Metodi Matematici e Statistici
Canale M-Z A.A 2024-2025
Appunti Semiseri

Disclaimer

Questo documento contiene gli appunti del corso di Metodi Matematici e Statistici del Corso di Laurea triennale in Informatica, canale M-Z, A.A. 2024-2025. Questi appunti devono considerarsi “semiseri”, informali, per cui per ogni approfondimento o trattazione formale si rimanda ai libri di testo che trovate nella sezione “Testi di riferimento” del Syllabus.

Questo documento ovviamente sarà pieno di errori, per cui non esitate a mandarmi la lista di tutte le cose inesatte che troverete sparse in giro.

Questo documento è ancora “in fieri”, per cui verrà periodicamente aggiornato in base al tempo che avrò a disposizione.

Avere a disposizione questo documento non vi autorizza a non venire a lezione, anche perché sarà ovviamente meno esaustivo di tutto il blabla che potrò raccontarvi in classe.

1 Eventi e Probabilità

In questa sezione introdurremo il concetto di *evento* e di *probabilità* .

1.1 Eventi

Un *evento* è una proposizione ben definita che può essere vera o falsa. Esempio:

- E = ”Lancio una moneta ed esce testa”
- E = ”Lancio un dado ed esce un numero pari”
- E = ”Supero l’esame”

Possiamo identificare dei tipi specifici di eventi

- Evento **certo**: si indica con Ω, la proposizione può essere solo vera;
- Evento **impossibile**: si indica con $\emptyset$, la proposizione può essere solo falsa;
- Due eventi A e B si dicono **incompatibili** se uno è necessariamente vero quando l’altro è falso
- Due eventi A e B si dicono **uguali** se quando A è vero anche B è necessariamente vero e viceversa. L’uguaglianza tra eventi si indica con $A = B$.

Possiamo anche introdurre alcune operazioni tra eventi. Assumiamo che A e B siano due eventi, allora

- A **implica** B ($A \subseteq B$): A vero implica B vero. Se $A \subseteq B$ e $B \subseteq A$ allora $A = B$;
- A^c è l’evento **contrario** di A se è falso quando A è vero e viceversa;
- **Unione** o **somma logica** di A e B ($A \vee B$, $A \cup B$) è vero se almeno uno tra A e B è vero;
- **Intersezione** o **prodotto logico** di A e B ($A \wedge B$, $A \cap B$) è vero se entambi gli eventi A e B sono veri. Se due eventi A e B sono incompatibili avremo che $A \cap B = \emptyset$.

Esercizio 1

Un congegno idraulico è costituito da tre valvole come in figura 1. Il congegno funziona se almeno uno dei due rami del circuito funzioni. Il ramo superiore funziona se le valvole A e B funzionano contemporaneamente, il ramo inferiore funziona se la valvola C funziona. Descrivere l’evento E = “Il congegno funziona”.

Introduciamo i tre eventi

1. F_A = “La valvola A funziona”
2. F_B = “La valvola B funziona”
3. F_C = “La valvola C funziona”

Introduciamo anche gli eventi

1. R_{sup} = “Il ramo superiore funziona” $= F_A \cup F_B$
2. R_{inf} = “Il ramo inferiore funziona” $= F_C$

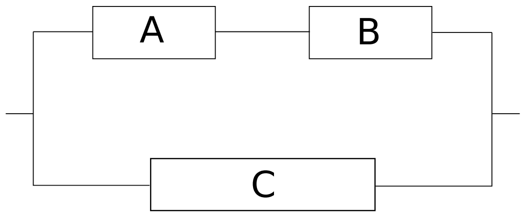

Figura 1: Esercizio 1

A questo punto l'evento E = "Il congegno funziona" si scriverà $E = R_{sup} \cap R_{inf} = (F_A \cup F_B) \cap F_C$.

1.2 Probabilità

In questa sottosezione proveremo a introdurre il concetto di probabilità. Daremo quattro definizione di probabilità.

1. Definizione Classica

Immaginiamo di avere n eventi elementari equiprobabili, conto quante volte h si verifica l'evento E. Allora la probabilità di E sarà

$$P(E) = \frac{h}{n} = \frac{\#\text{casi favorevoli}}{\#\text{casi totali}}. \tag{1}$$

2. Definizion Frequentista

Ho una sequenza di n istanze di un qualche fenomeno ripetibile. Conto quante volte k si verifica l'evento E che mi interessa. La probabilità dell'evento E sarà

$$P(E) = \frac{k}{n}. \tag{2}$$

OBS: più è grande n più è accurata come stima! Ma di questo parleremo poi.

3. Definizione soggettiva (B. De Finetti)

È una misura di fiducia del verificarsi di un evento. Prendiamo come esempio l'evento E = "passo l'esame". Non è possibile dare attribuire una probabilità né di tipo classica né frequentista.

Secondo questa definizione dirò che la probabilità che passi l'esame è del 70% se sono disposta a pagare 70 euro per vincerne 100 qualora dovesse passare l'esame.

Si può formalizzare questa cosa introducendo una quota $p \geq 0$ e una somma $S > 0$. Si dirà che si effettua una scommessa di quota p su un evento E se versando pS si riceve un importo S solo se E si verifica (scommessa vinta) e niente in caso contrario.

Introduciamo anche il guadagno G definito come segue

$$\begin{cases} G_1 = S - pS = S(1-p) & \text{se si verifica E} \\ G_2 = -Sp & \text{se non si verifica E.} \end{cases} \tag{3}$$

La scommessa si dice coerente se $G_1 G_2 \leq 0$, ovvero i due guadagni sono di segno discorde. E questo accade se $0 \leq p \leq 1$. In questo caso p sarà la probabilità di E.

OBS: Quando giocate la schedina alla SNAI fate esattamente questo. La partita del Catania è quotata $q > 1$. Voi andate a giocare una quota $\tilde{S}$, per cui se vincete ricevete $G_1 = q\tilde{S} - \tilde{S} = \tilde{S}(q-1)$. Se perdete, ovviamente, perderete $G_2 = -\tilde{S}$. Se confrontate quanto scritto prima si ha che $p = 1/q$ e $\tilde{S} = Sp$.

4. Definizione Assiomatica (Kolmogorov)

Questa è una definizione più formale di probabilità. Immaginiamo un qualche evento aleatorio e introduciamo tre elementi

- Ω l'insieme dei possibili esiti e $\mathcal{P}(\Omega)$ l'insieme delle sue parti;
- $\mathcal{A}$ la famiglia di sottoinsiemi di Ω che costituiscono gli eventi, $\mathcal{A} \subseteq \mathcal{P}(\Omega)$;
- P una misura della fiduciosità dell'evento della famiglia $\mathcal{A}$.

La terna $(\Omega, \mathcal{A}, P)$ si dice *spazio di probabilità*. Ci sono alcune proprietà degli eventi che è importante elencare:

1. $\emptyset, \Omega \in \mathcal{A}$, ovvero l'evento certo e l'evento impossibile sono eventi;
2. Se $A \in \mathcal{A}$ allora $A^c \in \mathcal{A}$, ovvero se esiste un evento esiste anche l'evento contrario;
3. Se ho tanti eventi $A_1, A_2, \ldots, A_n$, allora $\bigcup_{i=1}^n A_i \in \mathcal{A}$, ovvero l'unione di tanti eventi è ancora un evento.

Queste proprietà fanno sì che l'insieme $\mathcal{A}$ sia una σ-algebra.

A questo punto la probabilità sarà una funzione che va dalla famiglia degli eventi nell'intervallo $[0,1]$. Ovvero

$$P : \mathcal{A} \to [0,1].$$

Questa funzione, per essere coerente, deve godere (assiomaticamente) delle seguenti due proprietà:

1. $P(\Omega) = 1$;
2. $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ se $A_i \cap A_j = 0$.

Da questo si possono dimostrare le seguente proprietà

- $P(A^c) = 1 - P(A)$;
- $P(A) \leq P(B) \forall A, B \in \mathcal{A} : A \subseteq B$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(\bigcup_{i=1}^n) = \sum_{i=1}^n P(A_i) - \sum_{(i,j)\in I_2} P(A_i \cap A_j) + \sum_{(i,j,k)\in I_3} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$.

1.3 Probabilità condizionale

Prima di parlare di probabilità condizionale dobbiamo definire un evento condizionato. Un evento condizionato $A|B$ (A noto B o A dato B), con $B \neq \emptyset$ è un ente a tre valori

1. **Vero** se essendo vero B lo è anche A;
2. **Falso** se essendo vero B A è falso;

3. **Indeterminato** se A è falso.

