Probabilità e Processi Stocastici (455AA)

Lezione 6

Dario Trevisan

12/10/2023

Richiami

- Variabili aleatorie $X = \{X = x\}_{x \in E}$

Richiami

- Variabili aleatorie $X = \{X = x\}_{x \in E}$
- Densità discreta/continua

Richiami

- Variabili aleatorie $X = \{X = x\}_{x \in E}$
- Densità discreta/continua
- Variabili composte $g(X)$ e transformazione delle densità

Richiami

- Variabili aleatorie $X = \{X = x\}_{x \in E}$
- Densità discreta/continua
- Variabili composte $g(X)$ e transformazione delle densità
- Variabili congiunte e marginali

Richiami

- Variabili aleatorie $X = \{X = x\}_{x \in E}$
- Densità discreta/continua
- Variabili composte $g(X)$ e transformazione delle densità
- Variabili congiunte e marginali
- Formula di Bayes

$$p(Y = y | X = x) \propto p(Y = y) L(Y = y; X = x)$$

Richiami

- Variabili aleatorie $X = \{X = x\}_{x \in E}$
- Densità discreta/continua
- Variabili composte $g(X)$ e transformazione delle densità
- Variabili congiunte e marginali
- Formula di Bayes

$$p(Y = y|X = x) \propto p(Y = y)L(Y = y; X = x)$$

- Indipendenza (anche per più di due variabili)

$$p(X = x|Y = y) = p(X = x)$$

Section 1

Reti bayesiane

Reti Bayesiane

Presentiamo un metodo grafico per rappresentare le densità di variabili aleatorie (una specie di estensione dei diagrammi ad albero per variabili invece di eventi)

- Passo zero: si fissa un ordinamento tra le variabili (che corrisponde all'ordine in cui i sistemi di alternative vengono aggiunti nella costruzione del grafo ad albero): X_1, X_2, $\ldots$ X_k

Il diagramma è un grafo orientato su k nodi corrispondenti alle k variabili, viene costruito in k passi:

- nel primo passo si introduce solamente il nodo corrispondente alla variabile X_1;

Il diagramma è un grafo orientato su k nodi corrispondenti alle k variabili, viene costruito in k passi:

- nel primo passo si introduce solamente il nodo corrispondente alla variabile X_1;
- nel passo i-esimo, si introduce il nodo corrispondente alla variabile X_i, e si considera la densità di X_i condizionata a tutte le variabili inserite $X_1, \ldots X_{i-1}$,

$$P(X_i = x_i | I, X_{i-1} = x_{i-1}, \ldots, X_1 = x_1).$$

Il diagramma è un grafo orientato su k nodi corrispondenti alle k variabili, viene costruito in k passi:

- nel primo passo si introduce solamente il nodo corrispondente alla variabile X_1;
- nel passo i-esimo, si introduce il nodo corrispondente alla variabile X_i, e si considera la densità di X_i condizionata a tutte le variabili inserite $X_1, \ldots\ X_{i-1}$,

$$P(X_i = x_i | I, X_{i-1} = x_{i-1}, \ldots, X_1 = x_1).$$

- si individua un sottoinsieme $J \subseteq \{1, \ldots, i-1\}$ tale che la densità dipenda solo dalle $(X_j)_{j \in J}$, ossia, per ogni $x_1, \ldots, x_i$, valga

$$\begin{aligned} &P(X_i = x_i | I, X_{i-1} = x_{i-1}, \ldots, X_1 = x_1) \\ &\quad = P(X_i = x_i | I, X_j = x_j \text{ per ogni } j \in J). \end{aligned}$$

Il diagramma è un grafo orientato su k nodi corrispondenti alle k variabili, viene costruito in k passi:

- nel primo passo si introduce solamente il nodo corrispondente alla variabile X_1;
- nel passo i-esimo, si introduce il nodo corrispondente alla variabile X_i, e si considera la densità di X_i condizionata a tutte le variabili inserite $X_1, \ldots X_{i-1}$,

$$P(X_i = x_i | I, X_{i-1} = x_{i-1}, \ldots, X_1 = x_1).$$

- si individua un sottoinsieme $J \subseteq \{1, \ldots, i-1\}$ tale che la densità dipenda solo dalle $(X_j)_{j \in J}$, ossia, per ogni $x_1, \ldots, x_i$, valga

$$\begin{aligned} P(X_i = x_i | I, X_{i-1} = x_{i-1}, \ldots, X_1 = x_1) \\ = P(X_i = x_i | I, X_j = x_j \text{ per ogni } j \in J). \end{aligned}$$

- si inseriscono gli archi orientati (frecce) da ciascun nodo corrispondente alle variabili X_j, $j \in J$, verso il nodo corrispondente ad X_i.

Date k variabili aleatorie indipendenti $X_1, \ldots\ X_k$, l'algoritmo produce il diagramma:

- Si consideri una variabile aleatoria Λ tale che, condizionatamente ad essa, le variabili $T_1, \ldots, T_k$ sono indipendenti (un esempio concreto è $\Lambda = \lambda$ individua il parametro delle variabili T_i che hanno legge esponenziale).

