, a rigore si dovrebbe usare un simbolo (ad es. x_i) per il numero "i-esimo estratto in un particolare esperimento casuale", ed un altro simbolo (ad es. X_i) per la variabile aleatoria. In alcuni casi, quando il contesto è chiaro, preferiamo usare lettera minuscola latina x_i .

6.2 Stimatori

Definizione. Un **riassunto** o **statistica** o **stimatore** è una funzione dei dati campionari: $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$. Uno stimatore puntuale $\hat{\theta}$ fornisce un singolo valore come stima di un corrispondente parametro θ della popolazione.

Definizione. Uno stimatore $\hat{\theta}$ del parametro θ si dice **corretto** se la sua media coincide con θ stesso, cioè se $E(\hat{\theta}) = \theta$

Per chiarire: i parametri, quali μ, σ , sono costanti numeriche di solito incognite,

relative alla popolazione. Gli stimatori, quali la media campionaria

$$\bar{x} := \frac{1}{n} \sum_1^n x_k$$

o la varianza campionaria

$$\tilde{s}^2 := \frac{1}{n} \sum_1^n (x_k - \bar{x})^2$$

sono variabili aleatorie, perchè dipendono dalle n variabili aleatorie x_i . Ebbene, ogni stimatore avrà una sua distribuzione di probabilità. Ad esempio:

Teorema. Se x_1, \dots, x_n sono variabili aleatorie i.i.d. con media μ e varianza σ^2 ,

$$E(\bar{x}) = \mu; \quad Var(\bar{x}) \equiv (\sigma_{\bar{x}})^2 = \frac{\sigma^2}{n}$$

(la media campionaria è uno stimatore corretto della media vera μ). Se inoltre le variabili aleatorie x_i sono normali, allora $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$.

Dimostrazione: Si sa che la media di una somma è uguale alla somma delle medie. Inoltre, per variabili aleatorie indipendenti, la varianza di una somma è uguale alla somma delle varianze, perciò:

$$E[\bar{x}] = \mu, \quad Var[\bar{x}] = \frac{1}{n^2} n \sigma^2$$

Infine si sa che la somma di variabili aleatorie normali è normale. □

Definizione. Si dice **errore standard** la deviazione standard di \bar{x} :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

In sintesi: la media campionaria \bar{x} ha la stessa media della distribuzione originale, cioè della popolazione da cui proviene il campione, ma dispersione minore, pari a σ/\sqrt{n} . L'errore standard è importantissimo: misura la dispersione delle medie campionarie attorno a μ , pensando in linea teorica di fare ripetuti campionamenti. La dispersione è quindi inversamente proporzionale a \sqrt{n} , e quindi al crescere di n i valori delle corrispondenti medie campionarie tendono a concentrarsi attorno al loro valore medio.

Definizione. Si definisce **varianza campionaria corretta**

$$s^2 = \frac{n}{n-1} \tilde{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Per distinguerla dalla varianza campionaria \tilde{s}^2 , chiamiamo s^2 **varianza corretta**. Lo stimatore s^2 viene utilizzato per stimare la varianza σ^2 della popolazione X . Il fatto che si usi questo stimatore anzichè \tilde{s}^2 come potrebbe apparire più naturale, è dovuto al fatto che quest'ultimo non è uno stimatore corretto, mentre s^2 lo è. Infatti, si dimostra che $E(s^2) = \sigma^2$ e che $E(\tilde{s}^2) = \frac{n-1}{n}\sigma^2$. Per n grande non c'è differenza apprezzabile, ma per piccoli campioni essa è rilevante. $n - 1$ è detto “numero dei gradi di libertà” (GL):

Definizione. Il **numero dei gradi di libertà** è il numero di scarti linearmente indipendenti, cioè degli scarti il cui valore non dipende dal valore assunto dagli altri scarti.

Esercizio: Per la varianza campionaria il numero dei gradi di libertà è $n - 1$. Infatti, poichè $\sum_{i=1}^n (x_i - \bar{x}) = 0$, gli scarti sono indipendenti tranne l'ultimo, che si può ottenere come l'opposto della somma dei primi $n - 1$.

Ma non sempre il risultato è così ovvio: in questi casi il numero degli scarti indipendenti si calcola come numero degli scarti, diminuito del numero delle statistiche stimate sul campione e utilizzate per il calcolo degli scarti stessi. Nel caso della varianza: $n =$ numero di scarti; per calcolarli è necessario prima calcolare \bar{x} ; quindi gli scarti indipendenti sono $n - 1$.

6.3 Distribuzioni collegate alla normale

Vediamo ora due distribuzioni campionarie di notevole importanza in Statistica, entrambe collegate alla distribuzione normale.

Definizione. Se $Z \sim N(0, 1)$ è la variabile aleatoria normale standardizzata, la variabile aleatoria Z^2 è detta “**chi quadrato con 1 grado di libertà**”. Se $Z_i \sim N(0, 1)$, sono ν variabili aleatorie normali standardizzate indipendenti, la somma dei loro quadrati è detta “**chi quadrato con ν gradi di libertà:**”

$$Z^2 \simeq \chi^2(1), \quad \sum_{i=1}^{\nu} Z_i^2 \sim \chi^2(\nu).$$

Note. Per definizione i valori di χ^2 con ν gradi di libertà sono positivi o nulli, con una densità di probabilità

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{-1+\nu/2} e^{-x/2} & x > 0 \\ 0, & \text{altrove} \end{cases}$$

che ha un asintoto verticale per $x \rightarrow 0^+$ quando $\nu = 1$, mentre al crescere di ν crescono sia la media sia la dispersione:

$$E[\chi^2(\nu)] = \nu, \quad Var[\chi^2(\nu)] = 2\nu.$$

Per ν piccolo $f(x)$ ha il picco vicino all'origine ed è sempre più dispersa e più simmetrica al crescere di ν . Si dimostra, in legge, il seguente comportamento asintotico:

$$\text{per } \nu \rightarrow +\infty, \quad \sqrt{2\chi^2(\nu)} - \sqrt{2\nu - 1} \sim N(0, 1)$$

In pratica, se servono i quantili $\chi_\alpha^2(\nu)$ per $\nu > 30$, tutto è riportabile ai quantili ϕ_α della normale standard:

$$\chi_\alpha^2(\nu) \simeq \frac{1}{2}(\phi_\alpha + \sqrt{2\nu - 1})^2$$

[Remember: nel caso in cui la funzione distribuzione sia strettamente crescente, il quantile α -esimo di una variabile aleatoria è il punto nel quale la funzione distribuzione vale α : ad es.

$$P[N(0, 1) \leq \phi_\alpha] = \alpha.$$

Il quantile relativo ad $\alpha = 50\%$ è la mediana.]

Dalla definizione di $\chi^2(\nu)$ segue immediatamente che, se (X_1, X_2, \dots, X_ν) è un campione casuale estratto da una popolazione distribuita normalmente e con media μ e varianza σ^2 , allora la variabile aleatoria

$$Z^2 = \sum_{i=1}^{\nu} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

segue una distribuzione $\chi^2(\nu)$.

Definizione. Si definisce una variabile aleatoria T_ν detta "T di Student con ν gradi di libertà":

$$T_\nu = \frac{X}{\sqrt{Y}} \sqrt{\nu}, \quad \text{con } X \sim N(0, 1), \quad Y \sim \chi^2(\nu) \text{ indipendenti.}$$

La variabile aleatoria T_ν ha densità:

$$f(t) = \frac{\Gamma((\nu + 1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}, \quad t \in (-\infty, +\infty)$$

Ai nostri fini basterà ricordare che T_ν ha densità continua e pari rispetto all'origine, con media e varianza:

$$E(T_\nu) = 0, \quad Var(T_\nu) = \frac{\nu}{\nu - 2},$$

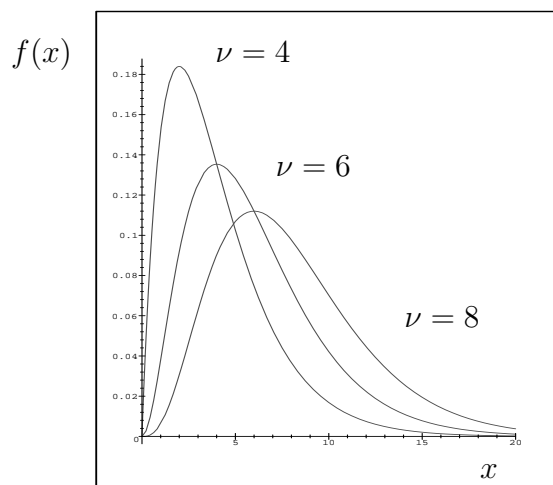


Figura 17: grafico della densità di probabilità di χ^2 con ν gradi di libertà.

e che converge in legge alla $N(0, 1)$ per $\nu \rightarrow \infty$. Ai fini applicativi, da $\nu = 30$ in poi la T_ν equivale in sostanza alla normale standard. Alla fine sono riportate due tavole delle distribuzioni $\chi^2(\nu)$ e T_ν con i valori più significativi ai fini delle applicazioni. In analogia con una terminologia già introdotta per la distribuzione $N(0, 1)$, le soluzioni x_α e t_α delle equazioni

$$[\chi^2(\nu)](x_\alpha) = \alpha \quad \text{e} \quad T_\nu(t_\alpha) = \alpha$$

saranno chiamate **quantili** relativi ad α (rispettivamente della distribuzione χ_ν^2 e della distribuzione di Student T_ν).

Ecco perchè sono utili il “chi quadro”, la “t di Student” e una ulteriore variabile aleatoria tabulata nei testi detta “F di Fisher”:

Teorema sul “chi quadro”. La varianza campionaria corretta, relativa ad n dati normalmente distribuiti $x_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$, è una variabile aleatoria proporzionale al chi quadro con $n - 1$ gradi di libertà:

$$s^2 \sim \frac{1}{n-1} \chi^2(n-1) \sigma^2.$$

Teorema sulla “t di Student”. Nella solita ipotesi che la popolazione sia distribuita normalmente con media μ e varianza σ^2 , lo scarto standardizzato stimato segue la distribuzione t di Student con $n - 1$ GL, cioè

$$\frac{\bar{x} - \mu}{s} \sqrt{n} \simeq t(n-1).$$

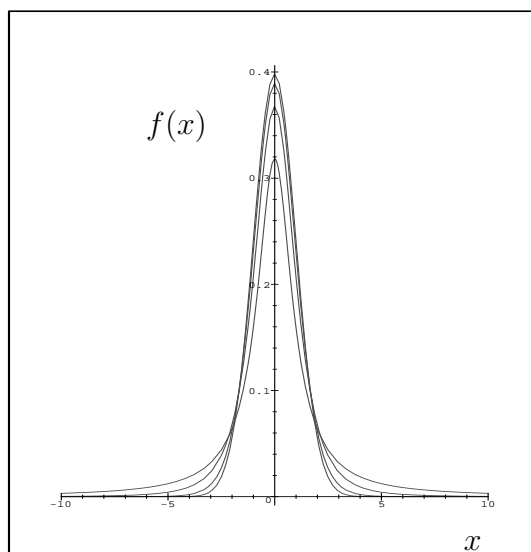


Figura 18: grafico della densità di probabilità della T di Student con ν gradi di libertà ($\nu = 1, \nu = 3, \nu = 10, \nu = 120$).

(Invece, come è noto, lo scarto standardizzato $z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$ è distribuito normalmente: si ricordi che \bar{x} è normale con media μ e varianza σ^2/n).

Teorema sulla “F di Fisher”. Disponendo di due stime indipendenti della varianza di una popolazione normalmente distribuita, il loro rapporto è distribuito secondo la legge F con $n_1 - 1$ ed $n_2 - 1$ GL, cioè:

$$\frac{s_1^2}{s_2^2} = \simeq F(n_1 - 1, n_2 - 1).$$

dove n_1, n_2 sono le lunghezze dei due campioni indipendenti.

6.4 Intervalli di fiducia

Come abbiamo già detto, uno stimatore è una variabile aleatoria che serve per stimare un parametro incognito θ della nostra popolazione. Ovviamente i parametri che ci interessa maggiormente stimare sono la media μ e la deviazione standard σ e ciò può essere fatto estraendo un campione e calcolando le stime puntuali \bar{x}, s . D’ora in poi, salvo avviso contrario, la popolazione sarà supposta normalmente distribuita. La domanda in questo ambito potrebbe essere: dato un particolare campione, quale intervallo del tipo $(\bar{x} - \delta, \bar{x} + \delta)$ conterrà la media incognita μ con probabilità del 95%? oppure del 99%?

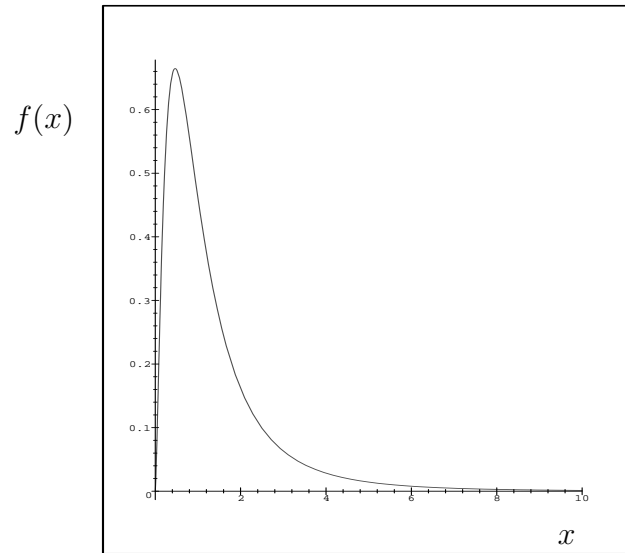


Figura 19: grafico della densità di probabilità della F di Fisher con $n_1 - 1$ ed $n_2 - 1$ gradi di libertà.

Figura 20: grafico della densità di probabilità della F di Fisher con $d_1 = n_1 - 1$ ed $d_2 = n_2 - 1$ gradi di libertà.

Definizione Si definisce **intervallo di fiducia** di livello $100(1-\alpha)\%$ per il parametro θ un intervallo (θ_1, θ_2) entro il quale il parametro assume i valori con una prefissata probabilità, cioè un intervallo (θ_1, θ_2) tale che

$$P(\theta_1 < \theta \leq \theta_2) = 1 - \alpha.$$

avendo indicato con $1 - \alpha$ la probabilità prefissata e con $\theta_i = f_i(X_1, \dots, X_n)$, $i = 1, 2$ variabili aleatorie funzioni del campione casuale.

La percentuale $100(1 - \alpha)\%$ viene detta **livello di fiducia**. Spesso, per semplicità, ci si riferisce al numero $1 - \alpha$, che ovviamente è sempre positivo, come al “livello di fiducia $1 - \alpha$ ”. In genere interessano piccoli valori di α , tipicamente $\alpha = 0.05$ oppure $\alpha = 0.01$. Il livello di fiducia nei due casi è quindi 95% , e 99% , rispettivamente.

Inoltre sono intervalli di fiducia con livello $1 - \alpha$ anche intervalli non necessariamente simmetrici: ad es. $(-\infty, \bar{x} + \delta']$ con la proprietà $P(\mu \in (-\infty, \bar{x} + \delta']) \geq 1 - \alpha$.

Se $P(\theta < \theta_1) = P(\theta > \theta_2) = \frac{\alpha}{2}$, l'intervallo di fiducia è detto **bilaterale**; se si ha $P(\theta > \theta_2) = \alpha$ o $P(\theta < \theta_1) = \alpha$ allora l'intervallo è detto **unilaterale sinistro** nel primo caso e **unilaterale destro** nel secondo.

Eseguito l'esperimento mediante il valore misurato (x_1, \dots, x_n) del campione si ricava l'intervallo numerico (θ_1^*, θ_2^*) con $\theta_i^* = f_i(x_1, \dots, x_n)$, $i = 1, 2$ che costituisce una stima per intervalli del parametro θ al livello di fiducia $(1 - \alpha)\%$.

Descriviamo ora i metodi per stimare la media μ di una popolazione che supporremo avere (come già più volte ribadito) una distribuzione normale. Tali metodi, ovviamente rigorosi solo per una popolazione normale, nella pratica corrente sono usati con maggior generalità, supportati in ciò, quando la dimensione del campione è sufficientemente grande, dal Teorema di Limite Centrale.

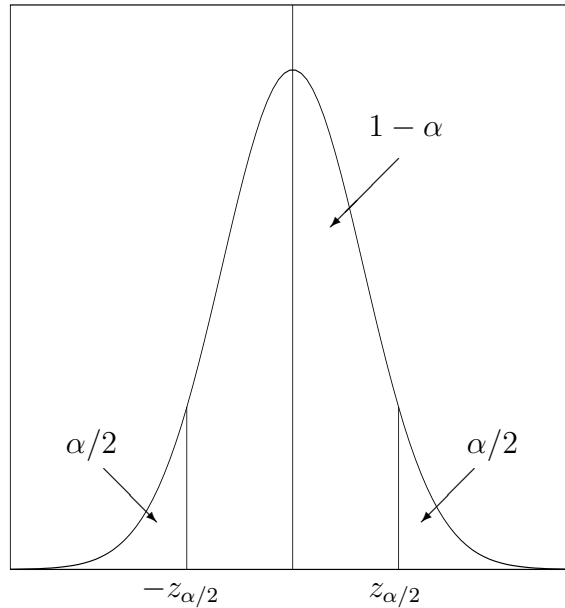
La stima per intervalli di fiducia della media μ di una popolazione normale viene ora affrontata considerando separatamente il caso in cui la varianza σ^2 è nota (anche se poco frequente) e quello in cui è incognita.

Intervallo di fiducia di una media, nota σ :

Con probabilità $100(1 - \alpha)\%$, l'intervallo di estremi $\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ contiene μ , cioè:

$$\mu = \bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}.$$

Dimostrazione. Come abbiamo già detto, lo stimatore che si usa per la media di una popolazione è la media campionaria \bar{X} . Sappiamo anche che



$$E(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

(Infatti, $\text{Var}(\bar{X}) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{\sigma^2}{n}$ per cui $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.)

