

Principi di Inferenza

Variabili aleatorie

Laura D'Angelo
Università Milano Bicocca

1

Variabili aleatorie

Consideriamo un esperimento aleatorio, e indichiamo con Ω lo spazio degli esiti.

Definizione

Una **variabile aleatoria** X è una funzione

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

che assegna un valore a ogni esito in Ω .

Esempio: lanciamo due dati $\implies \Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$

definiamo $X =$ somma dei due dadi.

Per esempio, se si realizza l'esito $(1, 1)$, allora $X = 2$.

Se si realizza uno tra gli esiti $\{(3, 6), (4, 5), (5, 4), (6, 3)\}$, allora $X = 9$.

Possiamo calcolare la probabilità che $X = 2$, che corrisponde alla probabilità dell'esito $(1, 1)$:

$$P(X = 2) = P(\text{l'esito è } (1, 1)) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

Mentre la probabilità che $X = 9$ è data da

$$P(X = 9) = P(\text{l'esito è } (3, 6) \text{ o } (4, 5) \text{ o } (5, 4) \text{ o } (6, 3)) = \frac{4}{36}$$

Le variabili aleatorie possono essere di due tipi:

- ▶ **discrete**: se i valori che la variabile può assumere sono numerabili/elencabili (e.g., $0, 1, 2, \dots$)
Per esempio, X nell'esempio della somma di due dati è discreta.
- ▶ **continue**: se la variabile può assumere un insieme continuo di valori, ad esempio tutti i valori in un intervallo (e.g., i numeri reali $\mathbb{R}$, o $(0, +\infty)$, o un intervallo (a, b)).

1.1

Variabili aleatorie discrete *e la distribuzione di probabilità*

Sia X una **variabile aleatoria discreta**.

In questo caso, i valori che X può assumere possono essere elencati: chiamiamo $x_1, x_2, \dots, x_n$ i valori distinti che può assumere.

Chiamiamo **supporto di X** , indicato con $\mathcal{X}$, l'insieme di questi valori.

- Nota: Il numero n di valori che può assumere X può anche **essere infinito**, per esempio, nel caso di conteggi (non c'è un estremo superiore, posso sempre “aggiungere 1 unità”).

Esempi di v.a. discrete:

- Nell'esempio precedente, X = somma di due dadi, $\mathcal{X} = \{2, 3, 4, \dots, 12\}$;
- Lancio un dado 5 volte e definisco X = numero di volte che esce 6 in 5 lanci, $\mathcal{X} = \{0, 1, 2, 3, 4, 5\}$;
- X = numero di clienti che entra in un negozio in un'ora, $\mathcal{X} = \{0, 1, 2, 3, \dots\}$ (nota: qui non c'è un estremo superiore!).

La distribuzione di probabilità (v.a. discreta)

Abbiamo visto che possiamo calcolare:

$$\left. \begin{array}{l} P(X = x_1) \\ P(X = x_2) \\ \vdots \\ P(X = x_n) \end{array} \right\} \quad \begin{array}{l} \text{chiamiamo questa funzione la} \\ \textbf{distribuzione di probabilità} \\ \text{della variabile aleatoria } X \end{array}$$

Definizione

Data una variabile aleatoria discreta X a valori in $\mathcal{X} = (x_1, \dots, x_n)$, la **distribuzione di probabilità** è la funzione

$$P(X = x_i) \quad \text{per ogni } x_i \in \mathcal{X}$$

► Per la distribuzione di probabilità vale

$$\sum_{x_i \in \mathcal{X}} P(X = x_i) = P(X = x_1) + P(X = x_2) + \dots + P(X = x_n) = 1$$

1.2

Variabili aleatorie continue

e la funzione di densità di probabilità

Esempio: consideriamo il **tempo di attesa** X di una persona alla fermata dell'autobus. Supponiamo, per semplicità, che l'autobus possa arrivare in un qualunque momento nei prossimi 10 minuti.

Quindi X può assumere valori come

$$0, \quad 1, \quad 1.5, \quad 2.37, \quad 5.821, \quad \dots$$

Se X può assumere qualunque valore in un intervallo, **ci sono infiniti valori possibili**. A differenza del caso dei dadi o del numero di clienti, **non possiamo elencare i possibili valori di X** .

Per esempio, se X è il tempo di attesa:

$$X = 2, \quad X = 2.1, \quad X = 2.01, \quad X = 2.001, \dots$$

Il **supporto** $\mathcal{X}$ di X , in questo caso, è un **intervallo reale**.

Per una variabile aleatoria discreta abbiamo visto che possiamo calcolare la probabilità di ciascun valore:

$$P(X = x_i) \quad \text{per ogni valore } x_i \in \mathcal{X}$$

e abbiamo chiamato questa funzione la *distribuzione di probabilità*.

Ma cosa succede se X è continua e **non possiamo elencare i valori del supporto**? Descrivere la distribuzione elencando la probabilità di ogni singolo valore non è la strada giusta.

Per una variabile aleatoria descriviamo la distribuzione attraverso una funzione, chiamata **densità di probabilità**, che indicheremo con $f_X(x)$.

La densità ci permette di calcolare la **probabilità che X assuma un valore all'interno di un intervallo**.

Definizione

Sia X una variabile aleatoria continua, a valori nel supporto $\mathcal{X} \subseteq \mathbb{R}$. La funzione **densità di probabilità**, $f_X(x)$, permette di calcolare la probabilità che X assuma valori in un intervallo $(a, b) \subseteq \mathcal{X}$ come:

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx$$

- ▶ La densità è sempre non-negativa: $f_X(x) \geq 0$
- ▶ L'area sotto tutta la densità è uguale a 1:

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

- ▶ Differenza importante: Per una variabile aleatoria discreta possiamo avere $P(X = x) > 0$. Per una variabile aleatoria continua, invece,

$$P(X = x) = 0 \quad \text{per ogni singolo valore } x.$$

Questo **non significa** che X non possa assumere il valore x ! Significa che la probabilità è associata agli **intervalli** di valori, e non ai singoli punti: $P(a \leq X \leq b)$.

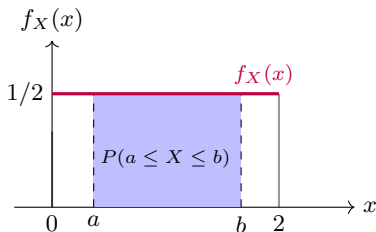
La funzione di densità: area sotto la curva

La funzione di densità $f_X(x)$ può essere pensata come una **curva**. La **probabilità** che X assuma un valore compreso tra a e b (ovvero l'integrale) corrisponde a calcolare l'**area sotto la curva tra a e b** .

Esempi:

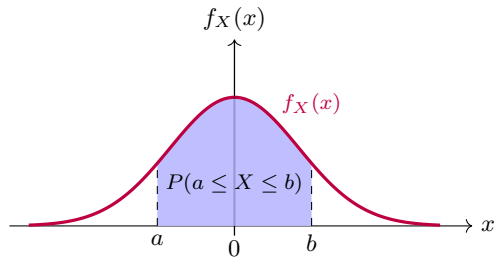
$f_X(x)$ funzione che assume il valore $1/2$ nell'intervallo $(0, 2)$, e 0 altrimenti.

Nota: l'area totale è $2 \times 1/2 = 1$



Curva “a campana” centrata in 0.

Il supporto sono tutti i numeri reali. L'area totale è pari a 1.



2

La funzione di ripartizione

Sia per variabili discrete sia continue, spesso siamo interessati una domanda del tipo:

“Qual è la probabilità che X non superi un certo valore x ?”

Definizione

La **funzione di ripartizione** di una variabile aleatoria X è la funzione

$$F_X(x) = P(X \leq x) \quad \text{per } x \in \mathbb{R}$$

- ▶ È definita **allo stesso modo** sia per v.a. discrete sia continue: cambia solo **come** si calcola.
- ▶ F_X è sempre **non-decrescente**: se x aumenta, $F_X(x)$ non può diminuire.
- ▶ Ai due estremi: $\lim_{x \rightarrow -\infty} F_X(x) = 0$ $\lim_{x \rightarrow +\infty} F_X(x) = 1$

Caso discreto: come si calcola F_X

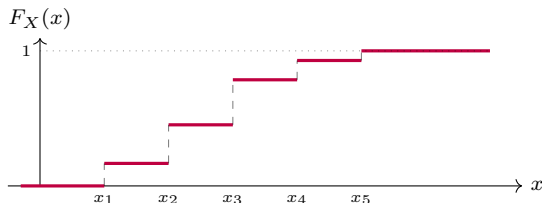
Se X è discreta con supporto $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$, per calcolare $F_X(x)$ basta **sommare le probabilità di tutti i valori $\leq x$** :

$$F_X(x) = \sum_{x_i \leq x} P(X = x_i)$$

Esempio: sia X = somma di due dadi. Vogliamo calcolare $F_X(4) = P(X \leq 4)$:

$$F_X(4) = P(X = 2) + P(X = 3) + P(X = 4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36}$$

- Conseguenza importante: F_X di una v.a. **discreta** è una funzione **“a gradini”**: fa un salto in corrispondenza di ogni $x_i \in \mathcal{X}$, l'altezza del salto è pari a $P(X = x_i)$, e resta costante tra un valore e l'altro.



