

Probabilità e Statistica

UniVR - Dipartimento di Informatica

Fabio Irimie

2° Semestre 2023/2024

Indice

1	Cos'è la probabilità e la statistica?	4
1.1	Popolazione, variabili e campione	4
1.2	Parametro e Stima	4
1.3	Variabili	4
2	Statistica descrittiva	5
2.1	Strumenti di sintesi	5
2.1.1	Tabelle di frequenza	5
2.1.2	Distribuzioni	5
2.1.3	Distribuzioni cumulative	5
2.1.4	Grafici	5
3	Frequenze	6
3.1	Frequenze campionarie	6
3.1.1	Frequenza assoluta	6
3.1.2	Frequenza relativa	6
3.2	Frequenze cumulative	6
3.2.1	Frequenza cumulativa assoluta	6
3.2.2	Frequenza cumulativa relativa	7
4	Statistica descrittiva	7
4.1	Indici statistici	7
4.1.1	Indici di posizione o centralità	7
4.1.2	Indici di dispersione	7
4.1.3	Indici di forma	8
4.1.4	Indici di posizione relativi	9
4.1.5	Box-Plot	9
4.2	Outliers	10
4.2.1	Outliers deboli	10
4.2.2	Outliers forti	10
5	Statistica descrittiva bivariata	10
5.1	Relazione tra 2 variabili	10
5.1.1	Correlazione	10
5.1.2	Regressione	11
5.1.3	Determinazione dei coefficienti della retta di regressione	12
5.2	Riassunto	12
6	Probabilità	13
6.1	Esperimenti aleatori	13
6.2	Spazio campionario ed eventi	14
6.3	Esperimenti	14
6.3.1	Esperimento 1: Lancio di un dado	14
6.3.2	Esperimento 2: Lancio di 2 dadi	14
6.3.3	Esperimento 3: Sesso dei nascituri	15
6.3.4	Caratteristiche degli esperimenti 1-3	15
6.3.5	Esperimento 4: Tempo di attesa	15
6.3.6	Esperimento 5: Misure	15
6.3.7	Tipi di esperimenti	16

6.3.8	Tipi di eventi	16
6.4	Spazio campionario e insieme degli eventi	17
6.4.1	Esempi	18
6.5	Probabilità degli esperimenti 1-2	18
6.5.1	Esperimento 1: Lancio di un dado	18
6.5.2	Esperimento 2: Lancio di 2 dadi	18
6.6	Definizione frequentista di probabilità	18
6.7	Definizione soggettiva di probabilità	19
7	Assiomi di Kolmogorov	19
7.1	Assiomi	19
7.1.1	Caso finito	19
7.1.2	Caso generale	20
8	Modelli probabilistici	20
8.1	Regola per l'unione e l'intersezione di eventi	20
8.2	Modello equiprobabile	20
8.3	Spazio di probabilità	21
8.3.1	Caso finito	21
8.3.2	Caso generale	22
8.4	Indipendenza di eventi	22
8.4.1	Proposizione	22
8.5	Probabilità condizionata	22
8.5.1	Esempio	23
8.5.2	Proposizione	24
8.5.3	Come determinare la probabilità condizionata	24
8.6	Probabilità a Priori e a Posteriori (Formula di Bayes)	24
8.6.1	Teorema delle probabilità totali	25
8.6.2	Dimostrazione	25
8.6.3	Teorema di Bayes	25
8.6.4	Dimostrazione	25
8.7	Probabilità a priori e a posteriori	25
8.7.1	Probabilità a priori	25
8.7.2	Probabilità a posteriori	26
8.8	Esercizi	26
8.8.1	Esercizio 1	26
9	Variabili aleatorie discrete	26
9.1	Distribuzione di una variabile aleatoria	27
9.2	Variabili aleatorie discrete	27
10	Schema di Bernoulli	29
10.1	Prove dicotomiche ripetute ed indipendenti	29
10.2	Conteggio del successo in ciascuna prova	30
10.3	Conteggio del successo in ciascuna delle n volte	30
10.3.1	Modello per il conteggio del successo	31
10.4	Processi stocastici (prove indipendenti ripetute infinite volte)	31
10.5	La variabile geometrica	32
10.6	Eventi rari: conteggio dei successi per un numero infinito di esperimenti	32

11 La variabile di Poisson	33
11.1 Eventi rari: il processo di Poisson	34
11.1.1 Conteggio del tempo del primo successo	34
12 Variabili aleatorie continue	34
13 Statistica inferenziale	35
13.1 Distribuzione esponenziale	35
14 Tipi di stima	36
14.1 Stima intervallare	36
14.2 Stima puntuale	36
14.2.1 Definizione	37
14.2.2 Stimatori	37
14.2.3 Stimatori di massima verosomiglianza	37
14.2.4 Bontà di uno stimatore	39

1 Cos'è la probabilità e la statistica?

La statistica è una scienza che si occupa di raccogliere, organizzare, analizzare e interpretare i dati. Nella statistica si cerca di estrapolare informazioni da esperimenti **aleatori** (esperimenti che non si possono ripetere esattamente allo stesso modo) e di prendere decisioni basate su queste informazioni. Ogni esperimento aleatorio ha bisogno di un **modello probabilistico** che ne descriva le caratteristiche principali.

1.1 Popolazione, variabili e campione

- **Popolazione:** tutti i possibili oggetti di un'indagine statistica
- **Individuo:** un singolo oggetto della popolazione
- **Variabile:** una qualsiasi caratteristica di un individuo della popolazione soggetta a possibili variazioni da individuo a individuo; è l'oggetto di interesse in uno studio
- **Range della variabile:** R_x è l'insieme di tutti i possibili valori che la variabile x può assumere
- **Campione:** un sottoinsieme rappresentativo della popolazione composto dalle variabili relative ad un sottoinsieme di individui
- **Realizzazione del campione di dimension n :** (post esperimento) le osservazioni del campione:

$$\underline{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$$

- **Range dei dati:** $\mathcal{R}_{\underline{x}}$ i valori che la variabile può assumere tra il minimo e il massimo

1.2 Parametro e Stima

- **Parametro:** una misura che descrive una proprietà dell'intera popolazione
- **Stima:** una misura che descrive una proprietà del campione e che fornisce informazioni sul parametro

1.3 Variabili

Le variabili possono essere di diverso tipo:

- **Variabili qualitative nominali:**
 - **Ordinali:** possono essere ordinate
 - **Non ordinali:** non possono essere ordinate

I valori che assumono si definiscono anche **modalità**

- **Variabili quantitative:** Sono valori numerici e si distinguono in:
 - **Aleatorie continue:** derivano da processi di misura e assumono i loro range (valori che possono assumere). Sono sottoinsiemi reali
 - **Aleatorie discrete:** derivano da processi di conteggio e assumono valori interi

2 Statistica descrittiva

Consiste nella raccolta, organizzazione, rappresentazione e analisi dei **dati**.

2.1 Strumenti di sintesi

2.1.1 Tabelle di frequenza

Sono tabelle di frequenze di individui con una certa caratteristica o aventi una caratteristica appartenente ad un certo intervallo.

- **Frequenza assoluta:** conteggio del numero di individui
- **Frequenza relativa:** percentuale del numero di individui
- **Frequenza cumulativa:** conteggio o percentuale del numero di individui fino ad un certo punto

2.1.2 Distribuzioni

Sono rappresentazioni del modo in cui diverse **modalità** si distribuiscono tra gli individui di una popolazione.

- **Caso discreto:** f : valore variabile \rightarrow frequenza relativa
- **Caso continuo o numerabile:** f : intervallo di valori variabile \rightarrow frequenza relativa

2.1.3 Distribuzioni cumulative

Sono distribuzioni che rappresentano la frequenza cumulativa di una variabile. Possono essere:

- **Caso discreto:** f : valore variabile \rightarrow frequenza cumulativa relativa
- **Caso continuo o numerabile:** f : intervallo \rightarrow frequenza cumulativa relativa

2.1.4 Grafici

Sono rappresentazioni grafiche delle distribuzioni. Possono essere:

- **Istogrammi:** è costituito da rettangoli, insistenti sulle classi della partizione, attigui le cui aree sono confrontabili con le probabilità.

$$\text{area rettangolo } i = h_i \cdot |\pi_i| \approx P_X(\pi) \approx f_i$$

$$h_i = \frac{f_i}{|\pi_i|} \quad \text{per ogni } i \in I$$

L'area del rettangolo che insiste sulla classe π_i della partizione è pari alla frequenza relativa della classe, quindi l'area totale è 1.

- **Diagrammi a barre:** rappresentano le frequenze di una variabile. Le barre sono separate e la loro altezza è proporzionale alla frequenza

- **Diagrammi a torta:** rappresentano le frequenze relative di una variabile
- **Boxplot:** rappresentano le frequenze di una variabile
- **Poligono di frequenza (ogiva):** è un grafico a linee continue che ha sull'asse delle ordinate le frequenze cumulative. Questo tipo di grafici è il più comune per rappresentare le frequenze cumulative.

