Elementi di probabilità e statistica

Luca Barra

Anno accademico 2022/2023

Indice

Capitolo 1

Probabilità

La probabilità è una parte recente della matematica (natta alla fine dell'800) che si basa sull'analisi ed è strettamente legata al gioco d'azzardo. Un evento probabilistico P è casuale e non prevedibile a priori, perchè la complessità delle leggi fisiche che lo regolano rende impraticabile una lettura deterministica.

1.1 Richiami di teoria degli insiemi

Definizione 1.1.1: Insieme

Un insieme è una collezione di oggetti che vengono detti elementi dell'insieme.
Dato S = $\{ x_1, x_2, x_3, ..., x_n\}$
$x \in S$ vuol dire che x appartiene all'insieme S.
$x \notin S$ vuol dire che x non appartiene all'insieme S.

Note:-

S può essere finito (se ha cardinalità non infinita), numerabile (se ha la stessa cardinalità dell'insieme dei numeri naturali) o più che numerabile (se ha la stessa cardinalità dell'insieme dei numeri naturali)

Operazioni:

- complementare: $A^c = x|x \notin A$;
- unione: A U B = $\{x|x \in A \vee x \in B$;
- intersezione: $A \cap B = \{x|x \in A \wedge x \in B\}$.

Definizione 1.1.2: Partizione

Dato un insieme S, una partizione di S è una collezione di insiemi $(A_i)_i^\infty = 1$ se:

- $A_i \cap A_j = \emptyset \ \forall A_i, A_j$;
- $\bigcup_{i=1}^{\infty} A_i = S$.

Leggi di De Morgan:

- $(\bigcup_{i=1}^{\infty} A_i)^c = \bigcap_{i=1}^{\infty} A_i^c$;
- $(\bigcap_{i=1}^{\infty} A_i)^c = \bigcup_{i=1}^{\infty} A_i^c$.

1.2 Modelli probabilistici

I modelli probabilistici sono oggetti matematici che vogliono essere rappresentazioni astratte di un esperimento probabilistico.

Definizione 1.2.1: Modello probabilistico

Un modello probabilistico è una coppia di oggetti che identifichiamo con $(\Omega, \mathbb{P})$.

Note:-

Ω è letto come spazio campionario ed è un insieme che contiene una rappresentazione di tutti i possibili esiti di un esperimento probabilistico. Questa rappresentazione non è univoca

Esempio 1.2.1

Esempio: lancio una volta un dado a 6 facce:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$
$$\Omega' = \{f_1, f_2, f_3, f_4, f_5, f_6\}$$

$\mathbb{P}$è la misura di probabilità ed è una funzione.

Proprietà della misura di probabilità

$\mathbb{P}$è positiva: $\forall A \in P(\Omega) \Rightarrow \mathbb{P}(A) \geqslant 0$
$\mathbb{P}$è finita: $\mathbb{P}(\Omega) = 1$
$\mathbb{P}$è additiva:

$$\begin{cases} A, B \in P(\Omega) con A \cap B = \emptyset \\ \mathbb{P}(AUB) = \mathbb{P}(A) + \mathbb{P}(B) \end{cases} \tag{1.1}$$

Per cui si giunge alla conclusione che Ω è l'evento certo e $\emptyset$ è l'evento impossibile.

Ma se A $\subset \Omega$ (contenuto propriamente) e A $\neq \Omega$, $\mathbb{P}$(A) = 1 è quasi certo. Se B $\subset \Omega$ e B $\neq \emptyset$, $\mathbb{P}$(B) = 0 è quasi impossibile.

Definizione 1.2.2: Legge uniforme discreta

Se lo spazio campionario è finito e l'esperimento probabilistico è equo (non truccato), $\mathbb{P}$(A) = (numero casi favorevoli) / (numero casi totali)

1.2.1 Altre proprietà

Si possono dedurre alcune proprietà. Siano A, B, C eventi (sottoinsiemi di Ω):

- se A $\subseteq$ B $\Rightarrow \mathbb{P}(A) \leqslant \mathbb{P}(B)$ (monotonia di $\mathbb{P}$);
- $\mathbb{P}$(A U B) = $\mathbb{P}$(A) + $\mathbb{P}$(B) - $\mathbb{P}(A \cap B)$;
- $\mathbb{P}$(A U B) $\leqslant$ $\mathbb{P}$(A) + $\mathbb{P}$(B);
- $\mathbb{P}$(A U B U C) = $\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(AUBUC)$.

