

## CALCOLO DELLA PROBABILITA'

### Importanza probabilità e definizioni

Occorre studiare la probabilità perché in molti esperimenti non è noto a priori l'esito, ma sono noti i possibili esiti di cui è possibile calcolare la probabilità. Un esperimento di cui a priori non si conosce l'esito è detto esperimento aleatorio. I possibili esiti di un esperimento aleatorio sono detti eventi elementari. L'insieme costituito da tutti gli eventi elementari è detto spazio campionario  $\Omega$ . I sottoinsiemi di  $\Omega$  sono gli eventi. Lo spazio campionario può essere:

- Discreto = quando è finito o infinito numerabile
- Continuo = quando è infinito non numerabile

### Linguaggio insiemistico

Dopo aver definito gli eventi è opportuno comprendere che significato hanno le varie relazioni insiemistiche dal punto di vista del "linguaggio degli eventi":

- Unione = l'esito dell'esperimento è un elemento di A oppure un elemento di B  $A \cup B$
- Intersezione = l'esito dell'esperimento è un elemento sia di A che di B  $A \cap B$
- Differenza = l'esito dell'esperimento è un elemento di A ma non di B  $A \setminus B$
- Eventi incompatibili = due eventi A e B sono incompatibili se sono disgiunti cioè non hanno elementi in comune ovvero la loro intersezione è l'insieme vuoto  $A \cap B = \emptyset$
- Implicazione logica = tutti gli elementi di B sono anche elementi di A  $A \subset B$

### Definizione informale e assiomatica della probabilità

Dopo aver definito un evento e aver studiato le relazioni insiemistiche occorre attribuire ora ad ogni evento la sua probabilità. Nel linguaggio comune la probabilità si esprime con percentuali. Informalmente la probabilità di un evento è un numero reale compreso tra 0 e 1 che misura quanto si ritiene probabile il verificarsi di tale evento.

Per dare una definizione rigorosa della probabilità fu necessario aspettare il 1930 quando Kolmogorov scoprì che le proprietà della probabilità si potevano ricavare da pochi assiomi, ovvero proprietà che si definiscono vere, e pose così le basi per il calcolo delle probabilità moderno.

Definiamo dunque assiomaticamente la probabilità: sia  $\Omega$  uno spazio campionario e  $\mathcal{F}$  una famiglia di suoi eventi, si chiama probabilità su  $\Omega$  una qualsiasi funzione  $\mathbb{P}: \mathcal{F} \rightarrow [0,1]$  con le seguenti proprietà:

- $\mathbb{P}(\Omega) = 1$
- Se  $\{A_n\}_n$  è una successione numerabile di eventi a due a due disgiunti allora:

$$\mathbb{P}\left(\bigcup_{j=1}^{+\infty} A_j\right) = \sum_{j=1}^{+\infty} \mathbb{P}(A_j)$$

Definiamo poi la  $\sigma$ -algebra ovvero una famiglia di sottoinsiemi di  $\Omega$  che ha delle proprietà di chiusura rispetto ad alle operazioni insiemistiche di unione numerabile di passaggio complementare. Assiomaticamente la  $\sigma$ -algebra è una famiglia  $\mathcal{F}$  se sono soddisfatte le seguenti proprietà:

- L'insieme vuoto appartiene ad  $\mathcal{F}$ :  $\emptyset \in \mathcal{F}$
- Se  $A \in \mathcal{F}$  allora anche il suo complementare appartiene ad  $\mathcal{F}$ :  $A^C \in \mathcal{F}$
- Se  $\{A_i\}_{i \in \mathbb{N}}$  è una collezione di insiemi di  $\mathcal{F}$  allora l'unione di tutti gli eventi appartiene ad  $\mathcal{F}$

Da quest'ultima ipotesi ne consegue che la probabilità  $\mathbb{P}$  deve soddisfare la richiesta di additività.

Con il solo utilizzo degli assiomi è possibile dimostrare le seguenti proprietà:

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- Se  $A \supseteq B$  allora  $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$  e in particolare  $\mathbb{P}(A) \geq \mathbb{P}(B)$

Definiamo ora la probabilità discreta: una misura  $\mathbb{P}$  su  $\Omega$  si dice discreta se esiste  $D \in \mathcal{F}$  al più numerabile tale che  $\mathbb{P}(D^c) = 0$ .

### Calcolo della probabilità

Se lo spazio campionario è discreto per conoscere la probabilità di ogni evento basta assegnare la probabilità agli eventi elementari. Infatti, se  $\Omega$  è discreto significa che i suoi elementi si possono elencare e ogni suo sottoinsieme  $A$  è un'unione finita di elementi. Quindi la probabilità di  $A$  è la somma delle probabilità degli elementi che lo compongono. Vediamo ora come assegnare la probabilità agli eventi elementari:

- Approccio classico = se lo spazio campionario è finito e ha  $n$  elementi e tutti i suoi elementi hanno la stessa probabilità, la probabilità si calcola come:

$$\mathbb{P}(A) = \frac{n^\circ \text{eventi favorevoli}}{n^\circ \text{di casi possibili}}$$

Questo approccio può essere usato quando i casi possibili sono un numero finito e quando i casi sono equiprobabili, se vengono meno queste due ipotesi non è più possibile calcolare la probabilità come rapporto tra casi favorevoli e casi possibili.