Ad esempio: A = "mangiare la pizza", B = "cenare fuori", $A|H$ = "mangiare la pizza sapendo di andare a cena fuori"

La probabilità condizionale di A noto H è

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

Due eventi A e B si dicono **indipendenti** se $P(A \cap B) = P(A)P(B)$ e, di conseguenza, $P(A|B) = P(A)$. OBS: quella sopra è condizione necessaria e sufficiente, ovvero se so che i due eventi sono indipendenti, allora la probabilità dell'intersezione sarà il prodotto delle probabilità. Se invece so che la probabilità dell'intersezione sarà il prodotto delle probabilità, allora posso dedurre che gli eventi sono indipendenti.

Siamo pronti per formulare il

Teorema delle probabilità totali

Siano $A_1, A_2, \dots, A_n \in \mathcal{A} : \sum_i A_i = \Omega, A_i \cap A_j = \emptyset \ \forall i \neq j, P(A_i) > 0 \ \forall i$, allora

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

Dimostrazione

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

Esercizio 2

Una stazione radio riceve segnale da due sorgenti A e B. La probabilità che il segnale dalla stazione A sia distorto è 0.1. La probabilita che il segnale dalla stazione B sia distorto è 0.2. Qual è la probaiblità di ricevere un segnale distorto?

Introduciamo i due eventi S_A = "Segnale dalla stazione A" e S_B = "Segnale dalla stazione B". Immaginiamo che il segnale sia equidistribuito tra A e B, per cui $P(S_A)P(S_B) = 0.5 > 0$. Inoltre $S_A \cup S_B = \Omega$. Inoltre $S_A \cap S_B = \emptyset$.

Possiamo utilizzare il teorema delle probabilità totali per valutare la probabilità dell'evento S = "Segnale distorto". Dal testo sappiamo che $P(S|S_A) = 0.1$ e $P(S|S_B) = 0.2$. Per cui avremo

$$P(S) = P(S|S_A)P(S_A) + P(S|S_B)P(S_B) = 0.1 \cdot 0.5 + 0.2 \cdot 0.5 = 0.15.$$

Teorema di Bayes

Dati due eventi $A, B \in \mathcal{A}$ tali che $P(A), P(B) > 0$, si ha che

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

Dimostrazione

Dato che $P(A \cap B) = P(B \cap A)$, sappiamo che $P(B|A) = \frac{P(A \cap B)}{P(A)}$, da cui $P(A \cap B) = P(A|B)P(B)$. Possiamo sostituire questa espressione nella formula per la probabilità condizionale e trovare il nostro risultato.

1.4 Richiami di calcolo combinatorio

Richiamiamo rapidamente alcuni concetti di calcolo combinatorio.

- r-disposizioni: r elementi ordinati scelti da un insieme di n elementi. Ne esistono n^r.
- r-disposizioni semplici: r elementi ordinati scelti da un insieme di n elementi senza ripetizioni. Ne esistono $D_{n,r} = \frac{n!}{(n-r)!}$. se $r = n$ abbiamo le permutazioni che sono $D_{n,n} = n!$.
- r-combinazioni: r elementi non ordinati scelti da un insieme di n elementi. Ne esistono $\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$.
- r-combinazioni semplici: r elementi non ordinati scelti da un insieme di n elementi senza ripetizioni. Ne esistono $C_{n,r} = \binom{n}{r} = \frac{(n)!}{r!(n-r)!}$.

È importante ricordare che dato un insieme di ne elementi e dati degli interi $n_1, n_2, \dots, n_k$ tali che $\sum_{i=1}^{k} n_i = n$, il numero di k partizioni con $n_1, n_2, \dots n_k$ elementi è

$$C^n_{n_1,n_2,\dots,n_k} = \frac{n!}{n_1!n_2!\dots n_k!}$$

1.5 Esercizi

Esercizio 1

Un'urna contiene 5 palline bianche, 6 nere, 4 rosse. Se ne estraggono due. Calcolare che siano dello stesso colore nel caso in cui a) l'estrazione sia istantanea; b) l'estrazione sia con reinserimento.

Svolgimento

Iniziamo contando il numero di palline, quindi poniamo $\text{n}_{\text{palline}} = 15$.

Per calcolare questa probabilità dobbiamo semplicemente fare casi favorevoli su casi totali. Quello che è diverso è che nel primo caso dobbiamo utilizzare le combinazioni semplici, nel secondo le combinazioni. Per cui

a)

Evento favorevole = "sono entrambe rosse" $\bigcup$"sono entrambe bianche"$\bigcup$ "sono entrambe nere"
Evento totale = "estraggo due palline di 15"

Il numero di possibili estrazioni è

$$\text{n}_{\text{totali}} = \binom{15}{2}$$

.

Usando le regole del calcolo combinatorio abbiamo che il numero di combinazioni favorevoli affinché entrambe le palline siano rosse è $\text{n}_{\text{rosse}} = \binom{4}{2}$. Ragionando analogamente abbiamo che il numero di combinazioni restituenti due palline bianche è $\mathrm{n}_{\text{bianche}} = \binom{5}{2}$, mentre quelle con due palline nere è $\mathrm{n}_{\text{rosse}} = \binom{6}{2}$.

Per cui avremo che il numero di casi favorevoli sarà

$$\mathrm{n}_{\text{favorevoli}} = \binom{5}{2} + \binom{6}{2} + \binom{4}{2}$$

Per cui la probaiblità sarà

$$p = \frac{\mathrm{n}_{\text{favorevoli}}}{\mathrm{n}_{\text{totali}}} = \frac{\binom{5}{2} + \binom{6}{2} + \binom{4}{2}}{(15//2)} = 0.2952$$

b)

Possiamo ripetere tutto per il caso con reinserimento. Quello che cambierà saranno le combinazioni danti vita agli eventi favorevoli Usando le regole del calcolo combinatorio abbiamo che il numero di combinazioni favorevoli affinché entrambe le palline siano rosse è $\mathrm{n}_{\text{rosse}} = 4^2$. Ragionando analogamente abbiamo che il numero di combinazioni restituenti due palline bianche è $\mathrm{n}_{\text{bianche}} = 5^2$, mentre quelle con due palline nere è 6^2, per cui avremo

$$\mathrm{n}_{\text{favorevoli}} = 5^2 + 4^2 + 6^2.$$

Mentre il numero di casi totali sarà

$$\mathrm{n}_{\text{totali}} = 15^2.$$

Pertanto la probabilità sarà

$$p = \frac{5^2 + 4^2 + 6^2}{15^2} = 0.3422$$

Esercizio 2

Dati i due eventi A = "Lo studente ha studiato bene" e B = "Lo studente ha passato l'esame" tradurre in simboli i seguenti eventi composti:

1. La probabilità che uno studente abbia studiato bene e passi l'esame è 0.4
2. La probabilità che uno studente che ha studiato bene passi l'esame è 0.8
3. La probabilità che uno studente che non ha studiato bene non passi l'esame è 0.9
4. La probabilità che uno studente che non ha studiato bene passi l'esame è 0.05
5. La probabilità che uno studente che non ha passato l'esame non avesse studiato è di 0.82

Noti $P(\mathrm{A}) = 0.5$ e $P(\mathrm{B}) = 0.45$ calcolare a) $P(\overline{\mathrm{B}}|\mathrm{A})$ e b) $P(\overline{\mathrm{A}} \cap \overline{\mathrm{B}})$

Svolgimento

1. $P(\mathrm{A} \cap \mathrm{B}) = 0.4$
2. $P(\mathrm{B}|\mathrm{A}) = 0.8$
3. $P(\overline{\mathrm{B}}|\overline{\mathrm{A}}) = 0.9$
4. $P(\mathrm{B}|\overline{\mathrm{A}}) = 0.05$
5. $P(\overline{\mathrm{A}}|\overline{\mathrm{B}}) = 0.82$

Calcoliamo adesso le probabilità richieste al punto

a) $P(\overline{\mathrm{B}}|\mathrm{A}) = 1 - P(\mathrm{B}|\mathrm{A}) = 0.2$

b) $P(\overline{\mathrm{A}} \cap \overline{\mathrm{B}}) = P(\overline{\mathrm{A}}|\overline{\mathrm{B}})P(\overline{\mathrm{B}}) = P(\overline{\mathrm{B}}|\overline{\mathrm{A}})P(\overline{\mathrm{A}}) = 0.45$

Esercizio 3

Una "roulette" semplificata consiste in 12 numeri classigicati rossi o neri in base al seguente schema

1	2	3	4	5	6	7	8	9	10	11	12
R	R	N	N	R	N	N	R	N	N	R	R

Siano dati i seguenti eventi:

A = "esce un numero pari" B = "esce un numero rosso" C = "esce un numero ≤ 6 D = "esce un numero ≤ 8

Stabilire se

1. A, B, C sono a due a due indipendenti
2. A, B, C è una famiglia di eventi indipendenti
3. A, B, D è una famiglia di eventi indipendenti

Svolgimento

Ricordiamo che se degli eventi sono indipendenti se il prodotto dell'intersezione è scrivibile come prodotto delle probabilità. Le probabilità le calcoliamo valutando dalla tabella qual è il rapporto casi favorevoli su casi totali. I casi totali sono 12. Ad esempio, i casi favorevoli per l'evento $\mathrm{B} \cap \mathrm{A}$, ovvero "esce un numero rosso pari" sarà $3/12$, in quanto gli unici numeri rossi pari sono 2, 8, 12.