Inoltre, siccome la popolazione è distribuita normalmente, anche \bar{X} è normale. Di conseguenza, la variabile aleatoria

$$\bar{Z} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

segue la distribuzione normale standardizzata, le cui probabilità possono essere desunte dalle tabelle statistiche della densità normale $\Phi(x)$. Determinando il quantile $z_{\frac{\alpha}{2}}$ soluzione dell'equazione

$$\Phi(z) = 1 - \frac{\alpha}{2},$$

ed essendo $\Phi(-z_{\frac{\alpha}{2}}) = 1 - \Phi(z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$, l'intervallo $[-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}]$ è tale che

$$P\left(-z_{\frac{\alpha}{2}} < \bar{Z} < z_{\frac{\alpha}{2}}\right) = P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Quindi, l'intervallo

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}\right],$$

che è *aleatorio* in quanto è tale il suo punto centrale \bar{X} , contiene con probabilità $1 - \alpha$ il valore vero μ . Eseguito l'esperimento, l'*intervallo osservato* si ottiene dall'*intervallo aleatorio* sostituendo alla media campionaria \bar{X} la media aritmetica \bar{x} dei valori osservati negli n esperimenti. Useremo quindi l'*intervallo osservato* per dare una stima di μ di livello di fiducia uguale ad $1 - \alpha$:

$$\mu \in \left[\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} \right].$$

□

Osserviamo che, se cresce il numero n degli esperimenti, l'ampiezza dell'intervallo diminuisce, e dunque la stima si fa più *informativa*: fare esperimenti è costoso, ma poi “ripaga”. Tuttavia, osserviamo anche che, poiché l'ampiezza dell'intervallo diminuisce in modo inversamente proporzionale a \sqrt{n} , il vantaggio che si ottiene aggiungendo via via nuovi dati diventa gradualmente sempre meno significativo. D'altra parte, se aumenta il grado di fiducia, diminuisce α e il quantile z cresce; di conseguenza cresce l'ampiezza dell'intervallo e quindi la stima diventa meno informativa.

Esempio: Con probabilità del 95% $\mu = \bar{x} \pm 1,96 \cdot \frac{\sigma}{\sqrt{n}}$. Analogamente, ad es., l'intervallo di confidenza con probabilità pari al 99% è $\mu = \bar{x} \pm 2,58 \cdot \sigma/\sqrt{n}$.

Intervallo di fiducia per una media con σ incognita.

Con probabilità 95%,

$$\mu = \bar{x} \pm t^* \cdot s/\sqrt{n}.$$

dove $t^* = t_{0,975}(n-1)$ è il valore della variabile aleatoria di Student, con $n-1$ GL, che esclude il 5% (2,5% per parte). Più in generale l'intervallo di fiducia al livello $1-\alpha$ per la media di una popolazione approssimativamente gaussiana è

$$\mu = \bar{x} \pm t_{1-\alpha/2}(n-1) \cdot s/\sqrt{n}.$$

Dimostrazione: Poiché è ignota $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, la standardizzazione dello scarto è ottenuta con

$$\frac{\bar{x} - \mu}{s_{\bar{x}}} = t, \quad \text{dove } s_{\bar{x}} = s/\sqrt{n}$$

Questa è precisamente la t di Student con $n-1$ GL. Allora dalle tavole si trova il numero $t^* = t_{0,975}(n-1)$ tale che

$$\begin{aligned} 95\% &= P(-t^* \leq t_{n-1} \leq t^*) = P(-t^* \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq t^*) \\ &= P(\bar{x} - t^* \cdot s/\sqrt{n} \leq \mu \leq \bar{x} + t^* \cdot s/\sqrt{n}). \end{aligned}$$

□

Esercizio: Nella produzione di semiconduttori non si può controllare esattamente la loro resistenza. Allora si scelgono a caso $n=5$ elementi, e le misure x_i , $i=1, \dots, 5$, sono 24, 26, 30, 28, 22.

Se non conosco il parametro σ della popolazione

$$\bar{x} = 26 \quad s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = 40/4 = 10 \quad s = 3.16$$

$$s/\sqrt{n} = 1.41 \quad \mu = 26 \pm t_{0.975}(4) \cdot 1.41 = 26 \pm 2.78 \cdot 1.4142 = 26 \pm 3.9.$$

Quindi l'intervallo di fiducia al livello $\alpha = 5\%$ è [22.1, 29.9].

Qualora la varianza fosse proprio $\sigma^2 = 10$,

$$\sigma = 3.16 \quad \sigma_{\bar{x}} = 3.16/\sqrt{5} = 1.41 \quad \mu = 26 \pm 1.96 \cdot 1.41 = 26 \pm 2.77.$$

Visto che è più accurato l'intervallo di confidenza? Si capisce: avendo un'informazione in più, che è σ , è meglio precisata la media.

Esercizio: Nelle misure fisiche si ottengono tanti dati in genere non coincidenti fra loro. Le $n = 100$ misurazioni della velocità della luce eseguite da Michelson nel 1879 diedero 299000 Km/sec più valori moderatamente fluttuanti aventi media e varianza campionaria:

$$\bar{x} = 852.40 \quad s^2 = 6242.67$$

Al livello $\alpha = 5\%$

$$\begin{aligned} \mu &= \bar{x} \pm t_{0.975}(99) \cdot s/\sqrt{100} = \\ &= 852.4 \pm (1.985)(79.01)/10 = 852.40 \pm 15.68 \end{aligned}$$

Perciò l'esperimento dava c nell'intervallo di fiducia [299836.72, 299868.08].

Esercizio: Un grossista di mele è disposto a pagare un prezzo-premio al produttore purchè il diametro medio superi 2.5 pollici. Egli estrae 12 mele a caso e constata i diametri:

$$2.9 \quad 2.8 \quad 2.7 \quad 3.2 \quad 2.1 \quad 3.1 \quad 3.0 \quad 2.3 \quad 2.4 \quad 2.8 \quad 2.4 \quad 3.4$$

La media e varianza campionarie:

$$\bar{x} = 2.758, \quad s^2 = 0.1554$$

In questo caso il compratore stabilisce un intervallo di fiducia "unilatero" del tipo $(\bar{x} - \delta, +\infty]$ perchè gli da problemi un diametro piccolo, non un diametro grande:

$$1 - \alpha = P[t(11) \leq t_\alpha(11)] = P\left[\frac{\bar{x} - \mu}{s/\sqrt{n}} \leq t_\alpha(11)\right] = P\left[\mu \geq \bar{x} - t_\alpha(11) \frac{s}{\sqrt{n}}\right]$$

Se si sceglie il livello, poniamo, $\alpha = 5\%$, l'estremo sinistro di fiducia è

$$\bar{x} - t_{0.95}(11) \cdot \sqrt{s^2/n} = 2.758 - (1.7959) \cdot (0.1137) = 2.758 - 0.204 = 2.554$$

Quale proporzione p , ($0 < p < 1$) di una popolazione ha una certa proprietà, conoscendo la proporzione \hat{p} in un campione di lunghezza n ? Ebbene per n grande (cioè, empiricamente, se $n \geq 30$, $np \geq 5$) si procede così:

Intervallo di fiducia di una proporzione per n grande.

Supponiamo che la proporzione osservata in un campione di lunghezza n sia \hat{p} . In condizioni di approssimazione della binomiale alla normale, con probabilità 95% la proporzione “vera” nella popolazione è

$$p = \hat{p} \pm 1.96\sqrt{\hat{p}\hat{q}/n}.$$

Con probabilità $1 - \alpha$ la proporzione “vera” nella popolazione è

$$p = \hat{p} \pm \phi_{1-\alpha/2}\sqrt{\hat{p}\hat{q}/n}.$$

Cenno di dim. Lo stimatore è $\hat{p} = \frac{Z_n}{n}$, dove $Z_n \sim B(n; p)$; allora

$$E\left(\frac{Z_n}{n}\right) = \frac{1}{n}np, \quad Var\left(\frac{Z_n}{n}\right) = \frac{1}{n^2}Var(Z_n) = \frac{1}{n}p(1-p) \equiv .$$

Chiamando $q = 1 - p$ e usando l'approssimazione della Binomiale alla Normale,

$$Z_n \sim N\left(p, \frac{1}{n}pq\right).$$

L'intervallo di fiducia per la proporzione incognita p al livello $1 - \alpha$ è determinato dalla uguaglianza:

$$\begin{aligned} 95\% &= P\left(\frac{Z_n}{n} - \delta \leq p \leq \frac{Z_n}{n} + \delta\right) = \\ &\simeq P\left(\delta + p \geq N\left(p, \frac{p(1-p)}{n}\right) \geq p - \delta\right) = \\ (\text{standardizzando}) &= P\left(\frac{\delta}{\sqrt{pq/n}} \geq N(0, 1) \geq -\frac{\delta}{\sqrt{pq/n}}\right). \end{aligned}$$

Per sfortuna c'è non solo l'incognita δ ma anche l'incognita p . Un procedimento è mettere, al posto di p il valore sperimentale dello stimatore \hat{p} ; di per sè non è rigoroso, ma si prova che questa sostituzione è giustificabile come ottima approssimazione, specialmente se \hat{p} non è troppo vicino a 0 o ad 1. Dunque

$$\delta/\sqrt{\hat{p}\hat{q}/n} = 1.96 \equiv \phi_{0.975}.$$

Generalizzando

$$1 - \alpha = P\left(\frac{Z_n}{n} - \delta \leq p \leq \frac{Z_n}{n} + \delta\right) \implies \delta = \phi_{1-\alpha/2} \cdot \sqrt{\hat{p}\hat{q}/n}$$

□

Esercizio: di stima di una proporzione: un sondaggio preelettorale su un campione casuale di $n = 150$ votanti, ne ha trovato 84 favorevoli al candidato X. Vogliamo un intervallo di fiducia al livello $1 - \alpha$, con $\alpha = 1\%$ per la proporzione favorevole ad X. La proporzione stimata è $\hat{p} = 0.56$ ed n è grande, dunque

$$p = \hat{p} \pm \phi_{0.995} \sqrt{\frac{\hat{p}\hat{q}}{n}} =$$

$$= 0.56 \pm (2.58) \sqrt{\frac{(0.56)(1 - 0.56)}{150}} = 0.56 \pm (2.58) \cdot (0.04) = 0.56 \pm 0.103$$

Ne risulta un intervallo da 0.457 a 0.663. Notare che l'intervallo contiene valori inferiori come superiori al 50%, perciò il candidato non può predire con fiducia 99% di vincere le elezioni.

Esercizio: Prima dell'esperimento (e quindi prima di conoscere \hat{p}), si può prevedere la dimensione n del campione ai fini della precisione richiesta. Ad es. un sondaggio preelettorale cerca la proporzione di favorevoli al candidato X. Esso deve garantire un errore di stima inferiore all'1%, avendo fissato la fiducia $1 - \alpha = 95\%$. Quale dimensione di campione si richiede a tal fine?

Imponiamo che l'errore di stima sia inferiore a 0.01:

$$\phi_{1-\alpha/2} \sqrt{\hat{p}\hat{q}/n} \leq 0.01$$

Si nota che la funzione $\hat{p} \rightarrow \hat{p}(1 - \hat{p})$ ha il suo massimo in 0.5, dove ha valore 0.25; e questo consente almeno di maggiorare l'incertezza:

$$\phi_{1-\alpha/2} \sqrt{\hat{p}\hat{q}/n} \leq \phi_{0.975} \sqrt{(0.5)(0.5)/n} \leq 0.01$$

$$\sqrt{n} \geq (100) \cdot (1.96)(0.5) \implies n \geq 10^4 \cdot (1.96)^2(0.5)^2 = 9604$$

E' necessario un campione abbastanza grande a causa sia dell'alta precisione richiesta, sia dell'alta fiducia.

Osservazione: Significato dell'intervallo di fiducia: ci attendiamo che 95 volte ogni cento prove l'intervallo di fiducia contenga il "vero" parametro della popolazione. E' improprio dire " μ ha una probabilità del 95% di essere compresa ...". Infatti μ è un parametro, una costante numerica priva di errore; ciò che è aleatorio è proprio l'intervallo di fiducia, che dipende dal campione estratto. E' falso dire: "il 95% delle mele ha un diametro compreso tra..." perch'è gli estremi di fiducia informano solo sul valor medio μ della popolazione: non è possibile per questa via fare previsioni sui valori di individui o di sottogruppi.

6.5 Test delle ipotesi

Definizione. Un **test statistico** è un procedimento che consente di rifiutare oppure di non rifiutare una ipotesi concernente un parametro della popolazione in esame. Praticamente è un dispositivo per misurare la discrepanza tra il parametro che ci si attende dall'ipotesi fatta, e quanto si osserva nel campione.

Procedura del test di una ipotesi

1) Si formula una ipotesi H_o (spesso detta "ipotesi nulla") su un parametro θ della popolazione: ad es.

$$H_o : \quad \theta = \theta_o$$

[tale è un'ipotesi semplice; essa è complessa quando include più valori del parametro, ad es. $\theta \leq \theta_o$]

2) si formula un'ipotesi alternativa H_A , che nel caso più semplice può essere del tipo

$$H_A : \quad \theta = \theta_A, \quad \text{dove } \theta_A > \theta_o$$

3) si sceglie (secondo il criterio sotto esposto) una soglia o **valore critico** c tale che $\theta_o < c < \theta_A$

4) se il valore $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$ osservato a partire dal campione è minore di c , si accetta (o, meglio, non si rifiuta) l'ipotesi H_o ; se il valore osservato è maggiore o uguale a c si rifiuta l'ipotesi H_o a favore dell'ipotesi alternativa H_A .

Definizione. Errore di prima specie: viene rifiutata H_o quando è vera:

$$\hat{\theta}(x_1, \dots, x_n) \geq c, \quad H_o \text{ è vera}$$

Errore di seconda specie: non viene rifiutata H_o quando è falsa:

$$\hat{\theta}(x_1, \dots, x_n) < c, \quad H_o \text{ è falsa.}$$

Si chiama α la probabilità di errore di prima specie:

$$\alpha = P(\bar{\theta} > c)|_{\theta=\theta_o}$$

dove il simbolo $P(\dots)|_{\theta=\theta_o}$ indica che la probabilità è calcolata in base alla distribuzione di probabilità implicita nell'ipotesi H_o .

Si chiama β la probabilità di errore di seconda specie:

$$\beta = P(\bar{\theta} \leq c)|_{\theta=\theta_A}.$$

Per enunciare come si fissa il valore critico c , diamo un esempio preliminare

Esercizio: La nascita di un bimbo o di una bimba hanno davvero uguale probabilità? Per saperlo si mette a test

l'ipotesi $H_o : \theta = 1/2$; contro l'ipotesi alternativa $H_A : \theta > 1/2$.

Usiamo un campione di $n = 3000$ nascite in una certa città: in esso contiamo 1578 bimbi. Una tale eccedenza di bimbi maschi (rispetto ai circa 1500 che ci si aspetta in base ad H_o) potrebbe essere dovuta semplicemente al "caso" (il campione può essere casualmente un po' anomalo). Con che criterio posso escludere che il campione sia "anomalo" rispetto a una popolazione binomiale con $\theta \leq 1/2$?

Detta X la variabile aleatoria Binomiale "numero di bimbi maschi fra $n = 3000$ nati", noi scegliamo la soglia c in modo che, se H_o è vera, la probabilità di un campione con $X \geq c$ non superi un prefissato α abbastanza piccolo. Scegliendo ad es. $\alpha = 1\%$ (o $\alpha = 5\%$), otteniamo c dall'equazione

$$P(X \geq c)|_{\theta=0.5} = \alpha = 0.01$$

cioè

$$c \text{ tale che } \sum_{k \geq c} \binom{3000}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{3000-k} = 0.01$$

Questo calcolo è lungo perché $n = 3000$ è troppo grande. Ma per il teorema di De Moivre - Laplace possiamo approssimare X con una variabile aleatoria normale di media $\mu = n\theta = 1500$ e varianza $\sigma^2 = n\theta(1 - \theta) = 750$. Quindi ricaviamo c dall'uguaglianza

$$0.01 \simeq P\left[N(1500, 750) = c - \frac{1}{2}\right] = 1 - \Phi\left(\frac{c - 0.5 - 1500}{\sqrt{750}}\right).$$

Consultando le tavole c è tale che:

$$\frac{c - 1500.5}{\sqrt{750}} \geq 2.32 \implies c = 1564,$$

Poiché $1578 > 1564$ rifiutiamo l'ipotesi H_o .

Osservazione: L'esempio proposto dà un senso probabilistico a questa domanda cruciale: il valore osservato è discosto dal valore atteso solo per le usuali leggi del caso, o si nota una discrepanza che ragionevolmente non può essere attribuita al solo caso ?

Ebbene: il "caso" vuol dire la distribuzione di probabilità implicita nell'ipotesi H_o . Prefissiamo allora una piccola probabilità α di avere campioni "anomali", "esageratamente discosti": secondo l'ipotesi di casualità H_o , α determina la soglia critica c ,

cioè il “quanto sono discosti”. E’ vero che nell’ipotesi H_0 potrà uscire un campione anomalo, ma solo 5 volte su cento (o una volta su cento...). Eccettuata questa piccola probabilità di errore (errore di prima specie), quando $\hat{\theta} \geq c$ possiamo tranquillamente rifiutare l’ipotesi H_0 . Riassumendo: il punto (3) della procedura del test si esplicita così:

Scelta del valore critico c :

3a) si fissa un “livello di significatività” α (ad es. $\alpha = 5\%$, o 1% , o 0.1%);

3b) si sceglie c tale che la massima probabilità di errore del I tipo sia uguale ad α :

$$\max_{\theta} P(X \geq c) |_{H_0} = \alpha$$

Osservazione: Il “**livello di significatività**” α è la massima probabilità di errore del primo tipo (massima rispetto ai parametri θ inclusi nell’ipotesi H_0). Ed $1 - \alpha$ è detto “**livello di protezione**” del test. Ad es. per affermare che un nuovo motore è meno inquinante di un altro si fa l’ipotesi nulla (= i due inquinamenti sono supposti equivalenti) e si cerca di smentirla sperimentalmente. Ma non basta osservare una qualunque discrepanza dal valore atteso. Si potrà dire che il nuovo motore è meno inquinante del vecchio solo se la discrepanza dell’osservato dall’atteso è “significativa”, nel senso che $\alpha = 1\%$ o 5% .

