Formula di Bayes

LEGGE delle PROBABILITA' COMPOSTE a più ALTERNATIVE

· mette in relazione un evento E (effetto) con altri eventi (cause)

→ quanto uno degli altri eventi intervenga nel determinare E

↳ quanto partecipa una certa causa nel determinare un effetto

Teorema :

Dati un evento E e un insieme finito (numerabile) di eventi incompatibili $\{A_r\}, r = 1,2,\dots,n$,

se $E \subset \underbrace{\bigcup_{r=1}^{n}}_{\text{unitoria}} \underbrace{A_r}_{\text{partizioni di A}}$ e $P(E) \neq 0$, si ha

$$\boxed{P(\underbrace{A_r \mid E}_{\text{dato}}) = \frac{P(A_r)P(E \mid A_r)}{\sum_{i=1}^{n} P(A_i)P(E \mid A_i)}} \qquad r = 1,2,\dots,n$$

dimostrazione :

· dalla legge delle probabilità composte ⟶ $P(A_r \mid E) = \frac{P(E \cap A_r)}{P(E)} = \frac{P(E \mid A_r)P(A_r)}{P(E)}$ $P(E) \neq 0$

⟶ $E \subset \bigcup_r A_r$: $E = E \cap \left(\bigcup_r A_r\right) = \bigcup_r (E \cap A_r)$

↳ se gli A_r sono incompatibili, allora $E \cap A_r$ sono incompatibili

↳ per la legge delle probabilità totali ⟶ $P(E) = P\left(\bigcup_r (E \cap A_r)\right) \overset{LPT}{=} \sum_{r=1}^{n} P(E \cap A_r) \overset{LPC}{=} \sum_{r=1}^{n} P(E \mid A_r)P(A_r)$

· una lettura della formula di Bayes può essere data in termini di ipotesi

⟶ se $A_1, \dots, A_n$ sono ipotesi diverse, allora $P(E \mid A_r)$ è la probabilità che si realizzi E sotto l'ipotesi A_r

↳ $P(A_r \mid E)$ è la probabilità che, verificatosi E si sia verificata A_r → quanto è plausibile l'ipotesi A_r data la realizzazione di E

DEFINIZIONI :

· eventi incompatibili → $E \cap F = \{\emptyset\}$

· Legge Probabilità Composte → $P(A \cap B) = P(B)P(A \mid B) = P(B)P(B \mid A)$

· Legge Probabilità Totali → A e B eventi compatibili : $P(A \cup B) = P(A) + P(B) - \underbrace{P(A \cap B)}_{\text{se eventi mutualmente esclusivi } = 0}$

· eventi indipendenti : $P(A \cap B) = P(A) \cdot P(B)$

Distribuzioni di probabilità

· distribuzione di probabilità = modello matematico che collega i valori di una variabile alle probabilità che tali valori possano essere osservati

⟶ Sono utilizzate per modellizzare il comportamento di un fenomeno di interesse in relazione alla popolazione di riferimento, ovvero alla totalità dei casi di cui lo sperimentatore osserva un dato campione

⟶ variabile di interesse = variabile ALEATORIA (casuale)

↳ variabile casuale X è una funzione definita sullo spazio campionario Ω che associa ad ogni elemento elementare ω_i, un unico numero $X(\omega_i) = x_i$ ⟶ realizzazione della v. casuale X

↳ "casuale": generata da un fenomeno di cui non siamo in grado di prevedere l'esito

NB: le variabili casuali non sono variabili ma funzioni

In base alla scala di misura della variabile di interesse *X*, possiamo distinguere due tipi di distribuzioni di probabilità:

1. **distribuzioni continue**: la variabile viene espressa su un scala continua (es: il diametro del pistone) (supp(x))
2. **distribuzioni discrete**: la variabile viene misurata con valori numerici interi (es: numero di elementi non conformi o difettosi in un circuito stampato) (spet(x))

Formalmente, le distribuzioni di probabilità vengono espresse da una legge matematica detta **funzione di densità di probabilità** (indicata con *f*(*x*)) o **funzione di probabilità** (indicata con *p*(*x*)) rispettivamente per le distruzioni continue o discrete.

