

Appunti di Probabilità e Statistica

4 ottobre 2022

Rosso Carlo

Contents

1	Introduzione	2
2	Statistica descrittiva	2
2.1	Dati univariati	2
	Statistiche elementari	2
	Statistiche d'ordine	2
	Statistiche della dispersione dei dati	2
	Disuguaglianza di Chebyshev	3
	Statistiche per la distribuzione dei dati	3
2.2	Dati multivariati	4
3	Teoria delle probabilità	4
3.1	Proprietà fondamentali delle misure di probabilità	5
3.2	Probabilità condizionali	6
	Formula di Bayes	6
4	Variabili aleatorie	7
4.1	Variabili aleatorie discrete	7
4.2	Distribuzioni notevoli	8
4.3	Valore medio	9
	Proprietà del valore medio (o atteso)	10
4.4	Disuguaglianza di Markov-Chebyshev	10
4.5	Teorema: legge dei piccoli numeri	11
4.6	Vettori aleatori discreti	11
	Proprietà della covarianza	13
4.7	Funzioni di ripartizione	13
	Proprietà della funzione di ripartizione	14
	Funzioni di ripartizione notevoli	14

1 Introduzione

In questo corso studiamo:

- statistica descrittiva;
- spazi di probabilità;
- probabilità condizionali;
- variabili aleatorie;
- teoremi limite;
- statistica inferenziale;

Def. 1.1 (Statistica descrittiva) *descrivere e sintetizzare i dati raccolti in un campione.*

Def. 1.2 (Statistica inferenziale) *trovare conclusioni su una popolazione a partire da un campione.*

2 Statica descrittiva

2.1 Dati univariati

Successioni finite di numeri reali. Sia $(x_i)_{i \in \{1, \dots, n\}}$ l'insieme ordinato dei dati, si chiama campione di numerosità n .

Statistiche elementari

Def. 2.1 (Media campionaria) $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$. Ci siano k valori distinti, sia f_j il numero di occorrenze di v_j , tale che $i \neq j \iff v_i \neq v_j$, allora la media campionaria è:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k f_i v_i \quad (1)$$

NB I valori che hanno frequenza massima si dicono valori modali. Se esiste un solo valore con frequenza massima, esso è chiamato moda campionaria.

Def. 2.2 (Mediana campionaria) Sia σ una permutazione degli indici $\{1, \dots, n\}$ tale che $x_{\sigma(1)} \leq \dots \leq x_{\sigma(n)}$. La mediana campionaria è:

$$m := \begin{cases} x_{\sigma(\frac{n+1}{2})} & \text{se } n \text{ dispari} \\ \frac{1}{2}(x_{\sigma(\frac{n}{2})} + x_{\sigma(\frac{n}{2}+1)}) & \text{se } n \text{ pari} \end{cases} \quad (2)$$

Statistiche d'ordine

Il minimo è:

$$\min\{x_i : i \in \{1, \dots, n\}\} = x_{\sigma(1)} \quad (3)$$

Il massimo è:

$$\max\{x_i : i \in \{1, \dots, n\}\} = x_{\sigma(n)} \quad (4)$$

Statistiche della dispersione dei dati

Def. 2.3 (Varianza campionaria)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^k (x_i^2 - \bar{x}^2) \quad (5)$$

Def. 2.4 (Deviazione standard campionaria)

$$s = \sqrt{s^2} \quad (6)$$

NB Siano $a, b \in \mathbb{R}$, $(s_y)^2$ la varianza campionaria di $(y_i)_{i \in \{1, \dots, n\}}$ e $(s_x)^2$ la varianza campionaria di $(x_i)_{i \in \{1, \dots, n\}}$. Allora:

$$s_y^2 = a^2 \cdot s_x^2 \iff s_y = |a| \cdot s_x \quad (7)$$

NB la devianza standard ha la medesima unità di misura dei dati.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - \frac{n}{n-1} \bar{x}^2 \quad (8)$$

Varianza e deviazione permettono di stimare la proporzione dei dati che sono "vicini" (o "lontani") dalla media campionaria.

Disuguaglianza di Chebyshev

se $s > 0$, allora $\forall \alpha > 0$:

1.

$$\frac{\#\{i \in \{1, \dots, n\} : |x_i - \bar{x}| < \alpha s\}}{n} \geq 1 - \frac{1}{\alpha^2} \quad (9)$$

Utile per $\alpha > 1$. Percentuale di dati che sono "vicini" alla media campionaria, ovvero che sono compresi nell'intervallo $]\bar{x} - \alpha s, \bar{x} + \alpha s[$.

2.

$$\frac{\#\{i \in \{1, \dots, n\} : |x_i - \bar{x}| \leq \alpha s\}}{n} > 1 - \frac{1}{\alpha^2} \quad (10)$$

Utile per $\alpha > 1$.

3.

$$\frac{\#\{i \in \{1, \dots, n\} : |x_i - \bar{x}| > \alpha s\}}{n} \leq \frac{1}{\alpha^2} \quad (11)$$

Utile per $\alpha > 1$. Percentuale di dati che sono "lontani" alla media campionaria, ovvero che sono compresi nell'intervallo $\mathbb{R} \setminus]\bar{x} - \alpha s, \bar{x} + \alpha s[$.

4.

$$\frac{\#\{i \in \{1, \dots, n\} : x_i - \bar{x} > \alpha s\}}{n} < \frac{1}{1 + \alpha^2} \quad (12)$$

Utile per $\alpha < 1$. Percentuale di dati che sono "lontani" e maggiori della media campionaria, ovvero che sono compresi nell'intervallo $]\bar{x} + \alpha s, +\infty[$.