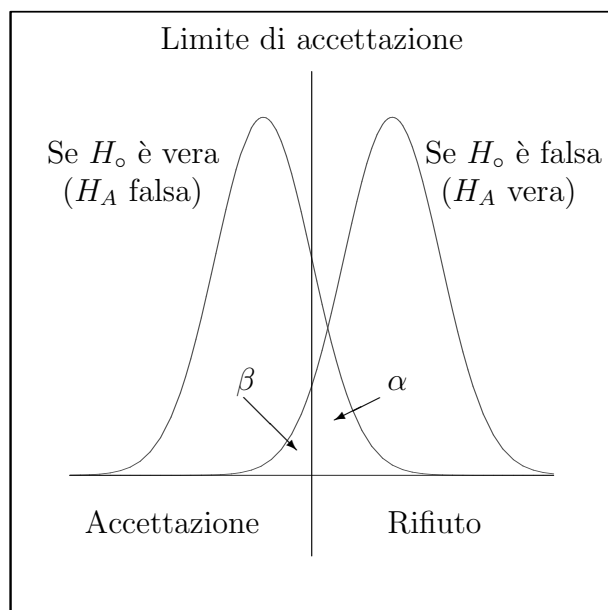
Osservazione: Quando l’esito dell’esperimento è $\hat{\theta} < c$, allora si accetta l’ipotesi H_0 . Ma c’è una probabilità di errore di seconda specie chiamata β in letteratura statistica.

Se l’ipotesi alternativa è semplice ($H_A : \theta = \theta_A$) allora β è il numero : $\beta = P(\bar{\theta} \leq c) |_{\theta=\theta_A}$. Il numero $1 - \beta$ è la “**potenza**” del test.

Osservazione: α e β **antagoniste**. Non si possono rendere contemporaneamente piccole le probabilità α e β . Per convincersene: disegnare il grafico di due densità di probabilità gaussiane, la prima con media θ_0 , la seconda con media $\theta_A > \theta_0$. Ovvio che α è l’area sottesa al primo grafico in $[c, +\infty)$. Invece β è l’area sottesa al secondo grafico in $(-\infty, c]$. Più α è piccolo, più c è spostato a destra, più è grande di conseguenza β . Analogamente, più β è piccolo, più grande risulta α . Significatività e potenza di un test sono antagoniste.

In pratica: si fissa α , si determina c , e infine si calcola β quando si ha interesse anche per la potenza. Se β è giudicato troppo grande, o si trova un compromesso alzando α , oppure si amplia il campione (aumentare n è l’unico modo di abbassare insieme α e β).

Osservazione: In letteratura si pone attenzione anche al “*valore p*”: se la regione di rigetto è del tipo $(X \geq k)$ il “*p-value*” è definito come il più piccolo livello a



cui un test di questa forma respingerebbe l'ipotesi H_0 . Nell'esempio sopra esposto, dove si osserva il valore $k = 1578$,

$$p = P(X \geq 1578)|_{H_0} = P(N(0,1) \geq \frac{1578 - 1500}{\sqrt{750}}) = 1 - 0.9978 = 0.22\%$$

Significa: questo esperimento respinge l'ipotesi H_0 anche al livello di significatività 0.22%. Se ad es. il dato sperimentale fosse stato $k = 1560$, esso almeno respingerebbe l'ipotesi nulla al livello $P(X \geq 1560) = P(N(0,1) \geq 2.19) = 1.4\%$ (e naturalmente anche ad ogni livello meno fine, come $\alpha = 2\%$, $\alpha = 5\%$, ecc.).

Osservazione: Se l'ipotesi alternativa è complessa (ad es. **ad una coda**, cioè $H_A : \theta > \theta_0$; oppure **a due code**, cioè $H_A : \theta$ diverso da θ_0) allora $\beta = \beta(\theta)$ non è un singolo numero ma una funzione. La funzione $\beta = \beta(\theta)$ è la curva operativa caratteristica e la funzione $\theta \rightarrow 1 - \beta(\theta)$ è detta curva di potenza. Ad esempio nei valori alternativi $\theta = 0.52$ e $\theta = 0.51$ la potenza del test (=probabilità di respingere effettivamente H_0 quando è falsa) è rispettivamente:

$$\begin{aligned} 1 - \beta(0.52) &= 1 - P(X \leq 1564)|_{\theta=0.52} = \\ &= 1 - P(N(0,1) \leq \frac{1564 - 3000 \cdot (0.52)}{\sqrt{3000 \cdot (0.52)(0.48)}}) = 44.44\% \\ 1 - \beta(0.51) &= 1 - P(X \leq 1564)|_{\theta=0.51} = \\ &= 1 - P(N(0,1) \leq \frac{1564 - 3000 \cdot (0.51)}{\sqrt{3000 \cdot (0.51)(0.49)}}) = 10.8\% \end{aligned}$$

Così la curva di potenza del test è tanto migliore quanto più è concentrata la sua zona di pendenza (il che avviene al crescere di n , fissato α).

6.6 Test eseguiti su un solo campione

Questi test sono fatti su un singolo campione, tratto da una popolazione gaussiana.

Test su una media. Per un campione x_1, \dots, x_n di legge normale consideriamo il test dell'ipotesi

$$H_0 : \mu = \mu_0 \text{ contro } \mu > \mu_0$$

Se il livello di significatività è α , una regione critica o di rigetto è:

$$t \geq t_{1-\alpha}(n-1), \quad \text{dove} \quad t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

Analogamente, se si vuole un test "bilatero" nel senso che

$$H_0 : \mu = \mu_0 \text{ contro } H_A : \mu \neq \mu_0,$$

una regione di rigetto è

$$\{ |t| \geq t_{1-\alpha/2}(n-1) \}$$

Dimostrazione: Se μ_0 è il valore della media, la statistica

$$t = \sqrt{n}(\bar{x} - \mu_0)/s$$

segue una legge di Student $t(n-1)$. Dunque, se vale l'ipotesi H_0 ,

$$\alpha = P\left(\frac{\bar{x} - \mu_0}{s/\sqrt{n}} \geq t_{1-\alpha}(n-1)\right)$$

In conclusione l'evento

$$\{t \geq t_{1-\alpha}(n-1)\}$$

è una regione di rigetto di livello α : si calcola t a partire dall'osservazione; se il valore sperimentale è maggiore o uguale al valore critico $t_{1-\alpha}(n-1)$ l'ipotesi è respinta al livello α ; altrimenti non c'è sufficiente evidenza per respingere H_0 .

È poi analoga la costruzione del test bilatero, come enunciato. \square

Esercizio: Il tempo medio di vita di una lampadina prodotta da una fabbrica era di 169 ore. Dopo avere installato un nuovo apparecchio nella catena di montaggio, si vuole sapere se si è allungato tale tempo medio. Si campionano 121 lampadine prodotte trovando tempi di vita con media e varianza campionarie:

$$\bar{x} = 171, \quad s^2 = 85$$

Si è allungata la vita media?

Si tratta di fare un test per l'ipotesi

$$H_o : \mu = \mu_o \text{ contro } \mu > 169$$

ad esempio con livello $\alpha = 5\%$. Allora il valore sperimentale

$$t = \frac{\bar{x} - 169}{s/\sqrt{n}} = \frac{2}{\sqrt{85}/\sqrt{121}} = 2.38$$

è superiore al valore critico

$$t_{0.95}(120) = 1.657 .$$

L'ipotesi che la vita media non sia aumentata è dunque respinta al livello $\alpha = 5\%$; anzi viene respinta anche al livello 1% , dato che $t_{0.99}(120) = 2.357$. Più precisamente il “valore p ” (=il più fine livello di significatività con cui H_o viene respinta in base a questi dati) è $0.0094 \equiv 0.94\%$ (calcolato con un software, perché le tavole della t di Student non danno questa informazione).

Test su una proporzione per n grande.

Supponiamo di voler saggiare il parametro p di una popolazione Bernoulliana. Si estrae il campione

$$x_i \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix} \quad i = 1, \dots, n$$

di rango n abbastanza grande. Sia

$$\hat{p} = Z_n/n, \quad \text{con } Z_n = x_1 + \dots + x_n$$

la proporzione stimata di “successi”. Allora, al livello α , il test bilatero

$$H_o : p = p_o \text{ contro } H_A : p \neq p_o$$

ha una regione di rigetto data da:

$$\left\{ \left| \frac{\hat{p} - p_o}{\sqrt{p_o(1-p_o)/n}} \right| \geq \phi_{1-\alpha/2} \right\}.$$

Analogamente si costruiscono i test a una coda.

Dimostrazione: Poiché $Z_n \sim B(np_o, np_o(1-p_o))$,

$$E\left[\frac{Z_n}{n}\right] = p_o, \quad \text{Var}\left[\frac{Z_n}{n}\right] = \frac{1}{n^2}np_o(1-p_o)$$

Standardizzando e applicando il teorema di De Moivre-Laplace

$$\frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim N(0, 1) \text{ per } n \rightarrow \infty$$

[Criterio pratico: $n > 30$ con np_0 o $nq_0 > 5$]. Dunque la regione di rigetto di livello α si ottiene così:

$$\begin{aligned} \alpha &= P[N(0, 1) \leq \phi_{\alpha/2} \text{ ovvero } N(0, 1) \geq \phi_{1-\alpha/2}] \\ &\simeq P\left[\left| \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \right| \geq \phi_{1-\alpha/2} \right] \end{aligned}$$

La regione di rigetto per il test a una coda si trova facilmente con analoghe considerazioni. \square

Esercizio: Per la legge 626 sulla sicurezza un'azienda deve garantire che i dipendenti lavorino esposti a certe superficie radianti meno del 3% del tempo lavoro. Si fanno 1000 controlli in istanti casuali e si trova il generico dipendente esposto 27 volte su 1000. Si può affermare che la proporzione di esposizione è inferiore al 3%?

Il test è a una coda:

$$H_0 : p = 0.03 \text{ contro } p < 0.03$$

Potremo rigettare $p = 0.03$ al livello α se la proporzione stimata \hat{p} è sufficientemente piccola:

$$\begin{aligned} \alpha &= P(N(0, 1) \leq \phi_\alpha) \simeq P\left[\frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \leq \phi_\alpha \right] \\ \hat{p} &\leq p_0 + \phi_\alpha \sqrt{p_0(1 - p_0)/n} = (0.03) + \phi_\alpha \sqrt{(0.03)(0.97)/1000} \end{aligned}$$

cioè, prendendo $\alpha = 5\%$,

$$\hat{p} \leq 0.03 + (-1.65)(0.00539) = 0.03 - 0.0089 = 0.0211$$

Poiché $\hat{p} = 0.027$ non è in tale regione di rigetto, l'azienda non può dimostrare che i dipendenti siano al sicuro. [Si noti infine che potremmo far perno sulla Z_n anziché su $\hat{p} = Z_n/n$: imporre $\alpha = P(Z_n \leq k) \simeq P(N(np, npq) \leq k + \frac{1}{2}) = \dots$ infine trovando $k \leq 20.65$. Così terremmo conto anche della (lieve) correzione $\pm \frac{1}{2}$ dovuta al passaggio dalla variabile aleatoria discreta alla continua].

Test su proporzioni mediante chi quadro.

Si è già illustrato il test su una proporzione: si tratta di fare un confronto tra proporzione attesa (ad es. $p = 1/2$ per le nascite di bimbi,...) e proporzione \hat{p} osservata.

Più in generale, possono esserci piú sottogruppi della popolazione da considerare. Inoltre limitiamo l'attenzione a grandi campioni in modo da applicare alla binomiale o alla multinomiale un'approssimazione di tipo normale. Ebbene, se con una sola proporzione indipendente è equivalente la scelta tra normale e $\chi^2(1)$, più generale si deve usare la χ^2 con tanti GL quante sono le proporzioni indipendenti:

Proposizione. Sia $H_o : \pi_1, \dots, \pi_r$ una ipotesi sulle proporzioni di r sottogruppi in una popolazione, in modo che $\pi_1 + \dots + \pi_r = 1$. Siano f_o le frequenze assolute osservate in un campione di dimensione n , in modo che $f_{o,1} + \dots + f_{o,r} = n$. Siano f_a le frequenze attese in base ad H_o , tali che $f_{a,1} = n\pi_1, \dots, f_{a,r} = n\pi_r$. Allora, per n grande, la statistica

$$\sum \frac{(f_o - f_a)^2}{f_a} = \chi_{r-1}^2$$

segue una distribuzione χ^2 con $r - 1$ GL.

Idea. Ci proponiamo di dare l'idea mediante il caso $r = 2$. Abbiamo:

$$f_o : x, n - x; \quad f_a : n\pi, n(1 - \pi).$$

Sostituendo e ricordando che in una distribuzione di Bernoulli $\sigma^2 = n\pi(1 - \pi)$, un breve calcolo mostra:

$$\sum \frac{(f_o - f_a)^2}{f_a} \equiv \frac{(x - n\pi)^2}{n\pi} + \frac{[n - x - n(1 - \pi)]^2}{n(1 - \pi)} = \frac{(x - n\pi)^2}{n\pi(1 - \pi)} = \frac{(x - \mu)^2}{\sigma^2}$$

che è a rigore la variabile aleatoria di Bernoulli standardizzata al quadrato. Per n grande possiamo interpolarla con la Z^2 , che equivale a $\chi^2(1)$ per definizione di chi quadro. \square

Esercizio: Nei piselli due geni controllano il colore (giallo o verde, con proporzione attesa 3/4 e 1/4) e la forma (liscio o non liscio con proporzione attesa 3/4 e 1/4). Se operano indipendentemente mi attendo:

$$(3/4)(3/4) = 9/16 \text{ dei semi lisci e gialli}$$

$$(3/4)(1/4) = 3/16 \text{ lisci e verdi, } 3/16 \text{ grinzosi e gialli, } 1/16 \text{ gr.e verdi}$$

In $n = 100$ semi osservo: 60 lisci e gialli, 21 lisci e verdi, 17 grinzosi e gialli, 2 grinzosi e verdi. Il modello genetico si attende rispettivamente:

$$f_{a,1} = 100 \cdot 9/16 = 56.25 \quad f_{a,2} = f_{a,3} = 18.75 \quad f_{a,4} = 6.25.$$

Abbiamo 4 scarti, di cui tre indipendenti:

$$\chi^2 = \frac{(60 - 56.25)^2}{56.25} + \frac{(21 - 18.75)^2}{18.75} + \frac{(17 - 18.75)^2}{18.75} + \frac{(2 - 6.25)^2}{6.25} = 3.57$$

che risulta minore di $\chi_{0.95}^2(3) = 7.81$. Non c'è motivo di respingere il modello mendeliano.

6.7 Test eseguiti su due campioni

Confronto tra 2 medie, con varianza incognita

Supponiamo di avere estratto due campioni, ciascuno da una popolazione normalmente distribuita: la prima con media μ_1 e varianza σ^2 , la seconda con media μ_2 e varianza uguale alla precedente. I due campioni sono indipendenti con lunghezze n_1, n_2 e con medie campionarie \bar{x}, \bar{y} . Spesso si vuole sapere se le due medie sono significativamente diverse, e quasi sempre è incognita la varianza comune alle due popolazioni, cioè σ^2 . A tal fine si mette a test l'ipotesi nulla

$$H_o : \mu_1 = \mu_2, \quad (\text{ossia } \mu_1 - \mu_2 = 0)$$

contro l'ipotesi alternativa

$$H_A : \mu_1 \text{ diverso da } \mu_2, \quad (\text{ossia } \mu_1 - \mu_2 \text{ diverso da } 0).$$

Confronto tra due medie da campioni gaussiani indipendenti L'ipotesi di tutti i nostri modelli che coinvolgono due o più popolazioni è l'omogeneità delle varianze. Detta σ^2 la varianza comune, essa è stimata dalla varianza mediata ("pooled variance"):

$$\bar{s}^2 = [(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2]/(n_1 + n_2 - 2).$$

Per costruire la t di Student a partire dalla variabile aleatoria normale $\bar{x} - \bar{y}$ ricordiamo:

$$Var[\bar{x}] = Var\left[\frac{1}{n_1}(x_1 + \dots + x_{n_1})\right] = \left(\frac{1}{n_1}\right)^2 \cdot n_1 \sigma^2 = \sigma^2/n_1$$

e analogamente $Var(\bar{y}) = \sigma^2/n_2$. Per l'indipendenza dei due campioni le due varianze si sommano:

$$\sigma_{\bar{x}-\bar{y}}^2 = Var(\bar{x} - \bar{y}) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Da ciò deriva lo stimatore corretto per $\sigma_{\bar{x}-\bar{y}}^2$. Ebbene, posto $s_{\bar{x}-\bar{y}} = \sqrt{\bar{s}^2/n_1 + \bar{s}^2/n_2}$, la differenza tra medie campionarie segue la t di Student con $\nu = n_1 + n_2 - 2$ gradi di libertà:

$$t(\nu) = [\bar{x} - \bar{y} - (\mu_1 - \mu_2)]/s_{\bar{x}-\bar{y}}$$

Nel confronto di due medie si può allora usare la t di Student come enunciato, sia impostando un test bilaterale ($H_A : \mu_1$ diverso da μ_2) sia impostando un test unilaterale ($H_A : \mu_1 > \mu_2$ o viceversa).