Esempio di **funzione di probabilità**

Esempio di **funzione di densità di probabilità**

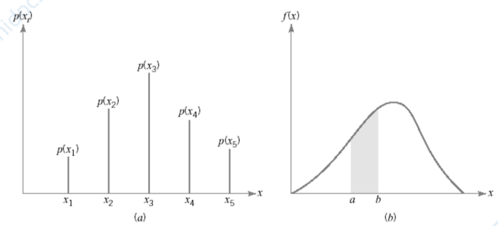


Figure 2-9 Probability distributions. (*a*) Discrete case. (*b*) Continuous case.

· mentre per una v. discreta è possibile elencare tutti i valori che può assumere, per una v. continua è necessario definire intervalli in cui suddividere i possibili valori della v.

N.B. Osserviamo infatti che se Ω è uno spazio di probabilità discreto, tutte le variabili casuali costruite su Ω saranno necessariamente discrete. Se invece Ω è uno spazio di probabilità continuo su di esso è possibile costruire sia variabile aleatorie continue che discrete (e ovviamente anche miste).

(DENSITÀ)

Variabili DISCRETE

· la funzione di probabilità di X è la f.ne p_k che ad ogni elemento associa la probabilità che la variabile discreta assuma valori uguali alle x_i (reali)

$$\longrightarrow \quad p_k = \begin{cases} P(X = x_i) & x \in \text{spett}(x) \\ 0 & x \notin \text{spett}(x) \end{cases}$$

$$\boxed{\text{C.N.S.}: \begin{cases} p_k \geq 0 \\ \sum_{i=1}^{k} p_k = 1 \end{cases}}$$

Variabili CONTINUE

· la funzione di densità di x non è una probabilità, ma è legata alla probabilità, perché rappresenta la probabilità che la v.a. continua X appartenga a un intervallo

$$\boxed{\text{C.N.S.}: \begin{cases} f_X(x) \geq 0 \\ \int_{\mathbb{R}} f_X(x)\,dx = 1 \end{cases}} \longrightarrow \text{probabilità evento certo } P(X \in \mathbb{R})$$

Funzione di ripartizione

· si vuole conoscere la probabilita' che X assuma un valore ≤ a un dato valore x

→ data la v.a. X si definisce funzione di ripartizione $F_X : \mathbb{R} \to [0,1]$

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

x = soglia

Proprieta'

· $0 \leq F_X(x) \leq 1$

· $\lim_{x\to-\infty} F_X(x) = 0$; $\lim_{x\to\infty} F_X(x) = 1$

· $F_X(x)$ e' monotona non decrescente: $x_1 < x_2 \to F_X(x_1) \leq F(x_2)$

· $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$

Variabili discrete

· $F_X(x)$ e' continua a destra → $\lim_{x\to x_0^+} F_X(x) = F_X(x_0)$

→ $\lim_{x\to x_0^-} F_X(x) \neq F_X(x_0)$ → $F_X(x)$ presenta dei punti di discontinuita' di 1° specie → salto a sinistra

↓

nel grafico pallino vuoto a sinistra

$$\boxed{F_X(x) = P(X \leq x) = \sum_{\substack{k\in\mathbb{Z} \\ x_k \leq x}} p_k}$$

Variabili continue

· $F_X(x)$ e' continua: $\lim_{x\to x^+} F_X(x) = \lim_{x\to x^-} F_X(x)$

→ $f_X(x) = F_X'(x)$

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

$$f(x) = \frac{d}{dx} F_X(x)$$

DISCRETE

- **Binomiale**: $X \sim Bin(n,p)$ ("si distribuisce in accordo con") (Bin di parametri n,p con $n \in \mathbb{N}$ e $p \in [0,1]$)

→ X = n° di successi su n prove indipendenti in cui il successo e' 1 evento di probabilita' p

→ $p_k = \binom{n}{x} p^x (1-p)^{n-x}$ (comb. semplici) $x \in \{0,1,\dots,n\} \subset \mathbb{Z}$ (x e' una var. discreta / deterministica)