- Approccio frequentista = si calcola come rapporto tra il numero di volte che si realizza un evento  $k$  e l'insieme dei tentativi  $n$ :

$$\mathbb{P}(A) = \frac{k}{n}$$

Questo approccio è valido se l'esperimento è ripetibile. Se un evento non viene osservato negli esperimenti fatti si attribuisce probabilità zero. Inoltre, bisogna tenere presente che l'idea frequentista non va bene se il futuro è "diverso" dal passato ovvero l'esperimento non si ripete sempre nelle medesime condizioni: in questo caso si parlerà di probabilità condizionata

- Se si vuole calcolare la probabilità di pescare  $m$  elementi con certe caratteristiche, indichiamo con  $k$  il numero degli elementi dell'insieme  $N$  che hanno queste caratteristiche, e  $t$  elementi con altre caratteristiche, si procede nel seguente modo:

$$\mathbb{P}(A) = \frac{\binom{k}{m} \binom{N-k}{t}}{\binom{N}{m+t}}$$

Ricordiamo come calcolare il coefficiente binomiale:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad n! = n(n-1)(n-2) \dots 1$$

### Probabilità condizionata

Dunque, la probabilità è uno strumento che viene usato quando si ha incertezza sull'esito degli esperimenti. Tuttavia, ci sono situazioni in cui abbiamo una conoscenza parziale dell'esito; questa conoscenza potrebbe alterare la probabilità che assegniamo a certi eventi (ad esempio il numero pescato non viene rimesso nel sacchetto). La probabilità che avvenga  $A$  sapendo che  $B$  avviene è calcolata considerando  $B$  come nuovo spazio campionario e si fa riferimento agli eventi elementari che appartengono sia ad  $A$  che a  $B$ . Dunque, dati due eventi  $A$  e  $B$  con  $\mathbb{P}(B) > 0$ , si definisce la probabilità di  $A$  dato  $B$ , o probabilità di  $A$  condizionata al verificarsi di  $B$ , il seguente numero:

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

$\mathbb{P}(A|B)$  risulta comunque essere una probabilità su  $\Omega$  come afferma in seguente teorema: data una probabilità  $\mathbb{P}$  su  $\Omega$  e fissato un evento  $B$  con  $\mathbb{P}(B) > 0$ , allora la funzione  $\mathbb{P}(\cdot|B)$  che all'evento  $A$  associa  $\mathbb{P}(A|B)$  è una probabilità su  $\Omega$ .

La probabilità condizionata gode delle seguenti proprietà:

- $\mathbb{P}(\emptyset|B) = 0$
- Se gli eventi sono a due a due disgiunti si ha:  $\mathbb{P}(A_1, A_2, \dots, A_n|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) + \dots + \mathbb{P}(A_n|B)$
- $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$
- Se gli eventi non disgiunti si ha:  $\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) - \mathbb{P}(A_1 \cap A_2|B)$

Dalla formula della probabilità condizionata ricaviamo anche un altro modo di calcolare la probabilità dell'unione di due eventi:

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

### Indipendenza di eventi

Due eventi  $A$  e  $B$  si dicono indipendenti se vale l'uguaglianza  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . È importante sottolineare che l'indipendenza non deve essere confusa con l'incompatibilità. Se due eventi  $A$  e  $B$  sono incompatibili, sapere che si è verificato  $A$  implica che  $B$  non può essersi verificato dunque sapere che si è verificato  $A$  cambia la probabilità di  $B$ .

### Teorema delle probabilità totali e formula di Bayes

Dato uno spazio campionario  $\Omega$ , si definisce partizione dello spazio campionario una famiglia di eventi  $B_1, B_2, \dots, B_n$  tali che la loro unione dia proprio lo spazio campionario  $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$  e l'intersezione a due a due di questi eventi sia l'insieme vuoto  $B_i \cap B_j = \emptyset \quad \forall i \neq j$ .

Enunciamo ora il teorema delle probabilità totali: dato uno spazio campionario  $\Omega$ , una sua partizione  $B_1, B_2, \dots, B_n$  e un evento  $A$ , vale la seguente formula:

$$\mathbb{P}(A) = \sum_{i=1, \dots, n} \mathbb{P}(A \cap B_i) = \sum_{i=1, \dots, n: \mathbb{P}(B_i) > 0} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Dimostrazione: poiché la famiglia  $B_1, B_2, \dots, B_n$  è una partizione vale che:

$$\sum_{i=1, \dots, n} \mathbb{P}(A \cap B_i) = \sum_{i=1, \dots, n: \mathbb{P}(B_i) > 0} \mathbb{P}(A \cap B_i)$$

Poiché  $\mathbb{P}(B_i) = 0$  implica che  $\mathbb{P}(A \cap B_i) = 0$ , dalla formula per la probabilità condizionata, se  $\mathbb{P}(B_i) > 0$  allora  $\mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i)\mathbb{P}(B_i)$ . □

Nel teorema delle probabilità totali si ha una partizione dello spazio campionario e si conoscono tutte le probabilità condizionate  $\mathbb{P}(A|B_i)$ , talvolta si potrebbe essere interessati invece a una partizione in particolare  $\mathbb{P}(B_k|A)$ . Per poter calcolare questo termine occorre utilizzare la formula di Bayes: sia  $\{B_i\}_{i=1}^n$  una partizione dello spazio campionario  $\Omega$  e sia  $A$  un evento con  $\mathbb{P}(A) > 0$ , allora per ogni  $k \in \{1, 2, \dots, n\}$  vale che:

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1, \dots, n: \mathbb{P}(B_i) > 0} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

Dimostrazione: per dimostrare questo teorema è sufficiente scrivere la definizione di  $\mathbb{P}(B_k|A)$  e applicare il teorema delle probabilità totali per ricavare  $\mathbb{P}(A)$ :

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1, \dots, n: \mathbb{P}(B_i) > 0} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

### Formule per il calcolo della probabilità

$$\mathbb{P}(\Omega) = 1$$

$$\mathbb{P}(\emptyset) = 0$$

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \quad \text{se A e B non sono disgiunti}$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \quad \text{se A e B sono disgiunti}$$

$$\mathbb{P}(A \cap B) = 0 \quad \text{se A e B sono disgiunti}$$

N.B.: per ricavare la probabilità  $\mathbb{P}(A) + \mathbb{P}(B)$  è fondamentale che i due eventi siano disgiunti  
 $A \cap B = \emptyset$

Mentre per ricavare la probabilità  $\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$  occorre che gli eventi siano a due a due disgiunti ovvero  $A \cap B = \emptyset, B \cap C = \emptyset, A \cap C = \emptyset$ , non è sufficiente che  $A \cap B \cap C = \emptyset$

N.B.: ricordarsi che può essere applicato de Morgan e quindi fare il complementare di tutto.