Esercizio: Si vogliono confrontare, al livello $\alpha = 1\%$, le altezze di certi cespugli in due aree A ed M . Il campione A di dimensione $n_1 = 20$ dà una media $\bar{x} = 1.686$ con varianza campionaria $s_x^2 = 0.2658$. Il campione M ha $n_2 = 26$, $\bar{y} = 1.3215$ ed $s_y^2 = 0.1407$. La varianza mediata è

$$\frac{19 \cdot (0.2658) + 25 \cdot (0.1407)}{20 + 26 - 2} = \frac{5.0495 + 3.5175}{44} = 0.1947$$

e la statistica $T_{sper.}$ è:

$$T_{sper.} = \frac{1.6860 - 1.3215}{\sqrt{(0.1947) \cdot (\frac{1}{20} + \frac{1}{26})}} = 2.778$$

Per fare un test bilaterale (quello con il minimo di assunzioni), la T sperimentale deve essere confrontata col quantile di ordine 0.995 della t di Student con 44 GL: è il quantile che esclude l'1% complessivamente, 0.5% a destra e 0.5% a sinistra. Nelle tavole non troviamo i 44 GL, ma solo 40 o 60 GL. La soluzione è avvalersi dei più vicini inferiori GL, in questo caso 40, che garantiscono un test più protetto (avremo un livello leggermente più fine dell'1%). Poiché il t sperimentale è superiore al t critico

$$T_{sper.} \equiv 2.778 \geq t_{0.995}(40) \equiv 2.7045,$$

l'ipotesi di differenza nulla $\mu_1 - \mu_2$ viene rifiutata con fiducia 99% a favore dell'ipotesi alternativa a due code $\mu_1 \neq \mu_2$.

Invece un test unilaterale pone come ipotesi alternativa $\mu_1 > \mu_2$. È come fare un'assunzione in più, giustificata ad es. dalle conoscenze o da decisioni del ricercatore: nessuno, nemmeno la statistica, insegna al ricercatore a formulare le ipotesi. Allora basta il confronto col quantile $t_{0.99}(40) = 2.4233$, il quale non a caso è meno esigente.

Osservazione: L'ipotesi che le 2 popolazioni abbiano la stessa varianza si controlla con un test F di Fisher sul rapporto tra le due varianze campionarie (v. sopra il teorema sulla F). Si mette al numeratore la varianza più grande e si effettua il test bilatero perché non c'è ragione per assumere che una varianza sia maggiore dell'altra.

Nell'esempio, non c'è motivo di rifiutare l'omogeneità delle varianze perché

$$s_y^2/s_x^2 = 0.2658/0.1407 = 1.8891$$

risulta minore di $F_{0.95}(n_1 - 1, n_2 - 1) \equiv F_{0.95}(19; 25) = 2.01$.

Osservazione: Più semplice è il confronto di due medie per dati appaiati ("paired values"), quindi da due campioni non indipendenti. Un esempio è rilevare dati

prima e dopo la somministrazione di un farmaco sullo stesso gruppo di n pazienti, per rispondere a domande del tipo: il farmaco ha realmente efficacia? Qui non c'è bisogno della statistica t “mediata”; basta saggiare la differenza di medie usando la varianza campionaria delle differenze $z_k = x_k - y_k$, che sono supposte gaussiane:

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{s_d^2/n}} \sim t(n-1), \quad s_z^2 = \sum_1^n \frac{1}{n-1} (x_k - y_k)^2.$$

7 PROPAGAZIONE DELL'ERRORE

La misurazione un'azione frequente e fondamentale per i lavori sviluppati in ambito scientifico. Spesso, inoltre, scienziati e ingegneri devono fare dei calcoli a partire da quantità misurate (ad esempio quando devono calcolare l'area di una superficie rettangolare moltiplicando la misura della base per quella dell'altezza oppure quando devono calcolare la densità di un oggetto dividendo la misura della massa per la misura del volume dell'oggetto stesso). Ogni volta che si effettua una misurazione si compiono degli errori (in quanto ogni procedura di misurazione contiene degli errori); inoltre quando si effettua un calcolo che utilizza delle quantità misurate, gli errori della misurazione producono errori nei valori calcolati. Si dice, quindi, che l'errore si è **propagato**. Esistono dei metodi che permettono di conoscere l'errore compiuto sul valore calcolato, conoscendo l'errore sulle singole misurazioni di partenza. Ad esempio se si conoscono gli errori compiuti sulle misurazioni della base e dell'altezza della superficie rettangolare allora tali metodi permettono di ottenere l'entità dell'errore sull'area. La presentazione di tali metodi, di cui si occupa la teoria della propagazione degli errori, sarà l'argomento di questo capitolo.

7.1 Errori di misurazione

Un geologo pesa una pietra su una bilancia per cinque volte consecutive ed ottiene le seguenti misurazioni in grammi: 151.5, 152.3, 150.4, 151.8, 150.1. Le cinque misurazioni sono tutte diverse e verosimilmente nessuna di esse sarà il valore vero della massa della pietra. La differenza tra un valore misurato ed il valore vero viene detto **errore** nel valore misurato. Ciascuna procedura di misurazione contiene diverse sorgenti di errore. Ad esempio, immaginiamo che le misurazioni del peso siano effettuate con una bilancia non elettronica; se la bilancia non è stata calibrata correttamente, allora ogni misurazione si discosterà dal valore vero di una quantità fissa. Inoltre, l'interpolazione tra i valori che si leggono sulla bilancia rappresenta un'altra fonte d'errore. La grandezza di questo secondo tipo di errore solitamente varia da una misurazione all'altra e può essere sia positiva che negativa. Per questo motivo è ragionevole pensare che gli errori dovuti all'interpolazione abbiano media nulla.

In generale, gli errori di misurazione sono composti da due parti, l'**errore sistematico** (o **distorsione** o **bias**) e l'**errore casuale**. Mentre la distorsione influenza ogni misurazione di una stessa quantità, l'errore casuale varia da misurazione a misurazione e la sua media, per un numero elevato di misurazioni, è nulla. Alcune cause d'errore possono contribuire sia all'errore sistematico che all'errore casuale. Si pensi, per esempio all'errore della parallasse (fenomeno per cui un oggetto sembra

spostarsi rispetto allo sfondo se si cambia il punto di osservazione). Se si guarda da punti distinti il quadrante della bilancia la grandezza dell'errore di parallasse dipende dalla posizione in cui si trova l'osservatore, rispetto al quadrante e poiché la posizione varia da osservatore ad osservatore, la parallasse contribuirà all'errore casuale. Inoltre, se l'osservatore si inclina da un lato piuttosto che da un'altro, allora la parallasse contribuirà all'errore sistematico.

Ogni misurazione può essere considerata come la somma del valore vero più il contributo delle due componenti di errore:

$$\text{Valore misurato} = \text{valore vero} + \text{distorsione} + \text{errore casuale}$$

Dal momento che una componente dell'errore casuale, appropriato usare un modello statistico per studiare l'errore di misura. Pertanto si modellerà ogni valore misurato come una variabile casuale proveniente da una popolazione composta da tutte le misurazioni. La media μ della popolazione rappresenta quella parte di misurazione che è comune ad ogni misurazione. Quindi il valore μ è la somma del vero valore e della distorsione. La deviazione standard σ della popolazione è la deviazione standard dell'errore casuale e rappresenta la variabilità dovuta al fatto che ogni misurazione presenta un valore diverso per l'errore casuale (si può dire intuitivamente che σ rappresenta la grandezza dell'errore casuale. Siamo interessati a due aspetti del processo di misurazione: l'**accuratezza** e la **precisione**. L'accuratezza è determinata dalla distorsione ossia dalla differenza tra la misura media μ ed il vero valore:

$$\text{Distorsione} = \mu - \text{valore vero}$$

Più è piccola la distorsione, più accurato sarà il processo di misurazione (se $\mu = 0$ il processo di misurazione è non distorto). La precisione si riferisce a quanto le misurazioni ripetute sulla stessa quantità tendono ad essere uguali. Se le misurazioni ripetute sono più o meno le stesse ogni volta, allora la precisione del processo di misurazione è alta, altrimenti se esse sono molto differenti tra loro la precisione è bassa. La precisione è quindi determinata dalla deviazione standard σ del processo di misurazione. Quanto più σ è piccolo, tanto più sarà preciso il processo di misurazione. Ci si riferisce a σ come all'**incertezza statistica** o più semplicemente all'**incertezza**.

Nella realtà, solitamente, non si conosce il vero valore che deve essere misurato e quindi in generale non si è in grado di dire nulla sulla distorsione. Mentre l'incertezza può essere stimata tramite misurazioni ripetute, per stimare la distorsione occorre conoscere altre informazioni sul valore vero. In generale, se X_1, \dots, X_n sono n misurazioni *indipendenti* di una stessa quantità ottenute tutte dallo stesso processo di misurazione allora:

- La deviazione standard campionaria s può essere utilizzata per stimare l'incertezza.
- Le stime dell'incertezza sono spesso approssimate, soprattutto quando sono effettuate su singoli campioni.
- Se il vero valore è noto, la media campionaria \bar{X} può essere utilizzata per stimare la distorsione (Distorsione = \bar{X} - valore vero).
- Se il vero valore non è noto, la distorsione non può essere stimata dalle misurazioni.

Da qui in avanti, si supporrà che la distorsione sia stata resa trascurabile (ad esempio con qualche processo di calibratura degli strumenti di misurazione). Le misurazioni saranno allora descritte nella forma:

$$\text{valore misurato} \pm \sigma \quad (14)$$

dove σ rappresenta l'incertezza del processo di misurazione.

7.2 Combinazioni lineari di misurazioni

È possibile trasformare le misurazioni ottenute facendone una combinazione lineare. L'incertezza influenza questo tipo di operazioni matematiche. Se X è una misurazione e c è una costante, allora

$$\sigma_{cX} = |c|\sigma_X. \quad (15)$$

Se X_1, \dots, X_n sono n misurazioni indipendenti e c_1, \dots, c_n sono n costanti, allora

$$\sigma_{c_1X_1 + \dots + c_nX_n} = \sqrt{c_1^2\sigma_{X_1}^2 + \dots + c_n^2\sigma_{X_n}^2}. \quad (16)$$

Tali proprietà si deducono dalle proprietà per le combinazioni lineari di variabili casuali indipendenti (vedi cap. 4).

Esempio: Il raggio di un cerchio misura 3.0 ± 0.1 cm. Si stimi la lunghezza della circonferenza e si trovi l'incertezza della stima.

Sia R il raggio del cerchio. Il suo valore misurato è 3.0 cm e l'incertezza è la deviazione standard di tale misurazione, cioè $\sigma_R = 0.1$ cm. La circonferenza è data da $C = 2\pi R$. L'incertezza in C è la deviazione standard di C , σ_C . Dato che 2π è una costante si ottiene, per la (15)

$$\sigma_C = 2\pi\sigma_R = (6.28)(0.1\text{cm}) = 0.63\text{cm}$$

Da cui si ha che la lunghezza della circonferenza misurerà 18.84 ± 0.63 cm.

Esempio: Un geometra sta misurando il perimetro di un terreno rettangolare. Misura i due lati adiacenti che sono 50.11 ± 0.05 m e 75.21 ± 0.08 m. Queste misurazioni sono indipendenti. Si stimi il perimetro del terreno e si calcoli l'incertezza di tale stima.

Siano $X = 50.11$ e $Y = 75.21$ le due misurazioni. Il valore del perimetro è dato da $P = 2X + 2Y = 250.64$ e, dall'eq. (16), l'incertezza in P è:

$$\sigma_P = \sigma_{2X+2Y} = \sqrt{4\sigma_X^2 + 4\sigma_Y^2} = \sqrt{4(0.05)^2 + 4(0.08)^2} = 0.19m \quad (17)$$

Il perimetro è dunque 250.64 ± 0.19 m.

7.2.1 Misure ripetute

Il modo migliore per ridurre l'incertezza è quello di utilizzare diverse misurazioni indipendenti e farne la media. Le misurazioni in questo caso formano un campione casuale semplice e la loro media è la media campionaria. Se X_1, \dots, X_n sono n misurazioni *indipendenti* ognuna con media μ e incertezza σ , allora la loro media campionaria \bar{X} è una misura avente media

$$\mu_{\bar{X}} = \mu$$

ed incertezza

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}. \quad (18)$$

Questo risultato ci dice che se si effettuano tante misurazioni indipendenti sulla stessa quantità, la media delle diverse misurazioni ripetute ha la stessa accuratezza, ma è più precisa di una singola misurazione.

Esempio: La lunghezza di un componente è stato misurato attraverso un processo, la cui incertezza è data da 0.05 cm. Se vengono effettuate 25 misurazioni indipendenti, quanto varrà tale incertezza? Quanto è precisa la media delle 25 misurazioni rispetto ad una singola misurazione?

L'incertezza della media delle 25 misurazioni è $0.05/\sqrt{25} = 0.01$ cm. L'incertezza di una singola misurazione è 0.05 cm. Quindi la media delle 25 misurazioni indipendenti è cinque volte più precisa di una singola misurazione.

Esempio: La massa di una pietra è misurata 5 volte da una bilancia, della quale non è nota l'incertezza. Le cinque misurazioni (in grammi) sono 21.10, 21.05, 20.98, 21.12 e 21.05. Si stimi la massa della pietra e si trovi l'incertezza nella stima.

Se \bar{X} è la media delle 5 misurazioni e s la deviazione standard campionaria, si ha $\bar{X} = 21.06$ g e $s = 0.0543$ g. Usando l'eq. (18) si ottiene $\bar{X} \pm \frac{\sigma}{\sqrt{5}}$ ma il valore di σ , ossia dell'incertezza, non è noto. Esso può essere comunque stimato con s . Quindi la stima della massa della pietra è $21.06 \pm 0.0543/\sqrt{5}$, ovvero 21.06 ± 0.02 g.

7.2.2 Misurazioni indipendenti con valori differenti per l'incertezza

A volte le misurazioni ripetute possono avere valori distinti per quanto riguarda l'incertezza. Ciò può accadere quando le misurazioni vengono effettuate attraverso strumenti diversi. In questo caso occorre combinare le misurazioni attraverso una media ponderata (anziché con la media campionaria semplice).

Se X e Y sono misurazioni *indipendenti* della stessa quantità, ma con valori di incertezza σ_X e σ_Y , rispettivamente, allora la media pesata di X e Y con l'incertezza più piccola è data da $c_{best}X + (1 - c_{best})Y$, dove

$$c_{best} = \frac{\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2} \quad 1 - c_{best} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2}$$

7.2.3 Combinazione lineare di misurazioni dipendenti

Se X ed Y sono misurazioni le cui incertezze sono σ_X e σ_Y e si vuole calcolare l'incertezza di $X + Y$ quando X e Y sono dipendenti allora l'incertezza della somma pu essere più grande o più piccola di quella che si otterrebbe nel caso in cui le misurazioni fossero indipendenti, e non può essere determinata soltanto conoscendo σ_X e σ_Y . La quantità che misura la relazione tra gli errori casuali di X e Y è denominata covarianza. In pratica, quando le misurazioni sono dipendenti, si riesce a calcolare un limite superiore dell'incertezza di una combinazione lineare di esse. Se X_1, \dots, X_n sono misurazioni e c_1, \dots, c_n sono costanti, allora

$$\sigma_{c_1X_1 + \dots + c_nX_n} \leq |c_1|\sigma_{X_1} + \dots + |c_n|\sigma_{X_n} \quad (19)$$

L'espressione nella parte a destra della disuguaglianza è una stima conservativa dell'incertezza in $c_1X_1 + \dots + c_nX_n$

Esempio: Un geometra sta misurando il perimetro di un terreno rettangolare. Misura i due lati adiacenti che sono 50.11 ± 0.05 m e 75.21 ± 0.08 m. Queste misurazioni non sono necessariamente indipendenti. Si trovi una stima conservativa dell'incertezza del perimetro del lotto.

Se indichiamo le due misurazioni con X_1 e X_2 , avremo che le incertezze sono $\sigma_{X_1} = 0.05$ e $\sigma_{X_2} = 0.08$ e il perimetro è dato da $P = 2X_1 + 2X_2$. Usando la disuguaglianza (19) si ottiene

$$\sigma_P = \sigma_{2X_1 + 2X_2} \leq 2\sigma_{X_1} + 2\sigma_{X_2} = 2(0.05) + 2(0.08) = 0.26 \text{ m} \quad (20)$$

l'incertezza del perimetro è inferiore o uguale a 0.26 m. Nell'esempio presentato nel paragrafo 7.2, l'incertezza risultava essere uguale a 0.19 m (ma in quel caso X e Y erano indipendenti).

7.3 Incertezza per funzioni di una misurazione

Se il raggio R di una circonferenza è 5.00 ± 0.01 cm, quanto vale l'incertezza dell'area $A = \pi R^2$? Ossia, in termini statistici, dato che la deviazione standard σ_R è 0.01 cm, è possibile calcolare la deviazione standard di A dove A è una funzione di R ? In generale: data una variabile casuale X , con deviazione standard nota σ_X e una funzione $U = U(X)$, come può essere calcolata la deviazione standard σ_U ?

Se X una misura la cui incertezza σ_X è piccola, e se U è una funzione di X , allora

$$\sigma_U \approx \left| \frac{dU}{dX} \right| \sigma_X \quad (21)$$

In pratica, si valuta la derivata $\frac{dU}{dX}$ per la misura osservata X . Questa è nota come la formula della **propagazione dell'errore**. Le funzioni non lineari sono generalmente distorte. Comunque se la misura X è non distorta e l'incertezza σ_X è piccola, la distorsione di U può essere ignorata.

Esempio: Il raggio R di una circonferenza è 5.00 ± 0.01 cm. Si stimi l'area del cerchio e si valuti l'incertezza di questa stima.