Statistiche per la distribuzione dei dati

Sia $k \in \{1, \dots, 100\}$:

Def. 2.5 (Percentile campionario k-esimo)

$$\bar{p}_k := \begin{cases} x_{\sigma(\frac{n+k}{100})} & \text{per eccesso, se } \frac{n+k}{100} \notin \mathbb{N} \\ \frac{1}{2}(x_{\sigma(\frac{n+k}{100})} + x_{\sigma(\frac{n+k}{100}+1)}) & \text{altrimenti} \end{cases} \quad (13)$$

NB $\bar{p}_{50} = \bar{m}$. Proporzioni dei dati che hanno valore inferiore o uguale a \bar{p}_k oppure che hanno valore superiore a \bar{p}_k :

•

$$\frac{\#\{i \in \{1, \dots, n\} : x_i \leq \bar{p}_k\}}{n} \geq \frac{k}{100} \quad (14)$$

•

$$\frac{\#\{i \in \{1, \dots, n\} : x_i > \bar{p}_k\}}{n} \leq 1 - \frac{k}{100} \quad (15)$$

2.2 Dati multivariati

Il campione di dati multivariati è una successione finita di vettori. Per semplicità tratteremo un campione bivariato, le generalizzazioni sono ovvie. Sia $((x_i, y_i))_{i \in \{1, \dots, n\}} \subset \mathbb{R}^2$ un campione bivariato. allora $(x_i)_{i \in \{1, \dots, n\}}$ e $(y_i)_{i \in \{1, \dots, n\}}$ sono campioni univariati.

Def. 2.6 (covarianza campionaria) La covarianza campionaria tra $(x_i)_{i \in \{1, \dots, n\}}$ e $(y_i)_{i \in \{1, \dots, n\}}$ è data da:

$$Cov_{x,y} := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (16)$$

NB Se $x_i = y_i \forall i \in \{1, \dots, n\}$, allora $Cov_{x,y} = s_x^2 = s_y^2$. La covarianza varia in \mathbb{R} .

Def. 2.7 (correlazione campionaria) La correlazione campionaria tra $(x_i)_{i \in \{1, \dots, n\}}$ e $(y_i)_{i \in \{1, \dots, n\}}$ è data da:

$$Corr_{x,y} := \frac{Cov_{x,y}}{s_x s_y} \quad (17)$$

Osservazioni:

1. dalla definizione otteniamo:

$$Corr_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (18)$$

Ovvero si cancella il prefattore $1/(n-1)$.

2. $Corr_{x,y} \in [-1, 1]$. In Particolare:

- $Corr_{x,y} = 1 \iff \exists a, b \in \mathbb{R} : a > 0, x_i = a \cdot y_i + b \forall i \in \{1, \dots, n\}$.
- $Corr_{x,y} = -1 \iff \exists a, b \in \mathbb{R} : a < 0, x_i = a \cdot y_i + b \forall i \in \{1, \dots, n\}$.

La correlazione quantifica la linearità della relazione tra i campioni.

Si ha la dipendenza massima per $Corr_{x,y} = \pm 1$. Si ha la dipendenza minima per $Corr_{x,y} = 0$.

3 Teoria delle probabilità

Def. 3.1 (Esperimento aleatorio) Un esperimento aleatorio è un'osservazione riguardo a un qualunque fenomeno il cui esito non è determinato con certezza priori.

Il metodo che meglio rappresenta il campione di dati completo è la scelta casuale di un sottoinsieme con data numerosità.

Def. 3.2 (Spazio campionario) Lo spazio campionario è l'insieme di tutti gli esiti possibili di un esperimento aleatorio, è un insieme non vuoto e si rappresenta con Ω .

Def. 3.3 (Sistema degli eventi) Il sistema degli eventi è l'insieme che contiene gli esiti possibili di un esperimento aleatorio, è un insieme non vuoto e si rappresenta con \mathcal{F} . Il sistema degli eventi è un sottoinsieme di $\Omega^{\mathbb{P}}$ dove \mathbb{P} è il prodotto cartesiano.

Def. 3.4 (Misura di probabilità) La misura di probabilità è una funzione $\mathcal{P} : \mathcal{F} \rightarrow [0, 1]$ tale che:

- $\mathcal{P}(\emptyset) = 0$
- $\mathcal{P}(\Omega) = 1$
- $\mathcal{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathcal{P}(A_i)$

La misura di probabilità è una funzione che associa ad ogni evento la probabilità che esso accada.

Def. 3.5 (Spazio di probabilità) Lo spazio di probabilità è una terna $(\Omega, \mathcal{F}, \mathcal{P})$ dove Ω è lo spazio campionario non vuoto, \mathcal{F} è una σ -algebra di Ω e \mathcal{P} è una misura di probabilità su \mathcal{F} .

Def. 3.6 (σ -algebra) \mathcal{F} si dice σ -algebra in Ω , $\Omega \neq \emptyset$, se:

- \mathcal{F} è chiuso sotto l'insieme vuoto
- \mathcal{F} è chiuso sotto l'insieme complementare
- \mathcal{F} è chiuso sotto l'insieme unione

Estremi della σ -algebra:

- σ -algebra triviale $:= \{\emptyset, \Omega\}$;
- σ -algebra totale $:= P(\Omega)$;

Ogni misura di probabilità su $P(\Omega)$ con Ω al più numerabile è determinata dalle probabilità dei singoli elementi di Ω (singoletti):

$$\mathcal{P}(A) = \sum_{\omega \in A} \mathcal{P}(\{\omega\}), \quad A \subset \Omega. \quad (19)$$

Una funzione $p : \Omega \rightarrow [0, 1]$ si dice densità discreta se:

$$\sum_{\omega \in \Omega} p(\omega) = 1 \quad (20)$$

Sia p una densità discreta su Ω . Allora

$$\mathcal{P}(A) = \sum_{\omega \in A} p(\omega), \quad A \subset \Omega. \quad (21)$$

definisce una misura di probabilità su $P(\Omega)$.

NB se Ω è numerabile, allora tutte le misura di probabilità sono di questa forma.