Per cui

1. $P(\mathrm{A} \cap \mathrm{B}) = \frac{1}{4}$, $P(\mathrm{A})P(\mathrm{B}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(\mathrm{A} \cap \mathrm{B})$. Quindi A e B sono indipendenti!

   Ripetiamo il ragionamento per gli altri eventi e troviamo che

   $P(\mathrm{A} \cap \mathrm{C}) = P(\mathrm{A})P(\mathrm{C}) = \frac{1}{4}$ e $P(\mathrm{B} \cap \mathrm{C}) = P(\mathrm{B})P(\mathrm{C}) = \frac{1}{3}$

   Deduciamo quindi che gli eventi sono a due a due indipendenti.

2. Sfruttando il risultato precedente possiamo affermare che A, B, C sono eventi a due a due indipendenti. Per cui per verificare se sono una famiglia di eventi indipendenti è necessario calcolare se $P(\mathrm{A} \cap \mathrm{B} \cap \mathrm{C}) = P(\mathrm{A})P(\mathrm{B})P(\mathrm{C})$. Sempre ussando la tabella troviamo che $P(\mathrm{A} \cap \mathrm{B} \cap \mathrm{C}) = \frac{1}{12}$ e $P(\mathrm{A})P(\mathrm{B})P(\mathrm{C}) = \frac{1}{8}$, per cui gli eventi non costituiscono una famiglia di eventi indipendenti!

3. Provate voi a verificare se A, B, D costituiscono una famiglia di eventi indipendenti!

Esercizio 4

Tutte le borse che si imbarcano su un aereo vengono passate al metal detector allo scopo di individuare eventuali ordigni. È noto che:

- La probabilità che una borsa con un ordigno faccia scattare l'alalrme è 0.99;
- La probabiità che una borsa senza ordigno faccia scattare l'allarme è 0.05;
- Una borsa ogni 5000 contiene un ordigno.

Calcolare:

1. Con che probabilità scatterà l'allarme;
2. Qual è la probabilità che una borsa che ha fatto scattare l'allarme contenga un ordigno;
3. Di quanto aumenta la probabilità che una borsa contenga l'ordigno sapendo che ha fatto scattare l'allarme?

Svolgimento

Per risolvere il problema dobbiamo prima identificare gli eventi che ci interessano. Chiamiamo

A = "La borsa contiene l'ordigno"

B = "La borsa ha fatto scattare l'allarme"

Per ipotesi noi conosciamo che:

- $P(\mathrm{A}) = 1/5000$
- $P(\mathrm{B}|\mathrm{A}) = 0.99$
- $P(\mathrm{B}|\overline{\mathrm{A}}) = 0.05$

Siamo pronti per rispondere.

1. Per rispondere a questa domanda utilizziamo il teorema delle probabilità totali. Per cui

$$P(\mathrm{B}) = P(\mathrm{B}|\mathrm{A})P(\mathrm{A}) + P(\mathrm{B}|\overline{\mathrm{A}})P(\overline{\mathrm{A}})$$

   Ricordando che $P(\overline{\mathrm{A}}) = 1 - P(\mathrm{A})$ e sostituendo i numeri troviamo $P(\mathrm{B}) = 0.0502$.

2. Per rispondere a questa domanda dobbiamo calcolare la probabilità condizionata $P(\mathrm{A}|\mathrm{B})$. Per farlo utilizziamo il teorema di Bayes e troviamo che

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|\mathrm{A})P(\mathrm{A})}{P(\mathrm{B})} = 0.0039$$

3. Noi conosciamo la probaiblità che una borsa contenga l'ordigno, ovvero $P(\mathrm{A})$. Conosciamo anche la probabilità che una borsa che ha fatto scattare l'allarme contenga l'ordigno, ovvero $P(\mathrm{A}|\mathrm{B})$. Ovviamente mi aspetto che la seconda sia più grande della prima. Mi calcolo il fattore di proporzionalità tra le due quantità

$$\gamma = \frac{P(\mathrm{A}|\mathrm{B})}{P(\mathrm{A})} = 19.5.$$

Questo significa che l'informazione che l'allarme è scattato ha fatto aumentare di quasi venti volte la probabilità che la bomba sia scattata!

2 Variabili aleatorie

Il concetto di variabile aleatoria generalizza quello di evento.

Matematicamente, dato uno spazio di probabilità $(\Omega, \mathcal{A}, \mathcal{P})$, definiamo **variabile aleatoria** un'applicazione $X : \Omega \to \mathbb{R}$ t.c. $\forall t \in \mathbb{R}$ l'insieme $\{\omega : X(\omega) \leq t\} \in \mathcal{A}$, ovvero è un evento possibile.

Da ora in poi useremo una notazione abbreviata per cui $\{X \leq t\} \equiv \{\omega : X(\omega) \leq t\}$.

In parole povere, una variabile aleatoria è una funzione che prende un evento, ovvero una proposizione, e vi associa un numero reale.

Esempio. Se considero il lancio di un dado ho sei possibili eventi a cui posso associare un numero reale:

$$\text{A} = \text{“Esce il numero 1”} \to 1$$

$$\text{B} = \text{“Esce il numero 2”} \to 2$$

$$\text{C} = \text{“Esce il numero 3”} \to 3$$

$$\text{D} = \text{“Esce il numero 4”} \to 4$$

$$\text{E} = \text{“Esce il numero 5”} \to 5$$

$$\text{F} = \text{“Esce il numero 6”} \to 6$$

L'utilità delle variabili aleatorie è legata al fatto che con i numeri reali sappiamo operare, per cui è più facile ottenere determinati risultati.

In realtà usiamo il concetto di variabile aleatoria ogni giorno senza rendercene conto. Se vi chiedessi di calcolare il valore medio ottenuto dal lancio di un dado voi mi rispondereste 3.5, ottenuto banalmente usando la formula per la media $(1+2+3+4+5+6)/6$. Nel darmi questa risposta voi avete inconsapevolmente associato all'evento “esce il numero i” il numero i.

Chiamiamo $A \in \mathbb{R}$ un generico sottoinsieme dei numeri reali. Definiamo quini **Legge o distribuzione di X**

$$P(X \in A) \ \ \forall A \in \mathbb{R}.$$

In sostanza, la legge di una variabile aleatoria (da ora in poi v.a.) mi dirà la probabilità che ogni evento si verifichi.

Ci sono due tipi di v.a.

- **Discreta**: L'insieme $X(\Omega)$ è finito o numerabile (ovvero il numero di esiti è discreto o numerabile) – Esempio: lancio di un dado;
- **Continua**: L'insieme $X(\Omega)$ è infinito (ovvero il numero di esiti è infinito) – Esempio: tempo decadimento sostanza radioattiva;

2.1 Variabili aleatorie discrete

Concentriamoci adesso sulle variabili aleatorie discrete. Sia X una v.a. discreta. Consideriamo $A \in \mathbb{R}$, sarà a sua volta un insieme discreto che si può scrivere come unione dei suoi elementi

a, ovvero $A = \bigcup_{a \in A}\{a\}$. Per cui avrò che

$$P(X \in A) = P(X \in \bigcup_{a \in A}\{a\}) = \sum_{a \in A} P(X = a).$$

Definisco **densità discreta** di X la funzione

$$p(x) = P(X = x).$$

La densità discreta gode di due proprietà:

1. $p(x) \geq 0 \ \ \forall x \in \mathbb{R}$;
2. $\sum_{x \in \mathbb{R}} p(x) = \sum_{i=1}^{\infty} p(x_i) = 1$.