L'area A è data da $A = \pi R^2$. La stima dell'area del cerchio è $A = \pi(5.00\text{cm})^2 = 78.5\text{cm}^2$. Calcoliamo la derivata $\frac{dA}{dR} = 2\pi R = 10\pi$ cm. Poiché $\sigma_R = 0.01$ cm, possiamo calcolare l'incertezza di A :

$$\sigma_A = \left| \frac{dA}{dR} \right| \sigma_R = (10\pi\text{cm})(0.01\text{cm}) = 0.31\text{cm}^2$$

La stima dell'area del cerchio è quindi 78.5 ± 0.3 cm².

Esempio: Il raggio R di una sfera è 3.00 ± 0.001 cm. Si stimi il volume della sfera e si valuti l'incertezza di tale stima.

Il volume V della sfera è dato da $V = 4\pi R^3/3$. La stima di V è $V = 4\pi(3.00\text{cm})^3/3 = 113.097$ cm³. Calcoliamo la derivata $\frac{dV}{dR} = 4\pi R^2 = 36\pi$ cm². Poiché $\sigma_R = 0.001$ cm, possiamo calcolare l'incertezza di V :

$$\sigma_V = \left| \frac{dV}{dR} \right| \sigma_R = (36\pi\text{ cm}^2)(0.001\text{ cm}) = 0.113\text{ cm}^3$$

La stima del volume della sfera è quindi 113.097 ± 0.113 cm³.

7.3.1 Incertezza relativa per funzioni di una misurazione

Se U è una misurazione, il cui vero valore è μ_U e la cui incertezza è σ_U , l'incertezza relativa di U è la quantità $\frac{\sigma_U}{\mu_U}$.

L'incertezza relativa è un numero puro, cioè senza unità di misura, ed è di solito espresso in percentuale. Poiché nella pratica μ_U non è nota, si stima l'incertezza relativa con $\frac{\sigma_U}{U}$.

Ci sono due metodi per approssimare l'incertezza relativa σ_U/U della funzione $U = U(X)$:

1. Calcolare σ_U usando l'eq. (21) e quindi dividerla per U .
2. Calcolare $\ln U$ ed usare l'eq. (21) per trovare $\sigma_{\ln U}$ che è uguale a σ_U/U .

Entrambi i metodi valgono per ogni funzione U . La scelta di quale usare dipende da quale funzione sia più semplice da derivare tra U e $\ln U$.

Esempio: Trovare l'incertezza relativa nell'esercizio della sfera di raggio 3.00 ± 0.001 .

Si è visto che il volume della sfera è $113.097 \pm 0.113 \text{ cm}^3$. L'incertezza assoluta è dunque $\sigma_V = 0.113 \text{ cm}^3$ e l'incertezza relativa è $\sigma_V/V = 1.056/113.097 = 0.001$. Dunque il volume è $V = 113.097 \pm 0.1\%$

Se non si fosse già calcolata σ_V , sarebbe stato più semplice calcolare l'incertezza relativa calcolando l'incertezza assoluta di $\ln V$. Dato che $\ln V = \ln(\frac{3}{4}\pi) + 3 \ln R$, allora $d \ln V/dR = 3/R = 0.1$.

L'incertezza relativa in V è dunque

$$\frac{\sigma_V}{V} = \sigma_{\ln V} = \left| \frac{d \ln V}{dR} \right| \sigma_R = (0.1)(0.001) = 0.1\%$$

7.4 Incertezze per funzioni di più misurazioni

Spesso si ha bisogno di stimare una funzione di diverse misurazioni. Per esempio, si misuri la massa m ed il volume V di una pietra e si calcoli la densità come $D = m/V$. Potrebbe essere necessario stimare l'incertezza di D quando sia m che V siano misurate con incertezza

Se X_1, X_2, \dots, X_n sono misure *indipendenti* le cui incertezze $\sigma_{X_1}, \sigma_{X_2}, \dots, \sigma_{X_n}$ sono piccole e se $U = U(X_1, X_2, \dots, X_n)$ è una funzione di X_1, X_2, \dots, X_n , allora

$$\sigma_U \approx \sqrt{\left(\frac{\partial U}{\partial X_1}\right)^2 \sigma_{X_1}^2 + \dots + \left(\frac{\partial U}{\partial X_n}\right)^2 \sigma_{X_n}^2} \quad (22)$$

Nella pratica le derivate parziali sono valutate nel punto (X_1, X_2, \dots, X_n) . L'equazione precedente (22) è nota con il nome di **formula della propagazione dell'errore di tipo multivariato**. È importante notare che essa è valida solo quando le misure X_1, X_2, \dots, X_n sono indipendenti

Esempio: I lati di un rettangolo sono $X = 2.0 \pm 0.1$ cm e $Y = 3.2 \pm 0.2$ cm. Si calcoli l'incertezza assoluta dell'area $A = XY$.

Le derivate parziali di A sono

$$\frac{\partial A}{\partial X} = Y = 3.2 \quad \frac{\partial A}{\partial Y} = X = 2.0$$

e quindi l'incertezza assoluta di A sarà, per l'eq. (22),

$$\sigma_A = \sqrt{(3.2)^2(0.1)^2 + (2.0)^2(0.2)^2} = \sqrt{0.1024 + 0.16} = 0.5122.$$

7.4.1 Incertezza per funzioni di misure dipendenti

Se X_1, X_2, \dots, X_n non sono misure indipendenti, l'incertezza di una funzione $U = U(X_1, X_2, \dots, X_n)$ può essere stimata solo se è nota la covarianza di ogni coppia (X_i, X_j) (la nozione di covarianza è stata introdotta nel paragrafo 4.3). In molte situazioni le covarianze non sono note; in questi casi si può allora calcolare una stima conservativa dell'incertezza di U .

Se X_1, X_2, \dots, X_n sono misure non indipendenti, le cui incertezze $\sigma_{X_1}, \sigma_{X_2}, \dots, \sigma_{X_n}$ sono piccole e se $U = U(X_1, X_2, \dots, X_n)$ è una funzione di X_1, X_2, \dots, X_n , allora una stima conservativa di σ_U è data da

$$\sigma_U \leq \left| \frac{\partial U}{\partial X_1} \right| \sigma_{X_1} + \dots + \left| \frac{\partial U}{\partial X_n} \right| \sigma_{X_n} \quad (23)$$

Nella pratica le derivate parziali sono valutate nel punto (X_1, X_2, \dots, X_n) . La disuguaglianza (23) vale in quasi tutte le situazioni pratiche; in linea di principio non dovrebbe valere se le derivate seconde di U sono molto grandi.

8 REGRESSIONE LINEARE SEMPLICE

Scopo del metodo di regressione è studiare come una data grandezza y (detta variabile regressa) dipende da una altra grandezza x assunta come indipendente (regressore); si parla di regressione di y su x . Si applica ai casi in cui ci si aspetta una dipendenza quantitativa (ad es. x = statura, y = peso corporeo).

Dall'analisi del grafico sparso è spesso possibile avere una rappresentazione intuitiva dell'andamento di una curva che passa abbastanza vicina ai dati. Una curva di questo tipo è detta *curva interpolatrice*. Se dalla figura i dati sembrano bene interpolati da una retta allora diremo che tra le variabili esiste una *relazione lineare* espressa attraverso una retta del tipo

$$y = \alpha + \beta x. \quad (24)$$

In altri casi invece si ha che la relazione che appare è una *relazione non lineare* che potrà essere espressa tramite una parabola di equazione

$$y = \alpha x^2 + \beta x + \gamma, \quad (25)$$

una curva logaritmica di equazione

$$y = \alpha \ln x + \beta, \quad (26)$$

etc..

I coefficienti α , β , γ , etc., sono incogniti. Descriviamo un modo per determinare tali coefficienti, una volta che sia stato scelto che tipo di relazione (lineare, parabolico, logaritmico, etc.) si ha tra le due variabili casuali, a partire dai dati $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ tratti da un campione di grandezza n . Per semplicità restringiamo la nostra analisi al caso della retta.

8.1 Il metodo dei minimi quadrati

Data la successione di punti $\{(x_k, y_k)\}_{k=1}^n$, nel piano bisogna intendersi su quale è la migliore curva interpolatrice di questi dati. Sono possibili diverse definizioni, una possibile definizione è la seguente:

Definizione. Sia data la successione di punti $\{(x_k, y_k)\}_{k=1}^n$. Assegnato il tipo di curva (ad es. lineare, parabolica, logaritmica, etc.) allora la *miglior curva interpolatrice* è la curva $y = f(x)$ che rende minima la quantità

$$\Delta = d_1^2 + d_2^2 + \dots + d_n^2 \quad (27)$$

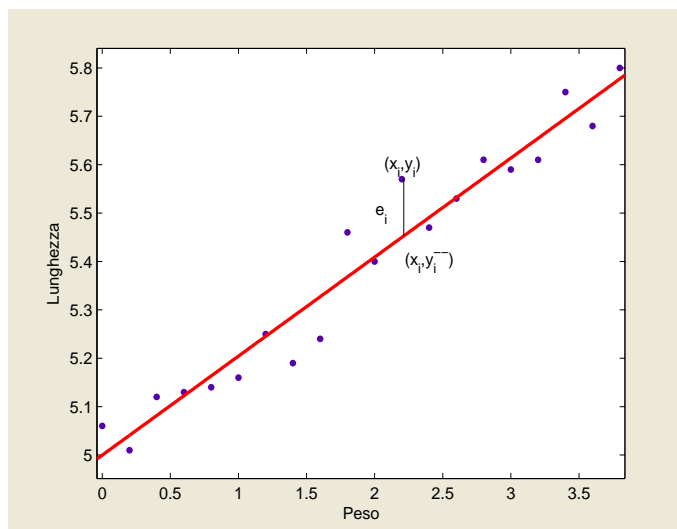


Figura 21: Curva dei minimi quadrati

dove $d_k = |y_k - f(x_k)|$ è la distanza tra il punto (x_k, y_k) ed il punto sulla curva corrispondente alla ascissa x_k . Questa curva viene comunemente chiamata *curva dei minimi quadrati*. Una retta con questa proprietà sarà quindi chiamata *retta dei minimi quadrati*, una parabola sarà chiamata *parabola dei minimi quadrati*, etc..

Determiniamo la retta dei minimi quadrati. Vale il seguente:

Teorema. Siano dati i punti $\{(x_k, y_k)\}_{k=1}^n$ non allineati su una retta parallela all'asse y . Essendo \bar{x} e \bar{y} le medie calcolate sul campione allora si prova che la *retta dei minimi quadrati* per il campione in esame ha equazione $y = \alpha + \beta x$ dove

$$\beta = \frac{\text{cov}(x, y)}{s_x^2}, \quad \alpha = \bar{y} - \beta \bar{x} \quad (28)$$

essendo

$$s_x^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2, \quad \text{cov}(x, y) = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y}). \quad (29)$$

Osserviamo che l'equazione della retta si può anche scrivere come

$$\frac{y - \bar{y}}{s_y} = \rho \frac{x - \bar{x}}{s_x}, \quad \text{dove } \rho = \frac{\text{cov}(x, y)}{s_x s_y} \quad (30)$$

è il *coefficiente di correlazione*.

9 ESERCIZI di STATISTICA MATEMATICA

1. Insiemi e probabilità.
2. Variabili aleatorie discrete.
3. Variabili aleatorie continue, approssimazione.
4. Calcolo di leggi, condizionamenti, sistemi di variabili aleatorie
5. Intervalli di fiducia, test di ipotesi, significatività e potenza.

9.1 Insiemi e Probabilità

1.1 - Due carte sono estratte senza rimpiazzo da un mazzo di 40 ben mescolato. Si calcoli la probabilità che esse siano la prima un asso e la seconda nè asso nè fante.
R.] 8.2%

1.2 - Un gene è composto di due alleli, ciascuno può essere di tipo A oppure a . Nella popolazione vi sono 3 tipi di individui: di tipo AA , Aa , e aa . Ciascun genitore trasmette al figlio uno dei due alleli scelto a caso. Sapendo che inizialmente le proporzioni dei tre tipi sono

$$AA : \frac{1}{3} \quad Aa : \frac{1}{5} \quad aa : \frac{7}{15}$$

quale sarà la proporzione del tipo AA alla generazione successiva?
R.] 18.78%

Soluzione. Sia F_A = “il primo dei due genitori trasmette A ”. Inoltre distinguiamo

$$B_1, B_2, B_3 = \text{“il primo genitore è } AA, \text{ oppure } Aa, \text{ oppure } aa\text{”}.$$

Infatti sappiamo che nei tre casi cambia la probabilità di trasmettere A :

$$P(F_A|B_1) = 1; \quad P(F_A|B_2) = \frac{1}{2}; \quad P(F_A|B_3) = 0.$$

Allora, per il teorema di probabilità totale,

$$\begin{aligned} P(F_A) &= P(F_A|B_1) \cdot P(B_1) + P(F_A|B_2) \cdot P(B_2) + P(F_A|B_3) \cdot P(B_3) = \\ &= 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{5} = \frac{13}{30} \end{aligned}$$

Allo stesso modo, l'evento $G_A =$ "il secondo dei due genitori trasmette A " ha una probabilità $P(G_A) = \frac{13}{30}$. In virtù dell'indipendenza,

$$P(\text{"un figlio sia"} AA) = P(F_A \cap G_A) = \frac{13}{30} \frac{13}{30} = \frac{169}{900}$$

□

1.3 - Un compilatore assegna ad ognuna delle variabili che intervengono in un programma una cella di memoria a caso, con indipendenza da una variabile all'altra. In caso di conflitto (cioè se due variabili sono assegnate alla stessa cella), l'operazione di assegnazione deve essere ripetuta. Se vi sono 100 celle di memoria e 4 variabili, qual è la probabilità che si verifichi un conflitto?

R.] 5.89%

1.4 - I componenti prodotti da una certa ditta presentano due tipi di difetti con percentuale del 2% e del 6% rispettivamente e con indipendenza. Qual è la probabilità che un componente presenti il difetto 1, sapendo che è difettoso?

R.] 25.38%

1.5 - Tre malattie A, B, C - e solo queste - causano un certo sintomo con probabilità $f_A = 9/10$, $f_B = 6/10$, $f_C = 4/10$. In Emilia d'estate un individuo è affetto da ciascuna malattia con probabilità $p_A = 0.1\%$, $p_B = 1\%$, $p_C = 5\%$. Sapendo che un paziente emiliano questa estate presenta il sintomo, qual è la probabilità che egli abbia la malattia B ?

R.] 22.3%

1.6 - Un dado a 4 facce è lanciato 3 volte. Qual è la probabilità di ottenere "quattro" almeno una volta?

R.] 57.81%

1.7 - Qual è la probabilità che almeno due fra 4 coetanei nati nella stessa stagione festeggino il compleanno nello stesso giorno? (una stagione = 92 giorni).

R.] 6.39%

1.8 - In quanti modi 10 persone possono sedersi su una panchina che ha solo 4 posti?

R.] 5040

1.9 - In uno scaffale ci sono 10 libri, 3 di matematica e 7 di fisica; si trovi la probabilità che i 3 libri di matematica si trovino insieme.

R.] 6.66%

1.10 - L'urna I contiene 3 palline rosse e 5 bianche, mentre l'urna II contiene 4 rosse e 2 bianche. Si sceglie una pallina a caso dall'urna I e la si mette, senza osservare il colore, nell'urna II : si estrae poi una pallina dall'urna II . Qual è la probabilità che la pallina così estratta sia bianca?

R.] 37.5%

1.11 - Una fabbrica produce componenti elettronici, che escono da due linee di produzione, A e B, nelle proporzioni del 35% e del 65% rispettivamente. La linea A ha una percentuale di pezzi difettosi del 10%, contro il 20% della linea B. Con quale probabilità un chip prodotto da quella fabbrica è difettoso?

R.] 16.50%

1.12 - La popolazione di una regione è affetta da virus Ebola con probabilità 1%. Il miglior test per il virus ha affidabilità 80% tanto sui sani quanto sui malati. Una persona è scelta casualmente e risulta positiva. Qual è la probabilità che sia effettivamente affetta da Ebola?

R.] 3.9%

1.13 - Uno studente è sottoposto a un quiz con 4 risposte possibili. Se ha studiato, egli risponderà certamente in maniera esatta, altrimenti sceglierà una risposta a caso tra le 4 disponibili. Supponiamo che abbia studiato con probabilità $1/2$ e che, sottoposto al quiz, abbia scelto la risposta esatta. Sulla base di ciò, qual è la probabilità che abbia studiato davvero?

R.] 80%

1.14 - Si abbiano tre scatole, indistinguibili una dall'altra, contenenti ciascuna due palline: una contiene due palline bianche (scatola 1), un'altra una pallina bianca ed una rossa (scatola 2), la terza due palline rosse (scatola 3). Scelta una scatola a caso, si estrae una pallina. La pallina è bianca. Ci si chiede: qual è la probabilità che la pallina sia stata estratta dalla scatola i ?

R.] $i = 1, p = 2/3; \quad i = 2, p = 1/3; \quad i = 3, p = 0$

Soluzione. Indicato con B l'evento "la pallina estratta è bianca" e con A_i l'evento "la pallina è stata estratta dalla scatola i ", ci interessa calcolare le probabilità $P(A_i|B)$. Osserviamo che si ha

$$P(A_i) = \frac{1}{3}, \quad i = 1, 2, 3; \quad P(B|A_1) = 1, \quad P(B|A_2) = \frac{1}{2}, \quad P(B|A_3) = 0.$$

Il fatto che le probabilità non condizionate $P(A_i)$ (probabilità a priori) siano tutte uguali a $\frac{1}{3}$ consegue ovviamente dal fatto che le tre scatole sono indistinguibili. Applicando il teorema di Bayes si ha quindi

$$\begin{aligned}
P(A_1|B) &= \frac{P(B|A_1) \cdot P(A_1)}{P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3)} = \\
&= \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{\frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3}} = \frac{2}{3}; \\
P(A_2|B) &= \frac{P(B|A_2) \cdot P(A_2)}{\frac{1}{2}} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}. \\
P(A_3|B) &= \frac{P(B|A_3) \cdot P(A_3)}{\frac{1}{2}} = \frac{0 \cdot \frac{1}{3}}{\frac{1}{2}} = 0.
\end{aligned}$$

Osserviamo che si trova confermato il fatto ovvio che $P(A_3|B) = 0$. Osserviamo anche come il verificarsi dell'evento B influisca sulle probabilità degli eventi A_i modificandone le probabilità. \square

9.2 Variabili Aleatorie Discrete

2.1 - Una compagnia ha un aereo di 19 posti e accetta 21 prenotazioni perché sa che il 10% dei prenotati non si presenta. Con quale probabilità almeno un passeggero resterà a terra?