### Formule per il calcolo della probabilità condizionata

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

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

$$\mathbb{P}(\emptyset|B) = 0$$

$$\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$$

$$\mathbb{P}(A_1, A_2, \dots, A_n|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) + \dots + \mathbb{P}(A_n|B) \quad \text{eventi sono a due a due disgiunti}$$

$$\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) - \mathbb{P}(A_1 \cap A_2|B) \quad \text{eventi non disgiunti}$$

### Alte formule per il calcolo della probabilità

$$\mathbb{P}(A) = \sum_{i=1, \dots, n} \mathbb{P}(A \cap B_i) = \sum_{i=1, \dots, n: \mathbb{P}(B_i) > 0} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1, \dots, n: \mathbb{P}(B_i) > 0} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

# VARIABILI ALEATORIE

## Variabili aleatorie

Definiamo variabile aleatoria  $X$  una funzione che ha come dominio lo spazio campionario  $\Omega$  e come codominio  $\mathbb{R}^n$ , in formule:  $X: \Omega \rightarrow \mathbb{R}^n$ . È dunque una funzione che fa corrispondere ad ogni valore nel dominio un valore nel codominio. Le caratteristiche interessanti di una variabile aleatoria sono due:

- L'insieme dei suoi possibili valori = nel nostro caso sono numeri reali e dunque si le variabili aleatorie si chiamano numeriche. Le variabili aleatorie numeriche possono essere suddivise in due tipi:
  1. Discrete = una variabile aleatoria si dice discreta se l'insieme dei valori che  $X$  può assumere è un insieme infinito oppure infinito numerabile
  2. Assolutamente continue = una variabile aleatoria si dice discreta se l'insieme dei valori che  $X$  può assumere è un insieme infinito continuo
- Usare le variabili aleatorie permette di lavorare con uno spazio più semplice, modellizzare esperimenti multipli contemporaneamente e lavorare con le "leggi" e non con lo spazio campionario. Definiamo dunque la legge: data una variabile aleatoria  $X$  la misura di probabilità  $\mathbb{P}_X$  definita di seguito si dice legge:

$$\mathbb{P}_X(A) := \mathbb{P}_X(X^{-1}(A)), \quad A \subseteq \mathbb{R}^n$$

Data una variabile aleatoria  $X$  a valori in  $\mathbb{R}$  la funzione  $F_X: \mathbb{R} \rightarrow [0,1]$  definita di seguito si dice funzione di ripartizione di  $X$ :

$$F_X(t) := \mathbb{P}_X(X^{-1}((-\infty, t])) \equiv \mathbb{P}_X((-\infty, t]), \quad \forall t \in \mathbb{R}$$

La funzione di ripartizione  $F_X$  ha le seguenti proprietà:

- È non decrescente, quindi ha limiti da destra e da sinistra
- $\lim_{s \rightarrow t^+} F_X(s) = F_X(t)$
- $\lim_{t \rightarrow +\infty} F_X(t) = 1$
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$

## VARIABILI ALEATORIE DISCRETE

### Densità di una variabile aleatoria discreta

Definiamo ora la densità  $f_X$  di una variabile discreta come una funzione che associa ad ogni valore di  $x$  la probabilità che  $X$  assuma quel valore. Se  $x$  non è uno dei valori possibili allora la densità è nulla  $f_X = 0$ . In generale dunque ciò che interessa di una variabile aleatoria è l'insieme dei valori possibili che può assumere e la densità  $f_X$ . L'insieme dei valori possibili può essere:

- Finito
- Numerabile = in questo caso viene indicato come successione

La densità di una variabile discreta ha le seguenti proprietà:

- Assume solo valori positivi o uguali a 0
- La somma dei suoi valori è 1

### Valore atteso di una variabile aleatoria discreta

Sia  $X: \Omega \rightarrow \mathbb{R}^n$  una variabile aleatoria discreta avente come immagine in  $\mathbb{R}^n$  l'insieme  $V$  (ovvero l'insieme dei valori che  $X$  può assumere). Se la serie  $\sum_{v \in V} |v| f_X(v)$  converge allora si chiama valore atteso di  $X$  il numero reale:

$$\mathbb{E}(X) = \sum_{v \in V} v f_X(v)$$

La  $\mathbb{E}$  sta per expectation, ovvero attesa. Il valore atteso nel caso in cui  $V$  sia finito si calcola sommando su tutti i valori possibili il prodotto tra il valore  $x_i$  e la sua probabilità, è dunque una media pesata dei valori dove il peso di un valore è la sua probabilità:

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

Vediamo ora una serie di proprietà (dedotti da alcuni teoremi) del valore atteso:

- Se  $a, b$  sono numeri reali:  $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
- Se  $X, Y$  sono due variabili aleatorie:  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

### Varianza di una variabile aleatoria

Data una variabile aleatoria  $X$  a valori in  $\mathbb{R}$  la sua varianza è il numero reale:

$$\text{Var}(X) = \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)^2\right)$$

La radice della varianza è detta deviazione standard o anche scarto quadratico medio. Per variabili discrete il calcolo si riduce a:

$$\text{Var}(X) = \sum_{x \in V} x^2 p_X(x) - \mathbb{E}(X)^2$$

Vediamo ora una serie di proprietà della varianza:

1.  $\text{Var}(X) \geq 0$
2.  $\text{Var}(X) = 0$  se e solo se  $X$  è una variabile aleatoria costante ovvero può assumere un solo valore
3.  $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$
4. Se  $a, b$  sono numeri reali:  $\text{Var}(aX + b) = a^2 \text{Var}(X)$
5. Se  $X, Y$  sono due variabili aleatorie:  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

Dove la covarianza di  $X$  e  $Y$  è:

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

La differenza tra la varianza e il valore atteso delle variabile aleatoria è che la varianza è un indice di dispersione mentre il valore atteso è un indice di posizione.