1.3 Probabilità condizionata

Definizione 1.3.1: Probabilità condizionata

La probabilità condizionata è la probabilità che riguarda eventi che includono informazioni parziali.

$\mathbb{P}(A|B)$ è la probabilità che si verifichi A sapendo che si è verificato B

Note:-

Per ciascun $A \subseteq \Omega$ la $\mathbb{P}(A|B)$ è $\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$. $\mathbb{P}$(B) è detto evento condizionante (deve essere diverso da zero), mentre $\mathbb{P}(A|B)$ è detto evento condizionato

Di seguito alcune formule che possono risultare utili con la probabilità condizionata.

Regola della moltiplicazione:
$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(C|A \cap B) * \mathbb{P}(B|A) * \mathbb{P}(A)$

Formula delle probabilità totali:
$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i) * \mathbb{P}(A_i) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i)$

Formula di Bayes:
$\mathbb{P}(A|B) * \mathbb{P}(B) = \mathbb{P}(B|A) * \mathbb{P}(A)$

1.3.1 Estrazioni senza reimbussolamento

Esempio 1.3.1 (Estrazioni senza reimbussolamento)
Esempio: si estraggono 3 carte (da un mazzo di 52), senza rimetterle nel mazzo. Calcolare la probabilità che nessuna di esse sia cuori. Ci sono due modi equivalenti di procedere.

In blocco: come se venissero estratte tutte e tre insieme.

$$\mathbb{P}(\mathrm{A}) = \frac{\binom{13}{0}\cdot\binom{39}{3}}{\binom{52}{3}}$$

Estrazioni successive: come se venisserò estratte una dopo l'altra.

$$\mathbb{P}(A_1) = \frac{39}{52}, \mathbb{P}(A_2|A_1) = \frac{38}{51}, \mathbb{P}(A_3|A_1 \cap A_2) = \frac{37}{50}$$

Dopo di chè si applica la regola della moltiplicazione.

1.4 Costruzioni di misure di probabilità su prodotti cartesiani

Se in un esperimento aleatorio una o più probabilità di eventi è truccata non si può applicare la legge uniforme discreta.

Esempio 1.4.1 (lancio di due dadi a 4 facce truccati)

- Dado 1: il 4 esce con probabilità $\frac{1}{2}$, gli altri numeri con probabilità $\frac{1}{6}$;
- Dado 2: 1 esce con probabilità $\frac{1}{2}$, gli altri numeri con probabilità $\frac{1}{6}$

Dado 1, $\Omega^1 = \{\ 1, 2, 3, 4\}$
$\mathbb{P}^1(\{1\}) = \mathbb{P}^1(\{2\}) = \mathbb{P}^1(\{3\}) = \frac{1}{6}$
$\mathbb{P}^1(\{4\} = \frac{1}{2}$

Dado 2, $\Omega^2 = \{ 1, 2, 3, 4\}$
$\mathbb{P}^2(\{4\}) = \mathbb{P}^2(\{2\}) = \mathbb{P}^2(\{3\}) = \frac{1}{6}$
$\mathbb{P}^2(\{1\}) = \frac{1}{2}$
Quindi $\Omega = \Omega^1 X \Omega^2$ e $\mathbb{P}(\{i, j\}) = \mathbb{P}^1(\{i\}) X \mathbb{P}^2(\{j\})$

1.5 Indipendenza

Definizione 1.5.1: Indipendenza

Due eventi A e B sono indipendenti se $\mathbb{P}(A|B) = \mathbb{P}(A)$, quindi $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

Corollario 1.5.1

Se A e B sono indipendenti allora anche le seguenti coppie sono indipendenti:

- A e B^c;
- A^c e B;
- A^c e B^c.

