Probabilità e Statistica

Isadora Antoniano-Villalobos
isadora.antoniano@unive.it

Laurea in Informatica
(Data science/ Tecnologie e scienze dell'informazione)
Università Ca' Foscari di Venezia

Anno academico 2024/2025

Variabili Casuali Continue
Testo Capitoli 7-10

Variabili aleatorie continue

Variabile aleatoria continua

Una **variabile aleatoria continua** X assume valori in un insieme continuo di punti (un sottoinsieme di $\mathbb{R}$ non numerabile).

Un modello probabilistico per X è un'assegnazione di probabilità ad ogni sottoinsieme di suoi possibili valori:

$$\mathbb{P}\left[X \in A\right] = \text{ area su A sottesa ad una curva.}$$

La curva è il grafico di una funzione $f(x)$ tale che:

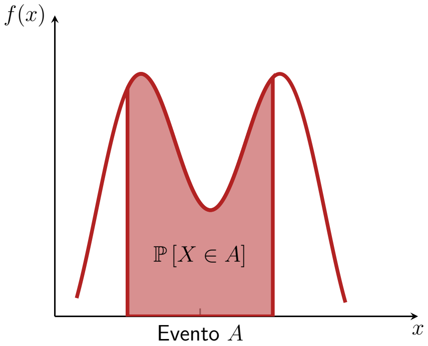
1. $f(x) \geq 0 \,, \quad \forall x \in \mathbb{R}$;
2. $\int_{\mathbb{R}} f(x)\, dx = 1$, cioè l'area totale sotto il grafico di $f(x)$ è 1.

Densità di probabilità

Una funzione $f(x)$ con le proprietà precedenti viene chiamata **densità di probabilità**.

Una volta assegnata una densità di probabilità alla variabile X, si può scrivere, per ogni evento A di $\mathbb{R}$:

$$\mathbb{P}[X \in A] = \int_A f(x)\,dx \ ,$$



Importante: $\mathbb{P}[X = x] = P(x) = 0, \ \forall x \in \mathbb{R}$.

Densità di probabilità

➔ **Esempio:** Sia X una variabile aleatoria continua con densità

$$f(x) = 2\mathrm{e}^{-2x}\mathbf{1}_{(0,+\infty)}(x).$$

- $f(x)$ è davvero una densità:
  1. $f(x) \geq 0, \ \forall x \in \mathbb{R}$,
  2. $\int_{\mathbb{R}} f(x)\,dx = \int_0^{+\infty} 2\mathrm{e}^{-2x}\,dx = 1.$
- Considerando due eventi: $A = (1, 2)$ e $B = (-1, 1)$, è possibile calcolare le probabilità:

$$\mathbb{P}[X \in A] = \mathbb{P}[X \in (1, 2)] = \int_1^2 2\mathrm{e}^{-2x}\,dx = \mathrm{e}^{-2} - \mathrm{e}^{-4}$$

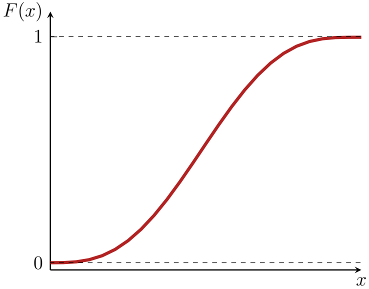
$$\mathbb{P}[X \in B] = \mathbb{P}[X \in (-1, 1)] = \int_{-1}^0 0\,dx + \int_0^1 2\mathrm{e}^{-2x}\,dx = 1 - \mathrm{e}^{-2}.$$

Funzioni di ripartizione e di sopravvivenza

Se X è una v.a. continua con densità $f(x)$, allora

$$F(x) = \int_{-\infty}^{x} f(t)\,dt; \quad \bar{F}(x) = \int_{x}^{\infty} f(t)\,dt = 1 - F(x)$$

Le funzionei di ripartizione e di sopravvivenza di una v.a. continua sono entrambe continua.



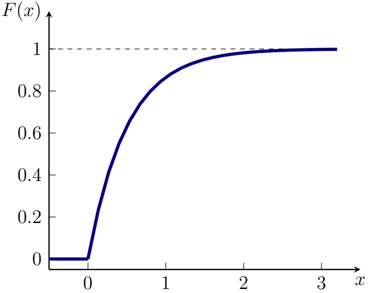
Dalla funzione di ripartizione si può risalire alla densità di probabilità della variabile in tutti i punti in cui $F(x)$ è derivabile:

$$f(x) = \frac{dF(x)}{dx},$$

Funzioni di ripartizione e di sopravvivenza

- **Esempio:** Per la variabile X con densità $f(x) = 2\mathrm{e}^{-2x}\mathbf{1}_{(0,+\infty)}(x)$, si ha:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \mathrm{e}^{-2x} & x \geq 0 \end{cases}$$



- **Esercizio:** trovare la densità di X a partire dalla funzione di ripartizione.

Valore atteso di una variabile aleatoria continua

Valore atteso

Sia X una v.a. continua con densità $f(x)$. Allora il **valore atteso** o **media** di X è:

$$\mathbb{E}[X] \quad = \quad \int_{\mathbb{R}} x f(x)\, dx.$$

➔ **Esempio:** per la variabile X con densità $f(x) = 2\mathrm{e}^{-2x}\mathbf{1}_{(0,+\infty)}(x)$, si ha

$$\mathbb{E}[X] = \int_0^{+\infty} 2x\mathrm{e}^{-2x}\, dx = \frac{1}{2}.$$

Valore atteso di $g(X)$

Sia $Y = g(X)$ una v.a. ottenuta trasformando la v.a. X tramite la funzione $g : \mathbb{R} \to \mathbb{R}$. Il valore atteso di Y si può calcolare anche senza conoscere direttamente la distribuzione di probabilità di Y:

- X discreta: $\mathbb{E}[Y] = \sum_i g(x_i) p_i$.
- X continua: $\mathbb{E}[Y] = \int_{\mathbb{R}} g(x) f(x)\, dx$.