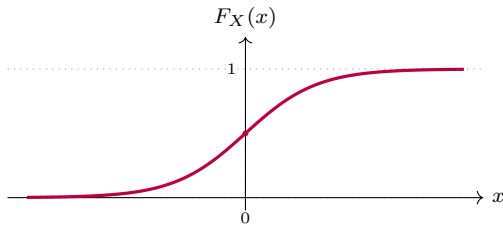
Caso continuo: come si calcola F_X

Se X è continua con densità $f_X(x)$, per calcolare $F_X(x)$ dobbiamo **sommare (integrare)** la densità da $-\infty$ fino a x :

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt$$

In pratica $F_X(x)$ è l'**area sotto la curva della densità**, calcolata da $-\infty$ fino al punto x .

- ▶ A differenza del caso discreto, F_X di una v.a. **continua** non ha "salti": è una curva continua che cresce con gradualità da 0 a 1.
- ▶ Relazione utile con la densità: **$f_X(x) = F'_X(x)$** : la densità è la derivata della funzione di ripartizione.



3

Indipendenza di variabili aleatorie

Quando due variabili X e Y sono definite sullo stesso esperimento, le descriviamo **insieme**.

Caso discreto

Distribuzione congiunta: $P(X = x, Y = y)$ per ogni coppia (x, y) .

Le **marginali** si ottengono sommando sull'altra variabile:

$$P(X = x) = \sum_y P(X = x, Y = y), \quad P(Y = y) = \sum_x P(X = x, Y = y).$$

Caso continuo

Densità congiunta: $f_{X,Y}(x, y) \geq 0$, con $P((X, Y) \in A) = \iint_A f_{X,Y}(x, y) dx dy$.

Le marginali si ottengono integrando sull'altra variabile:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx.$$

Esempio: congiunta e marginali (caso discreto)

Siano X e Y due variabili con valori in $\{0, 1\}$ e distribuzione congiunta

	$Y = 0$	$Y = 1$	$P(X = x)$
$X = 0$	0.3	0.2	0.5
$X = 1$	0.1	0.4	0.5
$P(Y = y)$	0.4	0.6	1

- I totali di riga e di colonna sono le distribuzioni **marginali**.
- Le marginali non bastano a ricostruire la congiunta: servono le probabilità delle coppie.

Indipendenza di due variabili aleatorie: idea

Consideriamo due variabili aleatorie X e Y definite sullo stesso esperimento.

Intuitivamente, X e Y sono **indipendenti** se conoscere il valore di una non ci dà alcuna informazione sull'altra.

Ricordiamo che due eventi A e B sono indipendenti se $P(A \cap B) = P(A) P(B)$. Per le variabili aleatorie chiediamo la stessa cosa per *tutti* gli eventi che riguardano X e Y :

Definizione

X e Y sono **indipendenti** se, per ogni coppia di intervalli (o insiemi) A e B ,

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

Esempi intuitivi:

- ▶ Il risultato del primo e del secondo lancio di un dado: **indipendenti**.
- ▶ L'altezza e il peso di una persona: **non** indipendenti (persone più alte tendono a pesare di più).

Indipendenza nel caso discreto

Se X e Y sono discrete, la definizione si riduce a una condizione sui singoli valori:

Indipendenza (v.a. discrete)

X e Y sono indipendenti se e solo se

$$P(X = x, Y = y) = P(X = x) P(Y = y) \quad \text{per ogni } x \in \mathcal{X}, y \in \mathcal{Y}$$

Esempio (indipendenti): lanciamo due dadi. X = risultato del primo, Y = risultato del secondo:

$$P(X = 1, Y = 2) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = P(X = 1) P(Y = 2)$$

e lo stesso vale per ogni coppia di valori.

Esempio (non indipendenti): X = risultato del primo dado, Z = **somma** dei due dadi.

Se $X = 1$ la somma non può essere 12:

$$P(X = 1, Z = 12) = 0 \neq \frac{1}{6} \times \frac{1}{36} = P(X = 1) P(Z = 12)$$

Basta **un solo** caso in cui la relazione non vale per concludere che X e Z **non** sono indipendenti.

Indipendenza nel caso continuo

Per due v.a. continue X e Y con densità $f_X(x)$ e $f_Y(y)$, l'indipendenza si esprime in termini di densità: la densità congiunta è il **prodotto** delle due densità.

Indipendenza (v.a. continue)

X e Y sono indipendenti se e solo se la densità congiunta si fattorizza:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \text{per ogni } (x,y)$$

Come conseguenza, le probabilità sui rettangoli si **moltiplicano**:

$$P(a \leq X \leq b, c \leq Y \leq d) = P(a \leq X \leq b) \times P(c \leq Y \leq d)$$

Esempio: i tempi di attesa al bus di due persone diverse, X e Y , ciascuno uniforme su $(0, 10)$ e indipendenti tra loro. La probabilità che entrambi aspettino meno di 5 minuti è

$$P(X \leq 5, Y \leq 5) = P(X \leq 5) P(Y \leq 5) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Possiamo estendere la definizione di indipendenza a una sequenza di n variabili aleatorie:

- **Caso discreto:** Siano $(X_1, \dots, X_n)$ variabili aleatorie **discrete indipendenti**. La probabilità di una qualunque sequenza si ottiene **moltiplicando** le probabilità dei singoli eventi:

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

Esempio: lanciamo n volte una moneta, e sia $X_i = 1$ se all' i -esimo lancio esce testa, 0 altrimenti.

Le $(X_1, \dots, X_n)$ sono indipendenti, quindi la probabilità dell'evento "esce sempre testa"

$$P(X_1 = 1, \dots, X_n = 1) = \prod_{i=1}^n P(X_i = 1)$$

- **Caso continuo:** Analogamente, per n variabili $(X_1, \dots, X_n)$ **continue indipendenti**, la densità **congiunta** è data dal prodotto delle singole densità:

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

4

Valore atteso

Definiamo il valore atteso separatamente per variabili aleatorie discrete e continue. Per calcolarlo, nel primo caso useremo la distribuzione di probabilità, nel secondo la funzione di densità.

Definizione

Definiamo il **valore atteso** di una variabile aleatoria **discreta** X , con supporto $\mathcal{X} = \{x_1, \dots, x_n\}$ come

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

Idea: immaginiamo di ripetere un esperimento aleatorio infinite volte e ogni volta ci annotiamo il valore che assume la variabile X . $E[X]$ è la media di questi valori.

Esempio: lancio di un dado equo a 6 facce.

X = numero uscito dal lancio

$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \dots + 6 \times \frac{1}{6} = 3.5$$

Definizione

Definiamo il **valore atteso** di una variabile aleatoria **continua** X , con supporto $\mathcal{X}$ e funzione di densità $f_X(x)$, come

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$$

- Sia X una variabile aleatoria, e siano a e b due numeri reali.

Il valore atteso della **trasformazione lineare** $Y = aX + b$ è

$$E[Y] = E[aX + b] = aE[X] + b$$

Esempio: Sia X = numero di panini venduti da un bar in un giorno.

Sappiamo da dati storici che $E[X] = 120$

Per ogni panino, il bar guadagna 3 euro.

Ora, chiamiamo Y = guadagno risultante dalla vendita di panini in un giorno $\implies Y = 3X$

Il valore atteso di Y è $E[Y] = E[3X] = 3E[X] = 360$

- Siano X e Y due variabili aleatorie,

$$E[X + Y] = E[X] + E[Y]$$

- Siano X e Y due variabili aleatorie **indipendenti**,

$$E[XY] = E[X] E[Y]$$

5

Varianza

Definizione

Definiamo la **varianza** di una variabile aleatoria X come

$$\begin{aligned}\text{var}(X) &= E\left[(X - E[X])^2\right] \\ &= E[X^2] - E[X]^2\end{aligned}$$

Idea: La varianza di una variabile aleatoria misura quanta incertezza c'è nel valore che assume X .