3 Frequenze

Siano $\underline{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$ una realizzazione del campione di dimensione n e $\mathcal{R}_{\underline{x}}$ il range dei dati. Si dice **partizione** di $\mathcal{R}_{\underline{x}}$:

$$\pi = \{\pi_i\}_{i \in I}$$

La **classe i-esima** è l'elemento i -esimo della partizione

3.1 Frequenze campionarie

3.1.1 Frequenza assoluta

Si dice **frequenza assoluta** n_i per ogni $i \in I$ il numero di osservazioni che appartengono a π_i , cioè:

$$n_i = \text{card}(\tilde{x}_j \in \pi_i, \quad j = 1, \dots, n) \quad (\text{cardinalità})$$

$$0 \leq n_i \leq n, \text{ per ogni } i \in I \quad e \quad \sum_{i \in I} n_i = n$$

3.1.2 Frequenza relativa

Si dice **frequenza relativa** f_i per ogni $i \in I$ la percentuale delle osservazioni che appartengono a π_i , cioè:

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

$$0 \leq f_i \leq 1, \text{ per ogni } i \in I \quad e \quad \sum_{i \in I} f_i = 1$$

3.2 Frequenze cumulative

3.2.1 Frequenza cumulativa assoluta

Si dice **frequenza cumulativa assoluta** N_i il numero di osservazioni che appartengono alle classi π_h , con $h \leq i$, cioè:

$$N_i = \sum_{h=1}^i n_h$$

$$0 \leq N_i \leq n, \text{ per ogni } i \in I \quad e \quad N_i \leq N_j, \quad i < j$$

3.2.2 Frequenza cumulativa relativa

Si dice **frequenza cumulativa relativa** F_i della i -esima classe la somma delle frequenze relative delle classi π_h , con $h \leq i$, cioè:

$$F_i = \sum_{h=1}^i f_h = \frac{1}{n} N_i = \frac{1}{n} N_{i-1} + f_i$$

$$0 \leq F_i \leq 1, \text{ per ogni } i \in I \quad e \quad F_i \leq F_j, i < j$$

4 Statistica descrittiva

4.1 Indici statistici

Sono misure quantitative che forniscono informazioni sulla distribuzione di una certa caratteristica.

4.1.1 Indici di posizione o centralità

Forniscono informazioni del valore attorno al quale si posizionano i dati. Consentono di valutare l'ordine di grandezza della variabile aleatoria e aiutano a "localizzare" la distribuzione. Sono espressi nella stessa unità di misura della variabile.

Sia $\underline{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$ un campione di dimensione n .

- **Media campionaria:** è il valore medio dei dati (baricentro dei dati):

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n \tilde{x}_j$$

- **Moda campionaria:** m , valore che si ripete più frequentemente. Ci possono essere più valori modali.

Sia $\underline{y} = (y_1, \dots, y_n)$ il campione ordinato ($y_i \in \{\tilde{x}_1, \dots, \tilde{x}_n\}$ e $y_i \leq y_{i+1}$)

- **Mediana campionaria:** M : è il valore centrale del campione, una volta ordinato.

$$M = \begin{cases} y_{\frac{n+1}{2}} & \text{se } n \text{ è dispari} \\ \frac{1}{2}(y_{\frac{n}{2}} + y_{\frac{n}{2}+1}) & \text{se } n \text{ è pari} \end{cases}$$

4.1.2 Indici di dispersione

Forniscono informazioni su quanto i dati si disperdono attorno ad un valore centrale. Sono:

- **Range:** differenza tra il massimo e il minimo valore:

$$r = \max_{j \in \{1, \dots, n\}} \tilde{x}_j - \min_{j \in \{1, \dots, n\}} \tilde{x}_j$$

- **Scarto Quadratico Medio campionario:** misura la dispersione dei dati attorno alla media

$$s'^2 = \frac{1}{n} \sum_{j=1}^n (\tilde{x}_j - \bar{x})^2$$

- **Varianza campionaria:** misura la dispersione dei dati attorno alla media

$$s^2 = \frac{1}{n-1} \sum_{j=1}^n (\tilde{x}_j - \bar{x})^2$$

- **Deviazione standard campionaria:** misura la distanza dei dati attorno alla media

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (\tilde{x}_j - \bar{x})^2}$$

Per interpretare la deviazione standard si possono definire **valori usuali** di una variabile i valori del campione compresi tra:

- **Minimo valore "usuale":** media campionaria - 2 deviazioni standard
- **Massimo valore "usuale":** media campionaria + 2 deviazioni standard

4.1.3 Indici di forma

Sia $\underline{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$ un campione di dimensione n .

- **Asimmetria / Skewness:** misura la simmetria della distribuzione

$$\gamma_1 = \frac{1}{n-1} \sum_{j=1}^n \left(\frac{\tilde{x}_j - \bar{x}}{s} \right)^3$$

- $\gamma_1 > 0$: distribuzione asimmetrica a destra (con coda più lunga a destra)
- $\gamma_1 < 0$: distribuzione asimmetrica a sinistra (con coda più lunga a sinistra)
- $\gamma_1 = 0$: distribuzione simmetrica

- **Curtosi:** misura la "appuntitura" della distribuzione

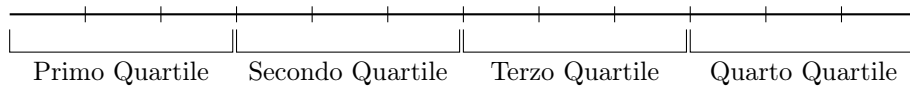
$$\gamma_2 = \frac{1}{n-1} \sum_{j=1}^n \left(\frac{\tilde{x}_j - \bar{x}}{s} \right)^4$$

- $\gamma_2 = 3$: curtosi della normale standard, (variabile di riferimento)
- $\gamma_2 > 3$: ci sono meno valori agli estremi di quanto aspettato, e di conseguenza si ha una minore dispersione dei dati. In tal caso la distribuzione risulta abbastanza appuntita
- $\gamma_2 < 3$: ci sono più valori agli estremi di quanto aspettato, e di conseguenza si ha una maggiore dispersione dei dati. In tal caso la distribuzione risulta piatta

4.1.4 Indici di posizione relativi

Rappresentano indici di posizione, ma non centrali, bensì indici di posizionamento relativo.

- **Percentili:** Se p è un numero tra 0 e 100, il **percentile di ordine p** (o p -esimo percentile, se p è intero) è il dato che delimita il primo $p\%$ dei dati (ordinati) dai rimanenti dati.
- **Quartili:** Valori che separano i dati in quattro parti, una volta ordinati.



$\underline{x} = (x_1, \dots, x_n)$ campione di dimensione n

$\underline{y} = (y_1, \dots, y_n)$ campione ordinato

Il primo quartile è il valore che separa il 25% inferiore dal 75% superiore dei dati.

$$Q_1 = \begin{cases} \frac{y_{\frac{n}{4}} + y_{\frac{n}{4}+1}}{2} & \frac{n}{4} \text{ intero} \\ y_{\lceil \frac{n}{4} \rceil} & \frac{n}{4} \text{ non intero} \end{cases}$$

Il secondo quartile è il 50-esimo percentile, ovvero la mediana. È il valore che separa il 50% inferiore dal 50% superiore dei dati.

$$Q_2 = M = \begin{cases} \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2} & \frac{n}{2} \text{ intero} \\ y_{\lceil \frac{n}{2} \rceil} & \frac{n}{2} \text{ non intero} \end{cases}$$

Il terzo quartile è il 75-esimo percentile, ovvero il valore che separa il 75% inferiore dal 25% superiore dei dati.

$$Q_3 = \begin{cases} \frac{y_{\frac{3n}{4}} + y_{\frac{3n}{4}+1}}{2} & \frac{3n}{4} \text{ intero} \\ y_{\lceil \frac{3n}{4} \rceil} & \frac{3n}{4} \text{ non intero} \end{cases}$$

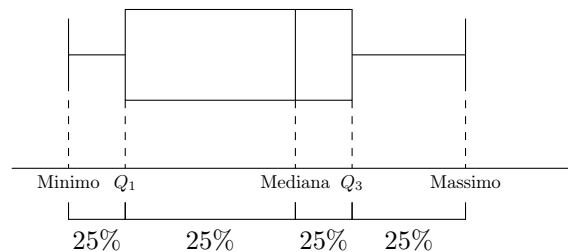
Lo scarto (o distanza interquartile) è la differenza tra il terzo e il primo quartile:

$$IR = Q_3 - Q_1$$

4.1.5 Box-Plot

Fornisce informazioni sulla forma della distribuzione:

Box-Plot



4.2 Outliers

Definizione 4.1

Gli **Outliers** sono valori estremi, insolitamente grandi o piccoli, rispetto al resto dei dati. La loro presenza potrebbe distorcere i risultati dell'analisi, e richiede pertanto un'analisi più accurata.

$$x \leq Q_1 - 1.5 \cdot IR \quad \text{oppure} \quad x \geq Q_3 + 1.5 \cdot IR$$

4.2.1 Outliers deboli

Si dicono outliers deboli:

$$Q_1 - 3 \cdot IR < x \leq Q_1 - 1.5 \cdot IR$$

oppure

$$Q_3 + 1.5 \cdot IR < x \leq Q_3 + 3 \cdot IR$$

4.2.2 Outliers forti

Si dicono outliers forti:

$$x \leq Q_1 - 3 \cdot IR$$

oppure

$$x \geq Q_3 + 3 \cdot IR$$

5 Statistica descrittiva bivariata

La statistica descrittiva bivariata si occupa di studiare la relazione tra due variabili.