☞ In particolare, la varianza di X è il valore atteso di una trasformata della X tramite la funzione $g(x) = (x - \mathbb{E}[X])^2$ e si può scrivere come

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \left[\mathbb{E}[X]\right]^2 .$$

Varianza di una variabile aleatoria continua

Varianza

Sia X una v.a. continua con densità $f(x)$. Allora la **varianza** di X è:

$$\text{Var}\left[X\right] = \int_{\mathbb{R}} (x - \mathbb{E}\left[X\right])^2 f(x)\, dx.$$

Una formula pratica per il calcolo è

$$\text{Var}\left[X\right] = \int_{\mathbb{R}} x^2 f(x)\, dx - \left[\mathbb{E}\left[X\right]\right]^2 \ .$$

Esempio: per la variabile già considerata in precedenza abbiamo:

$$\begin{aligned} \text{Var}\left[X\right] &= \int_0^{+\infty} 2x^2 \mathrm{e}^{-2x}\, dx - \left(\frac{1}{2}\right)^2 \\ &= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}. \end{aligned}$$

Moda di una variabile aleatoria

Moda

La **moda** di una variabile aleatoria continua X è il punto (o i punti) in cui la funzione di densità assume valore massimo.

Esempio:

- La variabile X con densità $f(x) = 2\mathrm{e}^{-2x}\mathbf{1}_{(0,+\infty)}(x)$, ha moda in 0.

La moda, così come il valore atteso di una v.a., è un **indice di posizione**

Quantili, quartili e mediana

Mediana

La **mediana** di una variabile aleatoria X è il minimo valore m per cui

$$F(m) = \mathbb{P}[X \leq m] \geq \frac{1}{2}.$$

- Per una v.a. continua (cioè con f.r. F continua) la mediana è l'unico punto m in cui

$$F(m) = \bar{F}(m) = 1/2$$

- **Esempio:** Per la variabile X con densità $f(x) = 2\mathrm{e}^{-2x}\mathbf{1}_{(0,+\infty)}(x)$, la mediana è tale che

$$1 - \mathrm{e}^{-2m} = \mathrm{e}^{-2m} = \frac{1}{2},$$

cioè $m = \log(2)/2$.

Quantili, quartili e mediana

Quantili

Fissato un valore $\alpha \in (0,1)$, il **quantile di livello α** di una variabile aleatoria X è il minimo valore q_α per cui

$$F(q_\alpha) = \mathbb{P}\left[X \leq q_\alpha\right] \geq \alpha.$$

- Per una v.a. continua (con f.r. F continua) il quantile di livello α è l'unico punto q_α in cui

$$F(q_\alpha) = \mathbb{P}\left[X \leq q_\alpha\right] = \alpha \text{ e } \bar{F}(q_\alpha) = \mathbb{P}\left[X \geq q_\alpha\right] \leq 1 - \alpha.$$

☞ **Importante:** Le proprietà e l'interpretazione delle costanti caratteristiche sono le stesse indipendentemente della natura discreta o continua della variabile casuale.

Distribuzione uniforme

Distribuzione uniforme

Immaginiamo che X sia una variabile che può assumere qualsiasi valore nell'intervallo (a, b), indifferentemente. La sua densità di probabilità è allora costante nell'intervallo (a, b) e nulla al di fuori:

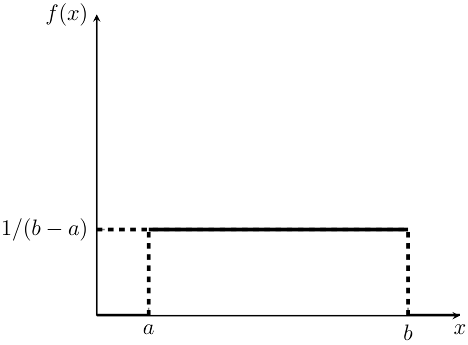
$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{altrove} \end{cases}$$

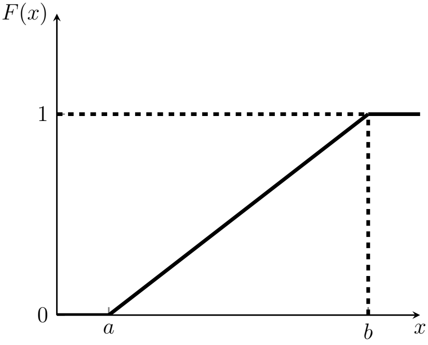
X è una variabile aleatoria **uniforme** e si scrive $X \sim \mathrm{U}\,(a, b)$.