Se $\text{var}(X) = 0$, allora X non è veramente aleatoria, e per ogni esperimento si verifica l'esito $X = E[X]$.

Esempio: Sia X una variabile aleatoria discreta con $\mathcal{X} = \{0, 1\}$ tale che

$$X = \begin{cases} 1 & \text{con probabilità } p \\ 0 & \text{con probabilità } (1 - p) \end{cases}$$

- ▶ il valore atteso di X è $E[X] = 1 \times p + 0 \times (1 - p) = p$
- ▶ il valore atteso di X^2 è $E[X^2] = 1^2 \times p + 0^2 \times (1 - p) = p$
- ▶ la varianza è $\text{var}(X) = E[X^2] - E[X]^2 = p - p^2 = p(1 - p)$

- Sia X una variabile aleatoria, e siano a e b due numeri reali.

La varianza della **trasformazione lineare** $Y = aX + b$ è

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

- Siano X e Y due variabili aleatorie **indipendenti**,

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

(la relazione non vale se le variabili non sono indipendenti!)

6

Covarianza e correlazione

Finora abbiamo descritto ogni variabile da sola (valore atteso, varianza). Spesso però ci interessa capire se due variabili X e Y tendono a muoversi insieme.

Idea: guardiamo se, quando X è sopra la sua media, anche Y tende a essere sopra la sua media (e viceversa, se quando X è sotto la sua media, anche Y tende a essere sotto).

Il prodotto $(X - E[X])(Y - E[Y])$ è

- ▶ **positivo** se X e Y sono entrambe sopra, o entrambe sotto, la propria media;
- ▶ **negativo** se una è sopra e l'altra è sotto.

Definizione

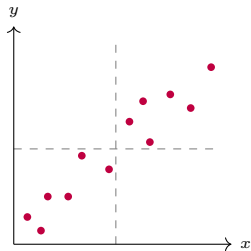
La **covarianza** tra X e Y è il valore atteso di questo prodotto:

$$\begin{aligned}\text{cov}(X, Y) &= E\left[(X - E[X])(Y - E[Y])\right] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

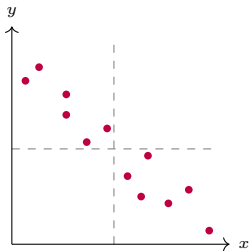
Esempio: X = altezza e Y = peso di una persona scelta a caso. Chi è più alto della media tende a essere anche più pesante della media: ci aspettiamo $\text{cov}(X, Y) > 0$.

Covarianza: cosa ci dice il segno

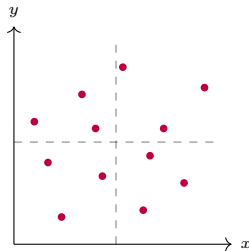
Ogni punto è una realizzazione della coppia (X, Y) ; le linee tratteggiate indicano le medie.



$$\text{cov}(X, Y) > 0$$



$$\text{cov}(X, Y) < 0$$



$$\text{cov}(X, Y) \approx 0$$

- ▶ $\text{cov} > 0$: i punti stanno soprattutto in alto a destra e in basso a sinistra: X e Y crescono **insieme**;
- ▶ $\text{cov} < 0$: i punti stanno soprattutto in alto a sinistra e in basso a destra: quando una cresce, l'altra **diminuisce**;
- ▶ $\text{cov} \approx 0$: nessuna tendenza (lineare) a muoversi insieme.

- La covarianza è **simmetrica**:

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

- La covarianza di una variabile con se stessa è la varianza,

$$\text{cov}(X, X) = \text{var}(X)$$

- Per **trasformazioni lineari**, con a, b, c, d numeri reali:

$$\text{cov}(aX + b, cY + d) = a c \text{cov}(X, Y)$$

- **Varianza di una somma**:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y)$$

Se X e Y sono **indipendenti** allora $\text{cov}(X, Y) = 0$, e ritroviamo la formula vista in precedenza:
 $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

Intuizione: se X e Y tendono a muoversi insieme ($\text{cov} > 0$), la varianza della somma è maggiore della somma delle varianze; se si muovono in direzioni opposte ($\text{cov} < 0$), le oscillazioni si compensano e la somma è meno variabile.

Dalla covarianza alla correlazione

La covarianza dipende dall'**unità di misura**.

Se X è l'altezza in metri e $X' = 100 X$ è la stessa altezza in centimetri, allora

$$\text{cov}(X', Y) = 100 \text{ cov}(X, Y)$$

anche se la relazione tra le due variabili è *esattamente la stessa*. Quindi il valore numerico della covarianza da solo **non ci dice se il legame è forte o debole**.

Per risolvere il problema **dividiamo per le deviazioni standard**, $\sigma_X = \sqrt{\text{var}(X)}$ e $\sigma_Y = \sqrt{\text{var}(Y)}$:

Definizione

Il **coefficiente di correlazione** tra X e Y è

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

La correlazione è un numero **senza unità di misura**, che vale sempre

$$-1 \leq \rho_{X,Y} \leq 1$$

Come si interpreta la correlazione

- ▶ $\rho = 1$: relazione lineare **crescente** perfetta ($Y = aX + b$ con $a > 0$);
- ▶ $\rho = -1$: relazione lineare **decrescente** perfetta ($Y = aX + b$ con $a < 0$);
- ▶ $\rho = 0$: **nessuna relazione lineare** (X e Y si dicono **incorrelate**);
- ▶ valori intermedi: più $|\rho|$ è vicino a 1, più la relazione lineare è forte.

Attenzione

La correlazione misura solo il legame **lineare**, e non implica una relazione di causa-effetto: due variabili possono essere fortemente correlate senza che una influenzi l'altra.

Queste correlazioni vengono dette **spurie**: qualche esempio

<https://www.tylervigen.com/spurious-correlations>

Indipendenza $\Rightarrow$ covarianza nulla, ma non viceversa

$$X, Y \text{ indipendenti} \Rightarrow \text{cov}(X, Y) = 0 \quad \text{ma} \quad \text{cov}(X, Y) = 0 \not\Rightarrow X, Y \text{ indipendenti}$$

Il motivo: la covarianza misura solo il legame **lineare**, mentre l'indipendenza esclude **qualsiasi** tipo di legame.

Controesempio: X assume i valori $-1, 0, 1$ con probabilità $1/3$ ciascuno, e $Y = X^2$.

- ▶ $E[X] = 0$ e $E[XY] = E[X^3] = 0$, quindi $\text{cov}(X, Y) = E[XY] - E[X]E[Y] = 0$;
- ▶ ma Y è **completamente determinata** da X ! Per esempio $P(X = 1, Y = 0) = 0 \neq \frac{1}{3} \times \frac{1}{3} = P(X = 1)P(Y = 0)$.

7

Distribuzioni fondamentali

Abbiamo visto come descrivere una variabile aleatoria attraverso la sua **distribuzione di probabilità** o **funzione di densità**. Esistono alcune distribuzioni e densità particolarmente importanti, che ricorrono frequentemente nelle applicazioni.

In particolare, vedremo:

- ▶ nel caso delle **variabili aleatorie discrete**:
 - ▶ la **distribuzione uniforme discreta**;
 - ▶ la **distribuzione di Bernoulli**;
 - ▶ la **distribuzione binomiale**;
 - ▶ la **distribuzione geometrica**;
 - ▶ la **distribuzione di Poisson**;
- ▶ nel caso delle **variabili aleatorie continue**:
 - ▶ la **distribuzione uniforme continua**;
 - ▶ la **distribuzione normale**;
 - ▶ la **distribuzione esponenziale**;
 - ▶ la **distribuzione gamma**;
 - ▶ la **distribuzione t di Student**.

Queste distribuzioni ci permetteranno di **modellare fenomeni casuali** che incontriamo in molte applicazioni economiche e statistiche.

7.1

Distribuzioni fondamentali

Variabili aleatorie discrete

La distribuzione uniforme discreta

La **distribuzione uniforme discreta** descrive una variabile che può assumere un numero finito di valori, **tutti ugualmente probabili**.

Definizione

Sia X una variabile aleatoria che assume i valori $\{a, a+1, \dots, b\}$, ciascuno con la stessa probabilità:

$$P(X = k) = \frac{1}{b - a + 1} \quad k = a, a+1, \dots, b.$$

Diremo che X ha **distribuzione uniforme discreta** su $\{a, \dots, b\}$, e scriviamo $X \sim \text{Unif}\{a, \dots, b\}$.