- Si consideri una variabile aleatoria Λ tale che, condizionatamente ad essa, le variabili $T_1, \ldots, T_k$ sono indipendenti (un esempio concreto è $\Lambda = \lambda$ individua il parametro delle variabili T_i che hanno legge esponenziale).
- La densità congiunta ha la forma

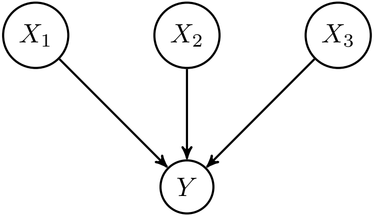
$$P(\Lambda, T_1, T_2, T_3, T_4) = P(\Lambda)P(T_1|\Lambda)P(T_2|\Lambda)P(T_3|\Lambda)P(T_4|\Lambda).$$

- Si consideri una variabile aleatoria Λ tale che, condizionatamente ad essa, le variabili $T_1, \ldots, T_k$ sono indipendenti (un esempio concreto è $\Lambda = \lambda$ individua il parametro delle variabili T_i che hanno legge esponenziale).
- La densità congiunta ha la forma

$$P(\Lambda, T_1, T_2, T_3, T_4) = P(\Lambda)P(T_1|\Lambda)P(T_2|\Lambda)P(T_3|\Lambda)P(T_4|\Lambda).$$

- La rete bayesiana costruita inserendo prima la variabile Λ e poi le rimanenti è la seguente:

Si considerino k variabili X_1, ..., X_k indipendenti tra loro (rispetto all'informazione iniziale) e sia $Y = g(X_1, \ldots, X_k)$ (ad esempio $Y = X_1 + \ldots + X_k$ nel caso di variabili a valori in $\mathbb{R}$). La rete bayesiana è rappresentata in figura.



Si consideri la rete bayesiana rappresentata in figura. Condizionando rispetto ad Y, si ottiene che X e Z sono indipendenti. Questo è un semplice esempio di *catena di Markov*.

$$X \longrightarrow Y \longrightarrow Z$$

Section 2

Cenni ai metodi numerici

Dalla teoria alla pratica

- In teoria abbiamo tutti gli strumenti per affrontare problemi concreti:

Dalla teoria alla pratica

- In teoria abbiamo tutti gli strumenti per affrontare problemi concreti:
  a. Si definisce un *modello* con parametri Θ e quantità osservabili X (una opportuna rete bayesiana)

Dalla teoria alla pratica

- In teoria abbiamo tutti gli strumenti per affrontare problemi concreti:
  a. Si definisce un *modello* con parametri Θ e quantità osservabili X (una opportuna rete bayesiana)
  b. Si stabiliscono probabilità *a priori* per Θ (ad es. uniformi) e *verosimiglianza* $L(\Theta = \theta; X = x) = P(X = x | \Theta = \theta)$

Dalla teoria alla pratica

- In teoria abbiamo tutti gli strumenti per affrontare problemi concreti:
  a. Si definisce un *modello* con parametri Θ e quantità osservabili X (una opportuna rete bayesiana)
  b. Si stabiliscono probabilità *a priori* per Θ (ad es. uniformi) e *verosimiglianza* $L(\Theta = \theta; X = x) = P(X = x | \Theta = \theta)$
  c. Si aggiorna il modello incorporando i dati *osservati* (Bayes)

Dalla teoria alla pratica

- In teoria abbiamo tutti gli strumenti per affrontare problemi concreti:
  a. Si definisce un *modello* con parametri Θ e quantità osservabili X (una opportuna rete bayesiana)
  b. Si stabiliscono probabilità *a priori* per Θ (ad es. uniformi) e *verosimiglianza* $L(\Theta = \theta; X = x) = P(X = x | \Theta = \theta)$
  c. Si aggiorna il modello incorporando i dati *osservati* (Bayes)
  d. Si stimano i parametri Θ (ad es. mediante MAP o MLE)

Dalla teoria alla pratica

- In teoria abbiamo tutti gli strumenti per affrontare problemi concreti:
  a. Si definisce un *modello* con parametri Θ e quantità osservabili X (una opportuna rete bayesiana)
  b. Si stabiliscono probabilità *a priori* per Θ (ad es. uniformi) e *verosimiglianza* $L(\Theta = \theta; X = x) = P(X = x | \Theta = \theta)$
  c. Si aggiorna il modello incorporando i dati *osservati* (Bayes)
  d. Si stimano i parametri Θ (ad es. mediante MAP o MLE)
  e. Si fanno *previsioni* su eventuali variabili non osservate usando

Dalla teoria alla pratica

- In teoria abbiamo tutti gli strumenti per affrontare problemi concreti:
  a. Si definisce un *modello* con parametri Θ e quantità osservabili X (una opportuna rete bayesiana)
  b. Si stabiliscono probabilità *a priori* per Θ (ad es. uniformi) e *verosimiglianza* $L(\Theta = \theta; X = x) = P(X = x | \Theta = \theta)$
  c. Si aggiorna il modello incorporando i dati *osservati* (Bayes)
  d. Si stimano i parametri Θ (ad es. mediante MAP o MLE)
  e. Si fanno *previsioni* su eventuali variabili non osservate usando
- Il problema è che nella pratica la numerosità sia dei dati osservati sia dei parametri è elevata e questo si richiede di calcolare densità di probabilità su spazi di *dimensione elevata*, che diventa impraticabile analiticamente.

Densità coniugate

Per certe funzioni di verosimiglianza, si possono introdurre delle densità a priori (*coniugate*) particolarmente convenienti perché la densità a posteriori è nella stessa classe.