A partire dalla densità si può trovare la funzione di ripartizione:

$$F(x) = \int_a^x \frac{1}{b-a} dy = \frac{x-a}{b-a}; \quad x \in (a, b)$$

Distribuzione uniforme





E' facile verificare che

$$\mathbb{E}[X] = \frac{a+b}{2}, \qquad \mathsf{Var}[X] = \frac{(b-a)^2}{12} \; .$$

Distribuzione normale

Distribuzione normale

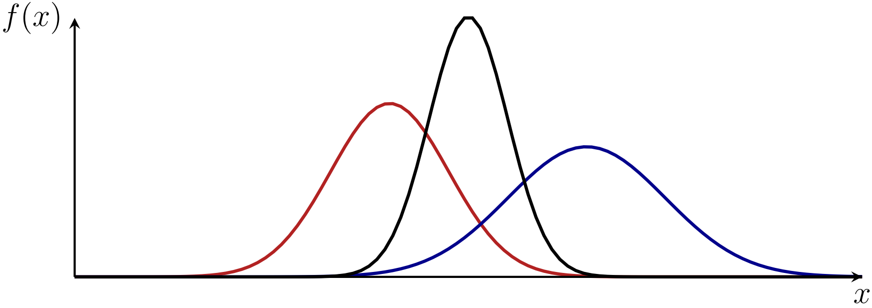
Una variabile aleatoria X ha **distribuzione normale** o gaussiana, indicato $X \sim \mathsf{N}\left(\mu, \sigma^2\right)$, se la sua funzione di densità ha la forma

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad x \in \mathbb{R},\ \mu \in \mathbb{R},\ \sigma > 0.$$

- **Johann Friedrich Carl Gauss (1777–1855):** matematico, astronomo e fisico tedesco, che ha dato contributi determinanti in analisi matematica, teoria dei numeri, statistica, calcolo numerico, geometria differenziale, geodesia, geofisica, magnetismo, elettrostatica, astronomia e ottica. Talvolta definito *il più grande matematico della modernità*.(Wikipedia)

Carl Friedrich Gauss

Distribuzione normale



Si può dimostrare che

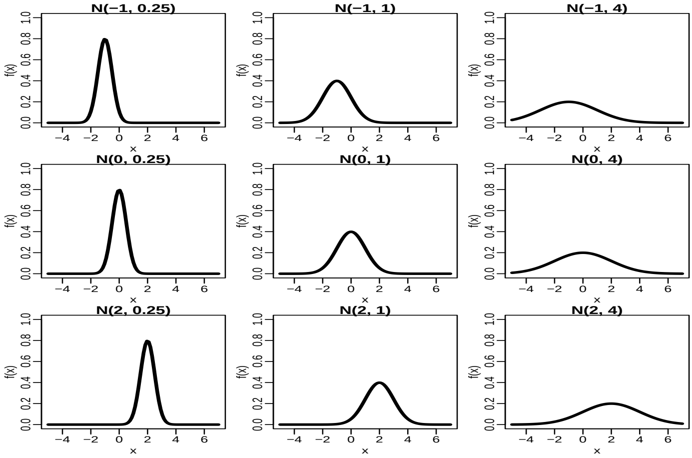
$$\mathbb{E}[X] = \mu, \qquad \mathsf{Var}[X] = \sigma^2 \ .$$

e se $X \sim \mathsf{N}(\mu, \sigma^2)$ allora, per qualsiasi a e b reali,

$$Y = aX + b \sim \mathsf{N}\left(a\mu + b, a^2\sigma^2\right)$$

Distribuzione normale

I parametri μ e σ^2 della normale non cambiano la forma a campana della densità ma determinano la sua posizione e concentrazione:



Distribuzione normale

La funzione di ripartizione di una v.a. normale non si può calcolare in forma esplicita. Ma utilizzando **R** (o un qualsiasi pacchetto statistico) si può oggi calcolare numericamente:

Esempio: $X \sim \mathsf{N}(-1, 4)$

➔ $\mathbb{P}[X \leq 1] = 0.8413$

```
> pnorm(q=1,mean=-1,sd=2)
[1] 0.8413447
```

➔ $\mathbb{P}[0 \leq X \leq 1] = 0.1499$

```
> pnorm(q=1,mean=-1,sd=2)-pnorm(q=0,mean=-1,sd=2)
[1] 0.1498823
```

Distribuzione normale

Standardizzazione

La variabile $Z \sim N(0,1)$ viene anche chiamata **normale standard** e la sua f.r. si denota con la Φ. Alla normale standard si riducono tutte le altre normali mediante la **standardizzazione**:

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

Una volta (e ancora adesso a volte...), per calcolare delle probabilità per qualsiasi variabile aleatoria normale si usavano le **tavole della distribuzione normale standard** (Le trovate nel vostro libro: date un'occhiata!). Queste bastano grazie alla standardizzazione:

$$\mathbb{P}\left[a < X < b\right] = \mathbb{P}\left[\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \ .$$

Esempio:

$$X \sim \mathsf{N}(-1,4) \quad \Rightarrow \quad Z = \frac{X+1}{2} \sim \mathsf{N}(0,1)$$

➔ $\mathbb{P}[X \leq 1] = \mathbb{P}\left[Z \leq \frac{1+1}{2}\right]$

```
> pnorm(q=1,mean=-1,sd=2)
[1] 0.8413447
> pnorm(q=(1+1)/2,mean=0,sd=1)
[1] 0.8413447
```

➔ $\mathbb{P}[0 \leq X \leq 1] = \mathbb{P}\left[\frac{0+1}{2} \leq Z \leq \frac{1+1}{2}\right]$

```
> pnorm(q=1,mean=-1,sd=2)-pnorm(q=0,mean=-1,sd=2)
[1] 0.1498823
> pnorm(q=(1+1)/2,mean=0,sd=1)-pnorm(q=(0+1)/2,mean=0,sd=1)
[1] 0.1498823
```