→ $\mathbb{E}[X] = np$

· **Bernoulliana**: binomiale con n = 1 : $X \sim Ber(p) \equiv Bin(1,p)$ ↳ n=1 ad ogni singola prova indipendente

→ $p_k = p^x (1-p)^{1-x} = \begin{cases} 1-p & x=0 \\ p & x=1 \end{cases}$

- **Geometrica**: $X \sim Geo(p)$ $p \in [0,1]$

→ X = prova corrispondente al primo successo (non puo' essere $\emptyset$)

= n° di prove necessarie per ottenere il primo successo

→ $p_k = P(X=x) = (1-p)^{x-1} p$ $x \in \mathbb{N} = \{1,\dots,n\}$

↳ sia S il successo con $P(S) = p \in [0,1]$

→ $\mathbb{E}[X] = \frac{1}{p}$

prove ripetute indipendenti
$\begin{cases} X \sim Bin(n,p) \\ X \sim Geo(p) \end{cases}$

- **Poisson**: $X \sim Pois(\lambda)$ $\lambda > 0$

assenza di memoria

$P(x > a+b \mid x > a) = P(x > b)$

→ X = n° di volte in cui si verifica un evento nell' unita' di tempo (tasso λ)

→ $p_k = e^{-\lambda} \frac{\lambda^x}{x!}$ $x \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ NB: $P(X=x) = 0$ $x < 0$

→ $\mathbb{E}[X] = \lambda$

CONTINUE

· $X \sim f_X$ → densita' continua = insieme delle funzioni continue su un dominio

→ sia f una f.ne continua $f: \mathbb{R} \to \mathbb{R}$ definiamo:

$supp(f) = \{x \in \mathbb{R} : f(x) \neq 0\} \in Dom(f)$ → prendo dal dominio solo i punti che <u>non</u> annullano $f(x)$

↳ sia $X \sim f_X$ → $supp(f_X) = supp(x)$

↓ valori per cui la densita' e' $\neq 0$ ↳ valori che x puo' assumere

- **Uniforme** : $X \sim Unif(a,b)$ $a<b$ → $(a,b) \subset \mathbb{R}$ e' un intervallo ; $f_X = K$

→ $X \sim f_X(x) = \frac{1}{b-a} \mathbb{1}_{(a,b)}(x)$, $x \in \mathbb{R}$ $= \begin{cases} 0 & x \notin (a,b) \\ \frac{1}{b-a} & x \in (a,b) \end{cases}$

→ $\mathbb{E}[X] = \frac{a+b}{2}$

- **Esponenziale** : $X \sim Exp(\lambda)$ $\lambda > 0$ usata per esprimere la vita residua

→ $X = f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x)$, $x \in \mathbb{R}$ $= \begin{cases} 0 & x \leq 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases}$

→ $\mathbb{E}[X] =$

- **Gamma** : $X \sim Gamma(\lambda, \nu)$ $\in \mathbb{R}^2_+$

→ $f_X(x) = \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x)$, $x \in \mathbb{R}$

→ $EX = \frac{\nu}{\lambda}$

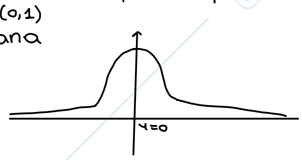
- **Normale** : $X \sim N(\mu, \sigma^2)$ $\mu \in \mathbb{R} ; \sigma^2 > 0$

↓ descrive comportamento delle variabili nell'avvicinarsi al valore medio

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

→ $\mathbb{E}[X] = \mu$ $Var = \sigma^2$

NB: Normale standard prevede $\mu = 0$ e $\sigma^2 = 1$
↳ $X \sim N(0,1)$
Gaussiana



Valore atteso (media)

· sia X la v.a su Ω , $x \in \Omega$ definiamo MEDIA di X il valore previsto che si potra' ottenere in un gran numero di prove

Discrete

· se la distribuzione e' finita, il valore atteso e' un numero reale dato dalla somma dei prodotti di ogni valore della variabile x_i per la rispettiva probabilita'

$\boxed{\mathbb{E}[X] = \sum_{i=1}^{n} x_i p_i}$ → somma pesata (puo' essere > o < di zero)