Def. 3.7 (Densità campionaria) Sia $(x_i)_{i \in 1, \dots, n}$ un campione di numerosità n a valori in X (as esempio, $X = \mathbb{R}$). La densità campionaria è una funzione $p : X \rightarrow [0, 1]$ tale che:

$$p(x) := \frac{\#i \in \{1, \dots, n\} : x_i = x}{n}, \quad x \in X. \quad (22)$$

Allora p è una densità discreta su X e la densità campionaria di $(x_i)_{i \in \{1, \dots, n\}}$.

$p(x)$ è la frequenza relativa del valore $x \in X$.

p induce una misura di probabilità su $P(X)$:

$$\mathcal{P}(A) := \sum_{x \in A} p(x), \quad A \subset X. \quad (23)$$

\mathcal{P} è detta misura di probabilità campionaria (o empirica).

3.1 Proprietà fondamentali delle misure di probabilità

Sia $(\Omega, \mathcal{F}, \mathcal{P})$ uno spazio di probabilità e siano $A, B \in \mathcal{F}$.

1. Se $A \subset B$, allora $\mathcal{P}(B \setminus A) = \mathcal{P}(B) - \mathcal{P}(A)$.
In particolare, $\mathcal{P}(A^c) = 1 - \mathcal{P}(A)$.
2. $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$.
In particolare, se $A \cap B = \emptyset$, allora $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B)$.
3. Se $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$, allora

$$\mathcal{P}(\cup_{n \in \mathbb{N}} A_n) \leq \sum_{n \in \mathbb{N}} \mathcal{P}(A_n). \quad (24)$$

In particolare, se $A_i \cap A_j = \emptyset \forall i \neq j$, allora $\mathcal{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathcal{P}(A_n)$.

4. Sia $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ tale che $A_i \cap A_j = \emptyset \forall i \neq j$ e $\cup_{n \in \mathbb{N}} A_n = \Omega$, allora:

$$\mathcal{P}(B) = \mathcal{P}\left(\sum_{n \in \mathbb{N}} B \cap A_n\right). \quad (25)$$

In particolare, $\mathcal{P}(B) = \mathcal{P}(B \cap A) + \mathcal{P}(B \cap A^c)$.

5. Sia $(A_n)_{n \in \mathbb{N}}$.

(a) se (A_n) è crescente, cioè $A_n \subset A_{n+1}, \forall n \in \mathbb{N}$, allora:

$$\mathcal{P}(\cup_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} \mathcal{P}(A_n). \quad (26)$$

(b) se (A_n) è decrescente, cioè $A_n \supset A_{n+1}, \forall n \in \mathbb{N}$, allora:

$$\mathcal{P}(\cap_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} \mathcal{P}(A_n). \quad (27)$$

3.2 Probabilità condizionali

Def. 3.8 (Probabilità condizionale) Sia $(\Omega, \mathcal{F}, \mathcal{P})$ uno spazio di probabilità e sia $B \in \mathcal{F}$ un evento tale che $\mathcal{P}(B) > 0$. Per ogni $A \in \mathcal{F}$, la probabilità

$$\mathcal{P}(A|B) := \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)} \quad (28)$$

si dice *probabilità condizionale di A dato B*.

NB $\mathcal{P}(B|B) = 1$ e $\mathcal{P}(B^c|B) = 0$.

Def. 3.9 (Proprietà delle probabilità condizionali) Sia $(\Omega, \mathcal{F}, \mathcal{P})$ uno spazio di probabilità.

1. Sia $B \in \mathcal{F}$ tale che $\mathcal{P}(B) > 0$. Allora la mappa $\mathcal{F} \ni A \mapsto \mathcal{P}(A|B) \in [0, 1]$ definisce una misura di probabilità su \mathcal{F} .

NB $(\Omega, \mathcal{F}, \mathcal{P}(\cdot|B))$ è uno spazio di probabilità.

Anche $(\Omega, \mathcal{F}_B, \mathcal{P}(\cdot|B))$ è uno spazio di probabilità, dove $\mathcal{F}_B := \{A \cap B \mid A \in \mathcal{F}\}$ è la σ -algebra indotta su B .

2. Siano $A_1, \dots, A_n \in \mathcal{F}$ eventi, tali che $\mathcal{P}(\cap_{i=1}^n A_i) > 0$. Allora:

$$\mathcal{P}(\cap_{i=1}^n A_i) = \mathcal{P}(A_1) \cdot \prod_{i=2}^n \mathcal{P}(A_i | A_1 \cap \dots \cap A_{i-1}). \quad (29)$$

In particolare, se $\mathcal{P}(A_1 \cap A_2) > 0$, allora: $\mathcal{P}(A_1 \cap A_2) = \mathcal{P}(A_1) \cdot \mathcal{P}(A_2|A_1)$. Rivediamo la divisione insomma, moltiplichiamo sopra e sotto per lo stesso numero. Utile per prendere a fattor comune qualcosa.

3. Sia $(B_i)_{i \in I}$, una partizione al più numerabile di $\Omega : \forall i \neq j, B_i \cap B_j = \emptyset, \cup_{i \in I} B_i = \Omega$, tale che $B_i \in \mathcal{F}, \mathcal{P}(B_i) > 0, \forall i \in I$. Allora, per ogni $A \in \mathcal{F}$:

$$\mathcal{P}(A) = \sum_{i \in I} \mathcal{P}(A|B_i) \cdot \mathcal{P}(B_i). \quad (30)$$

NB stiamo riscrivendo le proprietà fondamentali delle misure di probabilità utilizzando la nuova operazione: $\mathcal{P}(a|b)$, e si legge "probabilità di a intersecato b diviso probabilità di b".