Distribuzione normale

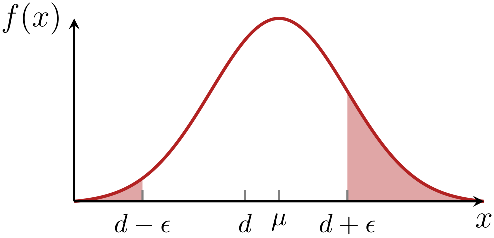
Esempio: Una macchina produce tubi di diametro X (in mm). Il diametro X potrebbe essere modellizzato con una distribuzione normale di media μ mm e varianza σ^2 mm^2. Supponiamo che per contratto i tubi debbano essere di diametro d mm più o meno un margine di specifica di ϵ mm. Si noti che d non è necessariamente uguale a μ perché la macchina potrebbe essere centrata su valori leggermente diversi.

- Calcolare la probabilità che un generico tubo sia difettoso (cioè fuori specifica, o non conforme).
  - ☞ Sia $p = \mathbb{P}[\text{"X è difettoso"}]$. Allora,

$$p = \mathbb{P}[(X < d - \epsilon) \cup (d + \epsilon < X)] = \mathbb{P}[X < d - \epsilon] + \mathbb{P}[d + \epsilon < X]$$
$$= \int_{-\infty}^{d-\epsilon} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx + \int_{d+\epsilon}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx.$$

Distribuzione normale

Dati μ, σ, T e ϵ siamo in grado di calcolare p (in R, per esempio, con il comando `pnorm`).



- Calcolare la probabilità che due tubi su 10 siano difettosi.

  ☞ Sia

  $$S_{10} = \text{“numero di tubi difettosi su 10”}$$

  la variabile che conta il numero di successi in 10 prove indipendenti in cui ogni prova può risultare un successo (difettoso o non conforme) con probabilità

  $$p = \mathbb{P}\left[\text{“X è difettoso”}\right]$$

Distribuzione normale

o un insuccesso (conforme) con probabilità $1 - p$. Allora

$$S_{10} \sim \text{Binom}\,(10, p)$$

e la probabilità che due tubi su 10 siano difettosi è data dalla formula della funzione di probabilità binomiale:

$$\mathbb{P}\,[S_{10} = 2] = \binom{10}{2} p^2 (1 - p)^8.$$

Esercizio: Fare i calcoli con $\mu = 3$, $\sigma = 0.09$, $d = 2.94$ e $\epsilon = 0.18$.

☞ $p = 0.0950$, $\mathbb{P}\,[S_{10} = 2] = 0.1827$

Approssimazione normale per la binomiale

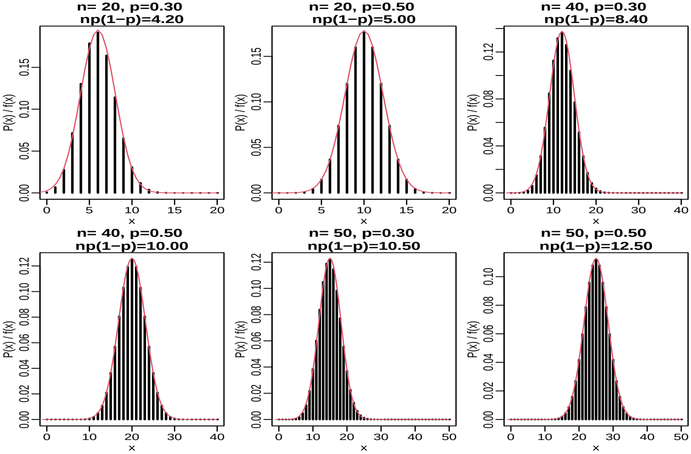
Quando n è "grande", allora la funzione di ripartizione di una v.a. binomiale di parametri n e p si può approssimare con la funzione di ripartizione di una normale di parametri $\mu = np$ e $\sigma^2 = np(1-p)$:

Approssimazione normale per la binomiale

$$\mathsf{Binom}\,(n,p) \approx \mathsf{N}\,(np, np(1-p))\,.$$

- L'approssimazione viene utilizzata nella pratica quando $np(1-p) \geq 10$.
- Questa è una conseguenza del **teorema del limite centrale** (che studieremo più avanti)

Approssimazione normale per la binomiale



Approssimazione normale per la binomiale

Esempio: Un certo virus danneggia un file con probabilità 0.35, indipendentemente dagli altri file. Il virus attacca una cartella con 2400 file. Qual è la probabilità che vengano danneggiati fra gli 800 e gli 850 file (inclusi)?

☞ Sia X la variabile che conta il numero di file danneggiati su 2400. Allora $X \sim \text{Binom}\,(2400, 0.35)$ e la probabilità richiesta è:

$$\mathbb{P}\left[800 \leq X \leq 850\right] = \sum_{k=800}^{850} \binom{2400}{k} 0.35^k (1-0.35)^{2400-k} = 0.632893.$$

Calcolata con **R**:

```
>  pbinom(850,2400,0.35)-pbinom(799,2400,0.35)
[1] 0.632893
```