Nell'ultima proprietà ho sfruttato il fatto che l'insieme è discreto per cui è equivalente sommare sui valori in $\mathbb{R}$ o sui numeri naturali. Da ora in poi le considererò interscambiabili.

Definisco **funzione di ripartizione** (da ora in poi f.r.) $\forall t \in \mathbb{R}$

$$F_X(t) = P(X \leq t) = \sum_{x \leq t} p(x).$$

Osservo che $F_X : \mathbb{R} \to [0, 1]$, ovvero associa una probabilità a ogni istanza della nostra variabile aleatoria.

Definisco **speranza matematica** la quantità

$$E[X] = \sum_{i=1}^{\infty} x_i p(x_i).$$

La speranza matematica gode di alcune proprietà:

1. $E[cX] = cE[X] \ \forall c \in \mathbb{R}$;
2. $E[X + Y] = E[X] + E[Y]$;
3. se $p(x \leq y) = 1 \Rightarrow E[X] \leq E[Y]$;
4. $|E[X]| \leq E[|X|]$;
5. Se X e Y sono v.a. indipendenti (le definiremo dopo) $E[YX] = E[X]E[Y]$

In realtà la speranza matematica si può definire se e solo se *esiste finita*, ovvero se e solo se vale la proprietà che $E[X] = \sum_{i=1}^{\infty} |x_i|p(x_i) < \infty$, dove $| \cdot |$ indica il valore assoluto.

Se considero una generica funzione $f : \mathbb{R} \to \mathbb{R}$ t.c. $f(X)$ è una v.a. a sua volta, se esiste finita (non ripeterò l'espressione, avete capito il concetto), posso calcolarne la speranza matematica e ottenere $E[f(X)] = \sum_{i=1}^{\infty} f(x_i)p(x_i)$.

Utilizzo quanto sopra per definire il **momento di ordine k** come il valore di aspettazione di X^k, ovvero

$$E[X^k] = \sum_{i=1}^{\infty} x_i^k p(x_i).$$

Possiamo anche definire il **momento centrato di ordine k** come

$$E[(X - E[X])^k] = \sum_{i=1}^{\infty} (x_i - E[X])^k p(x_i).$$

Il più importante tra i momenti centrati è il momento di ordine 2 noto anche come **varianza**, $\sigma_X^2 = Var(X) = E[(X - E[X])^2]$. La varianza misura quanto in media i dati sono lontani dalla speranza matematica, ovvero ci dà un'indicazsione di quanto siano "sparpagliati". Si indica con il simbolo σ^2 e gode delle seguenti proprietà

1. $\mathrm{Var}(X) = E[X^2] - E[X]^2$;
2. $\mathrm{Var}(aX) = a^2\ \mathrm{Var}(X)$;
3. $\mathrm{Var}(a + X) = \mathrm{Var}(X)$.

Definiamo **deviazione standardi** la radice quadrata della varianza $\sigma_X = \sqrt{\sigma_X^2}$.

Date due v.a. X, Y definiamo **covarianza tra X, Y**

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

La covarianza tiene conto di come cambia X relativamente a Y. In particolare se $\mathrm{Cov}(X, Y) = 0$ le due v.a. sono indipendenti.

Nota la covarianza, possiamo scrivere

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X, Y)$$

.

Esempio: X v.a. che segue una distribuzione discreta uniforme, ovvero assume uno degli n elementi in $A = \{1, 2, \ldots, n\}$ con ugual probabilità $p(x) = \frac{1}{n}$.

La funzione di ripartizione sarà una funzione discontinua (a scalini) definita così

$$F(t) = \begin{cases} 0 & \text{se } t < 1 \\ \frac{1}{n} & \text{se } 1 \leq t < 2 \\ \frac{2}{n} & \text{se } 2 \leq t < 3 \\ \ldots \\ 1 & \text{se} t \geq n \end{cases}$$

La speranza matemtica sarà

$$E[X] = \sum_{i \in A} i \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

Analogamente possiamo calcolare la varianza

$$\sigma_X^2 = \frac{1}{n} \sum_{i=1}^{n} i^2 - E[X]^2 = \frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2 - 1}{12}$$

Nelle prossime sezioni vedremo più in dettaglio alcune distribuzioni note che ci torneranno comode per alcune applicazioni.

2.1.1 Distribuzione binomiale

La descrizione binomiale descrive un fenomeno che avviene secondo uno schema successo/insuccesso in cui si considerano n prove **indipendenti** con due soli esiti.

- n è il numero di prove
- k è il numero di successi ($k \leq n$)
- p è la probabilità di successo
- $q = 1 - p$ è la probabilità di insuccesso

X è la v.a. che conta il numero di successi ottenuti. Sia $p(k) = P(X = k)$ la probabilità di ottenere k successi. Proviamo a calcolare questa probabilità a partire dalla definizione classica.

Per farlo dobbiamo prima introdurre gli n eventi

A$-i$ = "successo all'i-esima prova".

Introduciamo pure l'evento

A = "k successi nelle prime k prove"

Calcoliamo la $P(A) \equiv P(\mathrm{A}_1 \cap \mathrm{A}_2 \cap \cdots \cap \mathrm{A}_k \cap \overline{\mathrm{A}}_{k+1} \cap \cdots \cap \overline{\mathrm{A}}_n)$. Dato che gli eventi sono indipendenti avremo

$$P(\mathrm{A}_1 \cap \mathrm{A}_2 \cap \cdots \cap \mathrm{A}_k \cap \overline{\mathrm{A}}_{k+1} \cap \cdots \cap \overline{\mathrm{A}}_n) = P(\mathrm{A}_1)P(\mathrm{A}_2)P(\mathrm{A}_k)P(\overline{\mathrm{A}}_{k+1}) \dots P(\overline{\mathrm{A}}_n) = p^k(1-p)^{n-k}.$$

Notiamo che, data l'indipendenza degli eventi, la probabilità di ottenere k successi nelle prime k prove sarà uguale alla probabilità di ottenere k successi in un qualsiasi ordine. Per cui, per trovare la probabilità di ottenere esattamente k successi in n prove, ci basterà contare in quanti modi è possibile avere k successi in n prove e moltiplicarlo per la probabilità che abbiamo appena calcolato. In questo modo arriveremo a

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Diremo che una variabile aleatoria X segue una **legge binomiale** di parametri n e p se ha densità discreta

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{se} \quad x = 0, 1, \dots, n \\ 0 & \text{altrimenti} \end{cases}$$

In questo caso si scriverà $X \sim B(n, p)$.

Per calcolare la media conviene prima considerare un caso particolare, ovvero quello della legge di **Bernoulli**. La legge di Bernoulli non è altro che una binomiale con $n = 1$, ovvero descrive la probabilità di successo in un'unica prova (testa o croce?).

Chiamiamo $X_i \sim B(1, p)$ t.c

$$X_i = \begin{cases} 1 & \text{se successo all'iesima prova} \\ 0 & \text{altrimenti} \end{cases}$$

Da definizione si ha che

$$E[X_i] = \sum_{i=1}^{n} x_i p(x_i) = \binom{1}{1} 1 \cdot p + \binom{1}{0} 0 \cdot (1-p) = p$$

Sfruttando le proprietà della media possiamo quindi calcolare

$$E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$$

Possiamo operare analogamente per ricavare la varianza. Iniziamo valutando $\mathrm{Var}(X_i)$. Per farlo ci serve

$$E[X_i^2] = \sum_{i=1}^{n} x_i^2 p(x_i) = \binom{1}{1} 1^2 \cdot p + \binom{1}{0} 0^2 \cdot (1-p) = p$$

Da questo possiamo semplicemente calcolare

$$\mathrm{Var}(X_i) = E[X_i^2] - E[X_i]^2 = p - p^2 = p(1-p)$$

Sfruttando l'ipotesi che le prove debbano essere indipendenti, possiamo calcolare

$$\mathrm{Var}(X) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Prima di proseguire facciamo un paio di esempi per prendere familiarità.

Esercizio 1

Nella trascrizione di tre file si sono verificati 5 errori distribuiti randomicamente tra i tre file. Qual è la probabilità che in un singolo file a) ci sia un solo errore b) ci siano più di tre errori?