Definizione 1.5.2: Indipendenza a due a due

Data la collezione di eventi $(A_i)_{i=1}^n$ diciamo che gli eventi sono a due a due indipendenti se

$$\forall i \neq j, \mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$$

Definizione 1.5.3: Mutua indipendenza

Data la collezione di eventi $(A_i)_{i=1}^n$ diciamo che gli eventi sono mutuamente indipendenti se

$$\forall S \subset \{1, 2, 3, ..., n\}, \mathbb{P}(\bigcap_{i \in S} A_i) = \prod_{i \in S} \mathbb{P}(A_i)$$

Note:-

Due eventi mutuamente indipendenti sono anche a due a due indipendenti, ma due eventi a due a due indipendenti possono non essere mutuamente indipendenti

Capitolo 2

Variabili aleatorie

Una variabile aleatoria è un funzione di Ω in $\mathbb{R}$.

Esempio 2.0.1

Si è interessati a studiare i voti di maturità degli studenti del primo anno del CDL di informatica. Quindi si estrae uno studente (parte casuale) e lo si associa al suo voto di maturità (misurazione).

Le variabili aleatorie possono essere:

- discrete: se l'immagine è numerabile.
- continue: se l'immagine è sottoinsieme di $\mathbb{R}$

2.1 Variabili aleatorie discrete

L'esempio del paragrafo precedente è una variabile aleatoria discreta dato che il voto di maturità è espresso in centesimi, quindi con un intero.

$$X(\omega) = x$$

- X è la funzione che viene applicata.
- ω è il valore di Ω.
- x è il punto di $\mathbb{R}$

$Im(X) = \{x \in \mathbb{R} | \exists \omega \in \Omega \text{ t. c. } X(\omega) = x\}$

2.1.1 Probability mass-function (PMF)

Definizione 2.1.1: PMF

La funzione della massa di probabilità è la funzione così definita:
$\mathbb{P}_x : Im(X) = \mathbb{R}$
$x \longrightarrow \mathbb{P}(x) = \mathbb{P}(X = x) = \mathbb{P}(x \in \mathbb{R} | X(\omega) = x\})$

2.1.2 Variabili aleatorie note

Definizione 2.1.2: Bernoulli

Esperimento probabilistico: con esito binario, descritto da due etichette (S/F, 1/0, V/F, etc.)
Immagine: Im(x) = {0, 1}
PMF: $\mathbb{P}_X(x) = 1 - p$ se x = 0, $\mathbb{P}_X(x) = p$ se x = 1. $p \in \{0, 1\}$
Media:$E(x) = p$
Varianza: $E(x) = p * (1 - p)$

Definizione 2.1.3: Binomiali

Esperimento probabilistico: n prove identiche ripetute.
Immagine: Im(x) = $\{(w_1, w_2, ..., w_n), w_i \in \{s, f\}\}$
PMF: $\mathbb{P}_X(x) = \binom{n}{k} * p^k * (1 - p)^{n-k}$

- k = 0, 1, 2, ..., n (numero di successi).
- p = probabilità di successo.
- n = numero di prove.

Media: $E(x) = n * p$
Varianza = $var(x) = n * p * (1 - p)$

Definizione 2.1.4: Geometriche

Esperimento probabilistico: prove identiche ripetute fino a quando non si ottiene il primo successo.
Immagine: Im(x) = $\{1, 2, 3, ...\}$
PMF: $\mathbb{P}_X(x) = (1 - p)^{x-1} * p$
Media: $E(x) = \frac{1}{p}$
Varianza: $var(x) = \frac{1-p}{p^2}$

Definizione 2.1.5: Iper geometriche

Esperimento probabilistico: estrazioni senza reimbussolamento da una scatola composta da C elementi con una determinata caratteristica e N - C senza quella caratteristica.
Immagine: Im(x) = {0, 1, 2, ..., min{N, C}}
PMF: $\mathbb{P}_X(x) = \frac{\binom{C}{k} * \binom{N-C}{n-k}}{\binom{N}{n}}$

- $\binom{C}{k}$ = numero di possibili estrazioni di k elementi con C;
- $\binom{N-C}{n-k}$ = numero di possibili estrazioni di n-k elementi senza C;
- $\binom{N}{n}$ = numero di possibili estrazioni di n elementi da N.