R.] 36.47%

Soluzione. Sia $Z =$ “il n.o di passeggeri che si presentano fra i 21 che si sono prenotati”. Allora Z è binomiale con $n = 21$ e $p = \frac{9}{10}$.

$$P(Z = 20) + P(Z = 21) = \binom{21}{20} \left(\frac{9}{10}\right)^{20} \frac{1}{10} + \binom{21}{21} \left(\frac{9}{10}\right)^{21} = 0.3647$$

\square

2.2 - Con quale probabilità esce una o due volte “sette” in 5 lanci di una coppia di dadi?

R.] 56.26%

2.3 - Una coppia di dadi è lanciata 3 volte. Con quale probabilità la somma uscente sarà “cinque” nemmeno per una volta?

R.] 70.23%

2.4 - Un calcolatore è collegato a una rete che permette l'accesso ad un massimo di 20 persone. Collegati a questa rete vi sono i terminali di 22 operatori, ognuno dei quali, a un dato istante, richiede con probabilità $p = 0.8$ di essere connesso al calcolatore centrale. Qual è la probabilità che a un dato istante, la rete sia satura

(cioè che tutti i 20 accessi siano usati) e quindi la richiesta di collegamento venga negata?

R.] 4.8%

2.5 - Una fabbrica produce componenti elettronici, che escono da due linee di produzione, A e B nelle proporzioni del 30% e 70%. La Linea A ha una percentuale di pezzi difettosi del 10%, contro il 17% della linea B. Si considera una confezione di 10 chips di tale fabbrica: con quale probabilità la confezione contiene esattamente un chip difettoso?

R.] $\simeq 35\%$

Soluzione. $D =$ “un chip è difettoso”. Per il teorema di probabilità totale, rispetto alla partizione data A, B ,

$$\begin{aligned} P(D) &= P(D|A) \cdot P(A) + P(D|B) \cdot P(B) = \\ &= \frac{1}{10} \cdot \frac{3}{10} + \frac{17}{100} \cdot \frac{7}{10} = 0.03 + 0.119 = 0.149 \end{aligned}$$

Ora sia $Z =$ “n. di chip difettosi nell’ambito di 10”. Z è allora binomiale con $n = 10$ e $p = 0.149$.

$$P(Z = 1) = \binom{10}{1} (0.149)(0.851)^9 = 10 \cdot (0.149) \cdot (0.234) = 0.35$$

□

2.6 - Una compagnia di assicurazioni ha 3000 assicurati contro un dato rischio che ha probabilità 0.1% di colpire ogni singolo assicurato in un anno. Sapendo che il numero X di indennizzandi in un anno è di Poisson, che la compagnia indennizza ciascuno con 80000 Euro, che percepisce da ogni assicurato un premio annuale di 100 Euro, quali sono il valor medio e la varianza del beneficio annuale della compagnia? R.] $\mu = 60000, \sigma^2 = 192 \cdot 10^8$

Soluzione. Se ha 3000 assicurati con probabilità 1/1000 di incidente individuale all’anno, allora $X =$ “numero di infortunati all’anno” è di Poisson con media $\lambda = 3000 \cdot \frac{1}{1000} = 3$. Se l’indennizzo individuale è $8 \cdot 10^4$ Euro, se il premio annuale è 100 Euro, allora il beneficio annuale della compagnia è la variabile aleatoria

$$Y = 100 \cdot 3000 - X \cdot 8 \cdot 10^4$$

Essa ha valor medio $E(Y) = 3 \cdot 10^5 - E(X) \cdot 8 \cdot 10^4 = 300000 - 3 \cdot 80000$ Euro e varianza $Var(Y) = (8 \cdot 10^4)^2 Var(X) = 64 \cdot 3 \cdot 10^8$. □

2.7 - La memoria secondaria di un calcolatore è composta da 30 unità disco in ognuna delle quali sono archiviati 100 file. Durante l'esecuzione di un programma è necessario accedere a 40 di questi file, tutti diversi. Qual è la probabilità che sia necessario usare l'unità 1? (Cioè qual è la probabilità che tra i 40 file ve ne sia almeno uno contenuto nell'unità 1?)

R.] $1 - \frac{2900 \cdot 2899 \cdot \dots \cdot 2861}{3000 \cdot 2999 \cdot \dots \cdot 2961}$

Soluzione. Dire che il programma deve accedere al disco "1" equivale a dire che tra le 40 registrazioni ce n'è almeno una nel disco "1". È una ipergeometrica la variabile aleatoria $Z_1 =$ "numero di files necessari al programma che si trovano nell'unità 1". Infatti l'insieme dei 30 dischi può essere considerato come un'urna contenente $r = 100$ files del primo disco e $b = 2900$ files degli altri 29, e dalla quale si fanno $n = 40$ estrazioni.

$$\begin{aligned} P(Z_1 = k) &= \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{r+b}{n}} \implies P(Z_1 \geq 1) = 1 - P(Z_1 = 0) = \\ &= 1 - \frac{\binom{100}{0} \binom{2900}{40}}{\binom{3000}{40}} \\ &= 1 - \frac{2900! / (40! 2860!)}{3000! / (40! 2960!)} \end{aligned}$$

□

2.8 - Un canale di trasmissione dati può ricevere messaggi binari da due sorgenti diverse A e B con probabilità $\frac{1}{2}$ ciascuna. Ognuna delle due sorgenti produce messaggi in cui i bit successivi sono tra di loro indipendenti. Ma per la sorgente A i bit possono essere 1 o 0 con probabilità $\frac{1}{2}$, mentre per B il valore 1 si verifica con probabilità $\frac{1}{4}$ e 0 con probabilità $\frac{3}{4}$. Un messaggio di lunghezza 10 viene ricevuto e in esso si osservano 4 bit uguali a 1. Qual è la probabilità che si tratti della sorgente A ?

R.] 58%

Soluzione. Pongo $C =$ {il messaggio di $n = 10$ bits ha esattamente 4 "uno"}. Per la formula di Bayes,

$$P(A|C) = \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|B)P(B)}$$

Ma $P(C|A) = P(X_A = 4)$, dove

$X_A =$ {numero di "1" in $n = 10$ bits sapendo che $p = 1/2$ } $\sim Bin(n = 10, p = 1/2)$.

Così $P(C|B) = P(X_B = 4)$, dove

$X_B = \{\text{numero di "1" in } n = 10 \text{ bits sapendo che } p = 1/4\} \sim \text{Bin}(n = 10, p = 1/4)$.

$$\begin{aligned} \implies P(A|C) &= \frac{P(X_A = 4)P(A)}{P(X_A = 4)P(A) + P(X_B = 4)P(B)} = \\ &= \frac{\binom{10}{4}(1/2)^{10} \cdot \frac{1}{2}}{\binom{10}{4}(1/2)^{10} \cdot \frac{1}{2} + \binom{10}{4}(1/4)^4 \cdot (3/4)^6 \frac{1}{2}} = \\ &= \frac{0.00097}{0.00097 + (0.0039) \cdot (0.1779)} \simeq 0.58 \end{aligned}$$

□

2.9 - Per depurare un lago artificiale in cui si è rilevato un parassita, si esegue più volte un trattamento. Il trattamento riduce il numero medio di parassiti per litro, λ , portandolo a $\lambda/6$. Se inizialmente il numero di parassiti è una variabile aleatoria di Poisson di media 5, quanti interventi occorrono perché al termine ogni litro abbia parassiti con probabilità inferiore a 0.1% ?

R.] 5

Soluzione. Dopo 4 trattamenti $Z_4 = \text{"numero parassiti per litro"}$ è una variabile aleatoria di Poisson con media $\mu = 5 \cdot (1/6)^4$. Quindi

$$P(Z_4 = 0) = e^{-5 \cdot (1/6)^4} < 0.999.$$

Dopo il quinto trattamento si ha

$$P(Z_5 = 0) = e^{-5 \cdot (1/6)^5} > 0.999.$$

□

2.10 - Una moneta è lanciata 3 volte. Se X è il numero di teste che si verificano nei lanci, e se F indica la funzione distribuzione, quanto vale $F(2.9)$?

R.] 7/8

2.11 - In un'urna ci sono 5 palline bianche e 3 nere e si estraggono a caso e senza rimpiazzo due palline. Se X è il numero di bianche estratte ed f la sua funzione di probabilità, determinare $f(2)$.

R.] 5/14

2.12 - Una variabile aleatoria X discreta ha come valori possibili $x = 1, 2, 3, 4$, con funzione distribuzione

$$\text{per } x = 1, 2, 3, 4 \quad F(x) = \frac{1}{8}, \frac{3}{8}, \frac{3}{4}, 1$$

Detta μ la media, quanto vale la probabilità $P(X \leq \mu)$?

R.] 37.5%

Soluzione. Poiché $F(x) = \sum_{i \leq x} f(x_i)$,

$$f(4) = F(4) - F(3) = 1/4, \quad f(3) = F(3) - F(2) = 3/8,$$

$$f(2) = F(2) - F(1) = 2/8, \quad f(1) = 1/8$$

$$\Rightarrow \mu = \sum_i x_i f(x_i) = \frac{1}{8} \cdot 1 + \frac{2}{8} \cdot 2 + \frac{3}{8} \cdot 3 + \frac{1}{4} \cdot 4 = \frac{22}{8} = 2.75$$

$$\Rightarrow F(2.75) = F(2) = \frac{3}{8} = 0.375.$$

□

2.13 - Una variabile aleatoria X discreta ha i tre valori $x = 1, 2, 3$, con funzione distribuzione

$$\text{per } x = 1, 2, 3 \quad F(x) = \frac{1}{8}, \frac{3}{8}, 1$$

Qual è la varianza di X ?

R.] 1/2

2.14 - Se il numero di annegamenti in un anno è pari a 0.3 su centomila, si chiede la probabilità che in una città di duecentomila abitanti ci siano 3 o 4 annegamenti all'anno.

R.] 2.27%

2.15 - Un processo di produzione di viti è controllato ogni ora ispezionando n viti, scelte a caso tra quelle prodotte in quell'ora. Se una o più viti sono difettose, il processo è fermato ed esaminato. Il produttore vuole probabilità 95% di fermare il processo quando l'8% delle viti è difettoso. Quanto deve essere grande n ?

R.] $n \geq 36$

Soluzione. $X = \{\text{numero di difettose nell'ambito di } n \text{ viti sapendo che } p = \frac{8}{100}\}$. Il produttore vuole che

$$P(X \geq 1) \geq 95\% \quad \text{cioè } 1 - \binom{n}{0} \left(\frac{8}{100}\right)^0 \left(\frac{92}{100}\right)^n \geq \frac{95}{100}$$

$$\Leftrightarrow \left(\frac{92}{100}\right)^n \leq \frac{5}{100} \quad \Leftrightarrow n \log(0.92) \leq \log(0.05)$$

cioè $-n(0.0833) \leq -2.9957$, cioè $n \geq 35.96$

□

2.16 - Una scatola ha fondo quadrato di lato 1 metro, al centro del quale vi è un foro circolare di raggio 10 cm. Nella scatola sono gettate a caso e indipendentemente 10

palline di diametro piccolo (cioè $\ll 10$ cm). Con quale probabilità alla fine dei lanci si trovano nella scatola 7 palline?

R.] 0.28%.

2.17 - Un principiante di tiro al piattello lo colpisce con probabilità $2/9$. Qual è la probabilità che gli occorrono almeno 5 tiri per colpirlo la prima volta?

R.] $\simeq 36.6\%$

9.3 Variabili Aleatorie Continue, Approssimazione

3.1 - Si consideri la costante c tale che

$$f(x) = \begin{cases} cxe^{-2x} & \text{se } x \geq 0 \\ 0 & \text{altrove} \end{cases}$$

sia la funzione densità di una variabile aleatoria X . Determinare c .

R.] $c = 4$

Soluzione. Integrando per parti,

$$1 = c \int_0^{\infty} xe^{-2x} dx = c \frac{1}{2} \left[\frac{e^{-2x}}{-2} \right]_0^{\infty}$$

che vale $c/4$. □

3.2 - Si consideri la costante k tale che

$$F(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ k(1 - e^{-x})^2 & \text{se } x > 0 \end{cases}$$

sia la funzione di distribuzione di una variabile aleatoria X ; trovare c tale che $P(X > c) = 90\%$.

R.] 0.38

3.3 - Il numero di chilometri (misurato in migliaia) che può percorrere un certo tipo di gomme è una variabile aleatoria X con densità $f(x) = 0.05e^{-0.05x}$, se $x > 0$, e 0 altrove. Trovare la probabilità che le gomme durino almeno trentamila chilometri.

R.] 22%

Soluzione.

$$P(X \geq 30) = \int_{30}^{+\infty} \frac{5}{100} e^{-(5/100)x} dx = \frac{5}{100} \left[\frac{e^{-(5/100)x}}{-5/100} \right]_{30}^{+\infty}$$

cioè $e^{-15/10} = 0.22$ □

3.4 - Il voto ad una prova di ingresso è distribuito normalmente, e il miglior 10% dei candidati verrà assunto. Ad esame finito, il voto medio è stato 72 e la deviazione standard è stata 9. Qual è il voto minimo che un candidato deve ottenere per essere assunto?

R.] 84.

Soluzione. Essendo $\mu = 72$, $\sigma = 9$,

$$\begin{aligned} \frac{10}{100} &= P(X \geq c) = P\left(\frac{X - \mu}{\sigma} \geq \frac{c - \mu}{\sigma}\right) = P\left(N(0, 1) \geq \frac{c - 72}{9}\right) \\ \implies \frac{c - 72}{9} &= 1.28 \implies c = 72 + 9 \cdot (1.28) = 83.52 \end{aligned}$$

□

3.5 - Supponiamo che k sia la costante che rende densità di probabilità la funzione: $f(x) = kx$, se $0 \leq x \leq 3$, ed $f(x) = 0$ altrove. Trovare c tale che $P(c \leq X \leq 3 - c) = 95\%$.

R.] $c = 15/200$.

3.6 - Trovare c tale che la variabile aleatoria uniforme nell'intervallo $[-c, c]$ abbia varianza 1.

R.] $c = \sqrt{3}$.

3.7 - La stazione Radio Bruco trasmette il segnale orario allo scoccare di ogni ora. L'ascoltatore tipo sintonizza il proprio radiorecettore sulla stazione Radio Bruco a un istante uniformemente distribuito tra le ore 7 : 10 e le ore 19 : 30 nella giornata. Calcolare la probabilità che l'ascoltatore riceva il segnale orario entro 5 minuti dalla sintonizzazione su Radio Bruco (si adotti il minuto come unità di tempo).

R.] 8.1%

Soluzione. Dalle 7 : 10 alle 19 : 30 ci sono 740 minuti. La variabile aleatoria $X = \{\text{istante di sintonizzazione}\}$ è $X \sim U[0, 740]$, con densità di probabilità uguale a $1/740$ in tale intervallo. Il segnale orario è ai minuti $50 + k \cdot 60$, con $k = 0, 1, 2, \dots, 11$. La probabilità richiesta è la somma di 12 aree di piccoli rettangoli di base 5 e altezza $1/740$, cioè $(5/740) \cdot 12 = 0.081$. □

3.8 - Un segnale consiste in una parola di n bit, ciascuno dei quali può assumere i valori 0 oppure 1. Nel corso della trasmissione ogni bit con probabilità $p = 0.01$ può essere distorto. Qual è la probabilità che un segnale di 1000 bit contenga almeno 10 bit distorti?

R.] 56.36%

3.9 - Un segnale consiste in una parola di n bit, ciascuno dei quali può assumere i valori 0 oppure 1. Nel corso della trasmissione ogni bit con probabilità $p = 0.01$ può essere distorto. Per ridurre la distorsione si usa il seguente protocollo: ogni bit viene trasmesso 3 volte ed il vero valore viene deciso a maggioranza: il bit viene posto uguale ad A ($A = 0$ oppure 1) se vi sono almeno due valori A tra quelli ricevuti. Qual è la probabilità che un segnale di 1000 bit contenga bit distorti?