Svolgimento

Questo fenomeno rientra all'interno dello schema successo/insuccesso. Dato uno specifico file, il mio successo sarà trovare un errore. Dato che gli errori sono distribuiti randomicamente tra i tre file la mia probabilità di successo sarà $p = 1/3$. Il numero di prove, ovviamente, sarà il numero di errori, per cui $n = 5$.

La variabile aleatoria che descrive il numero di errori in un singolo file seguirà dunque una distribuzione binomiale di parametri n e p, ovvero $X \sim B(n,p)$. Per cui:

a) La probabilità che ci sia un solo errore è $P(X = 1)$.

$$P(X = 1) = \binom{5}{1} \left(\frac{1}{3}\right) \left(\frac{2}{3}\right)^4 = 0.3292$$

b) La probabilità che ci siano più di tre errori si riduce alla probabilità che ci siano o quattro o cinque errori, ovvero $P(X > 3) = P(X = 4) + P(X = 5)$. Sempre usando l'espressione della densità binomiale troviamo

$$P(X > 3) = \binom{5}{4} \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right) + \binom{5}{5} \left(\frac{1}{3}\right)^5 = 0.0453$$

Esercizio 2

Un segnale è costituito da 5 bits. La probabilità di ricevere un singolo bit distorto è $p = 0.1$. Qual è la probabilità di ricevere un segnale distorto?

Svolgimento

Ecco un altro problema che si inquadra nello schema successo/insuccesso. In questo caso il successo è osservare un bit distorto, la probabilità di successo è $p = 0.1$ mentre il numero di prove è il numero di bit, ovvero $n = 5$.

La variabile aleatoria che conta il numero di bit distorti seguirà pertanto una distribuzione di Bernoulli $X \sim B(n, p)$.

La probabilità che cerchiamo è che il segnale sia distorto, ovvero che almeno un bit sia distorto. Introduciamo l'evento S="segnale distorto" e avremo

$$P(S) = 1 - P(X = 0) = 1 - \binom{5}{0} 0.1^0 \, 0.9^5$$

2.1.2 Distribuzione multinomiale

È una generalizzazione della distribuzione binomiale al caso in uci ho n prove ripetute e indipentneti e m possibili esiti (esempio: il lancio del dado). Definiamo q_i la probabilità di ottenere l'i-esimo risultato. ovviamente $\sum_{i=1}^{m} q_i = 1$.

Introduciamo la variabile aleatoria Y_i che conta quante volte si è verificato l'iesimo risultato. Possiamo quindi definire $Y = (Y_1, Y_2, \ldots, Y_m)$, che sarà una v.a. *multivariata* che conta quante volte si è verificato ciascun esito. L v.a. Y è essenzialmente un vettore di m v.a.! La v.a. seguirà una **legge multinomiale** e si indicherà così $Y \sim B(n, q_1, q_2, \ldots, q_m)$.

Dato un generico vettore $\overline{\omega} = (\omega_1, \omega_2, \ldots, \omega_m)$ siamo interessati in $P(Y = \overline{\omega}) = P(Y_1 = \omega_1, Y_2 = \omega_2, \ldots, Y_m = \omega_m)$.

Ragionando in maniera analoga al caso della binomiale (salteremo la dimostrazione), si può vedere che

$$P(Y = \overline{\omega}) = \frac{n!}{\omega_1! \omega_2! \ldots \omega_m!} q_1^{\omega_1} q_2^{\omega_2} \ldots q_m^{\omega_m}$$

.

Ovviamente, se $m = 2$ ritroviamo la distribuzione binomiale.

Esempio

Calcolare la probabilità che lanciando un dado non truccato quattro volte esca tre volte 6 e una volta 2.

Svolgimento

La variabile che descrive questo evento segue una multinomiale in cui ogni risultato ha la stessa probabilità di accadere $p = 1/6$. La probabilità che cerchiamo, quindi sarà

$$P(Y = \{0, 1, 0, 0, 0, 3\}) = \frac{4!}{0!1!3!0!0!0!3!} \frac{1}{6} \frac{1}{6^3} = \frac{4}{6^4}$$

OBS: dato che il dado ha tutti i risultati equiprobabili avremmo potuto ricavare lo stesso risultato utilizzando la definizione classica di probabilità!

2.1.3 Distribuzione di Poisson

Questa distribuzione si usa quando in uno schema successo/insuccesso si ha un alto numero di prove e una bassa possibilità di successo. Essa dipende da un unico parametro λ. La distribuzione di Poisson approssima la distribuzione binomiale per $n \gg 1$ e $p \ll 1$ e in questo caso $\lambda = np$.

Una v.a. X segue una legge di Poisson di parametro λ se ha distribuzione

$$p(x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!} & \text{se} \quad x = 0, 1, 2, \ldots, n \\ 0 & \text{altrimenti} \end{cases}$$

In questo caso si scriverà $X \sim \text{Pois}(\lambda)$.

Dimostriamo intanto che è effettivamente una densità

$$\sum_{x\in\mathbb{R}} p(x) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda}.$$

Nell'ultimo passaggio abbiamo risommato lo sviluppo in serie di McLaurin dell'esponenziale (in caso, cercate tra gli appunti di analisi 1!).

Adesso invece dimostriamo che questa distribuzione è esattamente la stessa che otterremmo da una binomiale nel limite $n \to \infty$ e $p \to 0$. Se $X \sim B(k, n)$. Poniamo $\lambda = np$, allora

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

Osserviamo che

- $\frac{n!}{n^k(n-k)!} = \frac{n(n-1)(n-2)\ldots(n-k)}{n^k} \xrightarrow[n\to\infty]{} 1$
- $\left(1-\frac{\lambda}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-\lambda}$ (anche qui, appunti di analisi 1!)
- $\left(1-\frac{\lambda}{n}\right)^k \xrightarrow[n\to\infty]{} 1$

Per cui avremo che

$$P(X = k) \xrightarrow[n\to\infty, p\to 0]{} \frac{\lambda^k}{k!}e^{-\lambda}$$

N.B. Anche se non abbiamo utilizzato esplicitamente, il limite $p \to 0$ è molto importante! Infatti questo limite permette di avere $\lambda \sim 1$, altrimenti divergerebbe al divergere di n!

Calcoliamo adesso la speranza matematica e la varianza

$$E[X] = e^{-\lambda}\sum_{x=0}^{\infty} x\frac{\lambda^x}{x!} = e^{-\lambda}\lambda\sum_{x=0}^{\infty}\frac{\lambda^{(x-1)}}{(x-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

Analogamente possiamo calcolare

$$\begin{aligned}E[X^2] =&e^{-\lambda}\sum_{x=0}^{\infty} x^2\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty} x(x-1+1)\frac{\lambda^x}{x!}\\ =&e^{-\lambda}\sum_{x=0}^{\infty} x\frac{\lambda^x}{x!} + e^{-\lambda}\sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x}{x!} = \lambda + e^{-\lambda}\lambda^2\sum_{x=0}^{\infty}\frac{\lambda^{(x-2)}}{(x-2)!}\\ =&\lambda + e^{-\lambda}\lambda^2 e^{\lambda} = \lambda+\lambda^2\end{aligned}$$

Di conseguenza possiamo calcolare la varianza usando

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

Per concludere enunciamo una proprietà importante delle v.a. di Poisson. Se $X \sim \mathrm{Pois}(\lambda)$ e $Y \sim \mathrm{Pois}(\mu)$ sono variabili indipendenti, allora $X + Y \sim \mathrm{Pois}(\lambda + \mu)$.

2.1.4 Distribuzione ipergeometrica

Questa distribuzione si usa per descrivere il caso di estrazioni senza rimpiazzo, ovvero il caso in cui le singole estrazioni non sono indipendenti ma sono influenzate dalle precedenti.

Immaginiamo di avere m oggetti di cui b hanno una data proprietà e r no. Voglio estrarre $n \leq b + r$ di questi oggetti e voglio calcolare la probabilità che abbiano la proprietà b. Sia X la v.a. che descrive questo processo, voglio calcolare $P(X = x)$ $\forall x = 0, 1, 2, \ldots, n$.