5.1 Relazione tra 2 variabili

- **Correlazione:** Associazione **lineare** tra 2 variabili. La forza dell'associazione è data dal **coefficiente di correlazione**.
- **Regressione:** dipendenza di una variabile (dipendente) da un'altra variabile (indipendente)

5.1.1 Correlazione

Sia $(\underline{x}, \underline{y}) = ((\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_n, \tilde{y}_n))$ un campione di dimensione n di due misure x ed y , con medie campionarie \bar{x} e \bar{y} , deviazioni standard campionarie (s_x, s_y) .

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n \tilde{y}_i$$

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\tilde{x}_i - \bar{x})^2}$$

$$s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}$$

Il coefficiente di correlazione campionario è definito come:

$$\rho_n \triangleq \frac{\sum_{i=1}^n (\tilde{x}_i - \bar{x})(\tilde{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\tilde{x}_i - \bar{x})^2 \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}}$$

Il risultato sarà un numero compreso tra -1 e 1 :

$$|\rho_n| \leq 1$$

Questo indice misura il grado di dipendenza lineare tra le due variabili.

$ \rho_n $	Grado di correlazione tra \underline{x} e \underline{y}
$\rho_n = -1$	massima correlazione lineare inversa
$-1 < \rho_n < 0$	correlazione inversa
$\rho_n = 0$	assenza di correlazione
$0 < \rho_n < 1$	correlazione diretta
$\rho_n = 1$	massima correlazione lineare diretta

Sono indici qualitativi:

ρ_n	Grado di correlazione tra \underline{x} e \underline{y}
$ \rho_n \leq 0.5$	scarsa correlazione
$0.5 < \rho_n \leq 0.75$	correlazione moderata
$0.75 < \rho_n \leq 0.9$	correlazione buona
$ \rho_n > 0.9$	correlazione molto buona

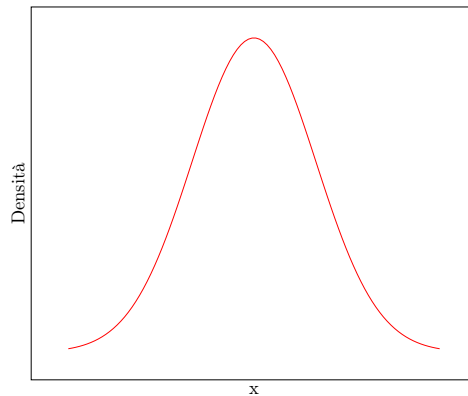
5.1.2 Regressione

La regressione lineare è un modello matematico che cerca di esprimere una variabile. Per ipotesi riteniamo che due variabili siano legate da una relazione del tipo $y = g(x)$

1. I dati accoppiati (x, y) costituiscono un campione di dati quantitativi
2. Dallo scatter plot possiamo ipotizzare che nella **popolazione** ci sia una relazione lineare del tipo:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

dove ε_i è l'errore casuale, con distribuzione a campana



3. Cerchiamo di individuare l'equazione della **curva di regressione relativa del campione**:

$$\hat{y}_i = a + bx_i$$

5.1.3 Determinazione dei coefficienti della retta di regressione

L'obiettivo è quello di determinare i coefficienti a e b in modo ottimale, affinché la retta di regressione $\hat{y}_i = a + bx_i$ sia il più possibile vicina ai punti (x_i, y_i) del campione.

Si determina quindi l'equazione generica della curva interpolante stimando i parametri in modo da rendere **minima** la distanza al quadrato dei punti osservati dalla curva.

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

Equazioni normali:

$$\begin{cases} \sum_{i=1}^n y_i = na + b \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 \end{cases}$$

5.2 Riassunto

- Dato **un campione**: abbiamo determinato una **stima** di alcuni parametri (media, deviazione standard, varianza, quartili, ...), una stima della distribuzione (frequenze relative) con grafici (istogramma [frequenza relativa], diagrammi [area = frequenza relativa], boxplot [quartili, outliers])
- Dati **due campioni**: abbiamo determinato una **stima** di alcuni parametri (media, deviazione standard, varianza, quartili, ...) ed una stima della distribuzione (frequenze relative) con grafici (scatter plot, retta di regressione, coefficiente di correlazione) e abbiamo fatto un **confronto**.

Abbiamo determinato una stima della **correlazione** e la retta di regressione lineare.

$$\rho_n = \text{coeff. di correlazione} \quad \rho_n \approx 1$$

Per capire se le informazioni tratte dal campione sono statisticamente significative si fa riferimento alla **statistica inferenziale**. Ma bisogna essere in grado di parlare di probabilità e di distribuzioni teoriche (modelli probabilistici).

6 Probabilità

La probabilità di un evento $A \in \mathcal{A}$ rappresenta una misura di quanto ci si aspetta che si verifichi l'evento A .

Calcolare le probabilità non significa "prevedere il futuro", ma trovare come distribuire un maggiore o minore **grado di fiducia** tra i vari possibili modi in cui si potrà presentare un certo fenomeno aleatorio.

Definizioni utili 6.1

*L'ipotesi dei modelli è lo spazio dei campioni **finito** $\Leftrightarrow \text{card}(\Omega) = n < \infty$*

Eventi equiprobabili:

$$P(\omega_i) = P(\omega_j), \quad i, j \in \{1, \dots, n\}$$

La probabilità di un evento $A \in \mathcal{A}$ si calcola come:

$$P(A) = \frac{\text{casi favorevoli ad } A}{\text{casi possibili}} = \frac{\text{card}(A)}{\text{card}(\Omega)}$$

6.1 Esperimenti aleatori

Un fenomeno **casuale**, o aleatorio, è un fenomeno **osservabile**, ma non prevedibile. Cioè conoscendo i dati iniziali e le leggi, non possiamo prevederne il risultato. Ciò che invece possiamo conoscere è l'insieme di tutti i possibili risultati.

- **Fenomeno deterministico:** Dati + Leggi = Conoscenza
- **Fenomeno non deterministico:** Dati + Leggi = Non Conoscenza

Alcuni esempi di esperimenti sono:

- Consideriamo tre figli di una stessa coppia. Controlliamo il sesso dei tre.
- Lancio un dado. Controllo il numero che esce.
- Lancio 2 dadi. Controllo i numeri che escono.
- Considero i piselli so che possono avere il baccello verde o giallo e il fiore bianco o viola. Ne estraggo uno a caso. Che caratteristiche ha?
- Sono ad un call center. Conto il numero di telefonate che arrivano in un intervallo di tempo
- Misuro all'altezza di un uomo di 40 anni italiano

6.2 Spazio campionario ed eventi

È l'insieme di tutti i possibili risultati di un esperimento casuale:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$$

Uno dei possibili risultati dell'esperimento si chiama **Evento elementare**:

$$\{\omega_i\}, \quad i = 1, \dots, n$$

L'**Evento** è un sottoinsieme dello spazio campione $A \subset \Omega$ in cui sono contenuti alcuni dei possibili eventi elementari, quelli favorevoli all'evento considerato.

6.3 Esperimenti

6.3.1 Esperimento 1: Lancio di un dado

Prendiamo in considerazione il lancio di un dado:

$$\text{Lo spazio dei campioni è: } \Omega = \{1, 2, 3, 4, 5, 6\}$$

I possibili eventi sono:

A = Il risultato del lancio è 1

B = Il risultato del lancio è dispari

C = Il risultato del lancio è maggiore di 4

D = Il risultato del lancio è dispari non maggiore di 4

E = Il risultato del lancio è pari

F = Il risultato del lancio è 7

G = Il risultato del lancio è tra 1 e 6

$$A = \{1\} \quad B = \{1, 3, 5\}$$

$$C = \{1, 2, 3, 4\} \quad D = \{1, 3, 5\} \cap \{1, 2, 3, 4\} = B \cap C = \{1, 3\}$$

$$E = \{2, 4, 6\} = \Omega \setminus B = \overline{\{1, 3, 5\}} = \overline{B}$$

$$F = \{7\} = \overline{\Omega} = \emptyset$$

$$G = \{1, 2, 3, 4, 5, 6\} = \Omega$$

6.3.2 Esperimento 2: Lancio di 2 dadi

Prendiamo in considerazione il lancio di 2 dadi:

$$\Omega_2 = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} = \{(1, 1), (1, 2), \dots, (6, 6)\}$$

L'evento: A = Esce almeno un 6 è:

$$A = \{(6, 1), (6, 2), \dots, (6, 6), (1, 6), (2, 6), \dots, (5, 6)\}$$

6.3.3 Esperimento 3: Sesso dei nascituri

Consideriamo 3 figli di una stessa coppia. Controlliamo il sesso dei tre.

Se considero una **singola nascita** lo spazio dei campioni è:

$$\Omega = \{M, F\}$$

Quindi si hanno due possibili eventi elementari:

$$\{M\}, \{F\}$$

Se invece considero **tre nascite** lo spazio dei campioni è:

$$\Omega_3 = \{(\omega_1, \omega_2, \omega_3) \mid \omega_i \in \Omega\}$$

quindi è costituito da tutte le **terne** ordinate di maschi e femmine.

1° Figlio	2° Figlio	3° Figlio
M	M	M
M	M	F
M	F	M
M	F	F
F	M	M
F	M	F
F	F	M
F	F	F

Ogni terna rappresenta un **evento elementare**.