Dimostrazione prop. 1: è conseguenza del fatto che la variabile aleatoria può solo essere maggiore o uguale a 0

Dimostrazione prop. 3:

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)^2\right) = \\ &= \mathbb{E}\left(X^2 + (\mathbb{E}(X))^2 - 2X\mathbb{E}(X)\right) = \\ &= \mathbb{E}(X^2) + \mathbb{E}\left((\mathbb{E}(X))^2\right) - 2\mathbb{E}(X\mathbb{E}(X)) = \\ &= \mathbb{E}(X^2) + (\mathbb{E}(X))^2 - 2\mathbb{E}(X)\mathbb{E}(X) = \\ &= \mathbb{E}(X^2) + (\mathbb{E}(X))^2 - 2(\mathbb{E}(X))^2 = \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \end{aligned}$$

Dimostrazione prop. 4:

$$\begin{aligned} \text{Var}(aX + b) &= \mathbb{E}\left((aX + b)^2\right) - (\mathbb{E}(aX + b))^2 = \\ &= \mathbb{E}(a^2X^2 + b^2 + 2abX) - (a\mathbb{E}(X) + b)^2 = \\ &= \mathbb{E}(a^2X^2) + \mathbb{E}(b^2) + \mathbb{E}(2abX) - a^2(\mathbb{E}(X))^2 - b^2 - 2ab\mathbb{E}(X) = \\ &= a^2\mathbb{E}(X^2) + b^2 + 2ab\mathbb{E}(X) - a^2(\mathbb{E}(X))^2 - b^2 - 2ab\mathbb{E}(X) = \\ &= a^2\mathbb{E}(X^2) - a^2(\mathbb{E}(X))^2 = \end{aligned}$$

$$= a^2 (\mathbb{E}(X^2) - (\mathbb{E}(X))^2) =$$

$$= a^2 \text{Var}(X)$$

Dimostrazione prop. 4:

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}((X + Y)^2) - (\mathbb{E}(X + Y))^2 = \\ &= \mathbb{E}(X^2 + Y^2 + 2XY) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 = \\ &= \mathbb{E}(X^2) + \mathbb{E}(Y^2) + \mathbb{E}(2XY) - (\mathbb{E}(X))^2 - (\mathbb{E}(Y))^2 - 2\mathbb{E}(X)\mathbb{E}(Y) = \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \end{aligned}$$

### Indipendenza variabili aleatorie discrete

Date  $n$  variabili aleatorie  $X_1, X_2, \dots, X_n$  essere si dicono indipendenti se l'uguaglianza:

$$\mathbb{P}(X_1 = v_1, X_2 = v_2, \dots, X_n = v_n) = \mathbb{P}(X_1 = v_1) \cdot \mathbb{P}(X_2 = v_2) \cdot \dots \cdot \mathbb{P}(X_n = v_n)$$

Per ogni scelta di  $v_1, v_2, \dots, v_n$ . La prima parte dell'uguaglianza significa la "probabilità che contemporaneamente  $X_1$  assume il valore  $v_1$ ,  $X_2$  assume il valore  $v_2$  e così via. Se  $X$  e  $Y$  sono variabili aleatorie indipendenti allora:

- Il valore atteso del prodotto è il prodotto dei valori attesi:  $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$
- La varianza della somma è la somma delle varianze:  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Formule valore atteso e varianza per variabili aleatorie discrete:

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

$$\text{Var}(X) \geq 0$$

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

se variabili aleatorie sono indipendenti

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

se variabili aleatorie sono indipendenti

## MODELLI DISCRETI DI VARIABILI ALEATORIE

### Bernoulli

Nel caso in cui gli esperimenti aleatori possono avere solo due esiti (esempio lancio della moneta oppure presenza o meno di una proteina oppure uomo sano o non sano) si utilizza la variabile aleatoria di Bernoulli. Si definisce dunque esperimento di Bernoulli un esperimento aleatorio che può avere solo due esiti. Una variabile aleatoria di Bernoulli  $\mathcal{B}(p)$  di parametro  $p$  è una variabile che può assumere solo i valori 0 e 1, e per cui la probabilità di assumere il valore 1 è pari a  $p$ , e quella di assumere il valore 0 è pari a  $1 - p$ . Il valore 0 è anche detto insuccesso mentre il valore 1 è detto successo. Ricordandoci che di una variabile discreta interessano i valori che può assumere e la funzione di densità definiamo l'insieme dei valori possibili è  $V = \{0, 1\}$ ;  $f_x(0) = 1 - p$  e  $f_x(1) = p$ . Studiamo ora alcune proprietà delle variabili aleatorie applicata alla bernoulliana:

- Valore atteso della bernoulliana:  $\mathbb{E}(X) = p$
- Varianza della bernoulliana:  $\text{Var}(X) = p(1 - p)$

Il massimo della varianza di una variabile di Bernoulli è  $\frac{1}{4}$  e lo si ottiene in corrispondenza a  $p = 1/2$ .

Nel caso in cui si ripetessero più esperimenti di Bernoulli si parlerebbe di processo di Bernoulli. Diamo ora una definizione formale di processo di Bernoulli: una sequenza finita o infinita di prove di Bernoulli, tutte con lo stesso parametro  $p$  e fra loro indipendenti, è detta processo di Bernoulli di parametro  $p$ . Dato un processo di Bernoulli si potrebbe voler calcolare:

- Se si fanno  $n$  prove, qual è la probabilità che si abbiano  $k$  successi? Per rispondere a questa domanda si definisce la variabile binomiale  $\mathcal{B}(n, p)$  ovvero una variabile che conta il numero di successi in  $n$  prove di un processo di Bernoulli di parametro  $p$ . La probabilità con cui si abbiano  $k$  successi in  $n$  prove si calcola come:

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

Studiamo ora alcune proprietà delle binomiale:

- Valore atteso della binomiale:  $\mathbb{E}(X) = np$
- Varianza della binomiale:  $Var(X) = np(1-p)$

Il massimo della varianza di una variabile binomiale è  $n/4$  e lo si ottiene in corrispondenza a  $p = 1/2$ .