- In n esperimenti indipendenti, ciascuno con probabilità di successo $Y = y$, il numero X di successi ha una densità binomiale

$$L(Y = y; X = k) = P(X = k | Y = y) = \binom{n}{k} y^k (1-y)^{n-k}.$$

Densità coniugate

Per certe funzioni di verosimiglianza, si possono introdurre delle densità a priori (*coniugate*) particolarmente convenienti perché la densità a posteriori è nella stessa classe.

- In n esperimenti indipendenti, ciascuno con probabilità di successo $Y = y$, il numero X di successi ha una densità binomiale

$$L(Y = y; X = k) = P(X = k | Y = y) = \binom{n}{k} y^k (1-y)^{n-k}.$$

- Se la densità di Y a priori è *Beta* di parametri α, $\beta > 0$, ossia

$$p(Y = y) \propto y^{\alpha-1}(1-y)^{\beta-1},$$

  allora la densità a posteriori avendo osservato $X = k$ successi è ancora Beta:

$$p(Y = y | X = k) \propto y^{\alpha+k-1}(1-y)^{\beta+n-k-1}.$$

Problema

Una routine probabilistica fornisce l'output desiderato con probabilità $p \in [0, 1]$ non nota. Ogni applicazione dell'algoritmo produce esiti tra di loro indipendenti. Si modellizza inizialmente p come una variabile uniforme su $[0, 1]$.

- Dopo aver effettuato 100 esperimenti, avendo osservato che l'output è corretto su 60 di questi, scrivere la densità a posteriori per p e riconoscere una densità beta. Determinare la stima di massimo a posteriori.

Problema

Una routine probabilistica fornisce l'output desiderato con probabilità $p \in [0, 1]$ non nota. Ogni applicazione dell'algoritmo produce esiti tra di loro indipendenti. Si modellizza inizialmente p come una variabile uniforme su $[0, 1]$.

- Dopo aver effettuato 100 esperimenti, avendo osservato che l'output è corretto su 60 di questi, scrivere la densità a posteriori per p e riconoscere una densità beta. Determinare la stima di massimo a posteriori.
- Come cambia la risposta se invece si effettuano 1000 esperimenti e si osserva un output corretto su 600 di questi?

Tecniche numeriche

Le densità coniugate non risolvono completamente il problema, in particolare se la numerosità delle osservazioni è molto grande.

Si ricorre a metodi numerici. Due approcci fondamentali:

- Approssimare tutta la densità a posteriori di Y

Tecniche numeriche

Le densità coniugate non risolvono completamente il problema, in particolare se la numerosità delle osservazioni è molto grande.

Si ricorre a metodi numerici. Due approcci fondamentali:

- Approssimare tutta la densità a posteriori di Y
- Determinare solamente la stima di massima verosimiglianza y_{MAP}

Integrazione numerica

Nel primo caso il problema è quindi di approssimare numericamente un integrale

$$P(Y \in U|I) = \int_U p(Y = y|I)dy.$$

- Negli esempi seguiamo un approccio elementare: calcolare $p(Y = y|I)$ su una griglia

Integrazione numerica

Nel primo caso il problema è quindi di approssimare numericamente un integrale

$$P(Y \in U|I) = \int_U p(Y = y|I)dy.$$

- Negli esempi seguiamo un approccio elementare: calcolare $p(Y = y|I)$ su una griglia
- Problema: se Y è a valori in $[0,1]^d \subseteq \mathbb{R}^d$ e il passo è δy, si devono calcolare circa

$$\frac{1}{\delta^d}$$

  valori (e d è molto grande).

Integrazione numerica

Nel primo caso il problema è quindi di approssimare numericamente un integrale

$$P(Y \in U|I) = \int_U p(Y = y|I)dy.$$

- Negli esempi seguiamo un approccio elementare: calcolare $p(Y = y|I)$ su una griglia
- Problema: se Y è a valori in $[0,1]^d \subseteq \mathbb{R}^d$ e il passo è δy, si devono calcolare circa

  $$\frac{1}{\delta^d}$$

  valori (e d è molto grande).
- Approcci alternativi cercano di ottimizzare la scelta dei punti, anche mediante scelte (pseudo)-*casuali* (metodi Monte Carlo).

Metodi di ottimizzazione

Nel secondo caso il problema è di determinare il punto di massimo di una funzione

$$y_{\mathrm{MLE}} = \arg\max \left\{ L(Y = y; X = x) \right\}.$$

- Il problema generale è studiatissimo e vi sono molteplici algoritmi (Newton, ascesa gradiente...)

Metodi di ottimizzazione

Nel secondo caso il problema è di determinare il punto di massimo di una funzione

$$y_{\mathrm{MLE}} = \arg\max \{L(Y = y; X = x)\}.$$

- Il problema generale è studiatissimo e vi sono molteplici algoritmi (Newton, ascesa gradiente...)
- In R le funzioni `nlm()` e `optim()` permettono di applicare i principali.

Metodi di ottimizzazione

Nel secondo caso il problema è di determinare il punto di massimo di una funzione

$$y_{\mathrm{MLE}} = \arg\max \{L(Y = y; X = x)\} .$$

- Il problema generale è studiatissimo e vi sono molteplici algoritmi (Newton, ascesa gradiente...)
- In R le funzioni `nlm()` e `optim()` permettono di applicare i principali.
- Se la numerosità delle osservazioni X è troppo elevata, si ricorre a metodi probabilistici.