6.3.4 Caratteristiche degli esperimenti 1-3

- Lo spazio dei campioni è finito
- Gli eventi sono tutte le parti di Ω , cioè tutti i possibili sottoinsiemi di Ω

6.3.5 Esperimento 4: Tempo di attesa

Sono ad un call center e conto il numero di telefonate che arrivano in un intervallo di tempo.

Lo spazio dei campioni è:

$$\Omega = \{0, 1, 2, 3, \dots\}$$

Caratteristiche:

- Lo spazio dei campioni è **infinito numerabile**
- Gli eventi sono tutte le parti di Ω , cioè tutti i possibili sottoinsiemi di Ω

6.3.6 Esperimento 5: Misure

Misuro l'altezza di un uomo di 40 anni italiano.

Lo spazio dei campioni è:

$$\Omega \subseteq \mathbb{R}$$

Caratteristiche:

- Lo spazio dei campioni è un sottoinsieme di \mathbb{R} , quindi è **infinito non numerabile**
- Gli eventi sono tutti i sottointervalli di \mathbb{R} , le loro unioni e le loro intersezioni

6.3.7 Tipi di esperimenti

Gli esperimenti possono essere di diversi tipi:

- **Misure di conteggio**
- **Misure continue**

6.3.8 Tipi di eventi

- **Evento aleatorio:**
È un sottoinsieme dello spazio campionario, cioè $A \subset \Omega$, ad esempio il lancio di un dado un evento aleatorio potrebbe essere: "esce un numero pari"
- **Evento elementare:**
È un evento che contiene un solo elemento, cioè $A = \{\omega\}$, ad esempio il lancio di un dado ha come eventi elementari: "esce 1", "esce 2", "esce 3", "esce 4", "esce 5", "esce 6"
- **Eventi complementari:**
Sono eventi che si escludono a vicenda, ad esempio nel lancio di un dado: $E = \text{"esce un numero pari"}$ e $\bar{E} = \text{"esce un numero dispari"}$ sono eventi complementari
- **Eventi incompatibili:**
Sono eventi che non possono verificarsi contemporaneamente, ad esempio nel lancio di un dado: $E = \text{"esce un numero pari"}$ e $F = \text{"esce il numero 6"}$ sono eventi incompatibili
- **Eventi compatibili:**
Sono eventi che possono verificarsi contemporaneamente, ad esempio nel lancio di un dado: $E = \text{"esce un numero pari"}$ e $F = \text{"esce il numero 2"}$ sono eventi compatibili
- **Evento certo:**
È un evento che si verifica sempre, cioè $A = \Omega$, ad esempio il lancio di un dado ha sempre un risultato certo.
- **Evento impossibile:**
È un evento che non si verifica mai, cioè $A = \emptyset$, ad esempio il lancio di un dado non può avere come risultato 7.

6.4 Spazio campionario e insieme degli eventi

Definizione 6.1

Lo **spazio dei campioni** Ω è l'insieme di tutti i possibili esiti (risultati). La cardinalità di uno spazio dei campioni può essere finita, infinita numerabile e infinita non numerabile.

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\} \quad \text{oppure} \quad \Omega \subseteq \mathbb{R}$$

Definizione 6.2

L'**insieme degli eventi** \mathcal{A} è un insieme **finito** di parti di Ω tali che sia un'algebra, cioè tale che:

$$A_1. \quad \Omega \in \mathcal{A}$$

$$A_2. \quad \text{Unione di eventi è un evento}$$

$$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$$

$$A_3. \quad \text{se } A, B \in \mathcal{A}, \text{ allora } A \setminus B \in \mathcal{A}$$

$$\Downarrow$$

$$A \in \mathcal{A} \Rightarrow A^c = \Omega \setminus A \in \mathcal{A}$$

L'insieme degli eventi rappresenta tutti gli eventi che ci **interessati** rispetto all'esperimento preso in considerazione, e che **ben descrivono** l'esperimento stesso.

Definizione 6.3

σ -algebra \mathcal{F} è un insieme qualsiasi \mathcal{F} di parti di Ω tali che:

$$A_1. \quad \Omega \in \mathcal{F}$$

$$A_2^{\sigma}. \quad \text{sia } \{A_n\}_{n \in \mathbb{N}} \text{ con } A_n \in \mathcal{F}, \text{ allora } \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$$

$$A_3^{\sigma}. \quad \text{se } A, B \in \mathcal{F}, \text{ allora } A \setminus B \in \mathcal{F} \text{ Diremo } \textbf{evento} \text{ ogni sottoinsieme } A \in \mathcal{F}.$$

$$\Downarrow$$

$$A \in \mathcal{F} \Rightarrow A^c = \Omega \setminus A \in \mathcal{F}$$

6.4.1 Esempi

Esempio 6.1

Lancio il dado e controllo che numero esce

$$\begin{aligned}\mathcal{A} = \mathcal{P}(\Omega) &= \\ &= \left\{ \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \right. \\ &\quad \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}, \\ &\quad \{2, 3\}, \{2, 4\}, \{2, 5\}, \{2, 6\}, \\ &\quad \dots \\ &\quad \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{1, 2, 6\}, \dots \\ &\quad \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 3, 6\}, \dots \\ &\quad \{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 6\}, \dots \\ &\quad \left. \{1, 2, 3, 4, 5, 6\} = \Omega, \emptyset \right\}\end{aligned}$$

6.5 Probabilità degli esperimenti 1-2

6.5.1 Esperimento 1: Lancio di un dado

$$\begin{aligned}\Omega_1 &= \{1, 2, 3, 4, 5, 6\} \\ A &= \text{esce almeno un 6} \\ P(\{i\}) &= \frac{\text{casi favorevoli}}{\text{casi possibili}} = \frac{\text{card}(\{i\})}{\text{card}(\Omega)} = \frac{1}{6}, \quad i = 1, \dots, 6\end{aligned}$$

6.5.2 Esperimento 2: Lancio di 2 dadi

$$\begin{aligned}\Omega_2 &= \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} = \{(1, 1), (1, 2), \dots, (6, 6)\} \\ A &= \text{esce almeno un 6} \\ P(A) &= \frac{\text{casi favorevoli ad } A}{\text{casi possibili}} = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{11}{36}\end{aligned}$$

6.6 Definizione frequentista di probabilità

Definizioni utili 6.2

L'ipotesi dei modelli deve essere ripetibile all'esperimento, quindi bisogna avere tante prove ripetute (nelle stesse condizioni) ed indipendenti

La probabilità di un evento $A \in \mathcal{A}$, fatte n prove:

$$P(A) = \frac{\text{numero di occorrenze di } A}{n} = f_n(A)$$

Si basa sulla **legge empirica del caso** che sintetizza una regolarità osservabile sperimentalmente.

6.7 Definizione soggettiva di probabilità

È la misura del grado di fiducia che un individuo **coerente** assegna al verificarsi di un dato evento in base alle sue **conoscenze**

Probabilità di un evento $A \in \mathcal{A}$:

$$P(A) = \frac{\text{posta}}{\text{vincita}} = \frac{P}{V}$$

In breve, "se ci credo, pago"

7 Assiomi di Kolmogorov

L'impostazione assiomatica permette a Kolmogorov di non esplicitare esattamente come valutare la probabilità (lasciando quindi la libertà di seguire l'approccio più adatto al caso in esame), ma di limitarsi solo a indicare quali sono le regole formali che una misura di probabilità deve soddisfare per poter essere dichiarata tale.

7.1 Assiomi

7.1.1 Caso finito

Definizione 7.1

$$(\Omega, \mathcal{A}, P)$$

$$P_1. P(\Omega) = 1$$

$$P_2. \text{ sia } A, B \in \mathcal{A} \text{ disgiunti, t.c}$$

$$A \cap B = \emptyset$$

allora

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

(additività finita)

7.1.2 Caso generale

Definizione 7.2

$$(\Omega, \mathcal{A}, P)$$

$$P_1. P(\Omega) = 1$$

$$P_2^\sigma. \text{ sia } \{A_n\}_n, A_n \in \mathcal{F} \text{ disgiunti t.c.}$$

$$A_i \cap A_j = \emptyset, \quad i \neq j$$

allora

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

(σ -additività)

8 Modelli probabilistici

8.1 Regola per l'unione e l'intersezione di eventi

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità. Siano $A, B \in \mathcal{A}$ due eventi. Allora:

- **L'unione di A e B :**

$$P(A \cup B) = \begin{cases} P(A) + P(B) & \text{se } A \text{ e } B \text{ sono disgiunti} \\ P(A) + P(B) - P(A \cap B) & \text{caso generale} \end{cases}$$

- **L'intersezione di A e B :**

$$P(A \cap B) = \begin{cases} P(A)P(B) & \text{se } A \text{ e } B \text{ sono indipendenti} \\ P(A|B)P(B) = P(A)P(B|A) & \text{caso generale} \end{cases}$$

8.2 Modello equiprobabile

L'equiprobabilità è un modello in cui tutti gli eventi elementari hanno la **stessa probabilità** di verificarsi. Ciò implica che lo spazio dei campioni deve essere **finito**.