- Quando si ha per la prima volta il successo? Per rispondere a questa domanda si definisce la variabile geometrica  $\mathcal{G}(p)$  ovvero una variabile aleatoria che dice in quale prova di un processo di Bernoulli di parametro  $p$  si ha il primo successo. Per la geometrica si ha che:

$$f_x(k) = p(1-p)^{k-1}$$

$$\mathbb{P}(X \leq k) = 1 - (1-p)^k$$

Studiamo ora alcune proprietà delle binomiale:

- Valore atteso della geometrica:  $\mathbb{E}(X) = \frac{1}{p}$
- Varianza della geometrica:  $Var(X) = \frac{1-p}{p^2}$

La varianza di una variabile geometrica è decrescente rispetto a  $p \in [0,1]$

## Poisson

Parliamo ora di un'altra variabile aleatoria discreta ovvero la variabile di Poisson che si costruisce come limite in legge di una successione di variabili binomiali. Una variabile aleatoria  $X$  si dice di Poisson di parametro  $\lambda > 0$ ,  $X \sim \mathcal{P}(\lambda)$  se:

$$P(X = i) = \begin{cases} \frac{\lambda^i}{i!} e^{-\lambda} & i \in \mathbb{N} \\ 0 & \text{altrimenti} \end{cases}$$

Questo tipo di variabile aleatoria viene utilizzata per determinare ad esempio il numero di automobili che passa per un determinato incrocio in un determinato intervallo di tempo oppure il numero di persone che si reca in un negozio in un giorno feriali. Studiamo ora alcune proprietà della variabile di Poisson:

- Valore atteso:  $\mathbb{E}(X) = \lambda$
- Varianza:  $Var(X) = \lambda$

Un teorema afferma che se  $X_1, X_2, \dots, X_n$  sono variabili indipendenti con legge di Poisson allora vale che:

$$X_1 + X_2 + \dots + X_n = \mathcal{P}(\lambda_1 + \lambda_2 + \dots + \lambda_n)$$

Definiamo ora processo di Poisson: sia  $\Sigma$  uno spazio con misura  $\mu$ , una famiglia di variabili aleatorie  $\{X_\Delta\}_{\Delta \in \Sigma}$  si dice processo di Poisson di intensità  $\mu$  se e solo se valgono le seguenti proprietà:

- Se  $\{\Delta_i\}_{i=1}^n$  sono disgiunti allora  $X_{\Delta_1}, \dots, X_{\Delta_n}$  sono indipendenti
- $X_\Delta \sim \mathcal{P}(\mu(\Delta))$
- $\Delta \rightarrow X_\Delta$  è una misura di conteggio



Formule di Bernoulli  $\mathcal{B}(p)$ :

$$f_x(0) = 1 - p$$

$$f_x(1) = p$$

$$\mathbb{E}(X) = p$$

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

Formule binomiale  $\mathcal{B}(n, p)$ :

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

$$\mathbb{E}(X) = np$$

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

Formule geometrica  $\mathcal{G}(p)$ :

$$f_x(k) = p(1 - p)^{k-1}$$

$$\mathbb{P}(X \leq k) = 1 - (1 - p)^k$$

$$\mathbb{E}(X) = \frac{1}{p}$$

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

Formule Poisson  $\mathcal{P}(\lambda)$ :

$$P(X = i) = \begin{cases} \frac{\lambda^i}{i!} e^{-\lambda} & i \in \mathbb{N} \\ 0 & \text{altrimenti} \end{cases}$$

$$\mathbb{E}(X) = \lambda$$

$$\text{Var}(X) = \lambda$$

## VARIABILI ALEATORIE CONTINUE

### Densità variabili aleatorie continue

Studiamo adesso le variabili aleatorie continue. Una variabile aleatoria si dice discreta se l'insieme dei valori che  $X$  può assumere è un insieme infinito continuo. Sia  $X: \Omega \rightarrow \mathbb{R}$  una variabile aleatoria continua, se esiste una funzione integrabile  $f_x: \mathbb{R} \rightarrow [0, +\infty)$  tale che per ogni intervallo  $J \subseteq \mathbb{R}$ :

$$\mathbb{P}(X \in J) = \int_J f_x(x) dx$$

allora  $f_x$  è la funzione di densità continua della variabile aleatoria  $X$  e la  $X$  è detta variabile assolutamente continua. Data una variabile aleatoria, se per ogni intervallo reale  $I$  sappiamo calcolare  $\mathbb{P}(X \in I)$  diciamo che conosciamo la legge o distribuzione di  $X$ . Se le variabili  $X_1, X_2, \dots, X_n$  hanno la stessa legge si dice che sono identicamente distribuito. Studiamo ora alcune proprietà della variabili continue:

- Valore atteso:  $\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot f_x(x) dx$
- Varianza:  $\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$

## Indipendenza variabili aleatorie discrete

Date  $n$  variabili aleatorie continue  $X_1, X_2, \dots, X_n$  essere si dicono indipendenti se l'uguaglianza:

$$\mathbb{P}(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = \mathbb{P}(X_1 \in I_1) \cdot \mathbb{P}(X_2 \in I_2) \cdot \dots \cdot \mathbb{P}(X_n \in I_n)$$

per ogni scelta di  $I_1, I_2, \dots, I_n$  intervalli reali. Se  $X$  e  $Y$  sono variabili aleatorie continue indipendenti allora:

- Il valore atteso del prodotto è il prodotto dei valori attesi:  $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$
- La varianza della somma è la somma delle varianze:  $Var(X + Y) = Var(X) + Var(Y)$

Formule valore atteso e varianza per variabili aleatorie discrete:

$$\mathbb{P}(X \in J) = \int_J f_x(X) dx$$

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot f_x(x) dx$$

$$Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

## MODELLI CONTINUE DI VARIABILI ALEATORIE

### Uniforme

Una variabile aleatoria uniforme sull'intervallo  $(a, b)$  è la variabile continua che assume valori fra  $a$  e  $b$  e ha densità:

$$f_x(X) = \frac{1}{b-a} \quad \text{se } x \in (a, b)$$

$$f_x(X) = 0 \quad \text{se } x \notin (a, b)$$

Studiamo ora alcune proprietà della uniforme:

- Valore atteso della uniforme:  $\mathbb{E}(X) = \frac{a+b}{2}$
- Varianza della uniforme:  $Var(X) = \frac{(b-a)^2}{12}$

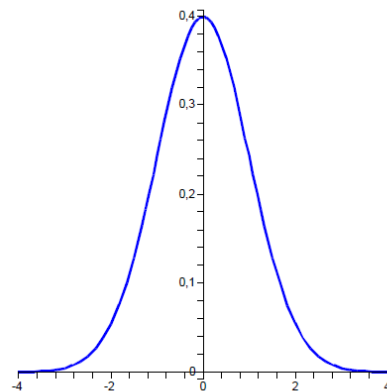
### Normale

Se si hanno misure effettuate da errore o grandezze misurate sugli individui di una popolazione come altezza o peso si utilizza la variabile normale. Una variabile normale standard è la variabile continua che assume valori in tutto  $\mathbb{R}$  e ha densità come nel grafico ovvero:

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

Non importa conoscere con esattezza il valore della densità, ciò che importa è invece conoscere il grafico e le sue proprietà:

- Il massimo è in  $x = 0$  ed è uguale a  $\frac{1}{\sqrt{2\pi}}$
- È simmetrica rispetto all'asse delle  $y$
- Va a zero a  $\pm\infty$
- $\mathbb{P}(|X| > 3) = 0,0027$



- $\mathbb{P}(|X| > 4) = 0,000063$
- $\mathbb{P}(-3 < x < 3) = 0,9973$

Studiamo ora alcune proprietà della normale:

- Valore atteso della normale:  $\mathbb{E}(X) = 0$
- Varianza della normale:  $Var(X) = 1$

Per determinare il valore della probabilità occorre usare le tavole. La riga ci permette di trovare il numero fino al primo decimale mentre la colonna ci permette di trovare il secondo decimale. Le probabilità non presenti in tabella si ottengono ragionando sul grafico come valori complementari.

Parliamo ora della variabile normale generiche: sia  $Z \sim \mathcal{N}(0,1)$  e siamo  $\mu, \sigma \in \mathbb{R}$ . La variabile aleatoria  $X = \sigma Z + \mu$  si indica con  $X \sim \mathcal{N}(\mu, \sigma^2)$  ed è detta variabile aleatoria normale di parametri  $\mu$  e  $\sigma^2$ . Studiamo ora alcune proprietà della normale generica:

- Valore atteso della normale generica:  $\mathbb{E}(X) = \mu$
- Varianza della normale generica:  $Var(X) = \sigma^2$

Le proprietà di un grafico riferito a una normale generica sono:

- Il massimo è in  $x = \mu$
- Altezza del massimo può anche superare 1 È simmetrica rispetto all'asse delle y
- Ha orma a campana
- L'area sotto la curva vale 1
- Se la varianza è minore di 1 la campana è stretta e alta
- Se la varianza è maggiore di 1 la campana è larga e bassa
- $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) = 0,9973$

Le tavole nella normale generalizzata sono riferite solo al valore Z dunque occorre effettuare l'operazione di standardizzazione ovvero sottrarre ad X il valore atteso e dividere per la radice della varianza:

$$Z = \frac{X - \mu}{\sigma}$$

Se  $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$  e  $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$  vale che:

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Se  $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$  e  $a, b \in \mathbb{R}$  vale che:

$$aX + b \sim \mathcal{N}(a\mu_1 + b, a^2\sigma_1^2)$$

$$\mathbb{P}(\mathcal{N}(0, 1) \leq 1.53) = 0.93699.$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

Formule dell'uniforme:

$$\begin{cases} f_x(X) = \frac{1}{b-a} & \text{se } x \in (a, b) \\ f_x(X) = 0 & \text{se } x \notin (a, b) \end{cases}$$

$$\mathbb{E}(X) = \frac{a+b}{2}$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

Formule della normale standard:

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

- Il massimo è in  $x = 0$  ed è uguale a  $\frac{1}{\sqrt{2\pi}}$
- È simmetrica rispetto all'asse delle  $y$
- Va a zero a  $\pm\infty$
- $\mathbb{P}(|X| > 3) = 0,0027$
- $\mathbb{P}(|X| > 4) = 0,000063$
- $\mathbb{P}(-3 < x < 3) = 0,9973$

$$\mathbb{E}(X) = 0$$

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

Formule della normale  $X \sim \mathcal{N}(\mu, \sigma^2)$ :

$$X = \sigma Z + \mu$$

$$\mathbb{E}(X) = \mu$$

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

- Il massimo è in  $x = \mu$
- Altezza del massimo può anche superare 1 È simmetrica rispetto all'asse delle  $y$
- Ha orma a campana
- L'area sotto la curva vale 1
- Se la varianza è minore di 1 la campana è stretta e alta
- Se la varianza è maggiore di 1 la campana è larga e bassa
- $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) = 0,9973$

$$Z = \frac{X - \mu}{\sigma}$$

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$aX + b \sim \mathcal{N}(a\mu_1 + b, a^2\sigma_1^2)$$

## TEOREMA CENTRALE DEL LIMITE

### Media campionaria

Date  $n$  variabili aleatorie  $X_1, X_2, \dots, X_n$  indipendenti e identicamente distribuiti, la variabile aleatoria è detta media campionaria:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Definita la media campionaria potremmo voler rispondere alla seguente domanda: posso dire che legge ha? In generale non è semplice conoscere la legge di  $\bar{X}_n$ , ma se  $n$  è grande e ci si accontenta di un'approssimazione, il teorema centrale del limite da una risposta.

### Teorema centrale del limite

Sia  $\{X_i\}_{i \geq 1}$  una successione di variabili aleatorie identicamente distribuite, tutte con valore atteso  $\mathbb{E}(X_i) = \mu$  e varianza finita  $Var(X_i) = \sigma^2 > 0$ . Per ogni  $t \in \mathbb{R}$  vale che:

$$\lim_{n \rightarrow \infty} \mathbb{P}(S_n^* \leq t) = \phi(t) \quad \text{con} \quad S_n^* = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

$S_n^*$  è la standardizzazione della media campionaria. In altre parole, se si vuole calcolare le probabilità relative a  $S_n^*$  si può utilizzare quelle relative alla normale  $\mathcal{N}(0,1)$  come approssimazione. Dunque:

$$S_n^* \approx \mathcal{N}(0,1)$$

$$\bar{X}_n \approx \mathcal{N}(\mu, \sigma^2)$$

$$X_1 + X_2 + \dots + X_n \approx \mathcal{N}(\mu, \sigma^2)$$

L'aspetto positivo è che questa approssimazione vale per qualsiasi legge, quello negativo è che in quanto approssimazione vale solo se  $n$  è sufficientemente grande. L'approssimazione diviene uguaglianza nel caso di variabili normali chiaramente:

$$S_n^* = \mathcal{N}(0,1)$$

$$\bar{X}_n = \mathcal{N}(\mu, \sigma^2)$$

$$X_1 + X_2 + \dots + X_n = \mathcal{N}(\mu, \sigma^2)$$

Il teorema del limite può essere applicato anche a Bernoulli. In questo caso ha validità quando  $np \geq 5$  e  $(1-p) \geq 5$ .