Metodi di ottimizzazione

Nel secondo caso il problema è di determinare il punto di massimo di una funzione

$$y_{\mathrm{MLE}} = \arg\max \left\{ L(Y = y; X = x) \right\}.$$

- Il problema generale è studiatissimo e vi sono molteplici algoritmi (Newton, ascesa gradiente. . .)
- In R le funzioni `nlm()` e `optim()` permettono di applicare i principali.
- Se la numerosità delle osservazioni X è troppo elevata, si ricorre a metodi probabilistici.
- A volte $Y = (Y_{\mathrm{par}}, Y_{\mathrm{hid}})$ e si richiede di massimizzare solo la verosimiglianza di Y_{par}. Necessità di ricorrere a metodi "misti" (ad esempio Expectation-Maximization).

Problema

Il numero di clienti che entrano in un negozio in un dato giorno è rappresentato da una variabile con densità Poisson di parametro $\Lambda > 0$ (inizialmente non noto). Si vuole stimare λ osservando gli ingressi in tre giorni consecutivi. Si suppone che a giorni diversi corrispondano numeri di ingressi indipendenti tra loro.

- Supponendo di avere osservato in sequenza $X_1 = 4$, $X_2 = 5$ e 3 ingressi nei tre giorni, fornire una stima di massima verosimiglianza per λ.

Problema

Il numero di clienti che entrano in un negozio in un dato giorno è rappresentato da una variabile con densità Poisson di parametro $\Lambda > 0$ (inizialmente non noto). Si vuole stimare λ osservando gli ingressi in tre giorni consecutivi. Si suppone che a giorni diversi corrispondano numeri di ingressi indipendenti tra loro.

- Supponendo di avere osservato in sequenza $X_1 = 4$, $X_2 = 5$ e 3 ingressi nei tre giorni, fornire una stima di massima verosimiglianza per λ.
- Supponendo che Λ sia a priori distribuito come una variabile esponenziale di parametro 1, avendo osservato la stessa sequenza 4, 5, 3 di ingressi, fornire la densità a posteriori e la stima di massimo a posteriori.

Problema

Il numero di clienti che entrano in un negozio in un dato giorno è rappresentato da una variabile con densità Poisson di parametro $\Lambda > 0$ (inizialmente non noto). Si vuole stimare λ osservando gli ingressi in tre giorni consecutivi. Si suppone che a giorni diversi corrispondano numeri di ingressi indipendenti tra loro.

- Supponendo di avere osservato in sequenza $X_1 = 4$, $X_2 = 5$ e 3 ingressi nei tre giorni, fornire una stima di massima verosimiglianza per λ.
- Supponendo che Λ sia a priori distribuito come una variabile esponenziale di parametro 1, avendo osservato la stessa sequenza 4, 5, 3 di ingressi, fornire la densità a posteriori e la stima di massimo a posteriori.
- Come cambiano le risposte se invece si osservano in sequenza $X_1 \leq 4$, $X_2 \leq 5$ e $X_3 \leq 3$ ingressi?

Section 3

Indicatori caratteristici

Come riassumere la legge di una variabile?

Indicatori di *posizione*, di *dispersione* e di *correlazione*.

- Funzione di ripartizione (CDF), funzione di sopravvivenza

Come riassumere la legge di una variabile?

Indicatori di *posizione*, di *dispersione* e di *correlazione*.

- Funzione di ripartizione (CDF), funzione di sopravvivenza
- Funzione quantile, mediana (quartili, decili, ecc.)

Come riassumere la legge di una variabile?

Indicatori di *posizione*, di *dispersione* e di *correlazione*.

- Funzione di ripartizione (CDF), funzione di sopravvivenza
- Funzione quantile, mediana (quartili, decili, ecc.)
- Valor medio

Come riassumere la legge di una variabile?

Indicatori di *posizione*, di *dispersione* e di *correlazione*.

- Funzione di ripartizione (CDF), funzione di sopravvivenza
- Funzione quantile, mediana (quartili, decili, ecc.)
- Valor medio
- Varianza e deviazione standard

Come riassumere la legge di una variabile?

Indicatori di *posizione*, di *dispersione* e di *correlazione*.

- Funzione di ripartizione (CDF), funzione di sopravvivenza
- Funzione quantile, mediana (quartili, decili, ecc.)
- Valor medio
- Varianza e deviazione standard
- Covarianza e correlazione.

Come riassumere la legge di una variabile?

Indicatori di *posizione*, di *dispersione* e di *correlazione*.

- Funzione di ripartizione (CDF), funzione di sopravvivenza
- Funzione quantile, mediana (quartili, decili, ecc.)
- Valor medio
- Varianza e deviazione standard
- Covarianza e correlazione.
- Momenti e funzione generatrice dei momenti

Come riassumere la legge di una variabile?

Indicatori di *posizione*, di *dispersione* e di *correlazione*.

- Funzione di ripartizione (CDF), funzione di sopravvivenza
- Funzione quantile, mediana (quartili, decili, ecc.)
- Valor medio
- Varianza e deviazione standard
- Covarianza e correlazione.
- Momenti e funzione generatrice dei momenti
- Funzione caratteristica

Come riassumere la legge di una variabile?

Indicatori di *posizione*, di *dispersione* e di *correlazione*.

- Funzione di ripartizione (CDF), funzione di sopravvivenza
- Funzione quantile, mediana (quartili, decili, ecc.)
- Valor medio
- Varianza e deviazione standard
- Covarianza e correlazione.
- Momenti e funzione generatrice dei momenti
- Funzione caratteristica
- Cenni all'entropia

Funzione di ripartizione e di sopravvivenza

Data X a valori in $\mathbb{R}$, si è spesso interessati a conoscere la probabilità che X assuma valori "grandi".