Definizione 8.1 (Proprietà della probabilità)

$$P_1. P(\Omega) = 1$$

$$P_2. \text{ Se } A, B \in \mathcal{P}(\Omega) \text{ tale che:}$$

$$A \cap B = \emptyset, \text{ allora}$$

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

\Downarrow

Definizione 8.2 (Probabilità uniforme)

$$\begin{aligned} 1 = P(\Omega) &= P(\{1, 2, 3, 4, 5, 6\}) = P(\cup_i \{\omega_i\}) = \\ &= \sum_i P(\{\omega_i\}) = \text{card}(\Omega) \cdot P(\{\omega_i\}) \end{aligned}$$

da cui:

$$P(\{\omega_i\}) = \frac{1}{\text{card}(\Omega)}, \text{ per ogni } \omega_i \in \Omega$$

Definizione 8.3 (Modello equiprobabile o uniforme)

È lo spazio di probabilità (Ω, \mathcal{A}, P) tale che:

M_1 . Ω è finito, cioè la cardinalità di Ω ($\text{card}(\Omega) \in \mathbb{N}$) è tale che:

$$\text{card}(\Omega) < \infty$$

M_2 . $\mathcal{A} = \mathcal{P}(\Omega)$ è l'insieme delle parti di Ω

M_3 . per ogni $\omega \in \Omega$

$$P(\{\omega\}) = \text{costante}$$

8.3 Spazio di probabilità

8.3.1 Caso finito

Definizione 8.4 (Spazio di probabilità)

Dato lo spazio dei campioni $\Omega = \{\omega_1, \dots, \omega_n\}$ e l'insieme degli eventi \mathcal{A} , la probabilità P è definita come:

P_1 . $P(\Omega) = 1$

P_2 . Se $A, B \in \mathcal{A}$, disgiunti, tale che:

$$A \cap B = \emptyset, \text{ allora}$$

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

(additività finita)

8.3.2 Caso generale

Definizione 8.5 (Spazio di probabilità)

Dato lo spazio dei campioni $\Omega = \{\omega_1, \dots, \omega_n\}$, oppure $\Omega \subseteq \mathbb{R}$ e la σ -algebra \mathcal{F} , la probabilità P è definita come:

$$P_1. \quad P(\Omega) = 1$$

P_2 . sia $\{A_n\}_n, A_n \in \mathcal{F}$, disgiunti, tale che:

$$A_i \cap A_j = \emptyset, \quad \text{per } i \neq j; \text{ allora}$$

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

(σ -additività)

8.4 Indipendenza di eventi

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità. Siano $A, B \in \mathcal{A}$ due eventi. Allora A e B si dicono **indipendenti** se:

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

In breve due eventi si dicono indipendenti se l'occorrenza di uno dei due non influenza la probabilità di occorrenza dell'altro.

8.4.1 Proposizione

Siano $A, B \in \mathcal{A}$ due eventi indipendenti. Allora:

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

L'intersezione:

$$P(A \cap B) = \begin{cases} P(A)P(B) & \text{se } A \text{ e } B \text{ sono indipendenti} \\ P(A|B)P(B) = P(A)P(B|A) & \text{caso generale} \end{cases}$$

8.5 Probabilità condizionata

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità **uniforme**. Siano $A, B \in \mathcal{A}$ due eventi. Prendiamo per ipotesi che $P(B) > 0$, quindi B è accaduto, cioè B diventa un evento certo $\Rightarrow \text{prob}(B) = 1$. Allora la relazione tra la probabilità P e la probabilità P_B sullo spazio ristretto a B è:

$$P_B(A) = \frac{\text{card}_E(A)}{\text{card}(\Omega_B)} = \frac{\text{card}(A \cap B)}{\text{card}(B)} = \frac{\text{card}(A \cap B)}{\text{card}(B)} \cdot \frac{\text{card}(\Omega)}{\text{card}(\Omega)} = \frac{P(A \cap B)}{P(B)}$$

Quindi:

Definizione 8.6

La **Probabilità condizionata a B** per ogni $A \in \Omega$ è:

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

8.5.1 Esempio**Esempio 8.1**

Consideriamo l'esperimento che consiste nel lancio di un dado equo. Consideriamo i due eventi:

$$A = \text{"esce un numero"} > 3 = \{4, 5, 6\}$$

$$B = \text{"esce un numero pari"} = \{2, 4, 6\}$$

Si lanci il dado una volta:

B1. Calcolare la probabilità dell'evento A

B2. Lanciato il dado una persona guarda il risultato e afferma che è uscito un numero pari, cioè è accaduto l'evento B. Sapendo questa informazione, come diventa la probabilità dell'evento A?

Il dado è equo, quindi si utilizza il modello equiprobabile.

B1. Calcolare $P(A)$:

Lo spazio dei campioni è:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

La probabilità di A è:

$$P(\{\omega\}) = \frac{1}{\text{card}(\Omega)} = \frac{1}{6}, \quad \omega \in \Omega$$

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{3}{6} = \frac{1}{2}$$

B2. Calcolare $P(A|B)$: probabilità dell'evento A sapendo che sia occorso B

Lo spazio dei campioni è:

$$\Omega_B = \{2, 4, 6\}$$

La probabilità di A sapendo che sia occorso B è:

$$P_B(\{\omega\}) = \frac{1}{\text{card}(\Omega_B)} = \frac{1}{3}, \quad \omega \in \Omega_B$$

$$P_B(A) = \frac{\text{card}(A)}{\text{card}(\Omega_B)} = \frac{2}{3}$$

Condizionare ad un evento significa costruire un nuovo spazio di probabilità ristretto all'evento condizionante:

Spazio dei campioni: $\Omega \rightarrow \Omega_B = B$

Misura di probabilità: $P \rightarrow P_B : P_B(\Omega_B) = \frac{\text{card}(\Omega_B)}{\text{card}(\Omega_B)} = 1 = P(\Omega)$

La relazione tra P e P_B :

$$P_B(A) = \frac{\text{card}_B(A)}{\text{card}(\Omega_B)} = \frac{\text{card}(A \cap B)}{\text{card}(B)} = \frac{\text{card}(A \cap B)}{\text{card}(\Omega)} \frac{\text{card}(\Omega)}{\text{card}(B)} = \frac{P(A \cap B)}{P(B)}$$

8.5.2 Proposizione

$P_B(\cdot) = P(\cdot|B)$ è una misura di **probabilità** su (Ω, \mathcal{A}) tale che:

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} \quad A \cap B \neq \emptyset$$

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} = 0 \quad A \cap B = \emptyset$$

$$P_B(\Omega) = P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = 1$$

$$P_B(B) = P(B|B) = \frac{P(B \cap B)}{P(B)} = 1$$

L'ultima proprietà è coerente con l'ipotesi che B è occorso e quindi è un evento certo.

8.5.3 Come determinare la probabilità condizionata

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità **uniforme** e sia $B \subset \Omega$ tale che $P(B) > 0$. Condizionare ad un evento B significa costruire un nuovo spazio di probabilità ristretto all'evento condizionante B : $(\Omega_B, \mathcal{F}_B, P_B)$

Spazio dei campioni: $\Omega \rightarrow \Omega_B = B$

Misura di probabilità: $P \rightarrow P_B : P_B(B) = 1$

Si utilizza lo spazio di partenza (Ω, \mathcal{A}, P) su cui definisco la misura condizionata, per ogni $A \subseteq \Omega$:

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

8.6 Probabilità a Priori e a Posteriori (Formula di Bayes)

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità. Si dice **partizione di Ω** un insieme di eventi $\{E_j\}_{1 \leq j \leq n}$ tali che:

1. sono disgiunti:

$$i, j \in I \quad \text{tale che} \quad i \neq j, \quad E_i \cap E_j = \emptyset$$

2. per ogni $i \in I$:

$$P(E_i) > 0$$

3. ricoprono tutto lo spazio:

$$\bigcup_{j=1}^n E_j = \Omega$$

8.6.1 Teorema delle probabilità totali

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità e sia $\{E_j\}_{1 \leq j \leq n}$ una partizione di Ω . Sia $B \in \mathcal{A}$ un evento. Allora:

$$P(B) = \sum_{j=1}^n P(B|E_j)P(E_j)$$

8.6.2 Dimostrazione

$$B = B \cap \Omega = B \cap \bigcup_{j=1}^n E_j = \bigcup_{j=1}^n (B \cap E_j)$$

$$P(B) = P\left(\bigcup_{j=1}^n (B \cap E_j)\right) = \sum_{j=1}^n P(B \cap E_j) = \sum_{j=1}^n P(B|E_j)P(E_j)$$

8.6.3 Teorema di Bayes

Definizione 8.7

Sia $\{E_i\}_{i \in I}$ una partizione di Ω (finita o numerabile) e sia $P(A) > 0$. Allora per un generico elemento della partizione E_n , con $n \in I$, si ha:

$$P(E_n|A) = \frac{P(A|E_n)P(E_n)}{\sum_{i \in I} P(A|E_i)P(E_i)}$$

8.6.4 Dimostrazione

$$P(E_n|A) = \frac{P(A \cap E_n)}{P(A)} = \frac{P(A|E_n)P(E_n)}{P(A)} = \frac{P(A|E_n)P(E_n)}{\sum_{i \in I} P(A|E_i)P(E_i)}$$

8.7 Probabilità a priori e a posteriori

8.7.1 Probabilità a priori

$P(A)$: è una probabilità da determinare senza altre informazioni. Non si hanno informazioni sufficienti per determinarla.

8.7.2 Probabilità a posteriori

$P(A|E_n)$: si hanno strumenti ed informazioni sufficienti per determinare queste probabilità. Se si conoscono delle altre informazioni, quindi se si restringe lo spazio campionario, si riescono a determinare.