Formula di Bayes

Siano $A, B \in \mathcal{F}$ eventi in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathcal{P})$ tale che $\mathcal{P}(A) > 0, \mathcal{P}(B) > 0$. Allora:

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(A)}{\mathcal{P}(B)} \cdot \mathcal{P}(B|A). \quad (31)$$

Ne deriva che

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(B|A) \cdot \mathcal{P}(A)}{\mathcal{P}(B|A) \cdot \mathcal{P}(A) + \mathcal{P}(A|B^c) \cdot (1 - \mathcal{P}(B))}. \quad (32)$$

Ci piacciono molto le quattro operazioni.

Def. 3.10 (Indipendenza di eventi) Siano A, B eventi in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathcal{P})$, allora A, B si dicono indipendenti se:

$$\mathcal{P}(A \cap B) = \mathcal{P}(A) \cdot \mathcal{P}(B). \quad (33)$$

Osservazioni:

- L'indipendenza di eventi dipende dalla misura di probabilità;
- se $\mathcal{P}(A) \in \{0, 1\}$ o $\mathcal{P}(B) \in \{0, 1\}$ allora A, B sono indipendenti. Ovvero se $A \in \{\emptyset, \Omega\}$, allora A, B sono indipendenti per ogni scelta di B ;
- Se $\mathcal{P}(A) > 0$ e $\mathcal{P}(B) > 0$ e $A \cap B = \emptyset$, allora A, B non sono indipendenti;
- Se A, B sono indipendenti, allora lo sono anche A, B^c e A^c, B^c ;
- Se $\mathcal{P}(B) > 0$, allora: A, B sono indipendenti $\iff \mathcal{P}(A|B) = \mathcal{P}(A)$; L'interpretazione sarebbe: se $\mathcal{P}(B) > 0$ e A, B sono indipendenti, allora sapere se si è verificato o meno B non cambia la valutazione di A .

Def. 3.11 (Indipendenza di famiglie di eventi) Sia $(\Omega, \mathcal{F}, \mathcal{P})$ uno spazio di probabilità e siano $A_1, \dots, A_n \in \mathcal{F}$, allora A_1, \dots, A_n si dicono indipendenti come famiglie se per ogni scelta di $\emptyset \neq J \subset \{1, \dots, n\}$ si ha:

$$\mathcal{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathcal{P}(A_j). \quad (34)$$

Ne deriva la proprietà: siano A_1, \dots, A_n eventi in $(\Omega, \mathcal{F}, \mathcal{P})$ allora A_1, \dots, A_n sono indipendenti come famiglia se e solo se per ogni scelta di eventi $B_1, \dots, B_n : B_i \in \{A_i, A_i^c\}$ si ha

$$\mathcal{P}(\cap_{i=1}^n B_i) = \prod_{i=1}^n \mathcal{P}(B_i). \quad (35)$$

Def. 3.12 (Modello probabilistico per n prove ripetute e indipendenti) Sia $q \in [0, 1]$ e sia $n \in \mathbb{N}$. Uno spazio di probabilità $(\Omega, \mathcal{F}, \mathcal{P})$ con eventi C_1, \dots, C_n si dice modello per n prove ripetute e indipendenti con probabilità di successo q se

C_1, \dots, C_n sono indipendenti come famiglia e $\mathcal{P}(C_i) = q$ per ogni $i \in \{1, \dots, n\}$.

4 Variabili aleatorie

Def. 4.1 (Variabili aleatorie) Sia $(\Omega, \mathcal{F}, \mathcal{P})$ uno spazio di probabilità, e sia $E \neq \emptyset$. Una variabile aleatoria X a valori in E è una funzione misurabile $X : \Omega \mapsto E$.

Una variabile aleatoria X è un modo per indurre una misura di probabilità sullo spazio misurabile di arrivo E a partire dalla misura di probabilità \mathcal{P} definita sull'insieme degli eventi Ω .

Sia $B \subset E$, l'anti-immagine di B sotto X è il sottoinsieme di Ω dato da $X^{-1} := \{\omega \in \Omega : X(\omega) \in B\}$.

Notiamo che, sia \mathcal{E} una σ -algebra in E , allora

$$\sigma(X) := \sigma_{\mathcal{E}}(X) := \{X^{-1}(B) : B \in \mathcal{E}\} \quad (36)$$

è una σ -algebra in Ω : la σ -algebra generate da X (rispetto a \mathcal{E}).

NB per semplicità pensare a \mathcal{E} come se fosse $\mathcal{P}(X)$, per quanto non esse parti di X per intero, ma un suo sottoinsieme.

4.1 Variabili aleatorie discrete

Sia una variabile aleatoria su $(\Omega, \mathcal{F}, \mathcal{P})$ a valori in E , l'immagine (insieme di arrivo di X) è il sottoinsieme di E dato da:

$$\text{Im}(X) := \{X(\omega) : \omega \in \Omega\} \subset E. \quad (37)$$

Def. 4.2 (Variabile aleatoria discreta) Una variabile aleatoria X a valori in E si dice discreta se $\text{Im}(X)$ è al più numerabile.

Def. 4.3 (Densità discreta di una variabile aleatori) La funzione:

$$p_X(x) := (P)(X = x), \quad x \in E, \quad (38)$$

si dice densità discreta di X .

NB p_X è una densità discreta:

$$\sum_{x \in E} p_X(x) = \sum_{x \in \text{Im}(X)} p_X(x) = 1. \quad (39)$$

Inoltre $\Omega = \cup_{x \in \text{Im}(X)} \{X = x\}$.

Def. 4.4 (Legge di X (o distribuzione di X))

$$\mathcal{P}(B) := \mathcal{P}(X^{-1}(B)), \quad B \in \mathcal{E}. \quad (40)$$

Definisco \mathcal{E} come l'insieme delle σ -algebre di E .

Per cui $\mathcal{P}_X(B) = \sum_{x \in B} p_X(x)$, $\forall B \subset \mathbb{R}$.