Definizione 2.1.6: Poisson

Esperimento probabilistico: osservo il verificarsi di une elemento.
Immagine: Im(x) = {0, 1, 2 ,...}
PMF: $\mathbb{P}_X(x) = \frac{\lambda^x}{x!} * e^{-\lambda}$
Media: $E(x) = \lambda$
Varianza: $var(x) = \lambda$

Media e varianza

Definizione 2.1.7: Media

La media delle variabili aleatorie discrete è la quantità:

$$E(x) = \sum_{x \in Im(x)} x * \mathbb{P}_X(x)$$

Corollario 2.1.1

Per la legge dei grandi numeri:

$$lim_{N \to \infty} \frac{k}{N} = \mathbb{P}(A)$$

Definizione 2.1.8: Momento di ordine k

Il momento di ordine k è la seguente quantità:

$$E(x^k) = \sum_{x \in Im(x)} x^k * \mathbb{P}_X(x)$$

Definizione 2.1.9: Varianza

La varianza di x è la seguente quantità:

$$var(x) = E[[x - E(x)]^2]$$

- $[x - E]^2$ = scarto quadratico medio.
- $\sqrt{var(x)}$ = deviazione standard.

Proprietà di media e varianza

- $E(aX + bY) = a * E(X) + b * E(Y)$.
- $var(x) = E(x^2) - E(x)^2$.
- $var(aX + b) = a^2 * var(x)$

2.1.3 PMF congiunte

Le PMF congiunte servono per considerare più variabili aleatorie sullo stesso esperimento.

Note:-
Dalle PMF congiunte si possono dedurre le PMF marginali

2.1.4 Indipendenza

Se le variabili aleatorie x e y sono indipendenti: $E(x * y) = E(x) * E(y)$ e $var(x + y) = var(x) + var(y)$

2.2 Variabili aleatorie continue

L'immagine di x è un sottoinsieme di $\mathbb{R}$ ed è più che numerabile. Esempio: lunghezza, altezza, tempo.

2.2.1 Probability density-function (PDF)

Definizione 2.2.1: PDF

La funzione di densità di probabilità è così definita:
$f_x : \mathbb{R} \rightarrow \mathbb{R}^+[0, +\infty)$
$\mathbb{P}(X \in A) = \int_a^b f_x(t)dt$

2.2.2 Variabili aleatorie note

Definizione 2.2.2: Esponenziale

L'esponenziale è la variante continua della geometrica. Ha media $E(x) = \frac{1}{\lambda}$ e varianza $var = \frac{1}{\lambda^2}$

Definizione 2.2.3: Gaussiana (normale)

PDF: $f(x) = \frac{1}{\sqrt{2 * \pi * \sigma}} * e^{-\frac{(x-\mu)^2}{2*\sigma}}$
Se σ è grande la curva è larga, se è piccolo la curva è stretta.

2.2.3 Media e varianza

Definizione 2.2.4: Media nella V. A. Continue

La media delle variabili aleatorie continue è la quantità:

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

Definizione 2.2.5: Varianza nelle V. A. Continue

La varianza delle variabili aleatorie continue è la quantità:

$$var(x) = \int_{-\infty}^{+\infty} (x - E(x))^2 * f(x)dx$$

2.3 Funzione di distribuzione cumulata (CDF)

Definizione 2.3.1: CDF

La funzione di distribuzione cumulata è così definita:

$$F(x) : \mathbb{R} \rightarrow \mathbb{R}$$

Nel caso discreto:

$$x \rightarrow F_x = \sum_{k \in Im(x)} p_x(k)$$

Nel caso continuo:

$$x \rightarrow F_x = \int_{-\infty}^{x} f_x(t)dt$$

Capitolo 3

Statistica descrittiva

La stastistica descrittiva si occupa di ricavare informazioni qualitative.
In statistica ci si riferisce sempre a una popolazione di interesse:

- con un censimento si raggiungono tutti gli individui;
- con un campionamento si raggiunge una parte della popolazione.

Note:-
Il campionamento deve essere rappresentativo, indipendente e di giusta taglia

3.1 Tipologie di dati

- Fattore: variabile che codifica l'appartenenza a un gruppo;
- Character data: identifica gli elementi, ma non viene usato per raggruppare.
- Quantitativi: possono essere discreti o continui.
- Data frame: matrice di dati.

3.2 Statistica descrittiva per dati quantitativi

3.2.1 Indici di posizione

- Media campionaria: la media aritmetica, è sensibile ai valori estremi.
- Mediana campionaria: lascia il 50% delle sue osservazioni alla sua sinistra e il 50% alla sua destra. Il vettore di elementi che si sta analizzando deve essere ordinato.
- Percentile campionario: è il valore che lascia a sinistra il p% delle osservazioni (è una generalizzazione della mediana).