Approssimazione normale per la binomiale

Possiamo anche approssimare X con una normale con parametri

$$\mu = 2400 \cdot 0.35 = 840$$

$$\sigma = \sqrt{2400 \cdot 0.35 \cdot (1 - 0.35)} = \sqrt{546} = 23.36664.$$

Otteniamo, con la cosiddetta **correzione per continuità**:

$$\begin{aligned}\mathbb{P}[800 \leq X \leq 850] &= \mathbb{P}[\mathbf{799.5} < X < \mathbf{850.5}] \\ &\approx \Phi\left(\frac{850.5 - 840}{23.36664}\right) - \Phi\left(\frac{799.5 - 840}{23.36664}\right) \\ &= \Phi(0.4493586) - \Phi(-1.73324) = 0.631887.\end{aligned}$$

Calcolata con **R**:

```
>  pnorm((850.5-840)/23.36664)-pnorm((799.5-840)/23.36664)
[1] 0.631887
```

Distribuzione gamma

Distribuzione gamma

Si dice che X ha **distribuzione gamma** di parametri $\alpha > 0$ e $\lambda > 0$, e si scrive $X \sim \text{Ga}\,(\alpha, \lambda)$, se la sua densità di probabilità è della forma

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \in (0, \infty) \\ 0 & \text{altrove} \end{cases}$$

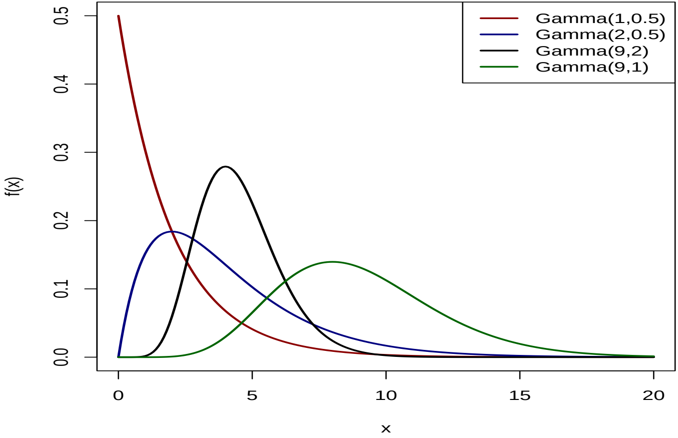
dove $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$ è la **funzione gamma**.

Si può dimostrare che

$$\mathbb{E}\,[X] = \frac{\alpha}{\lambda}, \qquad \text{Var}\,[X] = \frac{\alpha}{\lambda^2}\ .$$

Distribuzione gamma

La densità della gamma al variare di α e λ cambia sia forma che posizione:



Distribuzione gamma

Esempio: Se X è una v.a. gamma di media 4 e varianza 4, allora

$$\alpha = 4 \quad \text{e} \quad \lambda = 1 \ .$$

- Si può allora calcolare (integrando ripetutamente per parti)

$$\begin{aligned}\mathbb{P}[X < 4] &= \int_0^4 \frac{e^{-x}x^3}{6}\,dx \\ &= \frac{1}{6}\left[-(x^3+3x^2+6x+6)e^{-x}\right]_0^4 = \\ &= 0.567 \ .\end{aligned}$$

- In questo caso il calcolo esplicito è stato possibile poiché α è un numero intero e

$$\Gamma(4) = \int_0^\infty x^{4-1}e^{-x}\,dx = (4-1)! = 6 \ .$$

Distribuzione gamma

In generale, usando **R** si possono calcolare probabilità legate a variabili con distribuzione gamma:

- Per $X \sim \text{Ga}(1.5, 0.5)$, $\quad \mathbb{P}[X \leq 1] = 0.19874804$

```
> pgamma(q=1,shape=1.5,rate=0.5)
[1] 0.198748
```

- Per $X \sim \text{Ga}(7.5, 1)$, $\quad \mathbb{P}[X \geq 7.2] = 0.49543391$

```
> 1-pgamma(q=7.2,shape=7.5,rate=1)
[1] 0.4954339
```

Si può anche valutare la funzione gamma:

- $\Gamma(4) = 6 \quad \Gamma(7.5) = 1871.254$

```
> gamma(4)
[1] 6
> gamma(7.5)
[1] 1871.254
```

Distribuzione esponenziale

Distribuzione esponenziale

Un caso particolare di distribuzione gamma con $\alpha = 1$ è la **distribuzione esponenziale**, $X \sim \text{Exp}\,(\lambda)$.

La sua densità è

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \in (0, \infty) \\ 0 & \text{altrove} \end{cases}$$

e la sua funzione di ripartizione si può calcolare in forma esplicita:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Distribuzione esponenziale

Evidentemente,

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \text{e} \quad \mathsf{Var}[X] = \frac{1}{\lambda^2} \ .$$

La distribuzione esponenziale si usa per modellare tempi di attesa. Ad esempio:

1. il tempo che passa fra l'arrivo di un treno in stazione e il successivo;
2. la vita di un certo componente elettronico di un'automobile;
3. la lunghezza fra due difetti consecutivi di una fibra ottica.

Distribuzione esponenziale

Mancanza di memoria

Se $X \sim \text{Exp}(\lambda)$, allora

$$\begin{aligned}\mathbb{P}[X > t + s | X > t] &= \frac{\mathbb{P}[X > t + s]}{\mathbb{P}[X > t]} \\ &= \mathbb{P}[X > s]\ .\end{aligned}$$

Questa proprietà si dimostra tenendo conto che

$$\mathbb{P}[X > x] = e^{-\lambda x}\ , \quad x > 0\ .$$

- La distribuzione esponenziale è l'unica distribuzione continua con la proprietà di **mancanza di memoria** (come la geometrica fra le distribuzioni discrete).