### Disuguaglianza di Chebychev

Sia  $X$  una variabile aleatoria con  $\mathbb{E}(X) = \mu$  e  $Var(X) = \sigma^2$ . Sia  $\varepsilon$  un numero reale positivo prefissato. Vale la seguente disuguaglianza:

$$\mathbb{P}(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

O equivalentemente:

$$\mathbb{P}(|X - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

Se  $\varepsilon = \delta\sigma$  allora si ha che per la prima equazione:

$$\mathbb{P}(|X - \mu| \geq \delta\sigma) \leq \frac{1}{\delta^2}$$

Sciogliendo il modulo:

$$|X - \mu| \geq \delta\sigma \rightarrow X \geq \mu + \delta\sigma \quad X \leq \mu - \delta\sigma$$

Quindi si ottiene:

$$\mathbb{P}(X \geq \mu + \delta\sigma) + \mathbb{P}(X \leq \mu - \delta\sigma) \leq \frac{1}{\delta^2}$$

Se  $\varepsilon = \delta\sigma$  allora si ha che per la seconda equazione:

$$\mathbb{P}(|X - \mu| < \delta\sigma) \geq 1 - \frac{1}{\delta^2}$$

Sciogliendo il modulo:

$$|X - \mu| \leq \delta\sigma \rightarrow \mu - \delta\sigma < X < \mu + \delta\sigma$$

Quindi si ottiene:

$$\mathbb{P}(\mu - \delta\sigma < X < \mu + \delta\sigma) \leq 1 - \frac{1}{\delta^2}$$

Questa disuguaglianza non permette di prevedere esattamente il valore di  $X$  perché è sempre una variabile aleatoria ma mi permette di definire un intervallo di valori in cui è molto probabile che si trovi il valore che la variabile aleatoria  $X$  assumerà. Sottolineiamo che l'ampiezza dell'intervallo è stabilita da  $\sigma^2$ .

### La legge dei grandi numeri

Siano  $X_1, X_2, \dots, X_n$  variabili indipendenti e identicamente distribuite. Sia  $\mathbb{E}(X_i) = \mu$  e  $Var(X_i) = \sigma^2$  per ogni  $i$ . Allora per  $\varepsilon > 0$  vale:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0$$

La legge forte dei grandi numeri afferma invece che: siano  $X_1, X_2, \dots, X_n$  variabili indipendenti e identicamente distribuite. Sia  $\mathbb{E}(X_i) = \mu$  per ogni  $i$ . Allora per  $\varepsilon > 0$  vale:

$$\mathbb{P}\left(\left|\lim_{n \rightarrow +\infty} \bar{X}_n = \mu\right|\right) = 1$$

## AFFIDABILITA'

### Affidabilità

Supponiamo di avere un sistema costituito da vari sottosistemi. Si consideri una caratteristica Z che ciascun sottosistema può avere o non avere. La probabilità che il sistema abbia la caratteristica Z si chiama affidabilità del sistema. Ipotizziamo dunque di avere n sottosistemi di affidabilità  $a_1, a_2, \dots, a_n$  e supponiamo che gli eventi "l'elemento i-esimo ha la caratteristica Z" siano indipendenti. L'affidabilità si calcola in modo differente in funzione del tipo di relazione tra i sottosistemi:

- Sottosistemi in serie = quando il sistema S ha la caratteristica Z e se e solo se tutti i sottosistemi hanno la caratteristica Z. L'affidabilità  $a_s$  del sistema in serie si calcola come produttoria delle affidabilità dei sottosistemi:

$$a_s = \prod_{i=1}^n a_i$$

- Sottosistemi in parallelo = quando il sistema P ha la caratteristica Z se e solo se almeno uno dei sottosistemi ha la caratteristica Z. L'affidabilità  $a_p$  del sistema in serie si calcola come:

$$a_p = 1 - \prod_{i=1}^n (1 - a_i)$$

Iterando opportunamente queste due formule si possono costruire in maniera gerarchica le affidabilità di sistemi complessi.

# STATISTICA DESCRITTIVA

## Singole variabili e classi

La statistica descrittiva considera un insieme di dati e li elabora. I dati raccolti rappresentano la realizzazione di variabili aleatorie. Si distinguono in variabili continue e discrete. Spesso le singole variabili vengono raggruppate in classi ovvero intervalli contigui. Una volta raggruppati gli  $n$  dati in classi si definiscono le frequenze. Esistono diversi tipi di frequenze:

- Frequenza assoluta di una classe  $f_a$  = è il numero di osservazioni che ricadono in quella classe
- Frequenza relativa di una classe  $f_r$  = è la sua frequenza assoluta divisa per il numero totale di osservazioni
- Frequenza percentuale di una classe  $f_p$  = è la sua frequenza relativa moltiplicata per 100
- Frequenza cumulativa di una classe  $F_a, F_r, F_p$  = è la somma delle frequenze della classe stessa e di tutte quelle che la precedono

Se i valori diversi osservati in un esperimento non sono troppo numerosi si può scegliere tutte le classi come singoli valori.

Spesso per rappresentare le frequenze si utilizzano gli istogrammi. La scelta delle classi è arbitraria; talvolta può essere conveniente trattare classi di ampiezza diversa. In tal caso solitamente è l'area del rettangolo ad essere proporzionale alla frequenza.

## Indici di posizione

Definiamo ora alcuni indici di posizione:

- Media = si definisce media il numero:

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

La media di un insieme di dati si calcola utilizzando lo stimatore media campionaria; è quindi la stima puntuale del valore atteso della variabile aleatoria che modella la misura del dato in questione.