R.] 25.77%

Soluzione. Considerando il singolo bit, esso ha probabilità $p = \frac{1}{100}$ di essere distorto. Ma se viene trasmesso tre volte, si deve considerare la variabile aleatoria $X =$ “numero di trasmissioni distorte del bit su tre trasmissioni, sapendo che la probabilità di distorsione in una singola trasmissione è $p = \frac{1}{100}$ ”. Allora $X \sim Bin(n = 3, p = 0.01)$. In base al protocollo descritto, il rischio è una probabilità p' che dovrebbe risultare decisamente minore di p ; la probabilità di avere due bit distorti su 3, o 3 bit distorti su 3 è:

$$\begin{aligned} p' &= P(X = 2) + P(X = 3) = \binom{3}{2} \left(\frac{1}{100}\right)^2 \left(\frac{99}{100}\right) + \binom{3}{3} \left(\frac{1}{100}\right)^3 \\ &= 0.000297 + 0.000001 = 2.98 \cdot 10^{-4} \end{aligned}$$

La domanda chiede di trattare la variabile aleatoria binomiale $Y =$ “numero di bit distorti nell’ambito di $n = 1000$ bit, sapendo che $p' = 2.98 \cdot 10^{-4}$ ”. Questa binomiale è abbastanza prossima alla normale? No, perché si esigerebbe n tanto grande da rendere np' dell’ordine di alcune unità (qui la condizione pratica $np' \geq 5$ è lontana dall’essere verificata). Dunque la risposta è:

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y = 0) = 1 - \binom{n}{0} (p')^0 (1 - p')^n = \\ &= 1 - \binom{1000}{0} (1 - 2.98 \cdot 10^{-4})^{1000} = 1 - 0.7423 = 0.2577 \end{aligned}$$

Oppure si può approssimare la binomiale con la Poissoniana in questo modo:

$$P(Y \geq 1) \simeq P(Y_{Poiiss} \geq 1) = 1 - e^{-np'} = 1 - e^{-.298} = 0.2577.$$

□

3.10 - Un dado viene lanciato 900 volte e indichiamo con X il numero di volte in cui esce il 6. Sappiamo che esiste una partita di dadi truccati che producono il 6 con probabilità $2/9$. Per decidere se il dado è di questi ultimi usiamo questa procedura: lo lanciamo 900 volte e decidiamo che è truccato se il 6 esce almeno (\geq) 180 volte. Qual è la probabilità che un dado truccato venga effettivamente individuato?

R.] 95%.

Soluzione. Abbiamo due variabili aleatorie binomiali X, Y : entrambe sono del tipo {numero di “sei” uscenti in 900 lanci}; ma una nell’ ipotesi che $p = \frac{1}{6}$ (dado equo) e una nell’ipotesi che si abbia $p' = \frac{2}{9}$ (dado truccato).

$$X \sim Bin(n = 900, p = \frac{1}{6}), \quad \text{con } \mu = 150;$$

$$Y \sim Bin(n = 900, p' = \frac{2}{9}), \quad \text{con } \mu' = 200$$

La procedura descritta sceglie 180 come soglia di decisione statistica. Si domanda qual è la probabilità di accorgerci del trucco nell’ipotesi che il dado sia truccato. Cioè si domanda:

$$\begin{aligned} P(Y \geq 180) &\simeq P(N(np', np'q') \geq 179.5) = P\left(N(0, 1) \geq \frac{179.5 - 200}{\sqrt{200 \cdot 7/9}}\right) = \\ &= P(N(0, 1) \geq -1.6437) = P(N(0, 1) \leq 1.6437) = 0.949 \end{aligned}$$

Tale probabilità viene chiamata in statistica $1 - \beta =$ “potenza del test”. Invece $P(X \geq 180)$ viene chiamata $\alpha =$ “livello di significatività del test”. \square

3.11 - Un calcolatore esegue un milione di somme di numeri e in ognuna di queste addizioni si effettua un errore di arrotondamento; supponiamo che i singoli errori siano indipendenti e abbiano distribuzione uniforme su $[-\frac{1}{2}10^{-10}, \frac{1}{2}10^{-10}]$ (cioè supponiamo che la decima cifra decimale sia significativa). Qual è la probabilità che l’errore finale sia inferiore in valore assoluto a $\frac{1}{2}10^{-7}$? (cioè qual è la probabilità che la settima cifra decimale sia significativa?)

R.] 92%

Soluzione. Per $i = 1, 2, \dots, 10^6$, introduciamo la variabile aleatoria $X_i =$ “errore compiuto nell’ eseguire l’ i -esima somma”. La variabile aleatoria “errore dopo avere eseguito 10^6 somme” è $X = X_1 + X_2 + \dots + X_{10^6}$. Ora, per il teorema del limite centrale, X approssima una normale avente la stessa media e la stessa varianza. Andiamo a calcolarle:

$$\begin{aligned} E(X_i) &= 0, \quad \sigma^2 = Var(X_i) = 10^{10} \int_{-10^{-10}/2}^{+10^{-10}/2} x^2 dx = \\ &= \left[\frac{t^3}{3}\right]_{-10^{-10}/2}^{+10^{-10}/2} \cdot 10^{10} = \frac{1}{12}10^{-20} \end{aligned}$$

Poiché le X_i sono indipendenti, la varianza della somma è la somma delle varianze:

$$E(X) = 0, \quad Var(X) = n\sigma^2 = 10^6 \frac{1}{12}10^{-20} = \frac{1}{12}10^{-14}$$

Per il teorema di limite centrale e standardizzando i tre membri della disuguaglianza:

$$\begin{aligned} P\left(-\frac{1}{2}10^{-7} \leq X \leq \frac{1}{2}10^{-7}\right) &\simeq P\left(-\frac{1}{2}10^{-7} \leq N\left(0, \frac{1}{12}10^{-14}\right) \leq \frac{1}{2}10^{-7}\right) = \\ &= P\left(-\frac{1}{2}10^{-7}\left(\frac{1}{12} \cdot 10^{-14}\right)^{-1/2} \leq N(0, 1) \leq +\frac{1}{2}10^{-7}\left(\frac{1}{12} \cdot 10^{-14}\right)^{-1/2}\right) = \\ &= P(-1.73 \leq N(0, 1) \leq 1.73) = \end{aligned}$$

$$\Phi(1.73) - \Phi(-1.73) = \Phi(1.73) - (1 - \Phi(1.73)) = 0.96 - (1 - 0.96) = 0.92.$$

□

3.12 - Nella trasmissione di un'immagine ogni pixel resta integro con probabilità $p = 0.9984$. Un'immagine è composta da $512 \times 256 = 131072$ pixel. Qual è la probabilità che vi siano almeno (\geq) 200 pixel distorti?

R.] $\simeq 77\%$

3.13 - Il punteggio ottenuto dagli studenti alla prova scritta di un esame si può modellizzare come una variabile aleatoria normale di media 21 e varianza 8. Qual è la percentuale di studenti che hanno ottenuto un voto fra 16 e 27 (estremi inclusi)?

R.] $\simeq 94\%$

9.4 Calcolo di Leggi, Condizionamenti e Sistemi di Variabili Aleatorie

9.4.1 Calcolo della Legge di una Funzione di una Variabile Aleatoria

4.1 - Sia X una variabile aleatoria con densità

$$h(x) = \begin{cases} \frac{2}{3}xe^{-x^2/3} & \text{se } x > 0 \\ 0 & \text{altrove} \end{cases}$$

Se $Z = X^2$, calcolare la densità di Z nel punto $t = 4$.

R.] $f_Z(4) = 0.0878$

4.2 - Sia X una variabile aleatoria con densità $f_X(x) = \frac{1}{3}e^{-x/3}$ per $x > 0$, e nulla altrove. Se $W = e^{-2X}$, calcolare la densità di W nel punto $t = \frac{1}{10}$.

R.] $f_W(0.1) = 1.135$

4.3 - È data una variabile aleatoria $X \sim N(0, 1)$. Determinare la densità di probabilità $f_Y(y)$ della variabile aleatoria lognormale $Y = e^{-X}$.

R.] $f_Y(y) = (y\sqrt{2\pi})^{-1} \exp -(\log y)^2/2$

4.4 - In un certo istante l'ampiezza X di un segnale emesso da un generatore di segnali aleatori è una variabile aleatoria normale $N(0.2, 0.36)$. L'onda passa attraverso un selezionatore dando in uscita $Y = X^+$, dove $x^+ = x$ se $x > 0$ ed $x^+ = 0$ se $x < 0$. Trovare $P(-0.3 \leq Y \leq 0.3)$.

9.4.2 Calcolo della Legge di una Funzione di più Variabili Aleatorie

4.5 - Un componente elettronico è composto da due elementi in parallelo (il che significa che funziona se uno almeno dei due elementi è funzionante), ciascuno dei quali ha un tempo di vita esponenziale di media 8 giorni, con indipendenza. Con quale probabilità il componente durerà un tempo minore o uguale a 12 giorni?

R.] $(1 - e^{-3/2})^2 \simeq 60\%$

Soluzione. Le variabili aleatorie X, Y "tempo di vita" di ciascun componente hanno densità

$$f_X(x) = \frac{1}{8}e^{-x/8}\chi_{[0,+\infty)}(x) \quad f_Y(y) = \frac{1}{8}e^{-y/8}\chi_{[0,+\infty)}(y)$$

Sei due elementi sono in parallelo il tempo di vita del dispositivo è la variabile aleatoria $T = \max\{X, Y\}$. Allora, usando anche l'indipendenza,

$$\begin{aligned} P(T \leq t) &= P(X \leq t, Y \leq t) = P(X \leq t) \cdot P(Y \leq t) = \left(\int_0^t \frac{1}{8}e^{-x/8} dx\right)^2 = \\ &= ([-e^{-x/8}]_0^t)^2 = (1 - e^{-t/8})^2 \end{aligned}$$

Quindi

$$P(T \leq 12) = (1 - e^{-12/8})^2 = (1 - e^{-3/2})^2 \simeq 0.6035.$$

□

4.6 - Un componente elettronico è formato da tre elementi in serie, (cioè non funziona se uno almeno dei tre non funziona) aventi ciascuno un tempo di vita esponenziale di parametri $\lambda = 1/5, \mu = 2/5, \gamma = 1/10$ rispettivamente. I tre tempi di vita sono indipendenti. Indichiamo con T la variabile aleatoria "tempo di vita" del componente. Quanto vale $E(T)$?

R.] 1.43

Soluzione. Poiché i tre elementi sono in serie, T coincide con $\min\{T_1, T_2, T_3\}$, dove T_i è il tempo di vita dell' i -esimo elemento. Per definizione di minimo e $P(T \geq t) =$

$P(T_1 \geq t, T_2 \geq t, T_3 \geq t)$, e questo è uguale al prodotto delle tre probabilità in virtù della indipendenza

$$\begin{aligned} P(T \geq t) &= \int_t^\infty \frac{2}{10} e^{-2x/10} dx \int_t^\infty \frac{4}{10} e^{-4x/10} dx \int_t^\infty \frac{1}{10} e^{-x/10} dx \\ &= [-e^{-2x/10}]_t^\infty [-e^{-4x/10}]_t^\infty [-e^{-x/10}]_t^\infty = e^{(-7/10)t} \\ \implies P(T \leq t) &= 1 - e^{(-7/10)t} \implies f_T(t) = \frac{7}{10} e^{-7t/10}. \end{aligned}$$

Ma questa è una variabile aleatoria esponenziale e sappiamo che la media è l'inverso del parametro. Perciò: $E(T) = \frac{10}{7} = 1.43$ \square

4.7 - Un componente elettronico è formato da due elementi uguali in parallelo (cioè non funziona se ambedue hanno cessato di funzionare); ciascuno dei due a sua volta è formato da due elementi in serie (cioè non funziona se almeno uno dei due non funziona). Questi due elementi in serie hanno tempo di vita esponenziale di parametri rispettivamente $\lambda = 2/10$, $\mu = 1/10$ e si assume l'indipendenza. Qual è il tempo medio di vita del componente globalmente?

R.] 5

4.8 - Un componente elettronico è formato da due elementi in serie (cioè non funziona se almeno uno dei due non funziona), il primo dei quali ha un tempo di vita distribuito esponenzialmente con parametro $1/12$. Il secondo elemento è a sua volta formato da due elementi in parallelo (cioè non funziona se ambedue non funzionano), aventi tempo di vita esponenziale con parametro $5/12$ e $3/12$ rispettivamente. Si suppone l'indipendenza. Qual è il tempo medio di vita globale?

R.] 3.67

4.9 - Due numeri X ed Y sono scelti a caso e indipendentemente nell'intervallo $[0, 1]$. Indichiamo con Z la distanza fra loro. Calcolare $P(1/5 < Z < 2/5)$.

R.] 28%

Soluzione. Dalle variabili aleatorie $X \sim U[0, 1]$, $Y \sim U[0, 1]$, indipendenti, costruiamo geometricamente le proprietà della variabile aleatoria $Z = |X - Y|$. Nel piano cartesiano, nel quadrato di vertici $(0, 0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$ tracciamo la diagonale $x - y = 0$, ossia $z = 0$. La zona $\{Z \equiv |x - y| > t\}$ sarà delimitata dalle rette $\{|y - x| = t\} \equiv \{y = x \pm t\}$. In particolare la retta $y = x + t$ delimita dal basso il triangolo di vertici $(0, 1)$, $(0, t)$, $(1 - t, 1)$; l'altra delimita dall'alto il triangolo di vertici $(1, 0)$, $(t, 0)$, $(1, 1 - t)$. Ora l'unione di questi due triangoli equivale a un quadrato di lato $(1 - t)$:

$$P(|X - Y| > t) = P(Z > t) = (1 - t)^2.$$

Per calcolare la probabilità richiesta basta fare la differenza fra due aree del tipo $P(Z > t)$:

$$\begin{aligned} P\left(\frac{1}{5} < Z < \frac{2}{5}\right) &= P\left(Z \leq \frac{2}{5}\right) - P\left(Z \leq \frac{1}{5}\right) = 1 - P\left(Z > \frac{2}{5}\right) - \{1 - P\left(Z > \frac{1}{5}\right)\} \\ &= P\left(Z > \frac{1}{5}\right) - P\left(Z > \frac{2}{5}\right) = (4/5)^2 - (3/5)^2 = 7/25 = 0.28. \end{aligned}$$

□

4.10 - Un punto è scelto a caso nel piano con densità

$$f(x, y) = (2\pi)^{-1} \exp\left(-\frac{x^2 + y^2}{2}\right).$$

Indichiamo con Z la distanza del punto dall'origine. Calcolare $P(Z > 1/3)$.

R.] 94.6%

4.11 - Per ottenere il valore della potenza P dissipata su un resistore di resistenza nota 1.3 si effettuano 10 operazioni di misura in condizioni indipendenti della tensione ai capi di detto resistore e si modellano i risultati di tali operazioni come 10 variabili aleatorie V_i indipendenti e uniformemente distribuite tra -1.8 e $+1.8$. Si forma quindi una stima della potenza dissipata come segue:

$$\hat{P} = \frac{1}{10} \sum_{i=1}^{10} \frac{1}{1.3} V_i^2$$

Determinare il valore atteso $E(\hat{P})$ e l'accuratezza della stima, definita come la deviazione standard $\sqrt{\text{Var}(\hat{P})}$ della variabile aleatoria \hat{P} .

R.] $E(\hat{P}) = 0.83$; $\sigma_{\hat{P}} = 0.237$

Soluzione. Se $V_i \sim U[-1.8, 1.8]$ allora la densità di V_i è uguale a $\frac{1}{3.6} = 0.278$ in $[-1.8, 1.8]$ e zero altrove. Poiché in generale $E[g(X)] = \int_{-\infty}^{+\infty} g(x)f_X(x)dx$, abbiamo

$$\begin{aligned} E(V_i^2) &= \int_{-1.8}^{1.8} x^2(0.278)dx = 1.08 \\ \text{Var}(V_i^2) &= E[(V_i^2 - 1.08)^2] = 0.278 \int_{-1.8}^{1.8} (x^2 - 1.08)^2 dx = \\ &= 0.278 \left\{ \left[\frac{x^5}{5} \right]_{-1.8}^{1.8} + (1.166) \cdot (3.6) - (2.16) \cdot \left[\frac{x^3}{3} \right]_{-1.8}^{1.8} \right\} = 0.9335 \end{aligned}$$

Allora il valor medio è:

$$E(\hat{P}) = \frac{1}{10 \cdot (1.3)} \cdot 10 \cdot E(V_i^2) = (0.77) \cdot (1.08) = 0.83$$

La varianza è:

$$Var(\hat{P}) = \left(\frac{1}{10 \cdot (1.3)}\right)^2 \cdot 10 \cdot Var(V_i^2) = (0.006) \cdot 10 \cdot (0.933) = 0.056$$

Quindi l'accuratezza è $\sigma_{\hat{P}} = \sqrt{0.056} = 0.237$. □

4.12 - In due punti di un lago si misura l'intensità del suono causato da rumore di fondo generale (detto "rumore di ambiente"). Siano X, Y le due variabili aleatorie intensità del suono. Supponiamo che la loro legge congiunta sia continua con densità

$$f_{X,Y}(x, y) = \begin{cases} 4xe^{-x}ye^{-2y} & \text{se } x, y \geq 0 \\ 0 & \text{altrove} \end{cases}$$

Sia $U = \min(X, Y)$ l'intensità minima di rumore. Calcolare $P(U \geq 0.2)$.

R.] $\simeq 92\%$

4.13 - Si sa che un segnale X segue una legge normale $N(0, 1)$, ma non è osservato direttamente. Ciò che si osserva è una misurazione $Y = X + W$, dove W è un errore che segue una legge $N(0, \sigma^2)$ con $\sigma^2 = 0.1$, ed inoltre è indipendente da X . Calcolare la matrice di covarianza di (X, Y) .

R.] $\begin{pmatrix} 1 & 1 \\ 1 & 1.1 \end{pmatrix}$

9.4.3 Condizionamenti, Leggi Condizionate

4.14 - Calcolare $P(XY < \frac{1}{4} | X > 1/2)$, sapendo che X, Y sono variabili aleatorie indipendenti e uniformi in $[0, 1]$

R.] $\frac{1}{2} \log(2)$.

Soluzione. Nel quadrato $[0, 1] \times [0, 1]$, ci interessa il tratto del ramo di iperbole $xy = \frac{1}{4}$ corrispondente a $\frac{1}{2} \leq x \leq 1$. La probabilità dell'intersezione dei due eventi, $P(XY < \frac{1}{4}, X > 1/2)$, è l'area della superficie sottesa da tale tratto di grafico:

$$P\left(XY < \frac{1}{4} \mid X > 1/2\right) = \frac{1}{P(X > 1/2)} \cdot P\left(XY < \frac{1}{4}, X > 1/2\right) =$$

$$= \frac{1}{1/2} \int_{1/2}^1 dx \int_0^{\frac{1}{4x}} dy = 2 \cdot \frac{1}{4} \cdot [\log x]_{1/2}^1 = \frac{1}{2}(-\log \frac{1}{2})$$

□

4.15 - Un'apparecchiatura ha un tempo di vita Y che segue una legge esponenziale di parametro x che dipende dalla qualità di uno dei materiali impiegati. Ma nel processo di produzione x non è sotto controllo, e si assume distribuito come una variabile aleatoria esponenziale di parametro 2. Qual è la densità congiunta di (X, Y) ?