Se $X \sim \text{Unif}(\{a, \dots, b\})$, il **valore atteso** è

$$E[X] = \frac{a+b}{2}.$$

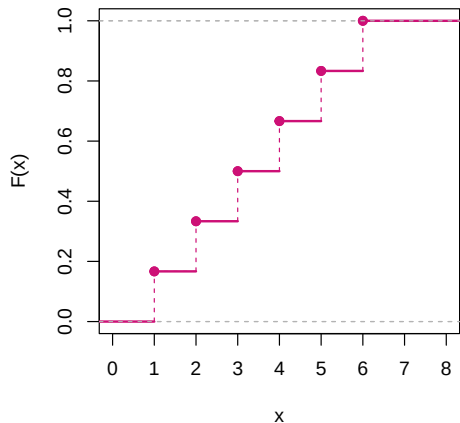
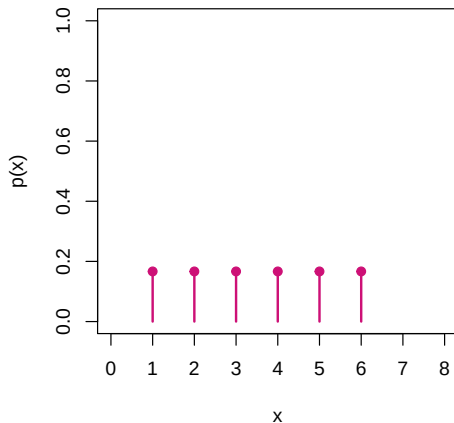
Esempio: Sia X l'esito del lancio di un dado a 6 facce equilibrato

$$X \sim \text{Unif}\{1, 2, 3, 4, 5, 6\}, \quad P(X = k) = \frac{1}{6} \quad \forall k; \quad \mathbb{E}[X] = \frac{1+6}{2} = 3.5.$$

La distribuzione uniforme discreta

Distribuzione di probabilità e funzione di ripartizione

Esempio: X = esito del lancio di un dado equilibrato a 6 facce



La **distribuzione di Bernoulli** descrive un esperimento aleatorio che può avere **due soli risultati possibili**.

Definizione

Sia X una variabile aleatoria che assume i seguenti valori:

$$X = \begin{cases} 1 & \text{se si verifica l'evento di interesse,} \\ 0 & \text{se non si verifica.} \end{cases}$$

Indichiamo con p la probabilità che si verifichi l'evento di interesse:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

Diremo che la variabile X ha **distribuzione di Bernoulli di parametro p** , e scriviamo $X \sim \text{Bernoulli}(p)$.

Esempio: Una fabbrica produce determinati prodotti. Vogliamo studiare se ciascun prodotto è difettoso: definiamo

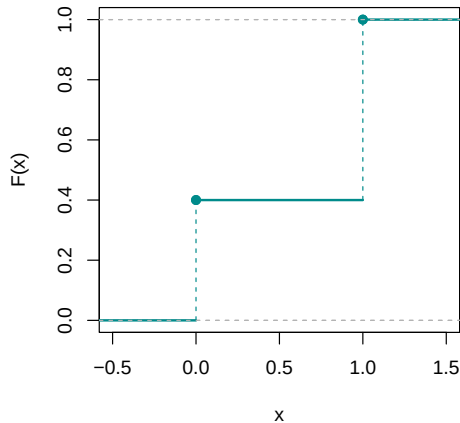
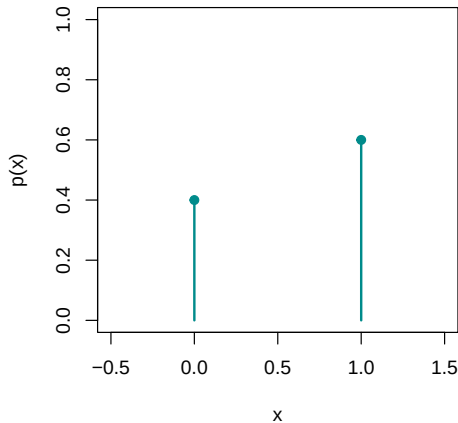
$$X = \begin{cases} 1 & \text{se il prodotto è difettoso,} \\ 0 & \text{se il prodotto non è difettoso.} \end{cases}$$

Se la probabilità che il prodotto sia difettoso è $p = 0.02$, allora $X \sim \text{Bernoulli}(0.02)$.

La distribuzione Bernoulli

Distribuzione di probabilità e funzione di ripartizione

Esempio: $X \sim \text{Bernoulli}(0.6)$



La distribuzione binomiale

Supponiamo di ripetere lo stesso esperimento di Bernoulli **n volte**, in condizioni analoghe e in modo indipendente.

Definiamo X come il **numero di volte** in cui si verifica l'evento di interesse:

X = numero di “successi” nelle n prove.

Quindi X può assumere i valori

$$0, 1, 2, \dots, n.$$

Definizione

Eseguiamo n prove di un esperimento con soli due risultati possibili: successo (1) o insuccesso (0).

Se ogni prova ha **probabilità p di successo** e le prove sono **indipendenti**, allora **X ha una distribuzione binomiale**:

$$X \sim \text{Bin}(n, p)$$

La probabilità di ottenere esattamente k successi è

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, \dots, n.$$

Esempio: distribuzione binomiale

Consideriamo di nuovo l'esempio della fabbrica. Scegliamo 10 prodotti e supponiamo che la probabilità che un prodotto sia difettoso sia $p = 0.02$.

Supponiamo inoltre che il fatto che un prodotto sia difettoso sia indipendente dagli altri prodotti.

Definiamo

X = numero di prodotti difettosi tra i 10.

Allora

$$X \sim \text{Bin}(10, 0.02).$$

Per esempio, la probabilità che ci siano esattamente 2 prodotti difettosi è

$$P(X = 2) = \binom{10}{2} (0.02)^2 (0.98)^8.$$

Bernoulli: una prova $\longrightarrow$ successo/non successo

Binomiale: n prove $\longrightarrow$ numero di successi

Sia $X \sim \text{Bin}(n, p)$,

- Il termine

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

dove $a! = a \times (a-1) \times (a-2) \times \cdots \times 2 \times 1$, è detto **coefficiente binomiale** e indica **in quanti modi possiamo scegliere k successi tra le n prove**. Si legge “ n su k ”.

- Il **valore atteso** di una variabile binomiale $X \sim \text{Bin}(n, p)$ è

$$E[X] = np$$

e rappresenta il numero medio di successi che ci aspettiamo di ottenere nelle n prove.

- La varianza di una variabile binomiale $X \sim \text{Bin}(n, p)$ è

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

e misura la variabilità del numero di successi intorno al suo valore atteso.

Una binomiale è la somma di n prove di Bernoulli indipendenti.

Siano $X_1, \dots, X_n$ i.i.d. con $X_i \sim \text{Bernoulli}(p)$ ($X_i = 1$ se la i -esima prova è un successo). Allora

$$X = X_1 + X_2 + \dots + X_n \sim \text{Bin}(n, p).$$

Valore atteso e varianza di una binomiale: calcolo a partire dalle Bernoulli.

Ogni X_i ha $E[X_i] = p$ e $\text{var}(X_i) = p(1 - p)$. Quindi:

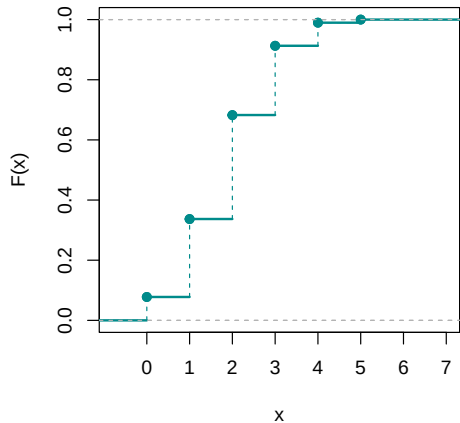
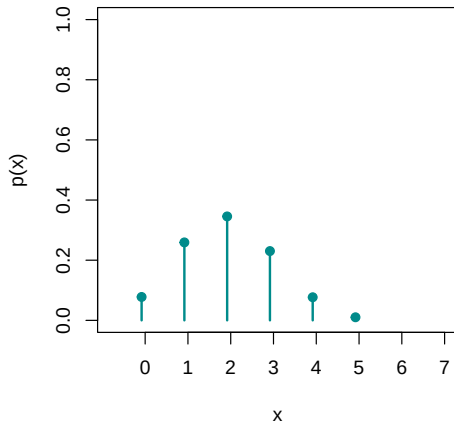
- ▶ $E[X] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = np$ (linearità del valore atteso);
- ▶ $\text{var}(X) = \text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) = np(1 - p)$ (le X_i sono indipendenti).