4.2 Distribuzioni notevoli

Def. 4.5 (Bernoulli) Una variabile aleatoria Y con densità (discreta) data da

$$p_Y(x) = \begin{cases} q & \text{se } x = 1 \\ 1 - q & \text{se } x = 0 \\ 0 & \text{altrimenti} \end{cases} \quad (41)$$

si dice variabile aleatoria di Bernoulli di parametro q .

In questo caso, si dice che Y ha una distribuzione di Bernoulli di parametro q , in simboli: $Y \sim \text{Ber}(q)$.

Def. 4.6 (Rademacher) Una variabile aleatoria Y con densità (discreta) data da

$$p_Y(x) = \begin{cases} q & \text{se } x = 1 \\ 1 - q & \text{se } x = -1 \\ 0 & \text{altrimenti} \end{cases} \quad (42)$$

si dice variabile aleatoria di Rademacher di parametro q .

In questo caso, si dice che Y ha una distribuzione di Rademacher di parametro q , in simboli: $Y \sim \text{Rad}(q)$.

Def. 4.7 (Binomiale) Una variabile aleatoria Y con densità data da:

$$p_Y(x) = \begin{cases} \binom{n}{x} \cdot q^x \cdot (1 - q)^{n-x} & \text{se } x \in \{0, \dots, n\} \\ 0 & \text{altrimenti} \end{cases} \quad (43)$$

si dice variabile aleatoria di parametri $n \in \mathbb{N}, q \in [0, 1]$.

In questo caso, si dice che Y ha distribuzione binomiale di parametri n, q , in simboli: $Y \sim \text{Bin}(n, q)$.

Def. 4.8 (Geometrica) Una variabile aleatoria con densità data da

$$p_Y(x) = \begin{cases} q \cdot (1 - q)^{x-1} & \text{se } x \in \mathbb{N}, \\ 0 & \text{altrimenti} \end{cases} \quad (44)$$

si dice variabile aleatoria geometrica di parametro $q \in [0, 1]$.

Si dice che Y ha distribuzione geometrica di parametro q , in simboli: $Y \sim Geo(q)$.

Def. 4.9 (Ipergeometrica) Una variabile aleatoria ha una distribuzione ipergeometrica di parametri $N \in \mathbb{N}$, $M \in \{0, \dots, N\}$, $n \in \{1, \dots, N\}$ se la sua densità è:

$$p_y(x) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} & \text{se } x \in \{0, \dots, \min(n, M)\} \\ 0 & \text{altrimenti} \end{cases} \quad (45)$$

In simboli: $Y \sim Iper(N, M, n)$.

Esempio di utilizzo: *numero di palline rosse in n estrazioni senza reinserimento da un'urna che contiene N palline di cui M rosse.*

Def. 4.10 (Poisson) Una variabile aleatoria ha una distribuzione di Posson di parametro $\lambda > 0$ se la sua densità è:

$$p_y(x) = \begin{cases} e^{-\lambda} \cdot \frac{\lambda^x}{x!} & \text{se } x \in \mathbb{N}_0 \\ 0 & \text{altrimenti} \end{cases} \quad (46)$$

In simboli: $Y \sim Poiss(\lambda)$.

Esempio di utilizzo: *numero di "arrivi" (richieste, clienti, ...) in un determinato intervallo temporale; λ rappresenta l'intensità.*

Def. 4.11 (Uniforme discreta) Una variabile aleatoria ha una distribuzione uniforme discreta su $A \subset \mathbb{R}$ finito, se la sua densità è:

$$p_y(x) = \begin{cases} \frac{1}{|A|} & \text{se } x \in A, \\ 0 & \text{altrimenti} \end{cases} \quad (47)$$

In simboli: $Y \sim Unif(A)$.

4.3 Valore medio

Quantità riasuntiva fondamentale, analoga alla media campionaria.

Def. 4.12 (Valore medio) Sia X una variabile aleatoria reale discreta su $(\Omega, \mathbb{F}, \mathbb{P})$ con densità discreta p_X . si dice che X ammette valore medio finito (o valore atteso finito) se

$$\sum_{z \in \mathbb{R}} |z| \cdot p_X(z) < \infty. \quad (48)$$

Il valore medio medio è dato da

$$\mathbb{E}[X] := \sum_{z \in \mathbb{R}} z \cdot p_X(z). \quad (49)$$

Se X ammette valore medio finito, allora

$$\mathbb{E}[X] := \sum_{z \in \mathbb{R}} z \cdot p_X(z) = \sum_{z \in Im(X)} z \cdot p_X(z) \quad (50)$$

Se $|Im(X)| = \infty$ (ma $Im(X)$ è numerabile) allora possiamo scrivere:

$$Im(X) = \{x_1, \dots\} \text{ con } x_i \neq x_j \text{ per } i \neq j,$$

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i \cdot p_X(x_i). \quad (51)$$

NB il valore medio dipende solo dalla distribuzione della variabile aleatoria (quindi attenzione alla variabile aleatoria che si intende usare).

Proprietà del valore medio (o atteso)

Siano X, Y variabili aleatorie reali discrete su $(\Omega, \mathcal{F}, \mathcal{P})$ tali che X, Y ammettano valore medio finito, allora:

- Monotonia: se $X \leq Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y]$;
- Stima fondamentale: $|\mathbb{E}[X]| \leq \mathbb{E}[|x| : x \in X]$;
- Linearità: per ogni $\alpha, \beta \in \mathbb{R}$:

$$\mathbb{E}[\alpha \cdot X + \beta \cdot Y] = \alpha \cdot \mathbb{E}[X] + \beta \mathbb{E}[Y]. \quad (52)$$

NB si può fare una mappa f dell'immagine di una variabile aleatoria X con un altro insieme ottenendo così un'altra variabile aleatoria Y la cui distribuzione dipende da X :

$$p_Y(y) := \sum_{x \in E: f(x)=y} p_X(x), \quad z \in \mathbb{R} \quad (53)$$

Magia: p_X è la densità discreta di X .