8.8 Esercizi

8.8.1 Esercizio 1

Una popolazione si compone per un 40% di fumatori (F) e per il restante 60% di non fumatori (N). Si sa che il 25% dei fumatori e il 7% dei non fumatori ha una malattia respiratoria cronica (M).

1. Calcolare la probabilità che un individuo scelto a caso sia affetto dalla malattia respiratoria
2. Se l'individuo scelto è affetto dalla malattia, calcolare la probabilità che sia un fumatore.

$$\Omega = \{F, F^c\} \quad \text{partizione}$$

$$M = \text{"Malattia respiratoria"}$$

$$P(M|F) = 0.25 \quad P(M|F^c) = 0.07$$

$$P(M) = P(M|F)P(F) + P(M|F^c)P(F^c) = 0.25 \cdot 0.4 + 0.07 \cdot 0.6 = 0.14$$

$$P(F|M) \stackrel{\text{Bayes}}{=} \frac{P(M|F)P(F)}{P(M)} = \frac{0.25 \cdot 0.4}{0.14} = 0.70$$

9 Variabili aleatorie discrete

Sia (Ω, \mathcal{F}, P) uno spazio di probabilità che modella un certo esperimento. Una **variabile aleatoria** è una funzione che associa un numero ad ogni esito possibile dell'esperimento. Formalmente:

$$X : \Omega \rightarrow \mathbb{R}$$

Sia \mathcal{B} l'insieme degli eventi su \mathbb{R} (intervalli, unione ed intersezione di intervalli), per ogni $A \subset \mathbb{R}$:

$$X^{-1}(A) \in \mathcal{F}$$

è l'inversa di X su A .

Le variabili aleatorie possono essere:

- **Discrete**: se assumono un numero finito o numerabile di valori. Ad esempio:

$$X \rightsquigarrow \{0, 1, 2, 3, 4, 5, 6\}$$

- **Continue**: se assumono valori in un intervallo reale. Ad esempio:

$$X \rightsquigarrow \mathbb{R}$$

9.1 Distribuzione di una variabile aleatoria

Sia (Ω, \mathcal{F}, P) uno spazio di probabilità che modella un certo esperimento e sia $X : \Omega \rightarrow \mathbb{R}$ una variabile aleatoria. Sia \mathcal{B} l'insieme degli eventi su \mathbb{R} (intervalli, unione ed intersezione di intervalli).

La probabilità di un evento nella funzione X prendendo in considerazione il lancio di un dado si indica come:

$$P(1) = P(X = 1) = P(\{\omega \in \Omega : X(\omega) = 1\}) = \frac{1}{6}$$

L'insieme $\{P(1), P(2), \dots, P(6)\}$ rappresenta le densità di probabilità, chiamato anche **distribuzione di X**

9.2 Variabili aleatorie discrete

Sia (Ω, \mathcal{B}, P) uno spazio di probabilità, si definisce **variabile aleatoria discreta** :

$$X : \Omega \rightarrow R_X = \{x_1, x_2, \dots, x_n\}$$

con $\text{card}(R_X)$ al più numerabile.

Definizione 9.1

Distribuzione o legge di X

$$P_X(\{x_i\}) = P(X^{-1}(\{x_i\})) = P(X = x_i) \quad \text{per ogni } x_i, i = 1, \dots, n$$

*La legge di una variabile aleatoria discreta è dunque caratterizzata da una **funzione di probabilità***

Esempio 9.1

Prendiamo in considerazione il lancio di 2 dadi: lancio i 2 dadi e ne sommo i valori.

$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$$

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

$$P(\omega) = \frac{1}{36}$$

$$X : \Omega \rightarrow \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\} \subset \mathbb{N}$$

$$(x, y) \rightarrow x + y$$

La legge di X è:

x_i	$X^{-1}(x_i)$	$P_X(x_i)$
2	$\{(1, 1)\}$	$1/36$
3	$\{(1, 2), (2, 1)\}$	$2/36$
4	$\{(1, 3), (2, 2), (3, 1)\}$	$3/36$
5	$\{(1, 4), (2, 3), (3, 2), (4, 1)\}$	$4/36$
6	$\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$	$5/36$
7	$\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$	$6/36$
8	$\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$	$5/36$
9	$\{(3, 6), (4, 5), (5, 4), (6, 3)\}$	$4/36$
10	$\{(4, 6), (5, 5), (6, 4)\}$	$3/36$
11	$\{(5, 6), (6, 5)\}$	$2/36$
12	$\{(6, 6)\}$	$1/36$

La rappresentazione grafica della distribuzione di X è:

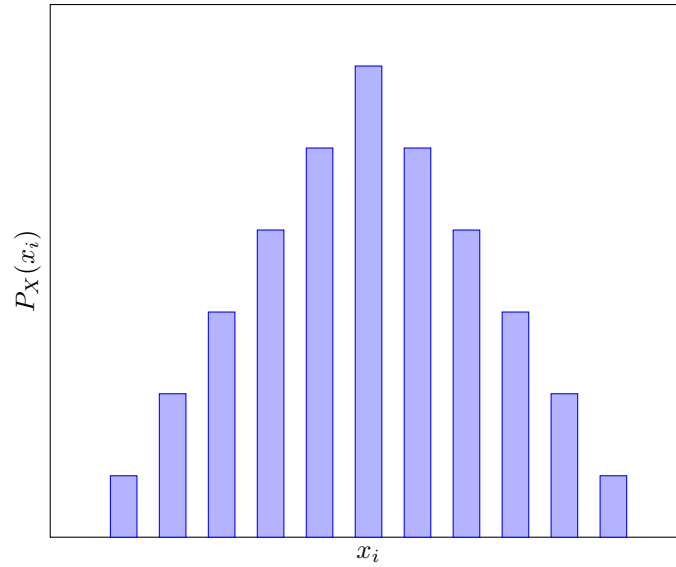


Figura 1: Distribuzione di X

cioè una distribuzione triangolare.

Definizione 9.2

Il **Valore atteso** $\mathbb{E}[X]$ è definito come:

$$\mathbb{E}[X] = \sum_{i=1}^n x_i P_X(x_i)$$

Si tratta di una media pesata dei valori assunti dalla variabile aleatoria con i pesi dati dalle probabilità che la variabile assuma quei valori.

Definizione 9.3

La **Varianza** di X è definita come:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{i=1}^n (x_i - \mathbb{E}[X])^2 P_X(x_i)$$

Esempio 9.2

Prendiamo in considerazione il lancio di una moneta:

$$\Omega = \{T, C\} \quad \mathcal{F} = \mathcal{P}(\Omega)$$

Si vuole calcolare sse esce testa, quindi X è la variabile aleatoria che rappresenta il lancio della moneta:

$$X : \Omega \rightarrow \{0, 1\} \quad \text{t.c.} \quad X(T) = 1, \quad X(C) = 0$$

Questa variabile viene chiamata **variabile aleatoria di Bernoulli** ed è una variabile **dicotomica** (solo 2 valori).

- **Moneta equa:**

$$P_1(\{T\}) = P_1(\{C\}) = \frac{1}{2}$$

$$\Downarrow$$

$$P_X(1) = P_1(X = 1) = P_1(\{T\}) = \frac{1}{2}$$

$$P_X(0) = \frac{1}{2}$$

- **Moneta non equa:**

$$P_2(T) = \frac{1}{3} \quad P_2(C) = \frac{2}{3}$$

$$P_X(1) = P_2(X = 1) = P_2(\{T\}) = \frac{1}{3}$$

$$P_X(0) = 1 - P_X(1) = \frac{2}{3}$$

10 Schema di Bernoulli

10.1 Prove dicotomiche ripetute ed indipendenti

Sia (Ω, \mathcal{L}, P) uno spazio di probabilità che modella un certo esperimento. Considero $A \in \mathcal{L}$ un successo e supponendo che

$$P(A) = p \in (0, 1)$$

ripeto un esperimento in modo indipendente

10.2 Conteggio del successo in ciascuna prova

Definizione 10.1 (MOLTO IMPORTANTE)

$X \sim \mathcal{B}(p)$ X è una variabile di Bernoulli

$p \equiv$ probabilità di successo

$$X : \Omega \rightarrow \{0, 1\} \subset \mathbb{N}$$

$$\text{successo} \rightarrow 1$$

$$\text{insuccesso} \rightarrow 0$$

La legge di X è:

$$p_x(1) = P(X = 1) = P(\text{"successo"}) = p$$

$$p_x(0) = P(X = 0) = P(\text{"insuccesso"}) = 1 - p_x(1) = 1 - p$$

Il valore atteso è:

$$\mathbb{E}[X] = p$$

cioè il successo.