Esempio: tra 10 prodotti con $p = 0.02$, il numero atteso di difettosi è $E[X] = 10 \times 0.02 = 0.2$.

La distribuzione binomiale

Distribuzione di probabilità e funzione di ripartizione

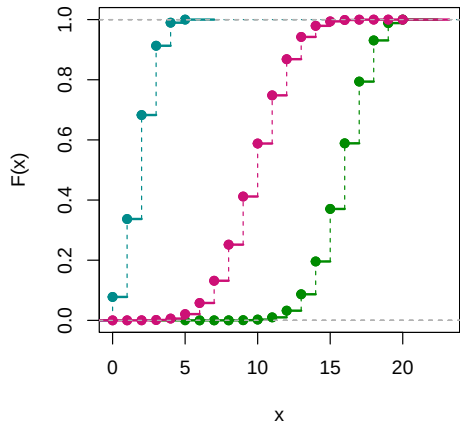
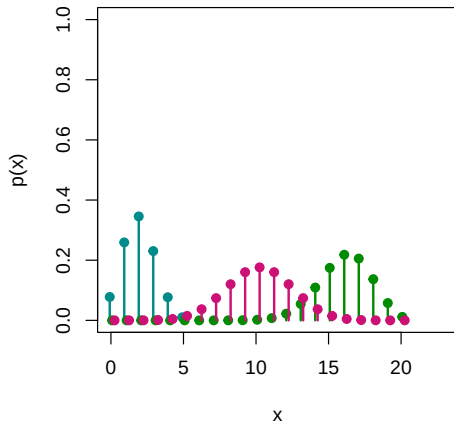
Esempio: $X \sim \text{Bin}(5, 0.4)$



La distribuzione binomiale

Distribuzione di probabilità e funzione di ripartizione: confronto

Esempio: $X_1 \sim \text{Bin}(5, 0.4)$, $X_2 \sim \text{Bin}(20, 0.5)$, $X_3 \sim \text{Bin}(20, 0.8)$



Consideriamo una successione di prove **indipendenti**, ciascuna con probabilità di successo p , ripetute **fino al primo successo**.

Definizione

Sia X il **numero di prove necessarie per ottenere il primo successo**. Allora X ha distribuzione geometrica di parametro $p \in (0, 1)$:

$$X \sim \text{Geom}(p)$$

La probabilità che servano esattamente k prove è

$$P(X = k) = (1 - p)^{k-1} p \quad k = 1, 2, 3, \dots$$

Se $X \sim \text{Geom}(p)$, il suo valore atteso e varianza sono:

$$E[X] = \frac{1}{p}, \quad \text{var}(X) = \frac{1-p}{p^2}.$$

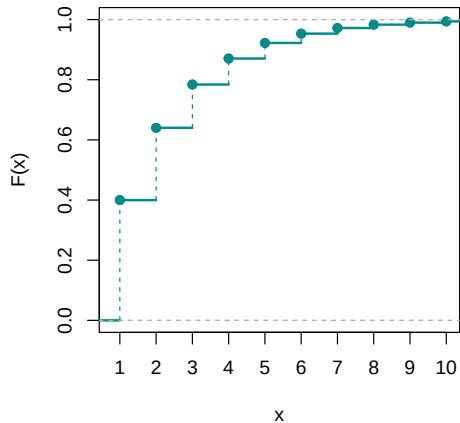
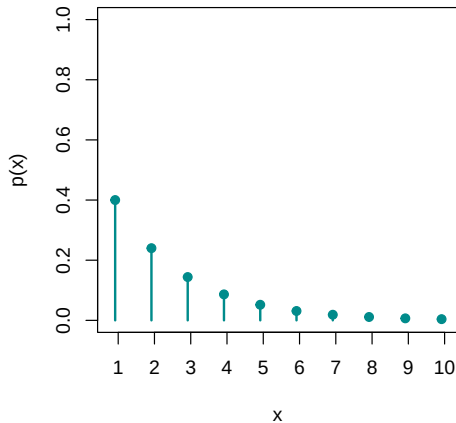
Esempio: Lanciamo una moneta equilibrata fino a ottenere testa. Se X è il numero di lanci necessari, $X \sim \text{Geom}(0.5)$, e

$$P(X = 3) = (0.5)^2(0.5) = 0.125 \quad (\text{cioè: croce, croce, testa}).$$

La distribuzione geometrica

Distribuzione di probabilità e funzione di ripartizione

Esempio: $X \sim \text{Geom}(0.4)$



La distribuzione di Poisson

La **distribuzione di Poisson** descrive il **numero di eventi** che si verificano in un intervallo fissato di tempo o spazio, quando gli eventi accadono con un **tasso medio costante** λ e in modo indipendente.

Definizione

Sia X il numero di eventi osservati in un intervallo, con tasso medio $\lambda > 0$. Allora X ha distribuzione di Poisson:

$$X \sim \text{Poisson}(\lambda)$$

La probabilità di osservare esattamente k eventi è

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad k = 0, 1, 2, \dots$$

Se $X \sim \text{Poisson}(\lambda)$, il valore atteso e la varianza sono uguali e sono pari al tasso λ

$$\mathbb{E}[X] = \text{var}(X) = \lambda$$

Esempio: In un bar entrano in media $\lambda = 10$ clienti all'ora. Il numero di clienti X in una data ora segue $X \sim \text{Poisson}(10)$, e

$$P(X = 2) = \frac{e^{-10} 10^2}{2!}.$$

Se X_1 e X_2 sono due variabili aleatorie **indipendenti** con distribuzioni di Poisson di parametri λ_1 e λ_2 rispettivamente, ovvero

$$X_1 \sim \text{Poisson}(\lambda_1) \perp\!\!\!\perp X_2 \sim \text{Poisson}(\lambda_2)$$

la loro **somma**

$$Y = X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

ovvero segue ancora una distribuzione di Poisson, di parametro $\lambda = \lambda_1 + \lambda_2$.

In generale, date n variabili aleatorie Poisson indipendenti, $X_i \sim \text{Poisson}(\lambda_i)$ per $i = 1, \dots, n$, la somma $Y = X_1 + X_2 + \dots + X_n$ segue una distribuzione di Poisson di parametro $\lambda = \lambda_1 + \dots + \lambda_n$

Esempio: un bar ha due porte d'ingresso: una che dà sulla strada principale, una sul vicolo laterale.

- Dalla porta principale il numero di clienti in un'ora (X_1) segue una distribuzione Poisson con $\lambda_1 = 10$;
- Dal vicolo il numero di clienti in un'ora (X_2) segue una distribuzione Poisson di parametro $\lambda_2 = 4$.

Inoltre, l'ingresso dei clienti da una porta è indipendente dall'ingresso dei clienti dall'altra.

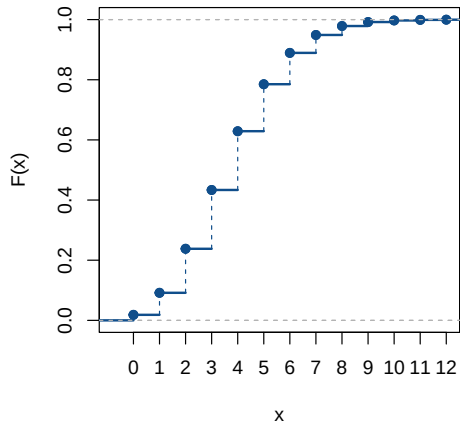
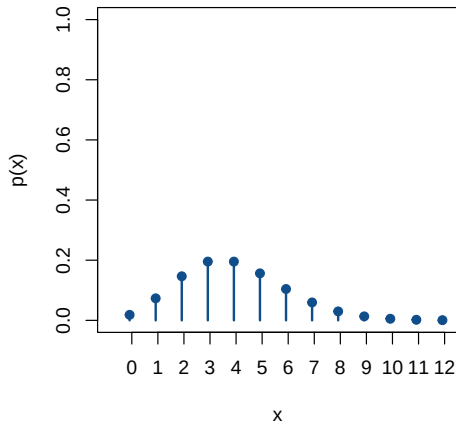
Noi siamo interessati solo a quanti clienti totali arrivano in un'ora. Il numero di clienti in un'ora è:

$$X_1 + X_2 \sim \text{Poisson}(10 + 4) = \text{Poisson}(14)$$

La distribuzione di Poisson

Distribuzione di probabilità e funzione di ripartizione

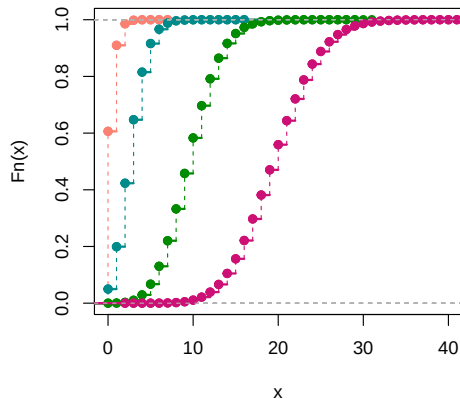
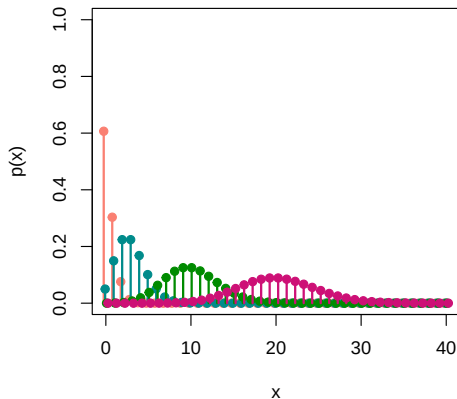
Esempio: $X \sim \text{Poisson}(4)$



La distribuzione di Poisson

Distribuzione di probabilità e funzione di ripartizione: confronto

Esempio: $X \sim \text{Poisson}(0.5)$, $X \sim \text{Poisson}(3)$, $X \sim \text{Poisson}(10)$, $X \sim \text{Poisson}(20)$



7.2

Distribuzioni fondamentali

Variabili aleatorie continue

La distribuzione **uniforme continua** descrive una variabile che può assumere qualunque valore in un **intervallo** (a, b) , con la **stessa densità** ovunque.

Definizione

X ha distribuzione uniforme su (a, b) , e scriviamo $X \sim \text{Unif}(a, b)$, se

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b, \\ 0 & \text{altrimenti.} \end{cases} \quad F_X(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \leq x \leq b, \\ 1 & x > b. \end{cases}$$

Valore atteso e varianza:

$$E[X] = \frac{a+b}{2}, \quad \text{var}(X) = \frac{(b-a)^2}{12}.$$

Esempio: il tempo di attesa all'autobus $X \sim \text{Unif}(0, 10)$ ha $E[X] = 5$ minuti e $P(X \leq 5) = \frac{5-0}{10-0} = \frac{1}{2}$.

La **distribuzione normale** (o **gaussiana**) è la distribuzione continua più importante: descrive moltissimi fenomeni naturali ed economici (es. errori di misura, caratteristiche di una popolazione).

Definizione

Una variabile aleatoria continua X ha **distribuzione normale** con parametri μ e σ^2 se la sua densità è

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}$$

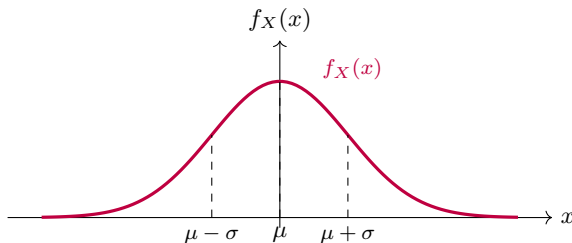
e scriviamo $X \sim N(\mu, \sigma^2)$.

I due parametri hanno un significato diretto:

- ▶ μ è il **valore atteso**: $E[X] = \mu$;
- ▶ σ^2 è la **varianza**: $\text{var}(X) = \sigma^2$.

La curva normale

La densità normale ha una caratteristica forma a **campana**, **simmetrica** attorno a μ .

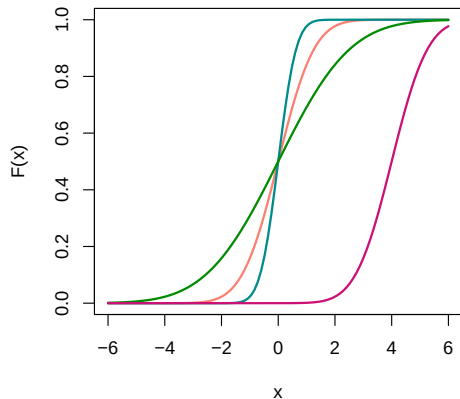
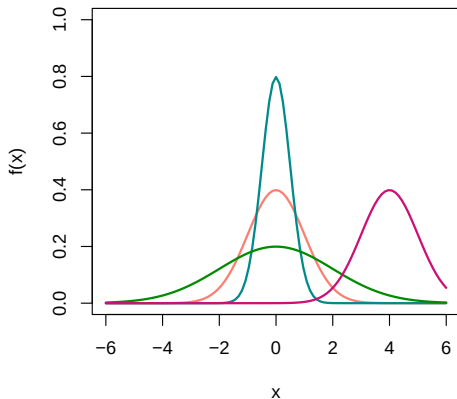


- ▶ μ determina il **centro** (posizione) della campana;
- ▶ σ^2 determina la **dispersione** (quanto è "larga" o "stretta" la campana): σ^2 piccolo $\rightarrow$ campana stretta e alta; σ^2 grande $\rightarrow$ campana larga e bassa.

La distribuzione normale

Densità e funzione di ripartizione: confronto

Esempio: $X \sim N(0, 1)$, $X \sim N(0, 0.25)$, $X \sim N(0, 4)$, $X \sim N(4, 1)$



Per una variabile $X \sim N(\mu, \sigma^2)$, la probabilità che X cada in un intervallo attorno a μ segue una regola semplice:

- ▶ $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$
- ▶ $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$
- ▶ $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997$

Esempio: Il punteggio X ottenuto in un test attitudinale da una popolazione di candidati segue $X \sim N(100, 15^2)$.

Circa il 68% dei candidati ottiene un punteggio tra 85 e 115; circa il 95% tra 70 e 130.

Definizione

La distribuzione normale con $\mu = 0$ e $\sigma^2 = 1$ è detta **normale standard**,

$$f_Z(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad z \in \mathbb{R}$$

e scriviamo $Z \sim N(0, 1)$.

Applicando la regola empirica per una normale standard, vediamo che

- ▶ $P(-2 \leq Z \leq 2) \approx 0.95$
- ▶ $P(-3 \leq Z \leq 3) \approx 0.997$

Ogni variabile normale può essere trasformata in una normale standard, cioè con $\mu = 0$ e $\sigma^2 = 1$.

► Sia $X \sim N(\mu, \sigma^2)$, allora

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

ha distribuzione normale standard, $Z \sim N(0, 1)$.

La standardizzazione è utile perché ci permette di calcolare probabilità per **qualsiasi** normale usando le **tavole della normale standard**.

Esempio: Sia X = rendimento annuo (%) di un titolo, $X \sim N(6, 4)$ (quindi $\sigma = 2$).

La probabilità che il rendimento superi il 10% si calcola standardizzando:

$$P(X > 10) = P\left(\frac{X - 6}{2} > \frac{10 - 6}{2}\right) = P\left(Z > \frac{10 - 6}{2}\right) = P(Z > 2)$$

Le tavole della normale standard

Per la normale non esiste una formula semplice per calcolare l'area sotto la densità. Le probabilità della normale standard per una serie di valori sono raccolte in tabelle dette **tavole** (o si calcolano a computer). Esempio di tavole: <https://users.dimi.uniud.it/~lorenzo.freddi/materialewebagraria/tavole.pdf>

Notazione

Indichiamo con $\Phi(z)$ la funzione di ripartizione della normale standard:

$$\Phi(z) = P(Z \leq z)$$

cioè l'**area a sinistra** di z sotto la densità normale standard.

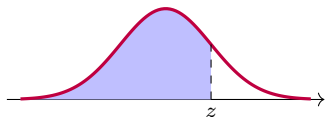
Come si legge la tavola (estratto): la riga dà l'intero e il primo decimale di z , la colonna il secondo decimale. Per esempio $z = 1.25$: riga 1.2, colonna 0.05.

z	0.00	0.01	0.02	0.03	0.04	0.05
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115

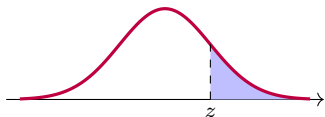
Quindi $P(Z \leq 1.25) = \Phi(1.25) = 0.8944$.

