

Probabilità e Statistica

UniVR - Dipartimento di Informatica

Fabio Irimie

2° Semestre 2023/2024

Indice

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

6.2.8	Tipi di eventi	15
6.3	Spazio campionario e insieme degli eventi	15
6.3.1	Esempi	16
7	Probabilità	17
7.1	Probabilità degli esperimenti 1-2	17
7.1.1	Esperimento 1: Lancio di un dado	17
7.1.2	Esperimento 2: Lancio di 2 dadi	17
7.1.3	TODO altri esperimenti	17
7.2	Definizione frequentista di probabilità	17
7.3	Definizione soggettiva di probabilità	18
8	Assiomi di Kolmogorov	18
8.1	Assiomi	18
8.1.1	Caso finito	18
8.1.2	Caso generale	19

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

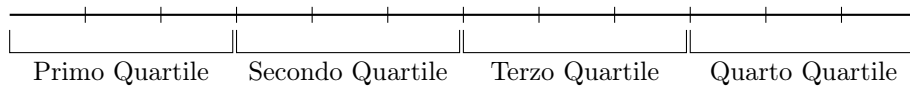
$$\gamma_2 = \frac{1}{n-1} \sum_{j=1}^n \left(\frac{\tilde{x}_j - \bar{x}}{s} \right)^4$$

- $\gamma_2 = 3$: 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.



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

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

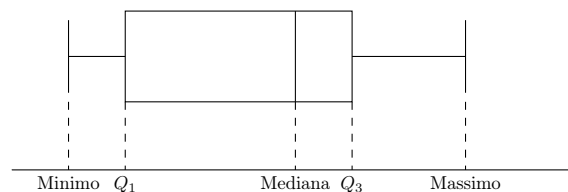
Lo scarto (o distanza interquartile) è la differenza tra il terzo e il primo quartile:

$$IR = Q_3 - Q_1$$

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.

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

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

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

4.2.3 Esempio

Prendiamo in considerazione l'altezza degli studenti

Indice	Valore
min	146
Q_1	163
$Q_2 = M$	168
Q_3	175
max	196
IR	12
$Q_1 - 1.5 \cdot IR$	2.5
$Q_3 + 1.5 \cdot IR$	6.5
$Q_1 - 3 \cdot IR$	1
$Q_3 + 3 \cdot IR$	8

5 Statistica descrittiva bivariata

5.1 Confronto tra due variabili

Prendiamo ad esempio l'età degli uomini e delle donne su una popolazione senza outliers.

Indice	Uomini	Donne
media	21.41	20.83
mediana	21.00	20.70
range	6.20	6.20
scarto quadratico medio	2.56	1.81
scarto quadratico medio	2.56	1.81
deviazione standard	1.60	1.35
asimmetria	0.66	1.09
curtosi	2.62	3.99

Indice	Uomini	Donne
min	19.10	19.10
Q_1	20.30	19.80
$Q_2 = M$	21.00	20.70
Q_3	22.40	21.40
max	25.30	25.30
IR	3.30	2.30
$Q_1 - 1.5 \cdot IR$	15.35	16.35
$Q_3 + 1.5 \cdot IR$	27.35	24.85
$Q_1 - 3 \cdot IR$	10.40	12.90
$Q_3 + 3 \cdot IR$	32.30	28.30

1. C'è **evidenza statistica** che le distribuzioni siano diverse?
2. Che l'età media sia uguale? Che quella delle donne sia minore?

5.2 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)

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

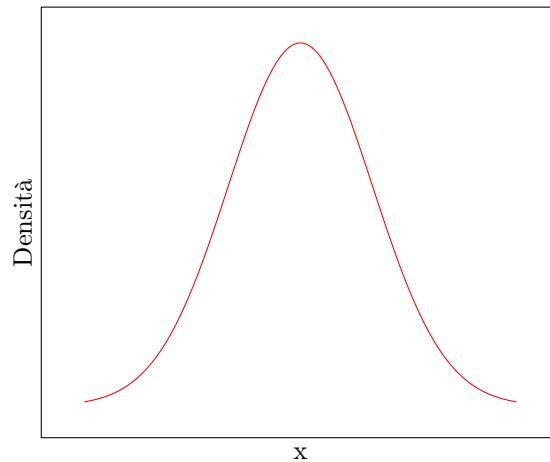
5.3 Regressione

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**:

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

5.3.1 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:

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

5.4 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

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.

6.2.1 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

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

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

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

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

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

6.2.2 Esperimento 2: Lancio di 2 dadi

Prendiamo in considerazione il lancio di 2 dadi:

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

6.2.3 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:

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

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**.

6.2.4 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

6.2.5 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

6.2.6 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

6.2.7 Tipi di esperimenti

Gli esperimenti possono essere di diversi tipi:

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

6.2.8 Tipi di eventi

- **Evento certo:**

È un evento che si verifica sempre, cioè $A = \Omega$, ad esempio il lancio di un dado ha sempre un risultato certo.

6.3 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:

1. $\Omega \in \mathcal{A}$

2. Unione di eventi è un evento

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

3. se $A, B \in \mathcal{A}$, allora $A \setminus B \in \mathcal{A}$

$$\Downarrow$$

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

Definizione 6.3

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

1. $\Omega \in \mathcal{F}$

2. sia $\{A_n\}$ *TODO*

6.3.1 Esempi**Esempio 6.1**

Lancio il dado e controllo che numero esce

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

7 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 7.1

*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:

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

7.1 Probabilità degli esperimenti 1-2

7.1.1 Esperimento 1: Lancio di un dado

$$\Omega_1 = \{1, 2, 3, 4, 5, 6\}$$

$$P(\{i\}) = \frac{\text{casi favorevoli}}{\text{casi possibili}} = \frac{\text{card}(\{i\})}{\text{card}(\Omega)} = \frac{1}{6}, \quad i = 1, \dots, 6$$

7.1.2 Esperimento 2: Lancio di 2 dadi

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

$$P(A) = \frac{\text{casi favorevoli ad } A}{\text{casi possibili}} = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{11}{36}$$

7.1.3 TODO altri esperimenti

7.2 Definizione frequentista di probabilità

Definizioni utili 7.2

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.

7.3 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"

8 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.

8.1 Assiomi

8.1.1 Caso finito

Definizione 8.1

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

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

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

$$A \cap B = \emptyset$$

allora

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

(additività finita)

8.1.2 Caso generale

Definizione 8.2

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

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

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

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

allora

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