Da ciò ne deriva che Z ammette valore medio finito se e solo se

$$\sum_{x \in \mathbb{R}} |g(x)| \cdot p_X(x) < \infty.$$

Analogamente al caso del valore atteso di X . Ripetiamo tutto allora:

$$\mathbb{E}[Z] = \sum_{x \in \mathbb{R}} g(x) \cdot p_X(x). \quad (54)$$

Abbiamo definito il valore medio, tanto vale definire anche la varianza.

Def. 4.13 (Varianza) Sia X una variabile aleatoria reale discreta su $(\Omega, \mathbb{F}, \mathbb{P})$ con densità discreta p_X , la varianza di X (se esiste) è data da

$$\text{var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] \quad (55)$$

Osservazioni:

- la varianza di X esiste finita se e solo se X^2 ammette valore atteso finito;
- $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$;
- $\text{var}(X) = \sum_{x \in \mathbb{R}} x^2 p_X(x) - (\sum_{x \in \mathbb{R}} x \cdot p_X(x))^2$.

Def. 4.14 (Indipendenze) Sia I un insieme qualsiasi di indici, e $\{X_i : i \in I\}$ una famiglia di variabili aleatorie a valori negli spazi $(E_i, \mathcal{E}_i), i \in I$. Si dice che le variabili aleatorie di tale famiglia sono indipendenti se, per ogni $J \subset I$ finito e per ogni scelta di $A_j \in \mathcal{E}_j, j \in J$, si ha:

$$P\left(\bigcap_{j \in J} \{X_j \in A_j\}\right) = \prod_{j \in J} P(\{X_j \in A_j\}). \quad (56)$$

Def. 4.15 (Valore atteso del prodotto) Siano X, Y due variabili aleatorie reali discrete su $(\Omega, \mathcal{F}, \mathcal{P})$ con valore medio finito, se X e Y sono indipendenti, allora:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \quad (57)$$

4.4 Disuguaglianza di Markov-Chebyshev

Sia X una variabile aleatoria reale discreta su $(\Omega, \mathcal{F}, \mathcal{P})$

- Markov: se $X \geq 0$ allora per ogni $\epsilon > 0$:

$$P(X \geq \epsilon) \leq \frac{\mathbb{E}[X]}{\epsilon}. \quad (58)$$

- Markov generalizzato: se $X \geq 0$ e $f : [0, \infty[\rightarrow [0, \infty[$ è crescente allora:

$$P(X \geq \epsilon) \leq \frac{\mathbb{E}[f(X)]}{f(\epsilon)}. \quad (59)$$

- Chebyshev: se X ammette varianza, allora $\forall \epsilon > 0$:

$$P(|X - \mathbb{E}[X]| \geq \epsilon) \leq \frac{\text{var}(X)}{\epsilon^2}. \quad (60)$$

Proprio come per la statistica descrittiva, la disuguaglianza di Chebyshev permette di stimare la probabilità di deviazioni dal valore atteso in termini della deviazione standard.

Def. 4.16 (Deviazione standard)

$$\omega(X)^2 := \text{var}(X). \quad (61)$$

$\omega(X)$ si definisce deviazione standard di X .

Grazie a Chebishev, si ha, per $\alpha > 0$:

$$P(|X - \mathbb{E}[X]| \geq \alpha \omega(X)) \leq \frac{\text{var}(X)}{\alpha^2 \omega(X)^2} = \frac{1}{\alpha^2}. \quad (62)$$

4.5 Teorema: legge dei piccoli numeri

Sia $(q_n)_{n \in \mathbb{N}} \subset [0, 1]$ tale che $\lim_{n \rightarrow \infty} n \cdot q_n = \lambda$ per una costante $\lambda > 0$. Sia p_n la densità discreta della $\text{Bin}(n, q_n)$ e sia p_λ la densità discreta della $\text{Poiiss}(\lambda)$. Allora:

$$\sum_{k=0}^{\infty} |p_n(k) - p_\lambda(k)| \rightarrow 0. \quad (63)$$

Stima dell'errore:

$$\sum_{k=0}^{\infty} |p_n(k) - p_\lambda(k)| \geq 2(n \cdot q_n^2 + |\lambda - n \cdot q_n|). \quad (64)$$

NB $\lambda \approx n \cdot q_n$, per un $n \gg 0$.

Inoltre $\mathbb{E}[\text{Poiiss}(\lambda)] = \lambda$.

4.6 Vettori aleatori discreti

Def. 4.17 (Vettore aleatorio) Siano X_1, \dots, X_n variabili aleatorie reali su $(\Omega, \mathcal{F}, \mathcal{P})$. Si definisce vettore aleatorio di dimensione n (o n -dimensionale) X tale che:

$$X = (X_1, \dots, X_n). \quad (65)$$

Def. 4.18 (Discreto) Sia X un vettore aleatorio di dimensione n .

Si dice discreto se $\text{Im}(X) := \{X(\omega) : \omega \in \Omega\}$ è al più numerabile.

NB sia $X = X_1, \dots, X_n$ un vettore aleatorio, allora X è discreto se e solo se X_1, \dots, X_n sono variabili aleatorie discrete.

Def. 4.19 (Densità discreta) Sia X un vettore aleatorio discreto di dimensione n su $(\Omega, \mathcal{F}, \mathcal{P})$. Si definisce densità discreta di X la funzione $p_X : \mathbb{R}^n \rightarrow [0, 1]$ tale che:

$$p_X(z) := \mathcal{P}(X_1 = z_1, \dots, X_n = z_n), \quad z \in \mathbb{R}^n. \quad (66)$$

Come al solito, $p_X(z) = 0$ se $z \notin \text{Im}(X)$.

Def. 4.20 (Densità congiunta e marginale) La densità discreta del vettore aleatorio X è detta densità congiunta di X_1, \dots, X_n (l'ordine ha importanza).

Le densità delle componenti X_1, \dots, X_n si dicono densità marginali del vettore aleatorio X .