Distribuzione esponenziale

Esempio: Il tecnico di un laboratorio dell'università in un'ora installa un certo *software* in media su 30 pc. Assumendo che il tempo di installazione su ogni pc segua una distribuzione esponenziale, vogliamo calcolare la probabilità che il tecnico impieghi più di 5 minuti per installare il *software* nel prossimo pc.

- X=tempo di attesa in ore $\sim \mathsf{Exp}\,(30)$

$$\Rightarrow \mathbb{P}\,[X > 1/12] = e^{-30/12}$$

- X=tempo di installazione in minuti $\sim \mathsf{Exp}\,(1/2)$

$$\Rightarrow \mathbb{P}\,[X > 5] = e^{-5/2}$$

Distribuzione esponenziale

Tasso di fallimento

Se X è una variabile casuale continua con funzione di densità $f_X(x)$ y funzione di sopravvivenza $\bar{F}_X(x)$, il **tasso di fallimento** di X è la funzione

$$r_X(x) = \frac{f_X(x)}{\bar{F}_X(x)}$$

- Il tasso di fallimento di una variabile esponenziale è costante. In effetti, se $X \sim \text{Exp}\,(\lambda)$, allora

$$r_X(x) = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda.$$

Questa è una conseguenza della mancanza di memoria e l'esponenziale è l'unica distribuzione continua con tasso di fallimento costante.

Distribuzione esponenziale

Esempio: Nel contesto del bilanciamento del carico della CPU, una tecnica che distribuisce i processi CPU spostandoli da workstation sovraccariche a quelle meno cariche in una rete, consideriamo tre concetti fondamentali legati a un processo CPU:

- Dimensione (o durata): la quantità totale di CPU richiesta da un processo, misurata in secondi o cicli CPU.
- Età: il tempo totale di CPU già utilizzato dal processo, anch'esso misurato in secondi o cicli CPU.
- Dimensione residua (o durata residua): la quantità di CPU ancora necessaria per completare il processo.

Supponiamo ora di avere due processi UNIX con età di 2 secondi e 100 secondi rispettivamente. Se assumiamo che le durate dei processi UNIX siano distribuite esponenzialmente, quale dei due processi è più probabile che richieda un tempo residuo maggiore per completarsi?

Distribuzione esponenziale

- Dato che le durate dei processi si assumono exponenzialmente distribuite, la durata per ognuno dei processi residua è indipendente della loro età e avrà, quindi, la stessa distribuzione. In altre parole, entrambi processi avranno, in media, la stessa durata residua. Per la distribuzione esponenziale, l'età del processo non ha nessun effetto sulla durata residua.

Distribuzione Pareto

- Abbiamo visto che la distribuzione esponenziale è l'unica ad avere un **tasso di fallimento costante**. In altre parole, per una varibile esponenziale, X, l'ulteriore tempo di attesa non dipende dal tempo precedentemente trascorso. Come già discusso, questo è un modello

- In alternativa, una variabile casuale potrebbe avere un **tasso di fallimento crescente**, se il rischio di fallimento aumenta con l'età del processo.Alcuni esempi sono:
  - Il tempo di vita di una macchina.
  - Il tempo di vita di un componente elettronico.
  - Il tempo di vita di una persona.

Distribuzione Pareto

- Finalmente, esistono fenomeni che si possono modellare attraverso variabili casuali con **tasso di fallimento decrescente**, in altre parole, il rischio di fallimento aumenta all'aumentare dell'età. Fenomeni di questo tipo sono in aparenza meno frequenti:
  - La durata di una relazione di lavoro o di amicizia.
  - Il tempo che una persona abita nella stessa casa, città o paese.
  - Il tempo di vita di un ristorante bar o negozio.
  - Vedremo in seguito che ce ne sono molti altri.

Distribuzione Pareto

Si dice che X ha **distribuzione Pareto** di parametr0 $0 < \alpha < 2$, e si scrive $X \sim \text{Pareto}\,(\alpha)$, se la sua funzione di sopravvivenza è della forma

$$\bar{F}_X(x) = \begin{cases} x^{-\alpha} & x > 1 \\ 1 & \text{altrove} \end{cases}$$

Distribuzione Pareto

La sua funzione di ripartizione è

$$F_X(x) = \begin{cases} 1 - x^{-\alpha} & x > 1 \\ 0 & \text{altrove} \end{cases}$$

e la sua funzione di densità si può calcolare dalla derivata di $F_X(x)$:

$$f_X(x) = F'_X(x) = \begin{cases} 0 & x \leq 1 \\ \alpha x^{-(\alpha+1)} & x > 1 \end{cases}$$

Si può dimostrare che,

$$\mathbb{E}[X] = \begin{cases} \infty & \text{se } \alpha \leq 1 \\ \frac{\alpha}{\alpha-1} & \text{se } \alpha > 1 \end{cases} \quad \text{e} \quad \text{Var}[X] = \frac{1}{\lambda^2}.$$

Distribuzione Pareto

- Il tasso di fallimento di una variabile Pareto è decrescente. In effetti, se $X \sim \text{Pareto}(\alpha)$, allora

$$r_X(x) = \frac{\alpha x^{-(\alpha+1)}}{x^{-\alpha}} = \frac{\alpha}{x}.$$

La distribuzione Pareto si usa per modellare tempi di attessa con rischio di fallimento decrescente, in altre parolle, in situazioni in cui più tempo passa, maggiore la probabilità di dover aspettare ancora tanto. Ma non solo! Alcuni esempi di variabili con distribuzione Pareto:

1. la grandeza dei un file in un sito web;
2. il grado (o numero di collegamenti) dei un nodi in una rete (abbastanza grande);
3. la durata di un flusso IP;
4. la richezza di una persona;
5. il costo di un disastro naturale;
6. il tempo di collegamento a una rete, di una chiamata telefonica, etc.