  Se mediana < media gli estremali sono piccoli.
  Se mediana > media gli estremali sono grandi.
  Se mediana = media si ha simmetria.

3.2.2 Indici di dispersione

- Varianza campionaria: s^2 è la media aritmetica degli scarti al quadrato.
- Deviazione standard campionaria: s è la radice quadrata della varianza campionaria.
- Z-Score: $z = \frac{x_i - x}{s}$. Standardizzazione dei valori, permette di confrontare dati diversi.
- CV (coefficiente di varianza/fano factor): $cv = \frac{s}{x}$.

Se $cv > 1$ si ha alta variabilità.
Se $cv < 1$ si ha bassa variabilità.
Se $cv = 1$ si ha una distribuzione esponenziale.

3.2.3 Indici di forma

- Skewness (indicedi asimmetria): è la media aritmetica degli z-score al cubo. Se i contributi dei vari z-score è prossima a 0 si ha la simmetria. Se il valore è maggiore o minore si ha asimmetria.

3.3 Rappresentazione grafica dei dati

3.3.1 Istogrammi

Solitamente si usano degli istogrammi per rappresentare la distribuzione:

- si divide l'asse delle x in classi;
- si contano le osservazioni in ogni classe;
- si costruisce un rettangolo in ciascuna classe, con altezza proporzionale al conteggio.

3.3.2 Legge del 3∇ per la gaussiana

Note:-
La gaussiana ha code "leggere". La maggior parte dei dati (0.99) si trova tra $\mu - 3\nabla$ e $\mu + 3\nabla$

3.3.3 Box-plot

Il box-plot è una rappresentazione grafica della distribuzione dei dati. La scatola contiene il 50% delle osservazioni, le restanti si trovano nei whiskers (che non sono troppo lontani dal centro), mentre i valori estremali (outliers) sono rappresentati come puntini e non vengono considerati in quanto troppo lontani dal centro.

3.4 Statistica descrittiva (univariata) per dati qualitativi

I dati qualitativi sono di tipo fattore, per cui non ha senso calcolare la media o altri valori numerici. Possono essere rappresentati come grafici a barre, a punti o a torta (pessima rappresentazione).

3.5 Statistica decrittiva (bivariata)

Nella statistica bivariata si considerano contemporaneamente due valori:

- Numero/Fattore;
- Numero/Numero;
- Fattore/Fattore.

3.5.1 Numero/Fattore

Si creano dei sottogruppi e si osserva la differenza nelle quantità.

3.5.2 Numero/Numero

Si creano delle relazioni di tipo funzione tra i valori numerici delle variabili. Generalmente le relazioni sono lineari con coefficiente angolare non nullo. Per fare ciò si usa il coefficiente di correlazione di Pearson:

$$\frac{1}{n-1} * \sum_{i=1}^{n} (\frac{x_i - x}{sx}) * (\frac{y_i - y}{sy})$$

- Se è circa 0: la nuvola di osservazioni è sparpagliata (nessuna correlazione lineare);
- Se è circa 1: c'è relazione lineare forte;
- Se è maggiore di 0: a valori grandi corrispondono valori grandi e viceversa (c'è relazione lineare);
- Se è minore di 0: a valori grandi corrispondono valori piccoli e viceversa (c'è relazione inversa).

Note:-

Tuttavia questo coefficiente soffre per valori grandi di segno opposto, quindi si ricorre al coefficiente di Spearman: le osservazioni non pesano in base alla grandezza, ma in base al segno

3.5.3 Fattore/Fattore

Si usano tabelle a doppia entrata. Se le variabili sono ordinabili si usa il coefficiente di Kendall per capire se gli individui si spostano all'interno dei gruppi secondo una direzione.

Capitolo 4

Statistica inferenziale

La stastistica inferenziale si occupa di trovare e stimare dei parametri. Fare "inferenza" significa estrarre dai dati alcune informazioni sull'intera popolazione di interesse. I dati sono soggetti a variabilità, per cui bisogna quantificare l'incertezza.

Un campione casuale è un vettore di variabili aleatorie indipendenti e identicamente distribuite.