La densità discreta congiunta determina le densità marginali:

$$p_{X_i}(x) = \sum_{z \in \mathbb{R}^n: z_i = x} p_X(z) = \sum_{z_j \in \mathbb{R}: z_j \neq i} p_X(z_1, \dots, z_{i-1}, x, z_{i+1}, \dots, z_n). \quad (67)$$

tale che $x \in \mathbb{R}$.

NB se le variabili aleatorie sono indipendenti, allora la loro distribuzione congiunta è determinata dalle distribuzioni marginali.

Def. 4.21 (Indipendenti) Siano X_1, \dots, X_n variabili aleatorie reali su $(\Omega, \mathcal{F}, \mathcal{P})$. Si dice che X_1, \dots, X_n sono indipendenti (come famiglia) se per ogni scelta di $x_1, \dots, x_n \in \mathbb{R}$ si ha:

$$\mathcal{P}(X_1 \geq x_1, \dots, X_n \geq x_n) = \prod_{i=1}^n \mathcal{P}(X_i \geq x_i). \quad (68)$$

NB per definizione, X_1, \dots, X_n sono indipendenti se e solo se gli eventi $\{X_1 \geq x_1\}, \dots, \{X_n \geq x_n\}$ sono indipendenti per ogni scelta di $x_1, \dots, x_n \in \mathbb{R}$.

Def. 4.22 (Distribuzione) Come al solito, se X_1, \dots, X_n sono variabili aleatorie discrete e p_X è la loro densità congiunta, allora:

$$\mathcal{P}(X_i \geq x_1, \dots, X_n \geq x_n) = \prod_{z \in \mathbb{R}: z_1 \geq x_1, \dots, z_n \geq x_n} p_X(z). \quad (69)$$

Modi diversi per scrivere che le componenti di un vettore aleatorio sono indipendenti: siano X_1, \dots, X_n variabili aleatorie reali discrete su $(\Omega, \mathcal{F}, \mathcal{P})$, allora sono equivalenti:

1. X_1, \dots, X_n sono indipendenti;
2. $p_{(X_1, \dots, X_n)}(z) = \prod_{i=1}^n p_{X_i}(z_i)$, $z \in \mathbb{R}^n$;
3. $\mathcal{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n \mathcal{P}(X_i \in B_i)$ per ogni scelta di $B_1, \dots, B_n \subset \mathbb{R}$.

Siano X, Y, Z variabili aleatorie reali indipendenti (come famiglia), se $g : \mathbb{R}^2 \rightarrow \mathbb{R}, h : \mathbb{R} \rightarrow \mathbb{R}$ allora $g(X, Y)$ e $h(Z)$ sono indipendenti.

Def. 4.23 (Covarianza) Siano X, Y , variabili aleatorie reali con valore atteso finito (cioè $\mathbb{E}[X^2] < \infty, \mathbb{E}[Y^2] < \infty$). Si definisce covarianza di X, Y , la funzione data da:

$$\text{cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \quad (70)$$

NB la covarianza dipende dalla distribuzione congiunta di X, Y .

In particolare, se X, Y sono variabili aleatorie discrete allora:

$$\text{cov}(X, Y) = \sum_{(x, y) \in \mathbb{R}^2} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) \cdot p_{(X, Y)}(x, y)$$

dove $p_{(X, Y)}$ è la densità congiunta di X, Y .

Proprietà della covarianza

Siano X, Y variabili reali con valore atteso finito, allora:

- Simmetria:

$$\text{cov}(X, Y) = \text{cov}(Y, X).$$

Perchè la somma e il prodotto sono commutativi;

- Bi-linearità: $\forall \alpha, \beta \in \mathbb{R}$:

$$\text{cov}(\alpha X + \beta Y, Z) = \alpha \text{cov}(X, Z) + \beta \text{cov}(Y, Z)$$

-

$$\text{var}(c \cdot X) = \text{cov}(c \cdot X, c \cdot X) = c^2 \text{var}(X) \quad \forall c \in \mathbb{R}.$$

- Indipendenza: se X, Y sono indipendenti, allora $\text{cov}(X, Y) = 0$.

NB due variabili aleatorie X, Y con $\text{cov}(X, Y) = 0$ si dicono *incorrelate* (incorrelate \neq indipendenti).

NB siano X, Y variabili aleatorie reali con valore atteso finito, allora:

$$\bowtie \text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

per la bi-linearità della covarianza.

Def. 4.24 (Correlazione) Siano X, Y variabili aleatorie reali con valore atteso finito, allora la correlazione di X, Y è definita come:

$$\rho(X, Y) := \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}. \quad (71)$$

NB il coefficiente di correlazione ($\rho(X, Y)$) non dipende dall'unità di misura per X, Y .

Caratteristiche:

-

$$\rho(\alpha X + \beta, Y) = \rho(X, Y) \quad \forall \alpha, \beta \in \mathbb{R}.$$

-

$$\rho(X, Y) \in [-1, 1].$$

1. $\rho(X, Y) = 1$ se X, Y sono direttamente proporzionali;
2. $\rho(X, Y) = -1$ se X, Y sono inversamente proporzionali;
3. $\rho(X, Y) = 0$ se X, Y sono incorrelate.

4.7 Funzioni di ripartizione

Def. 4.25 (Funzione di ripartizione) Sia X una variabile aleatoria su \mathbb{R} , definiamo la funzione $F_X : \mathbb{R} \rightarrow [0, 1]$ tramite:

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

ovvero, $F_X(x)$ è la probabilità che X sia in $(-\infty, x]$. La funzione F_X è detta funzione di ripartizione di X .

NB F_X dipende solo dalla distribuzione di X .