Calcolare le probabilità a partire dalle tavole

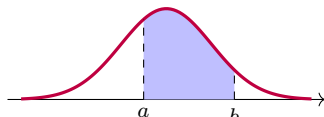
Poiché Z è continua, $P(Z < z) = P(Z \leq z)$: gli estremi non contano.



$$P(Z < z) = \Phi(z)$$



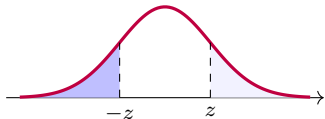
$$P(Z > z) = 1 - \Phi(z)$$



$$P(a < Z < b) = \Phi(b) - \Phi(a)$$

Valori negativi: la tavola a volte contiene solo $z \geq 0$. Per la **simmetria** della densità attorno a 0,

$$\Phi(-z) = 1 - \Phi(z) \quad \text{cioè} \quad P(Z < -z) = P(Z > z)$$



$$P(Z < -z) = P(Z > z)$$

Sapendo che $\Phi(1.25) = 0.8944$, $\Phi(0.5) = 0.6915$, $\Phi(1.5) = 0.9332$, $\Phi(1.96) = 0.9750$, calcolare

- ▶ $P(Z < 1.25) = \Phi(1.25) = 0.8944$
- ▶ $P(Z > 1.25) = 1 - \Phi(1.25) = 1 - 0.8944 = 0.1056$
- ▶ $P(Z < -1.25) = 1 - \Phi(1.25) = 0.1056$ (per simmetria, uguale a $P(Z > 1.25)$)
- ▶ $P(0.5 < Z < 1.5) = \Phi(1.5) - \Phi(0.5) = 0.9332 - 0.6915 = 0.2417$
- ▶ $P(-1.96 < Z < 1.96) = \Phi(1.96) - \Phi(-1.96) =$
 $= \Phi(1.96) - [1 - \Phi(1.96)] =$
 $= 2\Phi(1.96) - 1 = 2 \times 0.9750 - 1 = 0.95$

Le tavole esistono solo per la normale **standard**. Per $X \sim N(\mu, \sigma^2)$ si **standardizza** e ci si riconduce a Z :

1. **Standardizza**: sottrai μ e dividi per σ (la deviazione standard) tutti i termini della disuguaglianza (sia la v.a. sia il valore);

Esempio: $P(X < x) = P\left(\frac{X - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right)$

2. **Riscrivi** in termini di Φ ;

Esempio: $P\left(\frac{X - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$

3. La nuova formulazione è riscritta in termini di normale standard: usiamo le tavole seguendo le regole viste.

Esempio: punteggio in un test $X \sim N(100, 15^2)$, quindi $\mu = 100$, $\sigma = 15$.

- ▶ $P(X < 115) = \Phi\left(\frac{115 - 100}{15}\right) = \Phi(1) = 0.8413$
- ▶ $P(X > 130) = 1 - \Phi\left(\frac{130 - 100}{15}\right) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$
- ▶ $P(85 < X < 130) = P(X < 130) - P(X < 85) = \Phi\left(\frac{130 - 100}{15}\right) - \Phi\left(\frac{85 - 100}{15}\right) = \Phi(2) - \Phi(-1) = \Phi(2) - (1 - \Phi(1)) = 0.9772 - 0.1587 = 0.8185$

Finora, dato z , abbiamo cercato $\Phi(z) = P(Z \leq z) = p$. Spesso serve il problema **inverso**: data una probabilità p , trovare z_p tale che $\Phi(z_p) = p$.

Definizione

Il quantile di livello $p \in (0, 1)$ della normale standard è il valore z_p tale che

$$\Phi(z_p) = P(Z \leq z_p) = p.$$

- Sulle tavole: si cerca p *dentro* la tabella e si legge z su riga e colonna.
- Per simmetria: $z_{1-p} = -z_p$.

Quantili che useremo spesso:

p	0.90	0.95	0.975
z_p	1.28	1.645	1.96

- Se X e Y sono **indipendenti** e normali,

$$X \sim N(\mu_X, \sigma_X^2) \quad \text{e} \quad Y \sim N(\mu_Y, \sigma_Y^2),$$

allora anche la loro **somma** è normale:

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

- Perché la distribuzione normale è così importante?

Consideriamo un qualsiasi esperimento aleatorio e una variabile X associata all'esperimento, con

$$E[X] = \mu \quad \text{e} \quad \text{var}(X) = \sigma^2 > 0 \quad \text{finita.}$$

Se ripetiamo l'esperimento n volte in maniera **indipendente** e chiamiamo

X_1 = variabile associata al 1° esperimento,

X_2 = variabile associata al 2° esperimento,

$\vdots$

allora, per n **abbastanza grande**:

$$X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)$$

Questo vale (approssimativamente) **qualunque** sia la distribuzione di partenza di X !

La **distribuzione esponenziale** descrive il **tempo di attesa** di un fenomeno che “non invecchia” (ossia la distribuzione esponenziale è priva di memoria).

Cosa vuol dire? Esempio:

Immagina di essere al binario del treno alla stazione di Greco Pirelli e che l'**attesa del treno** segua una legge esponenziale: cosa significa?

- ▶ Ti siedi e aspetti 20 minuti. Guardi l'orologio. Il treno non è ancora arrivato.
- ▶ Qual è la probabilità che arrivi nei prossimi 5 minuti?
È esattamente la stessa identica probabilità che avevi nei primi 5 minuti appena arrivato al binario.
- ▶ Il passato non influisce sul futuro. Il sistema “non invecchia” e non accumula usura. Dal punto di vista statistico, ogni istante è come se fosse il primo.

Definizione

Una variabile aleatoria continua $X \geq 0$ ha **distribuzione esponenziale** di parametro $\lambda > 0$ se la sua densità è

$$f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0$$

e scriviamo $X \sim \text{Exp}(\lambda)$.

- Se $X \sim \text{Exp}(\lambda)$, il suo valore atteso e varianza sono:

$$E[X] = \frac{1}{\lambda}, \quad \text{var}(X) = \frac{1}{\lambda^2}.$$

- **Funzione di ripartizione:** ha una forma semplice, data da

$$P(X \leq x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

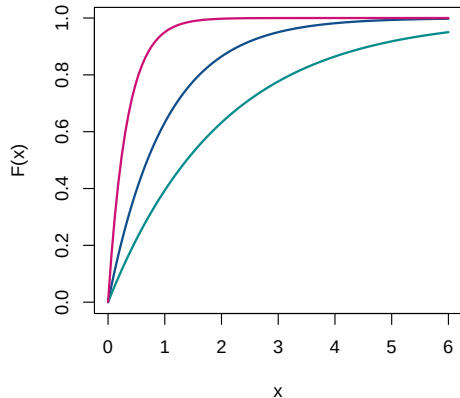
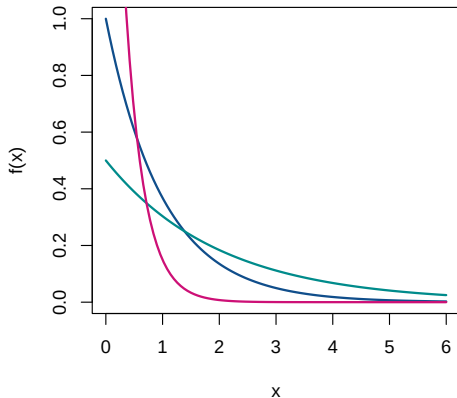
- **Proprietà di assenza di memoria:** il tempo già trascorso non influenza il tempo residuo di attesa.

$$P(X > s + t \mid X > s) = P(X > t) = e^{-\lambda t}$$

La distribuzione esponenziale

Densità e funzione di ripartizione: confronto

Esempio: $X \sim \text{Exp}(1)$, $X \sim \text{Exp}(0.5)$, $X \sim \text{Exp}(3)$



Definizione

Una variabile aleatoria continua $X \geq 0$ ha **distribuzione Gamma** con parametri $\alpha > 0$ e $\lambda > 0$, e scriviamo $X \sim \text{Gamma}(\alpha, \lambda)$, se

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad x \geq 0.$$

- Se $X \sim \text{Gamma}(\alpha, \lambda)$, il suo valore atteso e varianza sono:

$$E[X] = \frac{\alpha}{\lambda}, \quad \text{var}(X) = \frac{\alpha}{\lambda^2}.$$

- La funzione $\Gamma(\alpha)$ è detta funzione Gamma:

- se α è un numero intero, $\Gamma(\alpha) = (\alpha - 1)!$

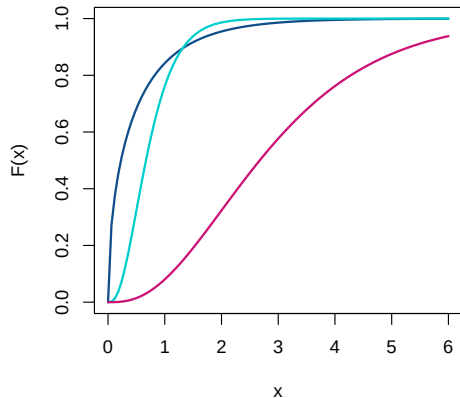
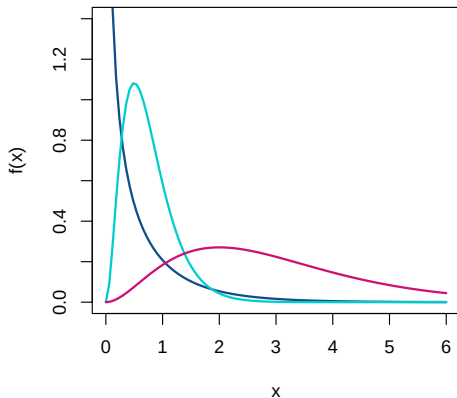
- se α è un numero reale positivo, $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$ (si calcola con un computer)

- La distribuzione esponenziale è un caso particolare: se $X \sim \text{Gamma}(1, \lambda)$, allora $X \sim \text{Exp}(\lambda)$
- Somma di v.a. Gamma indipendenti: Se $X_1, \dots, X_n$ sono variabili aleatorie indipendenti, ognuna con distribuzione $\text{Gamma}(\alpha_i, \lambda)$, allora la loro somma $X = X_1 + \dots + X_n$ segue la distribuzione $X \sim \text{Gamma}(\sum_{i=1}^n \alpha_i, \lambda)$ (segue anche che la somma di v.a. esponenziali indipendenti è gamma).

La distribuzione gamma

Densità e funzione di ripartizione: confronto

Esempio: $X \sim \text{Gamma}(0.5, 1)$, $X \sim \text{Gamma}(3, 4)$, $X \sim \text{Gamma}(3, 1)$



Definizione

Siano $Z_1, \dots, Z_k$ variabili aleatorie indipendenti, ciascuna con distribuzione $N(0, 1)$. La variabile

$$V = Z_1^2 + Z_2^2 + \dots + Z_k^2$$

ha distribuzione **chi-quadro con k gradi di libertà**, e scriviamo $V \sim \chi_k^2$.

- ▶ È un caso particolare della Gamma: $\chi_k^2 = \text{Gamma}\left(\frac{k}{2}, \frac{1}{2}\right)$.
- ▶ Di conseguenza $E[V] = k$ e $\text{var}(V) = 2k$.
- ▶ Il supporto è $(0, +\infty)$ (somma di quadrati).

Questa distribuzione diventerà importante quando stimeremo la media di una popolazione normale con **varianza non nota**.

Definizione

Siano $Z \sim N(0, 1)$ e $V \sim \chi^2_\nu$ **indipendenti**. La variabile

$$T = \frac{Z}{\sqrt{V/\nu}}$$

ha distribuzione **distribuzione t di Student con ν gradi di libertà**, e scriviamo $T \sim t_\nu$.

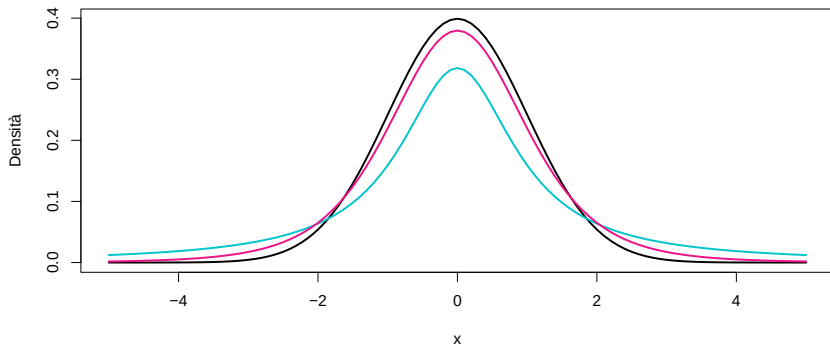
- ▶ Per $\nu \rightarrow \infty$, $V/\nu \rightarrow 1$ e quindi $T \rightarrow Z$: la t_ν tende alla normale standard.
- ▶ Momenti: $E[T] = 0$ per $\nu > 1$, $\text{var}(T) = \frac{\nu}{\nu - 2}$ per $\nu > 2$.

T di Student e normale standard

Come la normale standard, è **simmetrica attorno a 0**, ma ha **code più pesanti**: i valori estremi sono più probabili (quindi le code sono più alte).

- ▶ Per ν piccolo, le code sono molto pesanti;
- ▶ al crescere di ν , t_ν si **avvicina alla normale standard** $N(0, 1)$.

Figura: Normale standard ; t di Student con $\nu = 1$; t di Student con $\nu = 5$



8

Variabili indipendenti e identicamente distribuite

Molto spesso lavoriamo con più variabili $X_1, X_2, \dots, X_n$ che sono:

- **indipendenti**: il valore di una non influenza le altre;
- **identicamente distribuite**: hanno tutte la stessa distribuzione.

Le chiamiamo **i.i.d.** (*indipendenti e identicamente distribuite*).

Per variabili i.i.d. con media μ e varianza σ^2 , le proprietà precedenti danno

$$E\left[\sum_{i=1}^n X_i\right] = n\mu \qquad \text{Var}\left(\sum_{i=1}^n X_i\right) = n\sigma^2$$

Siano $X_1, \dots, X_n$ variabili i.i.d. con $E[X_i] = \mu$ e $\text{var}(X_i) = \sigma^2$.

Definizione

La **media campionaria** è

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

È anch'essa una variabile aleatoria. Usando le proprietà di valore atteso e varianza:

$$E[\bar{X}_n] = \frac{1}{n} n\mu = \mu, \quad \text{var}(\bar{X}_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

- ▶ In media $\bar{X}_n$ è centrata sul valore vero μ ;
- ▶ la sua variabilità **diminuisce** al crescere di n ;
- ▶ se le X_i sono normali, $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ esattamente.

Cosa succede alla media campionaria quando n cresce?

Legge dei grandi numeri

Siano $X_1, X_2, \dots$ i.i.d. con $E[X_i] = \mu$ (finito). Allora, per ogni $\varepsilon > 0$,

$$P(|\bar{X}_n - \mu| > \varepsilon) \longrightarrow 0 \quad \text{per } n \rightarrow \infty.$$

Si dice che $\bar{X}_n$ **converge in probabilità** a μ .

- ▶ Al crescere di n , la media campionaria si concentra attorno a μ (coerente con $\text{var}(\bar{X}_n) = \sigma^2/n \rightarrow 0$).
- ▶ Giustifica il fatto che possiamo usare la media campionaria per stimare la media “vera” incognita.
- ▶ Esempio: la frequenza relativa di teste in n lanci di una moneta si avvicina a p .

Teorema del limite centrale

La legge dei grandi numeri dice *dove* $\bar{X}_n$ va; il TLC dice *come* si distribuisce attorno a μ .

Teorema del limite centrale

Siano $X_1, X_2, \dots$ i.i.d. con $E[X_i] = \mu$ e $\text{var}(X_i) = \sigma^2 \in (0, \infty)$. Allora, per ogni $z \in \mathbb{R}$,

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) \longrightarrow \Phi(z) \quad \text{per } n \rightarrow \infty.$$

Per n abbastanza grande vale quindi, **approssimativamente**,

$$\bar{X}_n \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right).$$

Vale qualunque sia la distribuzione di partenza delle X_i (purché con varianza finita).

Dado equo, $n = 100$ lanci: $\mu = 3.5$, $\sigma \approx 1.71$, quindi $\text{sd}(\bar{X}_{100}) \approx 0.171$.

Vogliamo $P(\bar{X}_{100} > 3.7)$. Per il TLC, $\bar{X}_{100} \sim N(3.5, 0.171^2)$; standardizzando:

$$P(\bar{X}_{100} > 3.7) \sim P\left(Z > \frac{3.7 - 3.5}{0.171}\right) = P(Z > 1.17) = 1 - \Phi(1.17) = 1 - 0.8790 \sim 0.12.$$

Attenzione

Il TLC è un risultato **asintotico**: l'approssimazione è tanto migliore quanto più n è grande e quanto più la distribuzione di partenza è simmetrica. Per distribuzioni molto asimmetriche servono n più grandi.