La varianza si calcola come:

$$\text{Var}(X) = p(1 - p)$$

10.3 Conteggio del successo in ciascuna delle n volte

Si ripete l'esperimento dicotomico n volte in condizioni di indipendenza (lancio la moneta n volte). Chiamo X_1, X_2, \dots, X_n i risultati degli n lanci.

$$X_i \sim \mathcal{B}(p) \quad \text{per ogni } i = 1, \dots, n$$

Si è interessati all'evento:

$$A = \text{"numero di successi ottenuti negli } n \text{ lanci"} = \text{"numero di teste"}$$

La variabile che conta il numero di successi è:

$$X = X_1 + X_2 + \dots + X_n \equiv \text{"conta il numero di successi"}$$

$$X \sim \mathcal{B}(n, p) \quad \text{variabile binomiale}$$

La variabile binomiale conta il numero di successi con:

$$n \equiv \text{numero di prove}$$

$$p \equiv \text{probabilità di successo in ogni prova}$$

I valori che potrà assumere la variabile binomiale sono:

$$X \rightsquigarrow \{0, 1, 2, \dots, n\}$$

10.3.1 Modello per il conteggio del successo

Esito prova i-esima:

$$X_1 \sim \mathcal{B}(p) \quad P_X(x_i) = P_i^x (1-p)^{1-x_i}$$

perchè:

$$P_x(0) = P^0(1-p)^1 = (1-p) \quad \text{"insuccesso"}$$

$$P_x(1) = P^1(1-p)^0 = p \quad \text{"successo"}$$

Esito delle n prove: Vettore aleatorio X_1, X_2, \dots, X_n che assume valori:

$$(x_1, x_2, \dots, x_n) \quad x_j \in \{0, 1\}$$

Prove indipendenti (Distribuzione della variabile binomiale):

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &= P(X_1 = x_1)P(X_2 = x_2) \dots P(X_n = x_n) \\ &= P^{x_1}(1-p)^{1-x_1} P^{x_2}(1-p)^{1-x_2} \dots P^{x_n}(1-p)^{1-x_n} \\ &= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

dove $\sum_{i=1}^n x_i = k$ è il numero di successi ottenuti nei n lanci.

$$\begin{aligned} P_X(X = n) \binom{n}{k} &= \frac{n!}{k!(n-k)!} \quad \text{coefficiente binomiale} \\ &\quad \binom{n}{k} \cdot p^k (1-p)^{n-k} \end{aligned}$$

è la probabilità di ottenere k successi in n prove.

10.4 Processi stocastici (prove indipendenti ripetute infinite volte)

Il **processo** consiste nel ripetere infinite volte un esperimento dicotomico, infatti si dice **processo di Bernoulli** la variabile $\{X_n\}_{n \in \mathbb{N}}$ che fa riferimento ad una famiglia di variabili aleatorie indipendenti e identicamente distribuite:

$$X_i \sim B(p)$$

Il **processo binomiale** $\{Z_n\}_{n \in \mathbb{N}}$ fa riferimento ad una famiglia di variabili aleatorie ad incrementi indipendenti che modella il numero di successi in n prove all'aumentare del numero:

$$Z_0 = 0 \quad Z_n = \sum_{i=1}^n X_i \sim B(n, p)$$

10.5 La variabile geometrica

Definizione 10.2

*È il numero della prova in cui si ha il **primo successo***

$$T \sim G(p)$$

Il range, sapendo che a priori non si sa quando accade il primo successo, è:

$$\mathcal{R}_T = 1, 2, 3, \dots$$

Legge:

$$p_T(k) = p(1-p)^{k-1} \quad \text{per ogni } k = 1, 2, 3, \dots$$

Eventi equivalenti:

$$\{T = k\} = \{Z_{k-1}\} \cap \{X_k = 1\}$$

Se il primo successo si ha nella prova k -esima, allora fino alla prova $k-1$ non si hanno successi $\{Z_{k-1} = 0\}$ e nella prova k -esima si ha un successo $\{X_k = 1\}$.

Il valore atteso è:

$$\mathbb{E}[T] = \frac{1}{p}$$

La varianza si calcola come:

$$\text{var}(T) = \frac{1-p}{p^2}$$

Più la variabile è vicina a 1, più è probabile che il successo avvenga al primo tentativo.

10.6 Eventi rari: conteggio dei successi per un numero infinito di esperimenti

Supponiamo che il numero dei successi n cresca all'infinito e che la probabilità p_n di successo vada a zero, ma in modo tale che:

$$np_n \rightarrow \lambda$$

Sia dunque $Z_n \sim B(n, p_n)$

Definizione 10.3 (Poisson)

Conta il numero di eventi che avvengono in un intervallo finito di tempo $[0, t]$ **Poisson:**

$$Y \sim P(\lambda)$$

Range:

$$\mathcal{R} = \mathbb{N}$$

Legge:

$$p_y(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Definizione 10.4

Binomiale converge alla Poisson

$$\lim_{n \rightarrow \infty, p_n \rightarrow 0} B(n, p_n) = P(\lambda)$$

11 La variabile di Poisson

La variabile di Poisson di parametro $\lambda > 0$ modella il conteggio di successo in tempi continui.

Definizione 11.1

Variabile aleatoria di Poisson:

$$Y \sim P(\lambda)$$

Range:

$$\mathcal{R} = \mathbb{N}$$

Legge: per ogni $k = 0, 1, \dots$

$$p_y(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Il valore atteso è:

$$\mathbb{E}[Y] = \lambda$$

La varianza è:

$$\text{var}(Y) = \lambda$$

Poisson approssima la binomiale quando n è grande e p è piccolo.

11.1 Eventi rari: il processo di Poisson

Il processo di Poisson modella il conteggio dei successi in intervalli di tempi sotto l'ipotesi che la probabilità di un successo in uno specifico intervallo di tempo piccolo è molto piccola:

$$np_n = \lambda t$$

Il **processo di Poisson** $\{Y_t\}_{t \in \mathbb{R}_+}$ fa riferimento ad una famiglia di variabili aleatorie ad incrementi indipendenti che modella il numero di successi in un intervallo $[0, t]$ tale che $Y_0 = 0$ e

$$Y_t \sim P(\lambda t), \quad \forall t \geq 0$$

λ rappresenta un numero medio di successi atteso per unità di tempo:

$$\mathbb{E}[Y_t] = \lambda t \quad \Rightarrow \quad \frac{\mathbb{E}[Y_t]}{t} = \lambda$$

11.1.1 Conteggio del tempo del primo successo

Osservando un esperimento modellato da un processo di Poisson $\{Y_t\}_t$ di parametro λ . Si vuole modellare il tempo in cui si ha il primo successo:

Tempo del primo successo: T con $\mathcal{R}_T = \mathbb{R}_+$

Legge di T :

$$F_T(t) = 1 - e^{-\lambda t}, \quad f_T(t) = \lambda e^{-\lambda t}, \quad t \in \mathbb{R}_+$$

Esponenziale di parametro λ $T \sim \exp(\lambda)$

12 Variabili aleatorie continue

Sia (Ω, \mathcal{F}, P) uno spazio di probabilità, la variabile aleatoria continua viene indicata come:

$$X : \Omega \rightarrow R_X \subset \mathbb{R}$$

La legge di X (o distribuzione):

$$P_X(A) = P(X^{-1}(A)) = P(X \in A) \quad \text{per ogni } A \subset \mathcal{B}$$

dove \mathcal{B} è l'insieme degli eventi su \mathbb{R} (intervalli, unione ed intersezione di intervalli).

La legge di una variabile aleatoria continua è caratterizzata da una **densità di probabilità**:

$$f_X : \mathbb{R} \rightarrow \mathbb{R}$$

$$P_X(A) = \int_A f_X(x) dx$$

Proprietà di f_X :

$$0 \leq f_X(x) \leq 1 \quad x \in \mathbb{R}$$

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

13 Statistica inferenziale

La probabilità e la statistica sono due approcci complementari allo stesso problema che descrive fenomeni non deterministici. La statistica inferenziale introduce un modello probabilistico per descrivere la popolazione tramite una variabile aleatoria:

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \mu, \sigma^2 \text{ incogniti}$$

Gli individui sono le "realizzazioni" della X .

Esempio 13.1

$X = \text{"altezze degli italiani maschi di 40 anni"}$

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

Prendo $X_1 \sim X \rightsquigarrow \text{Mario}, X_1(10) = x_1$ (altezza di mario)

Campione: $\underline{X} = (X_1, X_2, \dots, X_n)$ $X_i \sim X \sim \text{Popolazione}$

Le X_i sono indipendenti

Dati: $X_1(\omega) = x_1, X_2(\omega) = x_2, \dots, X_n(\omega) = x_n$

Dati: $\underline{x} = (x_1, x_2, \dots, x_n)$

una possibile realizzazione del campione aleatorio (X_1, X_2, \dots, X_n)

Si parte da una teoria, cioè un modello con **parametri liberi** e si aggiungono dei dati sperimentali. Ci si chiede se si può fare delle inferenze su parametri.

- **Design:** stabilire gli esperimenti
- **Analisi:** analizzare i dati

13.1 Distribuzione esponenziale

È una variabile aleatoria continua:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$P(X \in [a, b]) = \int_a^b f_X(x) dx$$

$$f_X(x) = P(x \leq x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$P(X \leq x) = \int_0^x f_X(t) dt$$

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \mathbb{E}[X] = \int_0^\infty x f_X(x) dx$$

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

La successiva è una **funzione di sopravvivenza**, cioè fornisce la probabilità che un certo evento non si verifichi entro un certo tempo x :

$$S(x) = 1 - f(x) = e^{-\lambda x}$$

La seguente è la "perdita di memoria" della variabile aleatoria esponenziale:

$$P(X > x + x' | X > x) = P(X > x')$$

Esempio 13.2

Il tempo (in ore) necessario per riparare un macchinario è una variabile aleatoria esponenziale di parametro $\lambda = 1$

$X =$ Tempo necessario per la riparazione $X \sim \mathcal{E}(\lambda)$

$$f_X(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$\mathbb{E}[X] = \frac{1}{\lambda} = 1$$

$$\text{Var}(X) = \frac{1}{\lambda^2} = 1$$

1. Qual'è la probabilità che la riparazione superi le 2 ore?

$$P(X > 2) = \int_2^{+\infty} e^{-x} dx = e^{-2} \approx 0.135$$

2. Qual'è la probabilità che la riparazione richieda almeno 3 ore, sapendo che ne richiede più di 2?

$$P(X \geq 3 | X > 2) \underset{\text{assenza di memoria}}{=} P(X \geq 1)$$

14 Tipi di stima

Esistono 2 tipi di stima:

- **Stima puntuale**
- **Stima intervallare**

14.1 Stima intervallare

Si ottiene un intervallo di valori plausibili per il parametro θ e a ciascuno di questi intervalli si associa un livello di confidenza che θ vi appartenga.