Distribuzione Pareto

Code pesanti

Se $X \sim \text{Pareto}\,(\alpha)$, allora si dice che X ha le **code pesanti**. La definizione formale è più complessa ma intuitivamente questo vuole dire che la probabilità di osservare valori molto grandi decresce molto lentamente.

Esempi:

- La richezza: l'1% della popolazione mondiale accumula più richezza del restante 99%.
- La grandezza dei file: una piccola porzione dei file più pesanti in un sito web occupa più memoria dei restanti file messi insieme.
- La durata dei processi UNIX (nonostante in molti contesti vengano modellati come esponenziali): una piccola frazione dei processi più lunghi, occupano complessivamente più risorse nel CPU dei restanti processi messi insieme.

Distribuzione Pareto

Esercizio: Se invece di una variabile esponenziale, la durata, X di un processo UNIX si modella come una Pareto $(\alpha = 1)$, qual è la probabilità che un processo di età $x > 1$ soppraviva almeno il doppio?

- Dobbiamo calcolare

$$\P X \geq 2x | X > x = \frac{\bar{F}_X(2x)}{\bar{F}_X(x)} \frac{1/2x}{1/x} = 1/2$$

- In questo caso, quale dei due processi del esempio precedente è più probabile che richieda un tempo residuo maggiore per completarsi? Confrontare il risultato con quello ottenuto usando la distribuzione esponenziale come modello.

Distribuzioni congiunte

Densità congiunta

Densità congiunta

Si dice che X e Y sono **congiuntamente continue** se esiste una funzione $f : \mathbb{R}^2 \to \mathbb{R}$ integrabile e tale che:

1. $f(x, y) \geq 0$, $\forall (x, y) \in \mathbb{R}^2$,
2. $\iint_{\mathbb{R}^2} f(x, y) dx dy = 1$.

f è la **funzione di densità congiunta** di (X, Y) e si utilizza per calcolare probabilità riguardanti le due variabili congiuntamente.

Ad esempio,

- se $C \subset \mathbb{R}^2$,

$$\mathbb{P}\left[(X, Y) \in C\right] = \iint_C f(x, y) dx\, dy;$$

Densità congiunta

- in particolare, se $C = A \times B$, for $A, B \subset \mathbb{R}$,

$$\mathbb{P}\left[(X,Y) \in C\right] = \mathbb{P}\left[X \in A, Y \in B\right] = \int_B \left(\int_A f(x,y) dx \right) dy.$$

➔ Dalla densità congiunta si può ricavare la f.r. congiunta:

$$F(x,y) = \mathbb{P}\left[X \leq x, Y \leq y\right] = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t) ds\, dt.$$

➔ Viceversa, dalla f.r. congiunta si ottiene facilmente (in tutti i punti in cui F è derivabile)

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

Densità congiunta

Le **funzioni di densità marginali** si ricavano integrando la densità congiunta rispetto all'altra variabile:

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy$$

e

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)dx.$$

Densità congiunta

Esempio: La densità congiunta di X e Y è

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & x > 0,\ y > 0 \\ 0 & \text{altrove.} \end{cases}$$

- Allora:

$$\begin{aligned} \mathbb{P}\left[X > 1, Y < 1\right] &= \int_0^1 \left(\int_1^{+\infty} 2e^{-x}e^{-2y} dx \right) dy \\ &= \int_0^1 2e^{-2y}(-e^{-x})_1^{+\infty} dy = e^{-1} \int_0^1 2e^{-2y} dy \\ &= e^{-1}(-e^{-2y})_0^1 = e^{-1}(1 - e^{-2}), \end{aligned}$$

Densità congiunta

$$
\begin{aligned}
\mathbb{P}\left[X < Y\right] &= \iint_{\{(x,y):x<y\}} 2e^{-x}e^{-2y} dx\, dy \\
&= \int_0^{+\infty} \left(\int_0^y 2e^{-x}e^{-2y} dx \right) dy = \cdots = 1/3,
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}\left[X < 3\right] &= \int_0^3 e^{-x} \left(\int_0^{+\infty} 2e^{-2y} dy \right) dx \\
&= \int_0^3 e^{-x} dx = 1 - e^{-3}.
\end{aligned}
$$

Variabili aleatorie indipendenti

Variabili aleatorie indipendenti

Ricordiamo che due v.a. sono **indipendenti** se

$$\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A]\,\mathbb{P}[Y \in B], \quad \forall A, B \subset \mathbb{R}.$$

o, equivalentemente, se

$$F_{X,Y}(x,y) = F_X(x)F_Y(y), \quad \forall (x,y) \in \mathbb{R}^2,$$

Per le variabilil continue, si può dimostrare che questo è equivalente a dire che:

$$f(x,y) = f_X(x)f_Y(y), \quad \forall (x,y) \in \mathbb{R}^2$$

Variabili aleatorie indipendenti

Esercizio:

$$f(x,y) = \begin{cases} 24xy, & x, y \in (0,1),\ x+y \in (0,1) \\ 0 & \text{altrove.} \end{cases}$$

X e Y sono indipendenti?

Rappresentate graficamente il dominio di (X, Y) e provate a rispondere.