- Esempio: la quantità di denaro che un investitore potrebbe guadagnare (se positiva) o perdere (se negativa) in una fissata data futura, a seconda dell'andamento del mercato

Definiamo

Funzione di ripartizione e di sopravvivenza

Data X a valori in $\mathbb{R}$, si è spesso interessati a conoscere la probabilità che X assuma valori "grandi".

- Esempio: la quantità di denaro che un investitore potrebbe guadagnare (se positiva) o perdere (se negativa) in una fissata data futura, a seconda dell'andamento del mercato

Definiamo

1. la **funzione di ripartizione** (o funzione cumulativa, in inglese *cumulative distribution function*, CDF,) di X

$$\mathbb{R} \ni x \mapsto \mathrm{CDF}_X(x) = P(X \leq x).$$

Funzione di ripartizione e di sopravvivenza

Data X a valori in $\mathbb{R}$, si è spesso interessati a conoscere la probabilità che X assuma valori "grandi".

- Esempio: la quantità di denaro che un investitore potrebbe guadagnare (se positiva) o perdere (se negativa) in una fissata data futura, a seconda dell'andamento del mercato

Definiamo

1. la **funzione di ripartizione** (o funzione cumulativa, in inglese *cumulative distribution function*, CDF,) di X

$$\mathbb{R} \ni x \mapsto \mathrm{CDF}_X(x) = P(X \leq x).$$

2. la **funzione di sopravvivenza** (in inglese *survival function*) di X,

$$\mathbb{R} \ni x \mapsto \mathrm{SUR}_X(x) = P(X > x).$$

Come calcolarle

- Essendo $\{X \leq x\}$ e $\{X > x\}$ un sistema di alternative,

$$\mathrm{CDF}_X(x) + \mathrm{SUR}_X(x) = 1.$$

Come calcolarle

- Essendo $\{X \leq x\}$ e $\{X > x\}$ un sistema di alternative,
$$\mathrm{CDF}_X(x) + \mathrm{SUR}_X(x) = 1.$$

- Se la densità (discreta o continua) di X è nota, siccome $P(X \leq x) = P(X \in (-\infty, x])$, si trova
$$\mathrm{CDF}_X(x) = \begin{cases} \sum_{z \leq x} P(X = z) & \text{se } X \text{ ha densità discreta,} \\ \int_{-\infty}^{x} f(z) dz & \text{se } X \text{ ha densità continua.} \end{cases}$$

Come calcolarle

- Essendo $\{X \leq x\}$ e $\{X > x\}$ un sistema di alternative,

$$\mathrm{CDF}_X(x) + \mathrm{SUR}_X(x) = 1.$$

- Se la densità (discreta o continua) di X è nota, siccome $P(X \leq x) = P(X \in (-\infty, x])$, si trova

$$\mathrm{CDF}_X(x) = \begin{cases} \sum_{z \leq x} P(X = z) & \text{se } X \text{ ha densità discreta,} \\ \int_{-\infty}^{x} f(z)dz & \text{se } X \text{ ha densità continua.} \end{cases}$$

- Possiamo quindi intepretare (almeno nel caso continuo) la $\mathrm{CDF}_X(x)$ come l'area del sottografico della densità da $-\infty$ fino ad x.

Come calcolarle

- Essendo $\{X \leq x\}$ e $\{X > x\}$ un sistema di alternative,

$$\mathrm{CDF}_X(x) + \mathrm{SUR}_X(x) = 1.$$

- Se la densità (discreta o continua) di X è nota, siccome $P(X \leq x) = P(X \in (-\infty, x])$, si trova

$$\mathrm{CDF}_X(x) = \begin{cases} \sum_{z \leq x} P(X = z) & \text{se } X \text{ ha densità discreta,} \\ \int_{-\infty}^{x} f(z) dz & \text{se } X \text{ ha densità continua.} \end{cases}$$

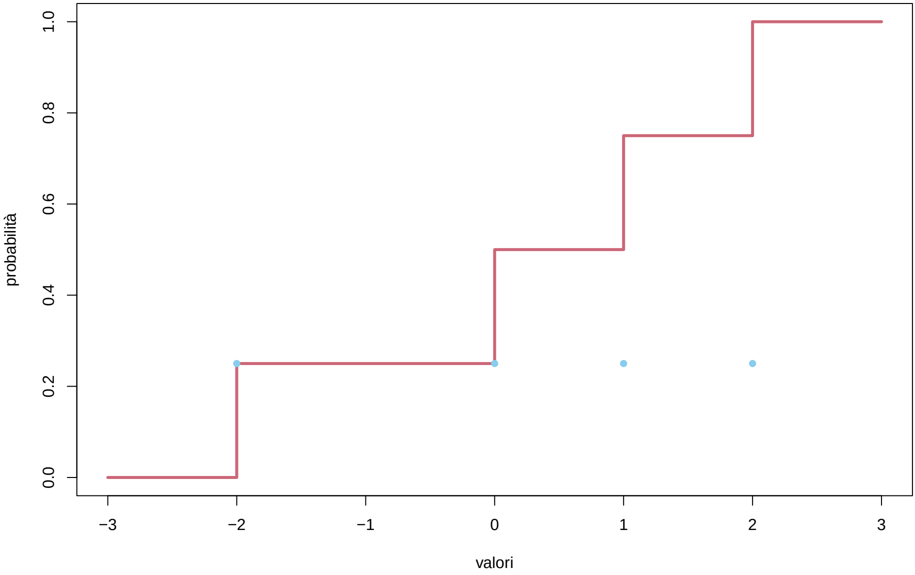
- Possiamo quindi intepretare (almeno nel caso continuo) la $\mathrm{CDF}_X(x)$ come l'area del sottografico della densità da $-\infty$ fino ad x.