Sostanzialmente dobbiamo contare quanti sono i casi favorevoli e quanti quelli possibili. Usando la combinatoria stimo:

1. In quanti modi posso scegliere x oggetti tra $b \to \binom{b}{x}$
2. In quanti modi posso scegliere $n - x$ oggetti tra $r \to \binom{r}{n-x}$
3. In quanti modi posso scegliere n oggetti tra $b + r \to \binom{b+r}{n}$

Raccogliendo tutto avremo che

$$p(x) = \begin{cases} \dfrac{\binom{b}{x}\binom{r}{n-x}}{\binom{b+r}{n}} & \text{if} \quad x = 0, 1, \ldots, n \\ 0 & \text{altrimenti} \end{cases}$$

Senza entrare troppo nel dettaglio dei conti si può dimostrare che

$$E[X] = \frac{nb}{b+r} \quad \mathrm{Var}(X) = \frac{nbr}{(b+r)^2}\frac{b+r-n}{b+r-1}.$$

2.1.5 Distribuzione geometrica

Una variabile aleatoria X segue una legge geometrica di parametro p, con $0 \leq p \leq 1$ se ha densità

$$p(x) = \begin{cases} p(1-p)^{x-1} & \text{Se } \; x = 1, 2, \ldots, n \\ 0 & \text{altrimenti} \end{cases}$$

In questo caso si scriverà $X \sim \mathrm{Geom}(p)$.

Notiamo subito che, a differenza delle precedenti distribuzioni, questa densità non è definita per $x = 0$. In realtà è possibile definire una densità geometrica, in maniera del tutto equivalente, che contenga anche lo 0, ma per evitare confusione ci limiteremo a considerare questa.

Per prima cosa dimostriamo che questa è realmente una densità

$$\sum_{x=1}^{\infty} p(x) = \sum_{x=1}^{\infty} p(1-p)^{x-1} = p \sum_{j=0}^{\infty} (1-p)^j = \frac{p}{1-(1-p)} = 1.$$

Questa distribuzione è intimamente legata alla distribuzione binomiale. Introduciamo T, v.a. che valuta il **tempo di primo successo** , ovvero il numero di tentativi necessario per osservare il primo successo in un processo binomiale. Vogliamo calcolare la probabilità $P(T = n)$. Per prima cosa osserviamo che $P(T > n-1) = P(T > n) + P(T = n)$. Pertanto, $P(T = n) = P(T > n-1) - P(T > n)$. Per valutare queste due probabilità, introduco $X^{(n)} \sim B(n, p)$.

Allora

$$P(T > n) = P(X^{(n)} = 0) = \binom{n}{0} (1-p)^n$$

Analogamente avremo

$$P(T > n-1) = (1-p)^{n-1}$$

Per cui, raccogliendo tutto assieme ritroviamo

$$P(T = n) = (1-p)^{n-1} - (1-p)^n = (1-p)^{n-1}(1-(1-p)) = p(1-p)^{n-1}.$$

La proprietà sicuramente più importante di cui gode la legge geometrica è la **proprietà di mancanza di memoria**. Ovvero, l'occorrere del primo successo è indipendente dalla storia precedente (esistono i numeri ricorrenti al lotto?).

In formule, questa proprietà si esprime così:

$$P(T = m + n \mid T > n) = P(T = m) \quad \forall n, m > 0 \in \mathbb{N}.$$

Per dimostrarla basta utilizzare alcune proprietà delle probabilità condizionali che abbiamo già incontrato in precedenza, ovvero:

$$P(T = n+m \mid T > n) = \frac{P(T = n+m \; , T > n)}{P(T > n)} = \frac{P(T = m+n)}{P(T > n)} = \frac{p(1-p)^{m+n}}{(1-p)^n} = p(1-p)^m.$$

Al volo, segnaliamo che per questa distribuzione si ha

$$E[X] = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

2.2 Variabili aleatorie continue

Nel caso in cui $X(\Omega)$ sia continuo la v.a. si dice **continua**. Nel caso di v.a. continue, la funzione di ripartizione sarà una funzione

$$F_X : \mathbb{R} \to [0,1], \text{ t.c. } F_X(t) = P(X \leq t).$$

La funzione di ripartizione gode delle seguenti quattro proprietà

1. $0 \leq F_X(t) \leq 1 \ \forall t \in \mathbb{R}$;
2. $F_X(t_1) \leq F_X(t_2) \ \forall t_1, t_2 \in \mathbb{R}, \text{ t.c. } t_1 \leq t_2$ (monotonia);
3. $\lim_{t\to-\infty} F_X(t) = 0, \ \lim_{t\to\infty} F_X(t) = 1$;
4. $\lim_{\varepsilon\to 0^+} F_X(t+\varepsilon) = F_X(t)$ (continuità a destra).

È importante notare che in generale la funzione di ripartizione può non essere continua. Sia t un punto di discontinuità, data l'ipotesi di monotonicità, sappiamo che esistono finiti i seguenti limiti

$$\lim_{\varepsilon\to 0^-} F_X(t+\varepsilon) = F_X(t^-) \qquad \lim_{\varepsilon\to 0^+} F_X(t+\varepsilon) = F(t^+).$$

In questo caso avremo che

$$P(X = t) = F_X(t^+) - F_X(t^-).$$

OBS: se la funzione di ripartizione è continua, la probabilità che X assuma un singolo valore t è zero. Questo non deve sorprenderci, perché lavorando nel continuo quello che è importante non è il singolo valore quanto gli intervalli. In generala, se X è una v.a. continua, varrà la seguente catena di uguaglianze:

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

Una v.a. si dice **assolutamente continua** se ammette **densità** $f(x)$ tale che

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

con

$$f(x) \geq 0 \ \forall x \in \mathbb{R}, \text{ t.c. } \int_{-\infty}^{\infty} f(x)dx = 1.$$

OBS (puramente matematica): la densità non è unica, dato che è definita a meno di un sottoinsieme di misura nulla!

Se esiste finita, ovvero se $\int_{-\infty}^{\infty} |x| f(x)dx < \infty$, posso definire la speranza matematica di una v.a. continua come

$$E[X] = \int_{-\infty}^{\infty} x f(x)dx.$$

Analogamente, se esiste finita posso definire la varianza di una v.a. continua

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x)dx < \infty.$$

Se esistono finiti, posso definire tutti i momenti centrati e non di ogni ordine analogamente a quanto fatto per le variabili discrete semplicemente sostituentdo $\sum_{x\in\mathbb{R}} \mapsto \int_{-\infty}^{\infty} dx$ e $p(x) \mapsto f(x)$.

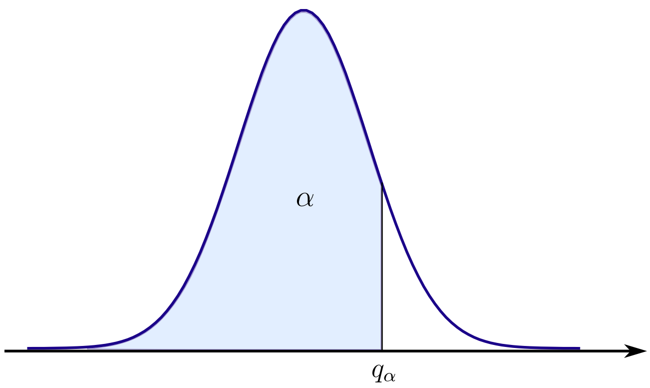

Figura 2: Definizione grafica di quantile di ordine α

Definisco **quantile di ordine** α la quantità

$$q_\alpha = \sup\{r \in \mathbb{R} : F_X(r) = \alpha\}.$$

In parole semplici, il quantile di ordine α è quel valore possibile della variabile aleatoria che divide l'area sottesa alla distribuzione in due parti di ampiezza α (a sinistra) e $1-\alpha$ (a destra). Va da sé che $F_X(q_\alpha) = \alpha$.

2.2.1 Distribuzione uniforme

Una v.a. segue una legge uniforme in $[a, b] \in \mathbb{R}$ se ha densità

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se} \quad x \in [a, b] \\ 0 & \quad \text{altrimenti.} \end{cases}$$

In questo caso si scriverà $X \sim U([a, b])$. Notiamo che questa distribuzione descrive una variabile aleatoria che può assumere tutti i valori nell'intervallo $[a, b]$ con uguale probabilità.

Verifichiamo che sia effettivamente una densità:

$$\int_{-\infty}^{\infty} dx f(x) = \frac{1}{b-a}\int_a^b dx = 1.$$

Possiamo calcolare, usando la definizione, la funzione di ripartizione $F_X(t) = \int_{-\infty}^{t} f(x)dx = \int_a^t f(x)dt$ e trovare

$$F_X(t) = \begin{cases} 0 & \text{se } t < a \\ \frac{t-a}{b-a} & \text{se } t \in [a, b] \\ 1 & \text{se } t > b \end{cases}$$

OBS: è una funzione continua in $\mathbb{R}$!