In particolare, se X è discreta con densità discreta p_X , allora:

$$F_X(x) = \sum_{y \in \mathbb{R}: y \leq x} p_X(y).$$

La funzione di ripartizione determina la distribuzione di una variabile aleatoria reale:

$$\mathcal{P}(X \in (-\infty, x]) = \mathcal{P}(X \leq x) = F_X(x) \quad \forall x \in \mathbb{R}.$$

Notiamo che:

$$\mathcal{P}(X \in (a, b]) = F_X(b) - F_X(a) \quad \forall a, b \in \mathbb{R}, a < b.$$

Inoltre, se X è continua in x , allora:

$$\mathcal{P}(X = x) = F_X(x) - \lim_{\epsilon \rightarrow 0+} F_X(x - \epsilon) = 0, \quad \forall x \in \mathbb{R}.$$

Proprietà della funzione di ripartizione

1. F_X è crescente: se $x_1 \geq x_2$, allora $F_X(x_1) \geq F_X(x_2)$;
2. F_X è continua: $\lim_{\epsilon \rightarrow 0} F_X(x + \epsilon) = F_X(x)$, per ogni $x \in \mathbb{R}$;
3. $\lim_{x \rightarrow \infty} F_X(x) = 1$;
4. $\lim_{x \rightarrow -\infty} F_X(x) = 0$;

Def. 4.26 Sia $F : \mathbb{R} \rightarrow [0, 1]$, allora si dice che F è una funzione di ripartizione se:

- F è crescente;
- F è continua;
- $\lim_{x \rightarrow \infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$.

NB sia F una funzione di ripartizione, allora esistono uno spazio di probabilità $(\Omega, \mathcal{F}, \mathcal{P})$ e una variabile aleatoria X su $(\Omega, \mathcal{F}, \mathcal{P})$ tali che F è la funzione di ripartizione di X , cioè:

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

Funzioni di ripartizione notevoli

Def. 4.27 (Uniforme) Siano $a, b \in \mathbb{R}, a < b$, definiamo $F_{Unif(a,b)} : \mathbb{R} \rightarrow [0, 1]$ tramite:

$$F_{Unif(a,b)}(x) := \begin{cases} 0 & \text{se } x < a, \\ \frac{x-a}{b-a} & \text{se } x \in [a, b), \\ 1 & \text{se } x \leq b. \end{cases}$$

La funzione $F_{Unif(a,b)}$ è detta funzione di ripartizione della distribuzione uniforme continua su (a, b) .

Def. 4.28 (Esponenziale) Sia $\lambda > 0$, definiamo $F_{Exp(\lambda)} : \mathbb{R} \rightarrow [0, 1]$ tramite:

$$F_{Exp(\lambda)}(x) := \begin{cases} 0 & \text{se } x < 0, \\ 1 - e^{-\lambda x} & \text{se } x \geq 0. \end{cases}$$

Allora $F_{Exp(\lambda)}$ è detta funzione di ripartizione della distribuzione esponenziale di parametro λ .

NB la distribuzione uniforme continua e la distribuzione esponenziale sono delle distribuzioni continue (anche assolutamente continue), nel senso che le loro funzioni di ripartizione sono continue.

Se X è una variabile aleatoria reale con distribuzione uniforme continua oppure esponenziale, allora per ogni $x \in \mathbb{R}$:

$$\mathcal{P}(X = x) = F_X(x) - \lim_{\epsilon \rightarrow 0+} F_X(x - \epsilon) = 0.$$

poiché, in questo caso, F_X è continua. In particolare, X non possiede una densità discreta (non sarebbe una variabile aleatoria).

Def. 4.29 (Assolutamente continua) Sia $F : \mathbb{R} \rightarrow [0, 1]$ una funzione di ripartizione, allora si dice assolutamente continua se esiste una funzione $f : \mathbb{R} \rightarrow [0, \infty)$ integrabile tale che:

$$F(x) = \int_{-\infty}^x f(t) dt, \quad \forall x \in \mathbb{R}. \quad (72)$$

In questo caso, f è detta densità continua di F .

Come sempre:

$$\lim_{x \rightarrow \infty} F(x) = \int_{-\infty}^{\infty} f(x) dx = 1.$$

Abbiamo già mostrato il caso in cui $F(x) = 0$

Def. 4.30 (Assolutamente continua) Sia X una variabile aleatoria reale su $(\Omega, \mathcal{F}, \mathcal{P})$, allora si dice che X è assolutamente continua se la sua funzione di ripartizione F_X è assolutamente continua, la densità continua di F_X si indica con f_X .

Def. 4.31 (Densità) Sia $f : \mathbb{R} \rightarrow [0, \infty)$ una funzione integrabile, tale che:

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

Allora f è detta densità di una variabile aleatoria assolutamente continua X .

NB Data una densità f , si può definire la funzione di ripartizione F :

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

Def. 4.32 (Normale standard) Sia $f_{N(0,1)}$ definita tramite:

$$f_{N(0,1)}(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \forall x \in \mathbb{R}. \quad (73)$$

allora, $f_{N(0,1)}$ è una densità. La funzione:

$$\Phi(x) := \int_{-\infty}^x f_{N(0,1)}(t) dt, \quad \forall x \in \mathbb{R}$$

è definita funzione di ripartizione della distribuzione normale standard.

Più in generale:

Def. 4.33 (Normale o gaussiana) Sia $\mu \in \mathbb{R}$ e $\sigma > 0$, allora la funzione $f_{N(\mu,\sigma)}$ definita tramite:

$$f_{N(\mu,\sigma)}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \forall x \in \mathbb{R}$$

è la densità della distribuzione normale o gaussiana di media μ e varianza σ^2 .

Come sempre arrivati a questo punto, definiamo il valore atteso, in questo caso di una distribuzione continua.

Def. 4.34 (Valore atteso) Sia X una variabile aleatoria reale su $(\Omega, \mathcal{F}, \mathcal{P})$ con densità continua f_X e sia $g : \mathbb{R} \rightarrow \mathbb{R}$ allora:

$$\mathbb{E}[g(x)] := \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx.$$

se e solo se l'integrale è ben definito.