Continue

· $\boxed{\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x)\, dx}$ → deve essere $< +\infty$ → altrimenti X non ammette valore atteso

Proprieta'

· linearita' → $\mathbb{E}g(X) = g(\mathbb{E}X)$

· media di costante e' la costante → $\mathbb{E}(a) = a$

Varianza

· indica quanto sono dispersi i valori della variabile aleatoria relativamente al suo valore medio

→ considero la v.a. $X-E(X)$ scarto → $Var(X) = E((X-E(X))^2) = E(X)^2 - (E(X))^2$

· non e' mai negativa e misura il grado di variabilita' di una distribuzione

→ tanto maggiore e' la varianza tanto piu' i valori sono dispersi

Proprieta' : non linearita' → $Var(aX) = a^2 Var(X)$

· $Var(X) = 0$ quando tutti i valori della variabile sono uguali

→ non c'e' variabilita' nella distribuzione

Variabile Degenere (DISCRETA)

· tutta la probabilita' e' associata ad 1 solo evento causale $\overset{\text{costante}}{a}$

→ $X \sim Deg(a)$

→ $P_X(x) = \begin{cases} 1 & X = a \\ 0 & \text{altrimenti} \end{cases}$ $\quad E[X] = a \quad Var(X) = 0$

Indipendenza

· siano X e Y indipendenti $P(X \leq x \overset{\cap}{,} Y \leq y) = F_{(X,Y)}(x,y) = \boxed{P(X \leq x) P(Y \leq y) = F_X(x) F_Y(y)}$

caso CONTINUO : $F_X(x) F_Y(y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x,y) = F_X'(x) F_Y'(y) = \underbrace{f_X(x) f_Y(y)}_{\text{densita' marginali}} = \underbrace{f_{(X,Y)}(x,y)}_{\text{densita' congiunta}}$

caso DISCRETO : $\underbrace{P(X = x, Y = y)}_{\text{densita' congiunta}} = \underbrace{P(X = x) \cdot P(Y = y)}_{\text{densita' marginali}}$ (no f.ne di ripartizione)

→ $X \perp Y$: $P(X \in A; Y \in B) = P(X \in A) \cdot P(X \in B)$

Identicamente distribuite

· hanno tutte la stessa distribuzioni di probabilita'

Inferenza statistica

· caratterizzare un parametro sconosciuto a partire da dati osservati

→ θ = parametro da stimare → $\hat{\theta}$ = stima di θ → $\hat{\Theta}$ = v.a. stimatore

stimatore = funzione che associa ad ogni possibile campione un valore del parametro da stimare

Famiglia di distribuzioni (di probabilità) → DENSITÀ

$f_X(x) = f_X(x,\theta)$ → θ = parametro che caratterizza X continua

$\mathcal{F} = \{f_X(x,\theta),\ x \in \mathbb{R} : \theta \in \mathbb{R}^m\}$ m = 1, 2

$P_k = P_k(\theta) = P(X = x_k)$ → θ = parametro che caratterizza X discreta

$\mathcal{F} = \{P_k(\theta),\ k \in \mathbb{Z} : \theta \in \mathbb{R}^m\}$ m = 1, 2

Stima di θ

· in generale: massimizzare l'affidabilità del risultato (probabilità)

→ massimizzare = trovare il MAX della funzione data dalla probabilità

tramite derivata 1° e 2°

→ $\max_\theta P(X \in I) = \max_\theta \int_I f_X(x,\theta)\,dx$

Proprietà desiderabili

CORRETTEZZA

· uno stimatore $\hat{\Theta}$ si dice corretto se: $\mathbb{E}[\hat{\Theta}] = \theta$

CONSISTENZA

· uno stimatore $\hat{\Theta}$ si dice consistente se: $\lim_{m \to \infty} Var(\hat{\Theta}) = 0$

numerosità campionaria

Variabile media campionaria

· definita dalla media aritmetica di v.a. indipendenti X_k, $k \in \mathbb{N}$

→ $\overline{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k$ → $\overline{X}$ dipende dalla numerosità campionaria