Variabili aleatorie indipendenti

Esempio: Un programma è formato da due blocchi che vengono processati sequenzialmente durante la compilazione. Ogni blocco viene processato in un tempo medio di 5 minuti e si assume che i tempi di compilazione dei due blocchi, X e Y, siano indipendenti e con distribuzione esponenziale.

☞ Per l'indipendenza, la densità congiunta di X e Y è

$$f(x,y) = f_X(x) f_Y(y) = \frac{1}{5} e^{-x/5} I_{(0,+\infty)}(x) \cdot \frac{1}{5} e^{-y/5} I_{(0,+\infty)}(y).$$

Variabili aleatorie indipendenti

Allora, la probabilità che l'intero programma venga compilato in meno di 12 minuti è:

$$
\begin{aligned}
\mathbb{P}\left[X+Y<12\right] &= \iint_{\{(x,y):x+y<12\}} f(x,y)\,dx\,dy \\
&= \int_0^{12}\left(\int_0^{12-x}\frac{1}{5}e^{-y/5}\,dy\right)\frac{1}{5}e^{-x/5}\,dx \\
&= \int_0^{12}\left(1-e^{-(12-x)/5}\right)\frac{1}{5}e^{-x/5}\,dx \\
&= \left(1-e^{-12/5}\right)-\frac{12}{5}e^{-12/5} = 0.69156.
\end{aligned}
$$

☞ Qual è la probabilità che il primo blocco del programma venga compilato in un tempo inferiore rispetto al secondo blocco?

Distribuzioni condizionate: variabili continue

Funzioni di densità condizionata

Siano X e Y v.a. continue con funzione di densità congiunta $f(x, y)$.
La **funzione di densità condizionata** di X dato $Y = y$ è

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, \quad \forall y \text{ t.c. } f_Y(y) > 0.$$

Allo stesso modo, la **funzione di densità condizionata** di Y dato $X = x$ è

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \quad \forall x \text{ t.c. } f_X(x) > 0.$$

☞ Allora,

$$\mathbb{P}\left[X \in A | Y = y\right] = \int_A f_{X|Y}(x|y)\, dx, \quad \mathbb{P}\left[Y \in B | X = x\right] = \int_B f_{Y|X}(y|x)\, dy.$$

Distribuzioni condizionate: variabili continue

- Se X e Y sono indipendenti, allora

$$f_{X|Y}(x|y) = f_X(x) \text{ e } f_{Y|X}(y|x) = f_Y(y).$$

Esempio: Sia $f(x,y) = \frac{15}{2}x(2-x-y)I_{(0,1)\times(0,1)}(x,y)$. Allora, dalla definizione,

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad \forall y \in (0,1).$$

- Calcoliamo prima la densità marginale di Y:

$$f_Y(y) = \int_0^1 \frac{15}{2}(2x - x^2 - xy)I_{(0,1)}(y)\, dx = \cdots = \frac{15}{2}\left(\frac{2}{3} - \frac{y}{2}\right) I_{(0,1)}(y).$$

Infine

$$f_{X|Y}(x|y) = \frac{x(2-x-y)I_{(0,1)}(x)}{\left(\frac{2}{3} - \frac{y}{2}\right)}, \quad \forall y \in (0,1).$$

Distribuzioni condizionate: variabili continue

☞ **Nota 1:** Come per le variabili discrete, anche per le continue, la **legge delle probabilità totali** ci permete di trovare la distribuzione marginale di una variabile, diciamo X a partire dalla marginale dell'altra, diciamo Y e la corrispondente distribuzione condizionata:

$$
\begin{aligned}
f_X(x) &= \int_y f_{X,Y}(x,y)dy \\
&= \int_y f_{X|Y}(x|y) f_Y(y)dy.
\end{aligned}
$$

☞ **Nota 2:** Allo stesso modo, tutte le proprietà delle costanti caratteristiche marginali, congiunte e condizionate, studiate per le variabili discrete rimangono valide per le variabili continue. L'unica differenza sta nel calcolo: le funzioni di probabilità vengono sostituite dalle funzioni di densità e le somme dalle integrali.

La distribuzione normale bivariata

Distribuzione normale bivariata

Siano $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ Cor $[X, Y] = \rho$ e

$$\text{Cov}\,[X, Y] = \sigma_{X,Y} = \rho\sigma_X\sigma_Y.$$

Allora, in termini matriciali, scriviamo

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \sigma_{X,Y} \\ \sigma_{X,Y} & \sigma_Y^2 \end{bmatrix}\right) \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$$

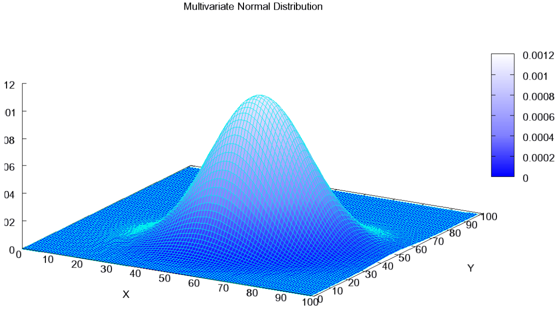
e la densità di $\mathbf{W} = (X, Y)$, posto $(x, y) = \mathbf{w}$, è

$$f(\mathbf{w}) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{w}-\boldsymbol{\mu})}.$$

La distribuzione normale bivariata

Svolgendo i calcoli, otteniamo:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\}.$$



La distribuzione normale bivariata