Usando la definizione possiamo calcolare anche la media

$$E[X] = \int_{-\infty}^{\infty} dx\ xf(x) = \int_a^b dx\ \frac{x}{b-a} = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

e la varianza

$$Var(X) = E[X^2] - E[x]^2 = \int_a^b dx\ \frac{x^2}{b-a} - \frac{(a+b)^2}{4} = \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

2.2.2 Legge normale standard

Eccoci alla distribuzione più importante di tutto il corso: la distribuzione normale standard. Una v.a. X segue una legge normale standard o gaussiana se ha densità

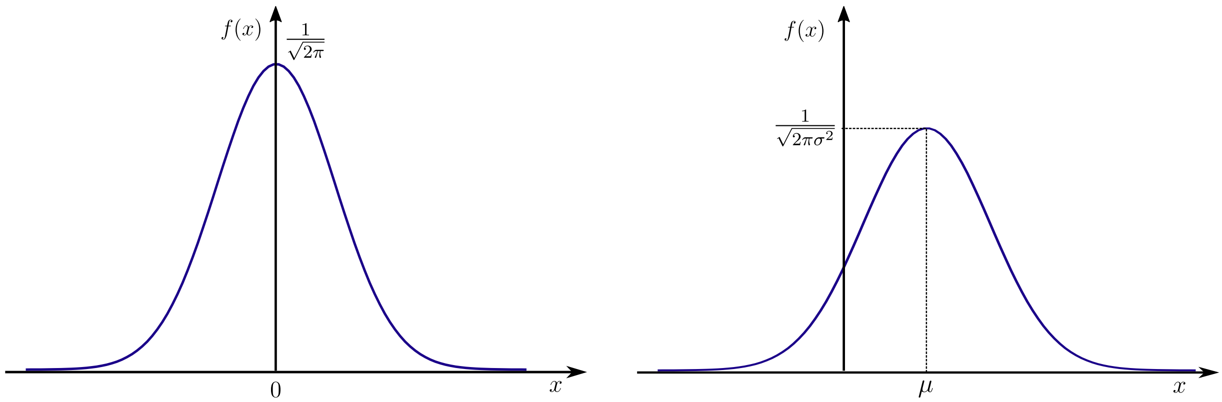

Figura 3: A sinistra, normale standard. A destra, normale di media μ e varianza σ^2

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

In questo caso si scriverà $X \sim N(0,1)$.

La distribuzione normale standard (pannello a sinistra in Fig. 3) è una distribuzione simmetrica rispetto allo 0. Per questo motivo si è ha che $E[X] = 0$. Proviamo a dimostrarlo

$$\begin{aligned} E[X] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\ x e^{-\frac{x^2}{2}} = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\ \frac{d}{dx} e^{-\frac{x^2}{2}} = -\frac{1}{\sqrt{2\pi}} \left[e^{\frac{x^2}{2}} \right]_{-\infty}^{\infty} \\ &= -\frac{1}{\sqrt{2\pi}} \left[\lim_{x\to\infty} \left(e^{-\frac{x^2}{2}} \right) - \lim_{x\to-\infty} \left(e^{-\frac{x^2}{2}} \right) \right] = 0. \end{aligned}$$

Analogamente si può dimostrare che $\mathrm{Var}(X) = 1$.

Dalla simmetria della distribuzione avremo che i quantili della normale standard,che da ora in poi indicheremo con ϕ_α sono simmetrici rispetto allo zero, ovvero $\phi_\alpha = -\phi_{1-\alpha}$.

Data $X \sim N(0,1)$ possiamo costruire una v.a. Y che segue una normale di parametri reali μ e $\sigma^2 > 0$ attraverso la relazione

$$Y = \sigma X + \mu$$

. In questo caso si scriverà $Y \sim N(\mu, \sigma^2)$.

La densità di tale variabile aleatoria sarà

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

Notiamo che, notele proprietà di media è varianza, è molto semplice calcolare media

$$E[Y] = E[\sigma X + \mu] = \sigma E[X] + \mu = \mu$$

e varianza

$$\text{Var}(Y) = \text{Var}(\sigma X + \mu) = \sigma^2 \text{Var}(X) = \sigma^2$$

Prima di concludere elenchiamo alcune proprietà interessanti (e utili) della distribuzione normale.

- In ottima approssimazione $f_Y(x) \approx 0$ quando $|x - \mu| > 3\sigma$. Questa si chiama **legge dei tre sigma**.
- Se $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$ sono indipendenti, allora $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- Se q_α è il quantile di ordine α di $Y \sim N(\mu_1, \sigma^2)$ e ϕ_α quello di $X \sim N(0, 1)$, allora $q_\alpha = \sigma\phi_\alpha + \mu$.

2.2.3 Legge esponenziale

Questa legge descrive la "durata di vita" di un fenomeno che "non invecchia", ossia di un fenomeno privo di memoria. Ad esempio, il tempo di decadimento di una particella radioattiva.

La densità di tale distribuzione è

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0. \end{cases}$$

Notiamo per prima cosa che è una densità:

$$\int_{\mathbb{R}} f(x)dx = \int_0^\infty \lambda e^{-\lambda x} dx = -\int_0^\infty \frac{d}{dx} e^{-\lambda x} = -\left[e^{-\lambda x}\right]_0^\infty = 1.$$

Integrando per parti si può trovare

$$E[X] = \frac{1}{\lambda}, \qquad \text{Var}(X) = \frac{1}{\lambda^2}.$$

È interessante calcolare la funzione di ripartizione

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

È utile introdurre la **funzione di sopravvivenza** $S(t) = 1 - F_X(t) = e^{-\lambda t}$. Notiamo che essa non è altro che la probaiblità di osservare il fenomeno oltre un tempo t, infatti $S(t) = 1 - F_X(t) = 1 - P(X \leq t) = P(X > t)$.

Un'altra osservazione che va fatta è che la distribuzione esponenziale può essere messa in relazione con la distribuzione geometrica. In un certo senso è la "versione continua" della distribuzione geometrica. È possibile ritrovare dall'una l'altra sostituendo $p \leftrightarrow \lambda$ e $n \leftrightarrow x$.

Proprio come la distribuzione geometrica, la distribuzione esponenziale gode della **proprietà di mancanza di memoria**, ovvero $P(X \geq x + t | X \geq x) = P(X \geq t) = S(t)$.

$$P(X \geq x + t | X \geq x) = \frac{P(X \geq x + t, X \geq x)}{P(X \geq x)} = \frac{P(X \geq x + t)}{P(X \geq x)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda x}} = e^{-\lambda t}.$$

2.2.4 Legge del χ^2

Una v.a. X segue una legge del χ^2 a n gradi di libertà, $X \sim \chi^2(n)$, se ha densità

$$f(x) = \begin{cases} \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} & \text{se } \; x > 0 \\ 0 & \text{altrimenti,} \end{cases}$$

dove $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx \; \forall \alpha > 0$ è la funzione Gamma di Eulero. Questa distribuzione è importante per applicazioni nei test d'ipotesi che vedremo in seguito. Pertanto elenchiamo alcune proprietà

- Se $X \sim N(0,1) \Rightarrow X^2 \sim \chi^2(1)$
- Se $X_1, X_2, \ldots, X_n \sim N(0,1) \Rightarrow \sum_{i=1}^n X_i^2 \sim \chi^2(n)$
- Se $X \sim \chi^2(n)$ e $Y \sim \chi^2(m)$, allora $X + Y \sim \chi^2(n+m)$.

2.2.5 Legge t di Student

Una v.a. X segue una legge ti di Student a n gradi di libertà, $X \sim T(n)$ se ha densità

$$f(x) = \frac{A_0}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}, \quad A_0 = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

Anche la legge t di Student ha un test d'ipotesi dedicato. Pertanto ricordiamo un paio di proprietà che ci torneranno comode:

- I quantili della t-Student si indicano con la lettera t e sono simmetrici, pertanto $t_\alpha = -t_{1-\alpha}$.
- Se $n \to \infty$, allora $T(n) \sim N(0,1)$. Nella pratica, basta $n > 30$.