4.1 Statistica inferenziale parametrica

Si suppone di conoscere il tipo di distribuzione delle variabili aleatorie, ma non si conosce il parametro. Per risolvere questo problema si usano degli stimatori, ossia delle funzioni calcolabili sul campione casuale. Possono essere puntuali (su un solo punto) o intervallari (su un intervallo).

4.1.1 Legge dei grandi numeri (per la media campionaria)

Tramite una dimostrazione si può capire che la media di una popolazione risulta uguale alla sua media campionaria. Inoltre, più si aumenta la taglia del campione, più si riducono le oscillazioni. Questo avviene perchè la media campionaria è uno stimatore corretto e consistente.

4.1.2 Legge dei grandi numeri (per la varianza campionaria)

Per la varianza, essendo direttamente correlata alla media, vale ciò che è stato detto nel precedente paragrafo.

4.1.3 TLC (per la media campionaria)

Il teorema centrale del limite spiega che la media campionaria standardizzata ha distribuzione molto vicina a quella di una normale (gaussiana), per campioni sufficientemente grandi (>30). Se i campioni sono piccoli (<30) si deve fare un'assunzione di normalità, altrimenti i dati non avrebbero senso.

4.2 Intervalli di confidenza

Essendo che, in statistica, non si ha una certezza assoluta si deve formare un intervallo [a, b] è specificare quanto si è sicuri del risultato.

Il livello di confidenza è la probabilità che l'intervallo [a, b] contenga il valore esatto del parametro incognito cercato ed è $1 - \alpha$ (tra 0.95 e 0.99). Se si aumenta il numero di osservazioni l'intervallo di confidenza aumenta.

4.2.1 Per medie

Per ottenere l'intervallo di confidenza si ha bisogno di una quantità pivotale:

- una variabile aleatoria;
- una funzione casuale;

- una funzione del parametro incognito;
- deve avere distribuzione nota.

La t di Student è una distribuzione di probabilità continua che funziona da stimatore. Essa è caratterizzata da un parametro chiamato grado di libertà. Se questo parametro è molto grande si può approssimare la t con una normale.

4.2.2 Per proporzioni

Si fa un'ipotesi sulla distribuzione del campione (solitamente una Bernoulli). Se il numero di osservazioni è minore di 30 non si fa nulla. Se è maggiore si calcola la quantità pivotale e si fissa $1 - \alpha = 0.95$.

4.2.3 Per differenze di medie

Ci sono due tipologie:

- si misura un parametro di interesse per individui appartenenti a diverse categorie, si è interessati alle differenze tra categorie. I campioni sono indipendenti dal punto di vista del dataset per cui si calcola l'intervallo di confidenza per differenza di medie (t.test);
- si misura lo stesso individuo prima e dopo per vedere la differenza. I campioni sono appaiati per cui si calcola l'intervallo di confidenza per la media della differenza (t.test(paired = true)).

4.3 Test di ipotesi

Un test di ipotesi è una delle possibili tecniche per controllare la variabilità nel contesto della statistica inferenziale.

Si ha un ipotesi di partenza (H_0) che per tutta la durata del test viene assunta come vera. Se i dati portano ad abbandonare H_0 in favore di un'ipotesi alternativa (H_1), essa diventa la nuova ipotesi di partenza.
Le ipotesi possono essere:

- semplici;
- composte: un sistema di ipotesi;
- bilaterali (two-sided): un unione di ipotesi;
- unilaterali (one-sided);

Procedimento di un test di ipotesi:

1. si parte con H_0 vera;
2. si considera una certa quantità aleatoria di cui si conosce il comportamento;
3. si prendono i dati (x1, x2, ..., xn);
4. si valuta la quantità aleatoris in corrispondenza di questi dati, sotto l'ipotesi H_0 (p-value);
5. se il valore ottenuto è plausibile non si rifiuta H_0, altrimenti si passa a una nuova ipotesi H_1.

α rappresenta la significatività del test e deve essere scelto piccolo (0.05 o 0.01).