R.] $f(x, y) = 2x \cdot e^{-x(2+y)}, \quad x, y > 0$

Soluzione.

$$\bar{f}_{Y|X}(y|X=x) = \begin{cases} xe^{-xy} & \text{se } y > 0 \\ 0, & y \leq 0 \end{cases} \implies$$

$$f(x, y) = f_X(x) \cdot \bar{f}_{Y|X}(y|X=x) = \begin{cases} 2e^{-2x}xe^{-xy} & \text{se } x > 0, y > 0 \\ 0 & \text{altrove} \end{cases}$$

da cui segue il risultato.

□

4.16 - La densità congiunta di (X, Y) è

$$f(x, y) = \begin{cases} 3xe^{-x(3+y)} & \text{se } x, y > 0 \\ 0 & \text{altrove} \end{cases}$$

Qual è la densità condizionata di X dato $Y = y$?

R.] $f_{X|Y}(x|Y=y) = x(3+y)^2e^{-x(3+y)}$

4.17 - In una fabbrica ci sono due linee di lavorazione. I pezzi prodotti dalla linea A hanno un tempo di vita esponenziale di parametro $\lambda = 1/5$; i pezzi prodotti dalla linea B hanno il tempo di vita esponenziale di parametro $\mu = 1/3$. Inoltre le linee A e B producono rispettivamente il 70% e il 30% dei pezzi. Un pezzo è scelto a caso e indichiamo con T il suo tempo di vita. Calcolare il valor medio $E(T)$.

R.] $22/5$

Soluzione. Una variabile aleatoria esponenziale di parametro λ ha densità uguale a $\lambda e^{-\lambda x}$ per $x \in (0, \infty)$ e uguale a 0 altrove. Per il teorema di probabilità totale applicato all'evento $(T \leq t)$

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(T \leq t|A) \cdot P(A) + P(T \leq t|B) \cdot P(B) \\ &= \left(\int_0^\infty \frac{1}{5} e^{-x/5} dx \right) \cdot \frac{7}{10} + \left(\int_0^t \frac{1}{3} e^{-x/3} dx \right) \cdot \frac{3}{10} = \end{aligned}$$

$$= \frac{7}{10}(1 - e^{-t/5}) + \frac{3}{10}(1 - e^{-t/3}).$$

$$\implies f(t) = F'(t) = \begin{cases} \frac{7}{10} \frac{1}{5} e^{-t/5} + \frac{3}{10} \frac{1}{3} e^{-t/3}, & \text{se } t > 0 \\ 0 & \text{altrove.} \end{cases}$$

Ricordando che la media di una variabile aleatoria esponenziale di parametro λ è $\frac{1}{\lambda}$,

$$E(T) = \frac{7}{10} \int_0^\infty \frac{1}{5} t e^{-t/5} dt + \frac{3}{10} \int_0^\infty \frac{1}{3} t e^{-t/3} dt = \frac{7}{10} \cdot 5 + \frac{3}{10} \cdot 3 = \frac{22}{5}.$$

□

4.18 - Ogni anno un tipo di macchina deve essere sottoposto ad alcuni arresti per manutenzione. Questo numero di arresti X è variabile aleatoria di Poisson con parametro y . Ma anche y è aleatorio (ad es. può dipendere dalla macchina) e assumiamo che esso segua una legge

$$f_Y(y) = 3e^{-3y}, \quad y > 0; \quad f_Y(y) = 0, \quad y < 0$$

Qual è la probabilità che una singola macchina sia sottoposta a 3 arresti in un anno?
R.] $\simeq 18\%$

4.19 - La densità di probabilità congiunta delle variabili aleatorie X ed Y è

$$f_{X,Y}(x, y) = \begin{cases} 8xy & \text{se } 0 \leq x \leq 1, 0 \leq y \leq x \\ 0 & \text{altrove.} \end{cases}$$

Determinare la media condizionata di Y sapendo che $X = \frac{1}{2}$.

R.] $\simeq 0.33$

4.20 - Si collaudano lampadine lasciandole accese fino al guasto. Il tempo di guasto T abbia la seguente densità:

$$f_T(t) = \frac{1}{5} e^{-t/5}, \quad t > 0; \quad = 0, \quad t < 0.$$

Si decide però di collaudare solo le lampadine sopravvissute dopo un rodaggio di tempo $t_0 = 1$. Qual è la media condizionata di T dato l'evento $B = \{T > 1\}$?

R.] 6

4.21 - Due giovani decidono di incontrarsi tra le 17 e le 18 con l'accordo che nessuno deve aspettare l'altro per più di 5 minuti. Supponiamo che gli orari X ed Y in cui arrivano siano indipendenti e casuali, variabili fra le 17 e le 18. Trovare la probabilità condizionata che i due giovani si incontrino, dato che lei arriva alle 17 : 30.

R.] $\frac{1}{6}$

4.22 - Un segnale X non è osservato direttamente, ma attraverso una misurazione disturbata Y . Tuttavia si sa che (X, Y) segue una legge normale bivariata con medie $E(X) = E(Y) = 0$ e con matrice di covarianza

$$C_{X,Y} = \begin{pmatrix} 1 & 1 \\ 1 & 1.1 \end{pmatrix}$$

Trovare la legge condizionata di X sapendo che $Y = \frac{11}{20}$,

R.] $N(0.5; 0.09)$.

9.5 Intervalli di Fiducia, Test di Ipotesi, Significatività

5.1 - Si misura la resistenza di una partita di semiconduttori saggiando un campione di dimensione $n = 7$. Si rileva $\bar{x} = 5.6$ ed $s^2 = 4$. Trovare un intervallo di fiducia al livello $1 - \alpha$ con $\alpha = 5\%$.

R.] [3.75, 7.45]

5.2 - In un paese si vuole stimare la proporzione θ di persone aventi una certa opinione. Per questo si intervistano a caso $n = 1000$ individui e si contano $Z_n = 470$ individui che hanno tale opinione. Trovare un intervallo di fiducia di livello $1 - \alpha$, con $\alpha = 5\%$, per la proporzione incognita θ .

R.] [0.439, 0.501]

5.3 - In una fabbrica si vuole controllare se lo standard di qualità (una proporzione $< 3\%$ di pezzi difettosi) è stato raggiunto. Si controllano 900 pezzi, 17 dei quali sono trovati difettosi. Al livello $\alpha = 0.05$ si può dire che sia stata raggiunta la qualità?

R.] sí, perché $0.019 < 0.027$

5.4 - L'altezza media delle reclute alla visita di leva nel 2000 era di 169 cm; 150 reclute sono scelte a caso nel 2010. I valori di media e varianza del campione sono

$$\bar{x} = 172, \quad s^2 = 126$$

Si può affermare (al livello $1 - \alpha$, con $\alpha = 0.01$) che l'altezza media delle reclute è aumentata?

R.] sí, perché $3.27 > 2.32$.

5.5 - L'etichetta di un liquore dichiara un contenuto di 730 ml. Un'associazione di consumatori decide di controllare: su 81 bottiglie provate si riscontra una media di $\bar{x} = 726$ ml, con una varianza campionaria $s^2 = 625$. I controllori riusciranno a dire con significatività $1 - \alpha = 95\%$ che le bottiglie contengono meno di quanto dichiarato?

R.] no, perché $-1.44 > -1.66$

Soluzione. I dati: $\bar{x} = 726$, $s = 25$, $n = 81$. Dunque i GL sono $\nu = 80$ e $\mu_0 = 730$ è il parametro da sottoporre a test.

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n) \sim N\left(\mu_0, \frac{n\sigma^2}{n^2}\right)$$

$$\implies T \equiv \frac{\bar{x} - 730}{s/\sqrt{n}} \sim t(80) \equiv \text{"t" di Student con 80 GL}$$

Per saggiare μ_0 dobbiamo confrontare il valore sperimentale T_{sper} di T (è un numero negativo) con il valore critico T_c di T che lascia fuori da $(T_c, +\infty)$ $\alpha = 5\%$ (il test è unilaterale). Dunque $T_c = -t_{0.95}^*(80)$. Si vuole fare un test per l'ipotesi $H_0 : \mu = \mu_0$ (le bottiglie contengono quanto dichiarato) contro $H_A : \mu < \mu_0$

$$\alpha = 0.05 = P\left(\left|\frac{\bar{x} - 730}{s/\sqrt{n}}\right| > t_{0.95}^*(80)\right) \implies t_{0.95}^*(80) = 1.66$$

Il test ha due esiti possibili:

$$\begin{cases} T_{sper} \leq T_c \implies \text{rigetto della ipotesi } H_0 : \mu = 730 \\ T_{sper} > T_c \implies \text{non rigetto della ipotesi } H_0 : \mu = 730 \end{cases}$$

Calcolando:

$$T_{sper} = \frac{\bar{x} - 730}{s/\sqrt{n}} = \frac{726 - 730}{25/9} = -1.44 > -1.66 = -t_{0.95}^*(80) = T_c$$

Tuttavia

$$P(t(80) \leq -1.44) = P(t(80) \geq 1.44) = 0.077 = 7.7\%$$

e i consumatori possono rigettare H_0 almeno con livello $\alpha = 7.7\%$. □

5.6 - Si sa che il numero di arresti per guasto in un reparto ogni settimana segue una legge di Poisson di parametro λ , ed inoltre i guasti in settimane successive possono considerarsi indipendenti. Indichiamo con X_1, \dots, X_{10} il numero di guasti in 10 settimane diverse e poniamo $\bar{X} = \frac{1}{10}(X_1 + \dots + X_{10})$. Determinare: 1) la soglia critica c tale che $\{\bar{X} \geq c\}$ sia una regione di rigetto di livello $\alpha = 0.05$ per il test dell'ipotesi $H_0 : \lambda = 2$ contro $H_A : \lambda = 3$; 2) la potenza $1 - \beta$ del test.

R.] $c = 2.7334$, $1 - \beta = 67\%$.

Soluzione. Le ipotesi del modello sono: $X_i \sim Pois(\lambda)$, $\lambda = E(X_i) = Var(X_i)$, $\bar{X} = \frac{1}{n^2} \cdot n\lambda = \frac{\lambda}{n}$. Le ipotesi del test sono: $H_o : \lambda = 2$, $H_A : \lambda = 3$. Per standardizzare $X_1 + \dots + X_n$ ricordiamo che ha media $n\lambda$ e varianza $n\lambda$; inoltre applicando il teorema del limite centrale:

$$Z = \frac{X_1 + \dots + X_n}{\sqrt{n\lambda}} \sim N(0, 1)$$

Dunque la soglia critica c_o fra $n\lambda = 10 \cdot 2$ ed $n\lambda = 10 \cdot 3$ è tale che

$$5\% = P(Z > c_o) \implies c_o = 1.64$$

La soglia critica c_o di Z determina allora la soglia di $X_1 + \dots + X_n$, sempre nella ipotesi H_o :

$$\frac{X_1 + \dots + X_{10} - 20}{\sqrt{20}} = 1.64 \implies X_1 + \dots + X_{10} = 20 + (1.64) \cdot \sqrt{20} = 27.334$$

da cui la soglia c richiesta: $c = \bar{X} = 2.7334$ Per trovare la potenza $1 - \beta$ ricordo che la probabilità β di errore di seconda specie è calcolata nella ipotesi H_A . Di nuovo mediante l' approssimazione alla normale:

$$\begin{aligned} 1 - \beta &= 1 - P(X_1 + \dots + X_{10} |_{\lambda=3} < 27.334) \simeq 1 - P(N(30; 30) < 27.33) = \\ &= 1 - P(N(0, 1) < \frac{27.33 - 30}{5.477}) = P(N(0, 1) < 0.487) = 67\% \end{aligned}$$

e questa è la potenza del test. □

5.7 - Si vuole saggiare l'efficacia di un farmaco anticolesterolo e per questo si misura il tasso di colesterolo in 24 pazienti prima e dopo la somministrazione del farmaco: x_i, y_i , con $i = 1, \dots, 24$. Dette $z_i = x_i - y_i$ le differenze, la media e la varianza campionaria sono:

$$\bar{z} = \bar{x} - \bar{y} = 8.21 \quad , \quad s_z^2 = 770.61$$

Il farmaco è efficace al livello $\alpha = 5\%$?

R.] no, perché $1.45 < 1.71$

5.8 - Un certo indice di inquinamento viene misurato appena prima e appena dopo l'adozione di targhe alterne. Prima si fanno $n_1 = 13$ misurazioni, con varianza $s_1^2 = 281$. Dopo si fanno $n_2 = 8$ misurazioni, con varianza $s_2^2 = 190$. Si deve rifiutare, al livello $\alpha = 5\%$, l'ipotesi di omogeneità delle varianze?

R.] no, perché $1.48 < 3.57$

5.9 - La concentrazione in mg/Kg di un idrocarburo è rilevata in due specie di pesci in un lago. $n_1 = 9$ pesci della prima specie danno media $\bar{x} = 60$ e varianza $s_x^2 = 108$. $n_2 = 16$ degli altri pesci danno media $\bar{y} = 35$ e varianza $s_y^2 = 89$. È significativa, al livello $\alpha = 5\%$, la differenza di concentrazione tra le due specie?
R.] sí perché $1.9636 > 1.7138$

Indice

| | | |
|----------|---|-----------|
| 1 | LA PROBABILITÀ MATEMATICA | 1 |
| 1.1 | Definizioni e Proprietà | 1 |
| 1.2 | Spazi di probabilità finiti | 5 |
| 1.3 | Spazi finiti equiprobabili | 6 |
| 1.4 | Calcolo combinatorio | 7 |
| 1.5 | Probabilità condizionata. Eventi indipendenti. | 13 |
| 2 | VARIABILI ALEATORIE | 20 |
| 2.1 | Definizioni e Proprietà | 20 |
| 2.2 | Variabili aleatorie discrete | 21 |
| 2.3 | Variabili aleatorie continue | 23 |
| 2.4 | Media e varianza | 25 |
| 3 | DISTRIBUZIONI PIÙ COMUNI | 29 |
| 3.1 | La distribuzione uniforme | 29 |
| 3.2 | Legge di probabilità di Bernoulli | 29 |
| 3.3 | Legge di probabilità binomiale | 31 |
| 3.4 | Legge di probabilità di Poisson | 34 |
| 3.5 | Altre leggi di probabilità discrete | 36 |
| 3.6 | Legge di probabilità normale o di Gauss | 40 |
| 3.7 | Teorema di approssimazione di De Moivre e Laplace | 44 |
| 3.8 | Legge di probabilità esponenziale e gamma | 46 |
| 4 | TRASFORMAZIONI DI VARIABILI ALEATORIE | 50 |
| 4.1 | Leggi congiunte di due variabili aleatorie | 50 |
| 4.2 | Indipendenza | 54 |
| 4.3 | Covarianza | 56 |
| 4.4 | Combinazioni lineari di variabili aleatorie | 59 |
| 4.5 | Applicazione alla gestione del portafoglio | 60 |
| 4.6 | Approssimazione | 61 |
| 4.7 | Condizionamenti, Leggi condizionali | 62 |
| 4.8 | Esempi di variabili aleatorie congiunte | 66 |
| 5 | STATISTICA DESCRITTIVA | 72 |
| 5.1 | Organizzazione e Descrizione dei Dati | 72 |
| 5.2 | Grandezze che sintetizzano i dati | 79 |

| | | |
|----------|--|------------|
| 6 | STATISTICA MATEMATICA | 94 |
| 6.1 | Popolazioni e Campioni | 94 |
| 6.2 | Stimatori | 95 |
| 6.3 | Distribuzioni collegate alla normale | 97 |
| 6.4 | Intervalli di fiducia | 100 |
| 6.5 | Test delle ipotesi | 108 |
| 6.6 | Test eseguiti su un solo campione | 112 |
| 6.7 | Test eseguiti su due campioni | 116 |
| 7 | PROPAGAZIONE DELL'ERRORE | 119 |
| 7.1 | Errori di misurazione | 119 |
| 7.2 | Combinazioni lineari di misurazioni | 121 |
| 7.2.1 | Misure ripetute | 122 |
| 7.2.2 | Misurazioni indipendenti con valori differenti per l'incertezza . | 123 |
| 7.2.3 | Combinazione lineare di misurazioni dipendenti | 123 |
| 7.3 | Incetezza per funzioni di una misurazione | 124 |
| 7.3.1 | Incetezza relativa per funzioni di una misurazione | 125 |
| 7.4 | Inceteezze per funzioni di piú misurazioni | 125 |
| 7.4.1 | Inceteezza per funzioni di misure dipendenti | 126 |
| 8 | REGRESSIONE LINEARE SEMPLICE | 127 |
| 8.1 | Il metodo dei minimi quadrati | 127 |
| 9 | ESERCIZI di STATISTICA MATEMATICA | 129 |
| 9.1 | Insiemi e Probabilità | 129 |
| 9.2 | Variabili Aleatorie Discrete | 132 |
| 9.3 | Variabili Aleatorie Continue, Approssimazione | 137 |
| 9.4 | Calcolo di Leggi, Condizionamenti e Sistemi di Variabili Aleatorie . . | 141 |
| 9.4.1 | Calcolo della Legge di una Funzione di una Variabile Aleatoria | 141 |
| 9.4.2 | Calcolo della Legge di una Funzione di piú Variabili Aleatorie | 142 |
| 9.4.3 | Condizionamenti, Leggi Condizionate | 145 |
| 9.5 | Intervalli di Fiducia, Test di Ipotesi, Significatività | 148 |