14.2 Stima puntuale

Si ottiene un singolo valore come stima per il valore di θ

14.2.1 Definizione

Definizione 14.1

Un **campione** (o campione casuale/aleatorio) di ampiezza n è una collezione di n variabili aleatorie X_1, \dots, X_n **indipendenti** e tutte con la stessa distribuzione $f(x; \theta)$, dove θ è un parametro incognito.

Una **statistica** è una variabile aleatoria ottenuta come funzione del campione, cioè:

$$T = T(X_1, \dots, X_n)$$

Esempio 14.1

Sia X_1, \dots, X_n un campione casuale. Le variabili aleatorie:

- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ (media campionaria)
- $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{X})^2$ (varianza campionaria)
- $V = \min\{X_1, \dots, X_n\}$ (minimo del campione)
- $W = \max\{X_1, \dots, X_n\}$ (massimo del campione)

sono tutte statistiche.

14.2.2 Stimatori

Uno **stimatore** (o statistica sufficiente) è una qualunque statistica T che non dipende dal parametro θ e viene usata per stimare θ .

Esempio 14.2

La media campionaria \bar{X} viene tipicamente usata come stimatore del valor medio di una popolazione. La varianza campionaria S^2 viene tipicamente usata come stimatore della varianza di una popolazione.

Una **Stima** è il **valore numerico** assunto dallo stimatore sui dati osservati x_1, \dots, x_n , cioè $\hat{\theta} = T(x_1, \dots, x_n)$

T si chiama stimatore quando:

$$\mathbb{P}_\theta(x_1, \dots, x_n \mid T(x_1, \dots, x_n) = \theta)$$

14.2.3 Stimatori di massima verosomiglianza

Sia X_1, \dots, X_n un campione casuale estratto da una popolazione con densità $f(x; \theta)$ (discreta o continua), che dipende dal parametro incognito θ .

Se interpretiamo:

$$f(x_1, \dots, x_n; \theta)$$

come la verosomiglianza (o plausibilità) che si realizzi la n -upla di dati x_1, \dots, x_n quando θ è il vero valore del parametro, sembra ragionevole adottare come stima

di θ il valore che rende massima la la verosomiglianza per i dati osservati. In altri termini la stima di massima verosomiglianza $\bar{\theta}_{mv}$ è definita come il valore di θ che rende massima la funzione $f(x_1, \dots, x_n; \theta)$ quando i dati osservati sono x_1, \dots, x_n

In breve si "sposta" θ in modo che sia sul massimo dei valori possibili. Si salva questo valore di θ e si chiama $\bar{\theta}$:

$$\bar{\theta} = \operatorname{argmax}_{\theta} f(x_1, \dots, x_n; \theta)$$

Esempio 14.3

Consideriamo gli esperimenti X_1, \dots, X_n distribuiti indipendentemente

$$X_i \quad f(x, \theta)$$

$$f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

$$f'(\theta) = \frac{d}{d\theta} f(\hat{x}_1, \dots, \hat{x}_n; \theta) = 0$$

$$f'(\theta) = 0 \quad \hat{\theta} \text{ t.c.}$$

$$\frac{d}{d\theta} f(x_1, \dots, x_n; \theta) = f'(x) \quad (\text{è una statistica})$$

$$\hat{\theta} \text{ t.c. } \frac{d}{d\theta} f(x_1, \dots, x_n; \hat{\theta}) = 0$$

$$\Theta = g(X_1, \dots, X_n) \text{ stimatore}$$

Esercizio 14.1

Stimatore di massima verosomiglianza per il parametro di una Bernoulli:

Sia X_1, \dots, X_n un campione casuale estratto da una popolazione con densità (discreta) di Bernoulli di parametro q , dove $q \in (0, 1)$ è incognito. Dal momento che la densità marginale delle X_i è data da

$$f(x; q) = q^x (1 - q)^{1-x} \quad x \in \{0, 1\}$$

la densità congiunta (e quindi la verosomiglianza) è data da:

$$f(x_1, \dots, x_n; q) = q^{\sum_{i=1}^n x_i} (1 - q)^{n - \sum_{i=1}^n x_i}$$

La verosomiglianza corrispondente risulta quindi pari a:

$$\log(f(x_1, \dots, x_n; q)) = \left(\sum_{i=1}^n x_i \right) \log(q) + \left(n - \sum_{i=1}^n x_i \right) \log(1 - q)$$

Derivando rispetto a q otteniamo:

$$\frac{d}{dq} \log(f(x_1, \dots, x_n; q)) = \frac{\sum_{i=1}^n x_i}{q} - \frac{n - \sum_{i=1}^n x_i}{1 - q}$$

Risolvendo $\frac{d}{dq} \log(f(x_1, \dots, x_n; q)) = 0$ rispetto alla variabile q , si ottiene

$$\hat{q} = \frac{1}{n} \sum_{i=1}^n x_i$$

che è un punto di massimo dal momento che:

$$\frac{d^2}{dq^2} \log(f(x_1, \dots, x_n; q)) \leq 0$$

Pertanto lo stimatore di massima verosomiglianza per il parametro q di una distribuzione di Bernoulli è $Q_{mv} = \bar{X}$ (stimatore di massima verosomiglianza).

Supponiamo di aver osservato il seguente campione di dati:

1 0 0 0 0 1 0 0 1

Allora otteniamo la stima di massima verosomiglianza

$$\hat{q} = \frac{1}{10}(1 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 1) = \frac{3}{10} = 0.3$$

(i dati sono stati generati con $\mathcal{B}(0.25)$) che sarebbe il calore vero di q)

14.2.4 Bontà di uno stimatore

Sia X_1, \dots, X_n un campione casuale estratto da una popolazione nota eccetto che per un parametro ignoto θ .

Come scegliamo uno stimatore $T = T(X_1, \dots, X_n)$ per θ ?

Come ne valutiamo la bontà?

Vogliamo minimizzare la deviazione dal valore reale del parametro e per farlo ci basiamo sui valori di $\mathbb{E}(T)$ e $Var(T)$.

Definizione 14.2 (Bias)

Sia $T = T(X_1, \dots, X_n)$ uno stimatore del parametro θ . Allora $b(T) = \mathbb{E}(T) - \theta$ è detto **bias** (o distorsione) di T come stimatore di θ .

Se esso è nullo, diciamo che T è uno stimatore corretto (o non distorto) di θ .

Esempio 14.4

Sia X_1, \dots, X_n un campione casuale estratto da una popolazione con media

μ . La media campionaria \bar{X} è uno stimatore corretto per il valor medio μ . Infatti, calcoliamo:

$$b(\bar{X}) = \mathbb{E}(\bar{X}) - \mu = \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}(X_i)}_{\mu \text{ per ogni } i} - \mu = \frac{1}{n} \cdot n\mu - \mu = \mu - \mu = 0$$

Definizione 14.3 (Errore quadratico medio)

Sia $T = T(X_1, \dots, X_n)$ uno stimatore del parametro θ . Allora $MSE(T) = \mathbb{E}[(T - \theta)^2]$ è detto errore quadratico medio dello stimatore T .

Possiamo esplicitare:

$$\begin{aligned} MSE(T) &= \mathbb{E}[(T - \theta)^2] = \mathbb{E}(T^2) - 2\theta\mathbb{E}(T) + \theta^2 = \\ &= \underbrace{\mathbb{E}(T^2) - \mathbb{E}(T)^2}_{=Var(T)} + \underbrace{\mathbb{E}(T)^2 - 2\theta\mathbb{E}(T) + \theta^2}_{=(\mathbb{E}(T) - \theta)^2} = \\ &= Var(T) + b(T)^2 \end{aligned}$$

- $Var(T)$ Misura la dispersione/variabilità dello stimatore e quindi la sua **precisione**.
- $b(T)^2$ Misura l'**accuratezza** dello stimatore.

Uno stimatore buono (o **efficiente**) controlla sia la varianza che la distorsione. Se T è corretto, allora $MSE(T) = Var(T)$.

Uno stimatore corretto, ma con varianza molto elevata può avere un errore quadratico medio maggiore di uno stimatore distorto, ma con varianza inferiore

Esempio 14.5

Lo stimatore di massima verosomiglianza S_{mv}^2 per la varianza di una popolazione normale **non** è corretto, mentre la varianza campionaria lo è.

Risulta però: $MSE(S_{mv}^2) \leq MSE(S^2)$.

Lo scopo della statistica è quello di trovare tra gli stimatori corretti (o asintoticamente corretti) quelli che hanno errore quadratico medio minore.