2.3 Cenni di v.a. multivariate

Tutti i discorsi presentati fin'ora si possono estendere al caso di più v.a. (ad esempio: la distribuzione multinomiale è una distribuzione multivariata!). Immaginiamo di avere due v.a. X e Y. Per comodità le immaginiamo discrete ma tutto quanto diremo si può estendere al caso continuo sostituendo alle sommatorie gli integrali e alle densità discrete quelle continue.

Definiamo:

densità marginali

$$p_X(x) = P(X = x), \quad p_Y(y) = P(Y = y);$$

densità conginuntà

$$p(x, y) = P(X = x, Y = y);$$

densità condizionale

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} \quad \text{se } p_Y(y) > 0$$

Dalla conoscenza della congiunta possiamo ricavarci tutto ciò che ci serve, infatti:

$$p_X(x) = \sum_{y\in\mathbb{R}} p(x,y) \quad \text{e } p_Y(y) = \sum_{x\in\mathbb{R}} p(x,y).$$

Iteressante notare che se X e Y sono indipendenti, allora $p(x, y) = p_X(x)p_Y(y)$ e, di conseguenza, $p_{X|Y}(x|y) = p_X(x)$.

Infine introduciamo il **coefficiente di correlazione lineare**

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}, \quad \text{t.c. } -1 \leq \rho \leq 1.$$

In base al valore di questo coefficiente possiamo dedurre se c'è correlazione (NON CAUSALITÀ!) tra le due variabili aleatorie. Se il coefficiente è positivo la variabili saranno correlate, se negativo saranno anticorrelate. In particolare, si avrà anche che

- Se $|\rho| < 0.3$ la correlazione (o anticorrelazione) è debole;
- Se $0.3 < |\rho| < 0.7$ la correlazione (o anticorrelazione) è moderata;
- Se $0.7 < |\rho|$ la correlazione (o anticorrelazione) è forte;

2.4 Esercizi

Esercizio 1

In un libro di 500 pagine sono distribuiti a caso 300 errori di stampa. Qual è la probabilità che una data pagina contenga almeno due errori.

Svolgimento

Questo problema si inquadra nello schema successo-insuccesso, dove il successo è quello di trovare un errore in una data pagina. Gli errori sono 300, il numero di tentativi che ho è $n = 300$. Essendo gli errori distribiti a caso tra le cinquecento pagine, la probabilità di trovare un errore in una data pagina sarà $p = 1/500$. Pertanto, la v.a. che descrive questo problema sarà $X \sim B(n, p)$ e la probabilità q di avere almeno due errori in una pagina sarà

$$q = P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 0.121769.$$

Avremmo potuto risolvere lo stesso esercizio in maniera del tutto analoga notando che $n = 300 \gg 1$ e $p = 1/500 \ll 1$, per cui lo stesso problema può essere descritto da una v.a. $Y \sim \mathrm{Pois}(\lambda)$, con $\lambda = np$. Usando questo approccio avremmo trovato che

$$q = P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 0.121901.$$

Osserviamo che i due numeri sono praticamente identici. Non sono esattamente uguali perché l'uguaglianza delle due distribuzioni vale strettamente solo nel limite $p \to 0$ e $n \to \infty$.

Esercizio 2

Lancio un dado tre volte. a) Qual è la probabilità di ottenere 6 almeno una volta? b) Qual è la probabilità che ciò avvenga al quinto tentativo? c) Quante volte devo lanciare il dado affinché abbia una probabilità del 90% di avere 6? d) Questo risultato sarebbe stato diverso se avessimo considerato un altro numero?

Svolgimento

a) Schema successo-insuccesso in cui il successo è rappresentato dall'ottenere 6. Ho $n = 3$ lanci a disposizione e una probabilità di successo $p = 1/6$. Pertanto posso descrivere questo problema usando $X \sim B(n,p)$ e la probabilità richiesta sarà

$$q = P(X > 0) = 1 - P(X = 0) = 0.4219$$

b) Il tempo di primo successo T è una v.a. che segue una legge geometrica, ovvero $T \sim \text{Geom}(p)$. Pertanto la probabilità richiesta è

$$q = P(T = 5) = p(1-p)^4 = 0.0803$$

c) Utilizzo sempre il fatto che il tempo di primo successo sia una v.a. che segue una distribuzione geometrica. La richiesta fatta dal problema, in altre parole, equivale a trovare un n tale per cui $P(T \leq n) = 0.9$. Pertanto

$$P(T \leq n) = 1 - P(T > n) = 1 - (1-p)^n > 0.9 \Rightarrow (1-p)^n < 0.1.$$

Per risolvere questa disequazione passiamo ai logaritmi

$$n \log(1-p) > \log(0.1) \Rightarrow n > \frac{\log(0.1)}{\log(1-p)} = 12.63.$$

Pertanto, avremo bisogno di lanciare il dado almeno 13 volte per avere una probabilità del 90% di ottenere 6.

d) Fintanto che il dado è equilibrato, il risultato di sopra non sarebbe cambiato se avessimo risolto il problema per un altro numero.

Esercizio 3

Su un tavolo ci sno due monete. Quando vengono lanciate, una moneta dà testa con $p = 0.5$, l'altra con $p = 0.6$. Una moneta scelta a caso viene lanciata. a) Qual è la probabilità che esca testa? b) Qual è la probabilità che, sapendo che è uscita testa, la moneta lanciata fosse quella equilibrata?

Svolgimento

a) Per risolvere questo esercizio utilizziamo il teorema delle probabilità totali.

Definiamo gli eventi

T = "Esce testa" A = "Lancio moneta equilibrata" B = "Lancio moneta truccata".

Dato che la scelta è casuale, la probabilità di lanciare la moneta equilibrata è identica a quella di lanciare la moneta truccata e quindi $P(A) = P(B) = 0.5$. Dato che l'intero spazio dell'eventi è costituito dal lancio di una o dell'altra moneta avremo che,

$$P(T) = P(T|A)P(A) + P(T|B)P(B) = 0.5 \cdot 0.5 + 0.5 \cdot 0.6 = 0.55$$

b) Usiamo il teorema di Bayes

$$P(A|T) = \frac{P(T|A)P(A)}{P(T)} = \frac{0.5 \cdot 0.5}{0.55} = 0.45454545.$$

Esercizio 4

Il numero di anni di funzionamento di una radio segue una exp(1/8). a) Qual è la probabilità che funzioni più di 10 anni? Qual è la probabilità che funzioni più di 10 anni se ne ha già funzionati 4?

Svolgimento

Questo problema è descritto da $X \sim \exp(1/8)$. Allora

a) $P(X > 10) = 1 - P(X \leq 10) = 1 - (1 - e^{-\frac{10}{8}}) = e^{-\frac{10}{8}} = 0.2865.$

b) Per la proprietà di assenza di memoria della distribuzione esponenziale, avremo che

$$P(X > 10|X > 4) = P(X > 6) = e^{-\frac{6}{8}} = 0.4724$$

Esercizio 5 L'altezza degli italiani segue una normale di media $\mu = 175cm$ e varianza $\sigma^2 = 8cm^2$. a) Qual è la percentuale di italiani con altezza superiore a $190cm$? b) Qual è la percentuale di italiani con altezza inferiore a $153cm$?

Svolgimento

Data $X \sim N(175, 8)$, avremo

a) $P(X > 190) = 1 - P(X \leq 190) = 1 - F_X(190) = 0.047.$

b) $P(X < 153) = F_X(153) = 0.0073.$

OBSNel calcolare queste probabilità, non è importante dove metto l'uguaglianza, tanto, essendo la normale una distribuzione continua, la $P(X = a) = 0$, per cui $P(X \geq 190) = P(X > 190)$.

Esercizio 6 Ad un esame la distribuzione dei voti segue una normale di medi $\mu = 24$ e deviazione standard è $\sigma = 4$. Calcolare a) la probabilità che il voto sia maggiore di 27; b) la probabilità che il voto sia minore di 22; c) la probabilità che il voto sia tra 23 e25; d) Il voto minimo riportato dal 70% degli studenti; e) Il voto massimo non superato dal 90% degli studenti.

Svolgimento

Per fare questo esercizio utilizziamo funzione di ripartizione di una v.a. $X \sim N(\mu, \sigma^2)$ e i quantili q_α.

a) $P(X > 27) = 1 - P(X \leq 27) = 1 - F(27) = 0.22$

b) $P(X < 22) = F(22) = 0.303$

c) $P(23 < X < 25) = P(X < 25) - P(X < 23) = F(25) - F(23) = 0.197$

d) $x_{\min} = q_{1-70\%}$

e) $x_{\max} = q_{90\%}$