- Se p-value $< \alpha$: si rifiuta H_0;
- Se p-value $> \alpha$: non si rifiuta H_0;

Capitolo 5

Appendice - Esempi noti

5.1 Paradosso del compleanno

Il paradosso afferma che la probabilità che almeno due persone in un gruppo compiano gli anni lo stesso giorno è largamente superiore a quanto potrebbe dire l'intuito. Si calcola inizialmente la probabilità che, nel gruppo, non esistano due compleanni uguali.

$$\mathbb{P}(A) = \frac{364!}{365^{n-1} \cdot (365-n)!}$$

Di questa probabilità si deve fare il complementare.

$$\mathbb{P}(A) = 1 - \frac{364!}{365^{n-1} \cdot (365-n)!}$$

5.2 Problema di Monty Hall

Il problema di Monty Hall è un famoso problema di teoria della probabilità, legato al gioco a premi statunitense Let's Make a Deal. Prende il nome da quello del conduttore dello show, Monte Halprin, noto con lo pseudonimo di Monty Hall. Il problema è anche noto come paradosso di Monty Hall, poiché la soluzione può apparire controintuitiva, ma non si tratta di una vera antinomia, in quanto non genera contraddizioni logiche.

Nel gioco vengono mostrate al concorrente tre porte chiuse; dietro ad una si trova un'automobile, mentre ciascuna delle altre due nasconde una capra. Il giocatore può scegliere una delle tre porte, vincendo il premio corrispondente. Dopo che il giocatore ha selezionato una porta, ma non l'ha ancora aperta, il conduttore dello show – che conosce ciò che si trova dietro ogni porta – apre una delle altre due, rivelando una delle due capre, e offre al giocatore la possibilità di cambiare la propria scelta iniziale, passando all'unica porta restante; cambiare la porta migliora le chance del giocatore di vincere l'automobile, portandole da 1/3 a 2/3.

Questo accade perchè inizialmente la probabilità che il concorrente abbia scelto una capra è maggiore rispetto alla probabilità cha abbia scelto l'automobile.

Capitolo 6

Appendice - Formule

6.1 Probabilità

Legge uniforme discreta:
$\mathbb{P}(A) = \frac{\#A}{\#\Omega}$

Probabilità condizionata:
$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Regola della moltiplicazione:
$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(C|A \cap B) * \mathbb{P}(B|A) * \mathbb{P}(A)$

Formula delle probabilità totali:
$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i) * \mathbb{P}(A_i) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i)$

Formula di Bayes:
$\mathbb{P}(A|B) * \mathbb{P}(B) = \mathbb{P}(B|A) * \mathbb{P}(A)$

6.2 Variabili aleatorie

Bernoulli

- Immagine: Im(x) = {0, 1};
- PMF: $\mathbb{P}_X(x) = 1 - p$ se x = 0, $\mathbb{P}_X(x) = p$ se x = 1. $p \in \{0, 1\}$;
- Media: $E(x) = p$;
- Varianza: $E(x) = p * (1 - p)$.

Binomiali

- Immagine: Im(x) = $\{(w_1, w_2, ..., w_n), w_i \in \{s, f\}\}$;
- PMF: $\mathbb{P}_X(x) = \binom{n}{k} * p^k * (1 - p)^{n-k}$;
- Media: $E(x) = n * p$;
- Varianza: $var(x) = n * p * (1 - p)$.

Geometriche

- Immagine: Im(x) = $\{1, 2, 3, ...\}$;
- PMF: $\mathbb{P}_X(x) = (1-p)^{x-1} * p$;
- Media: $E(x) = \frac{1}{p}$;
- Varianza: $var(x) = \frac{1-p}{p^2}$.

Iper geometriche

- Immagine: Im(x) = {0, 1, 2, ..., min{N, C}};
- PMF: $\mathbb{P}_X(x) = \frac{\binom{C}{k} * \binom{N-C}{n-k}}{\binom{N}{n}}$.

Poisson

- Immagine: Im(x) = {0, 1, 2 ,...};
- PMF: $\mathbb{P}_X(x) = \frac{\lambda^x}{x!} * e^{-\lambda}$;
- Media: $E(x) = \lambda$;
- Varianza: $var(x) = \lambda$.

Esponenziale

- Media: $E(x) = \frac{1}{\lambda}$;
- Varianza: $var = \frac{1}{\lambda^2}$.

Gaussiana

- PDF: $f(x) = \frac{1}{\sqrt{2 * \pi * \sigma}} * e^{-\frac{(x-\mu)^2}{2 * \sigma}}$;
- Media: μ;
- Varianza: σ^2.