- Per la funzione di sopravvivenza,

$$\mathrm{SUR}_X(x) = \begin{cases} \sum_{z > x} P(X = z) & \text{se } X \text{ ha densità discreta,} \\ \int_{x}^{+\infty} f(z) dz & \text{se } X \text{ ha densità continua,} \end{cases}$$

Esempio nel caso discreto:

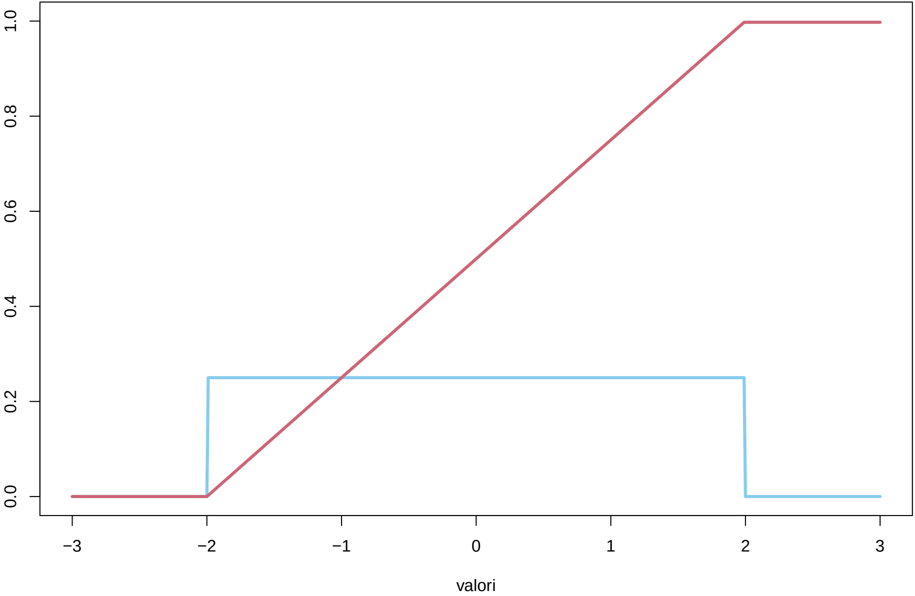
- Sia X uniforme discreta sui valori $E = \{-2, 1, 0, 2\}$.


probabilità
valori
0.0
0.2
0.4
0.6
0.8
1.0
−3
−2
−1
0
1
2
3

Esempio nel caso continuo

Si consideri una variabile aleatoria X avente densità uniforme continua nell'intervallo $[-2, 2]$, ossia per $x \in [-2, 2]$,

$$p(X = x) = 1/4.$$


1.0
0.8
0.6
0.4
0.2
0.0
−3
−2
−1
0
1
2
3
valori

Proprietà

1. vale $\mathrm{CDF}_X(x) \in [0,1]$, essendo una probabilità.

Proprietà

1. vale $\mathrm{CDF}_X(x) \in [0,1]$, essendo una probabilità.
2. la funzione $x \mapsto \mathrm{CDF}_X(x)$ è crescente (ma non strettamente): se $x < z$, per la monotonia della probabilità,

$$\mathrm{CDF}_X(x) \leq \mathrm{CDF}_X(z)$$

Proprietà

1. vale $\mathrm{CDF}_X(x) \in [0,1]$, essendo una probabilità.
2. la funzione $x \mapsto \mathrm{CDF}_X(x)$ è crescente (ma non strettamente): se $x < z$, per la monotonia della probabilità,

$$\mathrm{CDF}_X(x) \leq \mathrm{CDF}_X(z)$$

3. vale $\mathrm{CDF}_X(-\infty) = 0$ e $\mathrm{CDF}_X(+\infty) = 1$ (nel senso di limiti opportuni):

Proprietà

1. vale $\mathrm{CDF}_X(x) \in [0,1]$, essendo una probabilità.
2. la funzione $x \mapsto \mathrm{CDF}_X(x)$ è crescente (ma non strettamente): se $x < z$, per la monotonia della probabilità,

$$\mathrm{CDF}_X(x) \leq \mathrm{CDF}_X(z)$$

3. vale $\mathrm{CDF}_X(-\infty) = 0$ e $\mathrm{CDF}_X(+\infty) = 1$ (nel senso di limiti opportuni):
4. Nel caso di densità discreta, la CDF_X è una funzione costante a tratti, mentre nel caso di variabili con densità continua, la CDF_X è una funzione continua.

Proprietà

1. vale $\mathrm{CDF}_X(x) \in [0,1]$, essendo una probabilità.
2. la funzione $x \mapsto \mathrm{CDF}_X(x)$ è crescente (ma non strettamente): se $x < z$, per la monotonia della probabilità,
$$\mathrm{CDF}_X(x) \leq \mathrm{CDF}_X(z)$$
3. vale $\mathrm{CDF}_X(-\infty) = 0$ e $\mathrm{CDF}_X(+\infty) = 1$ (nel senso di limiti opportuni):
4. Nel caso di densità discreta, la CDF_X è una funzione costante a tratti, mentre nel caso di variabili con densità continua, la CDF_X è una funzione continua.
5. Per la funzione SUR, valgono proprietà analoghe: la funzione è decrescente e vale $\mathrm{SUR}_X(-\infty) = 1$ mentre $\mathrm{SUR}_X(+\infty) = 0$.