- Mediana = disponendo i dati in ordine crescente la mediana è il dato nella posizione centrale se  $n$  è dispari, oppure la media aritmetica dei due dati in posizione centrale se  $n$  è pari.
- Quantili, percentili e quartili = generalizzando il concetto di mediana cercando un valore  $q_p$  con la proprietà che almeno una frazione  $p$  dei dati sia non superiore a  $q_p$  ed almeno una frazione  $1 - p$  sia non inferiore a  $q_p$ . Si definisce dunque  $p$ -esimo quantile:
  - Se  $np$  non è intero allora  $q_p = x_{k+1}$
  - Se  $np = k$  con  $k$  intero allora  $q_p = \frac{x_k + x_{k+1}}{2}$

Il  $p$ -esimo quantile viene anche detto 100  $p$ -esimo percentile. Il 25°, 50°, 75° percentile vengono detti anche primo, secondo e terzo quartile e indicati con  $Q_1, Q_2, Q_3$ . Il secondo quartile coincide con la mediana

- Moda = è il valore o più in generale la classe in corrispondenza del quale si ha la popolazione più numerosa. Se vi è un solo punto dove la frequenza è massima, si dice che la distribuzione delle frequenze è unimodale, se vi è più di un massimo si dice che la distribuzione delle frequenze è plurimodale

## Indici di dispersione

Definiamo ora una serie di indici di dispersione:

- Range = se i dati sono  $x_1, x_2, \dots, x_n$  il range è il numero reale:

$$r = \max\{x_i: i = 1, \dots\} - \min\{x_i: i = 1, \dots\}$$

- Varianza = si definisce varianza:



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

Una definizione alternativa di varianza per un insieme di dati è la seguente:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2$$

## Indici di forma

- La **skewness**

$$\gamma_3 = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{\sigma} \right)^3$$

È una grandezza *adimensionale*. Può assumere valori sia positivi che negativi.

Se è negativa denota una *coda* verso sinistra.

Se è positiva denota una *coda* verso destra.

se la distribuzione è simmetrica, allora la skewness è nulla, ma l'inverso non è vero.

Per trasformazioni lineari  $y_i = ax_i + b$  la skewness non cambia:  $\gamma_3^y = \gamma_3^x$ .

- La **curtosi**

$$\gamma_4 = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{\sigma} \right)^4$$

È una grandezza adimensionale e non negativa. Misura (in un certo senso) l'appiattimento della distribuzione delle frequenze, poiché assegna un peso elevato agli scarti grandi: valori elevati della curtosi segnalano distribuzioni significativamente diverse da  $\bar{x}$  per grandi scarti, piccoli valori distribuzioni *appuntite* in corrispondenza di  $\bar{x}$ .

Per trasformazioni lineari  $y_i = ax_i + b$  la curtosi non cambia:  $\gamma_4^y = \gamma_4^x$ .

# STATISTICA INFERENZIALE

## Introduzione

Finora abbiamo visto che esistono modelli probabilistici che possiamo utilizzare per prevedere gli esiti di esperimenti aleatori. Naturalmente la previsione è di tipo probabilistico: diciamo infatti con quanta probabilità si verificano certi esiti.

Con la statistica inferenziale si utilizzano i dati di più esperimenti aleatori per dire qualcosa sul modello probabilistico che si suppone incognito. In tutti i problemi di stima dei parametri si seguono questi passi:

1. Scegliere un modello con un parametro incogniti
2. Fare  $n$  esperimenti
3. Raccogliere i dati e calcolare una funzione di questi dati
4. Usare questa funzione per stimare il parametro incognito

## Modello statistico parametrico

Diamo ora una serie di definizioni utili allo studio della statistica inferenziale. Definiamo un modello statistico in due diversi modi:

- Utilizzando la densità = un modello statistico parametrico è una famiglia di leggi di variabili aleatorie dipendenti da uno o più parametri  $\vartheta$ : lo indichiamo tramite la funzione di densità  $\{f(x; \vartheta): \vartheta \in \Theta\}$  dove  $\Theta$  è lo spazio dei parametri. In altre parole, il modello statistico indica la scelta di un tipo di variabile aleatoria.
- In generale = un modello statistico parametrico è una famiglia di leggi di variabili aleatorie dipendenti da uno o più parametri  $\vartheta$ : lo indichiamo tramite la legge  $\{\mathcal{L}_\vartheta: \vartheta \in \Theta\}$  dove  $\Theta$  è lo spazio dei parametri. In altre parole, la legge  $\mathcal{L}_\vartheta$  simboleggia la scelta di un tipo di legge (la bernoulliana, la binomiale, la poissoniana, normale etc)

## Stimatore

Definiamo inoltre il campione casuale di dimensione  $n$  una  $n$ -upla di variabili aleatorie  $X_1, \dots, X_n$ . Dato un campione casuale  $X_1, \dots, X_n$  si dice statistica di variabile aleatorie  $T$  funzione del campione casuale che non sia funzione di alcun parametro incognito. Assegnata la statistica  $T$  una volta estratto un particolare campione il numero  $\tau = t(x_1, \dots, x_n)$  si dice stima di  $g(\vartheta)$ . Uno stimatore si dice non distorto o corretto se  $\mathbb{E}_\vartheta(T) = g(\vartheta)$  ovvero quando il valore atteso della statistica è uguale alla sua stima. In generale se ciò non accade si definisce distorsione la differenza tra il valore atteso della statistica e la sua stima:  $\mathbb{E}_\vartheta(T) - g(\vartheta)$ . Spesso lo stimatore scelto dipende esplicitamente dal numero  $n$  ovvero dalla numerosità del campione. Una successione di stimatori si dice consistente se il limite per  $n$  che tende a infinito del valore atteso del quadrato della distorsione è nullo ovvero:

$$\lim_{n \rightarrow \infty} \mathbb{E}_\vartheta \left( (T_n - g(\vartheta))^2 \right) = 0$$

Il valore atteso del quadrato della distorsione  $\mathbb{E}_\vartheta \left( (T_n - g(\vartheta))^2 \right)$  è detto rischio quadratico medio. Per uno stimatore consistente vale che:

$$\mathbb{E}_\vartheta \left( (T_n - g(\vartheta))^2 \right) = \text{Var}(T_n) + \left( \mathbb{E}_\vartheta(T_n - g(\vartheta)) \right)^2$$

Se lo stimatore oltre a essere consistente è anche non distorto allora vale che:

$$\mathbb{E}_\vartheta \left( (T_n - g(\vartheta))^2 \right) = \text{Var}(T_n)$$