Un ulteriore esempio

Si consideri una variabile aleatoria X con densità esponenziale di parametro $\lambda > 0$.

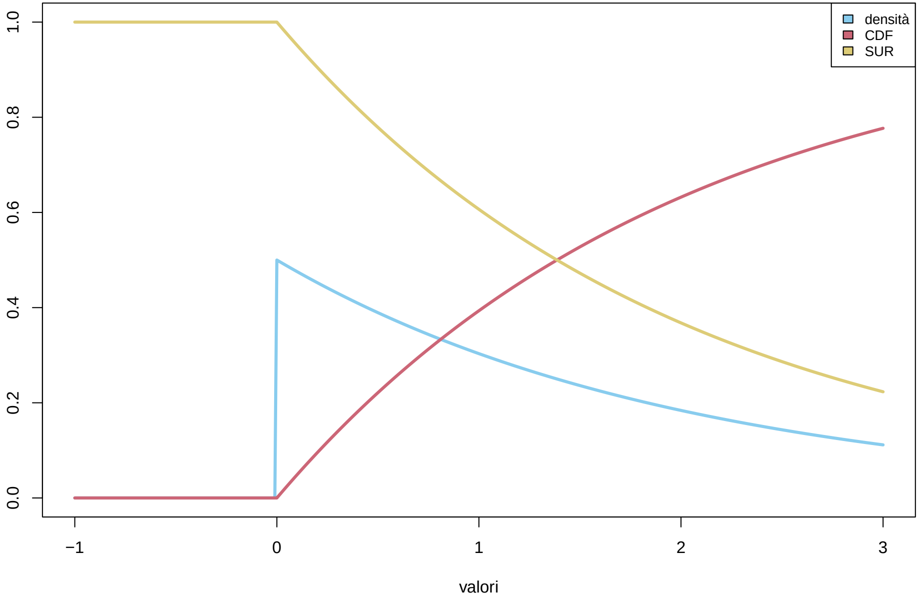


Figura 2: densità, CDF, e SUR, di una variabile X con densità esponenziale di

Dalla CDF alla densità

- Nel caso discreto, la densità è non nulla solo nei valori $x \in \mathbb{R}$ in cui la $\mathrm{CDF}_X(x)$ ha un salto, e il valore della densità in quel punto è proprio *l'ampiezza del salto*.

Dalla CDF alla densità

- Nel caso discreto, la densità è non nulla solo nei valori $x \in \mathbb{R}$ in cui la $\mathrm{CDF}_X(x)$ ha un salto, e il valore della densità in quel punto è proprio *l'ampiezza del salto*.

- Nel caso di densità continua, per invertire la formula
$$\int_{-\infty}^{x} p(X = z)dz = \mathrm{CDF}_X(x)$$
è sufficiente applicare il teorema fondamentale del calcolo integrale, e quindi derivare la CDF_X:
$$\frac{d}{dx}\mathrm{CDF}_X(x) = p(X = x).$$
(nei punti in cui CDF_X è derivabile)

Dalla CDF alla densità

- Nel caso discreto, la densità è non nulla solo nei valori $x \in \mathbb{R}$ in cui la $\mathrm{CDF}_X(x)$ ha un salto, e il valore della densità in quel punto è proprio *l'ampiezza del salto*.

- Nel caso di densità continua, per invertire la formula
$$\int_{-\infty}^{x} p(X = z)dz = \mathrm{CDF}_X(x)$$
è sufficiente applicare il teorema fondamentale del calcolo integrale, e quindi derivare la CDF_X:
$$\frac{d}{dx}\mathrm{CDF}_X(x) = p(X = x).$$
(nei punti in cui CDF_X è derivabile)

- Per la SUR_X, è sufficiente cambiare di segno alle quantità: nel caso continuo, si ottiene
$$-\frac{d}{dx}\mathrm{SUR}_X(x) = f(x).$$

Mediana

- La *mediana* di X è definita come un valore $\bar{x} \in \mathbb{R}$, se esiste, tale che

$$\mathrm{CDF}_X(\bar{x}) = \frac{1}{2}.$$

Mediana

- La *mediana* di X è definita come un valore $\bar{x} \in \mathbb{R}$, se esiste, tale che

$$\mathrm{CDF}_X(\bar{x}) = \frac{1}{2}.$$

- Significa che

$$P(X \leq \bar{x}) = P(X > \bar{x}) = \frac{1}{2},$$

  ossia $\bar{x}$ è scelta in modo che le due alternative $\{X \leq \bar{x}\}$ e $\{X > \bar{x}\}$ abbiano la stessa probabilità.

Mediana

- La *mediana* di X è definita come un valore $\bar{x} \in \mathbb{R}$, se esiste, tale che

$$\mathrm{CDF}_X(\bar{x}) = \frac{1}{2}.$$

- Significa che

$$P(X \leq \bar{x}) = P(X > \bar{x}) = \frac{1}{2},$$

  ossia $\bar{x}$ è scelta in modo che le due alternative $\{X \leq \bar{x}\}$ e $\{X > \bar{x}\}$ abbiano la stessa probabilità.

- La mediana è quindi un indicatore di "centralità" per X, alternativo alla moda.

Esempi

- se la densità di X è uniforme (diciamo su un insieme E con un numero pari di valori), significa che metà dei valori possibili di X sono $\leq \bar{x}$, mentre i rimanenti sono $> \bar{x}$. In questo caso invece una moda è uno qualsiasi dei valori di E.

Esempi

- se la densità di X è uniforme (diciamo su un insieme E con un numero pari di valori), significa che metà dei valori possibili di X sono $\leq \bar{x}$, mentre i rimanenti sono $> \bar{x}$. In questo caso invece una moda è uno qualsiasi dei valori di E.
- Sia X uniforme su

$$E = \{0, 0.1, 0.15, 0.3, 0.4, 0.7, 0.73, 0.9, 0.95, 1.1\}$$

Una mediana per X è $\bar{x} = 0.4$, ma anche un qualsiasi valore compreso tra 0.4 e 0.7 (escluso)

Esempi

- se la densità di X è uniforme (diciamo su un insieme E con un numero pari di valori), significa che metà dei valori possibili di X sono $\leq \bar{x}$, mentre i rimanenti sono $> \bar{x}$. In questo caso invece una moda è uno qualsiasi dei valori di E.

- Sia X uniforme su

$$E = \{0, 0.1, 0.15, 0.3, 0.4, 0.7, 0.73, 0.9, 0.95, 1.1\}$$

Una mediana per X è $\bar{x} = 0.4$, ma anche un qualsiasi valore compreso tra 0.4 e 0.7 (escluso)

- Nel caso di una variabile esponenziale di parametro λ,

$$1 - e^{-\lambda \bar{x}} = \frac{1}{2} \quad \leftrightarrow \quad \lambda \bar{x} = \log(2),$$

ossia

$$\bar{x} = \frac{\log(2)}{\lambda}.$$

Funzione quantile

Problemi della mediana (così definita) - non necessariamente la mediana esiste sempre - non è unicamente determinata - non è facile calcolarla

Per risolvere i primi due, si introduce una inversa generalizzata CDF_X, detta funzione *quantile* di X

$$q_X : (0,1) \to \mathbb{R}$$

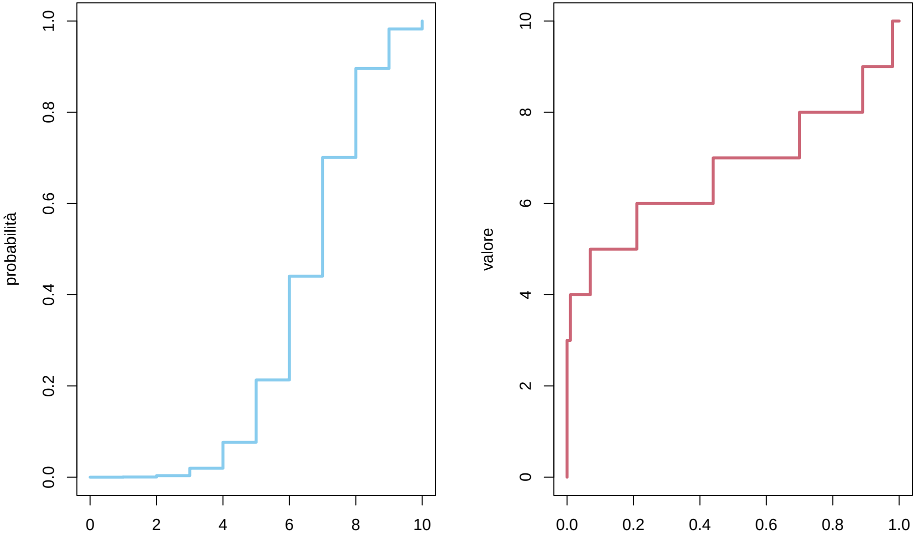
che ad ogni *livello* $\alpha \in (0,1)$ associa

$$q_X(\alpha) = \min \{ x \in \mathbb{R} \, : \, \mathrm{CDF}_X(x) \geq \alpha \} \, .$$

- se X ha densità continua, allora

$$\mathrm{CDF}_X(q_X(\alpha)) = \alpha.$$

Un esempio discreto



Mediana, quartili, decili ecc.

Si definiscono quindi

- la mediana di X come il valore $\bar{x} = q_X(1/2)$.

Mediana, quartili, decili ecc.

Si definiscono quindi

- la mediana di X come il valore $\bar{x} = q_X(1/2)$.
- i *quartili*, $q_X(\alpha)$ corrispondenti ad $\alpha \in \{1/4, 2/4, 3/4\}$ (detti il primo, secondo e terzo quartile),

Mediana, quartili, decili ecc.

Si definiscono quindi

- la mediana di X come il valore $\bar{x} = q_X(1/2)$.
- i *quartili*, $q_X(\alpha)$ corrispondenti ad $\alpha \in \{1/4, 2/4, 3/4\}$ (detti il primo, secondo e terzo quartile),
- i *decili* $q_X(\alpha)$ per $\alpha = k/10$,

Mediana, quartili, decili ecc.

Si definiscono quindi

- la mediana di X come il valore $\bar{x} = q_X(1/2)$.
- i *quartili*, $q_X(\alpha)$ corrispondenti ad $\alpha \in \{1/4, 2/4, 3/4\}$ (detti il primo, secondo e terzo quartile),
- i *decili* $q_X(\alpha)$ per $\alpha = k/10$,
- i *percentili* $q_X(\alpha)$ per $\alpha = k/100$