

Distribuzioni di probabilità e variabili aleatorie continue

Funzione di ripartizione → viene indicata con $F(x_0)$, esprime la probabilità che X non superi x_0 , come funzione di x_0 → $F(x_0) = P(X \leq x_0)$

Calcolo della probabilità in un intervallo on la funzione di ripartizione

Sia $a < b$. la probabilità che X assuma valori tra a e b è → $P(a < x < b) = F(b) - F(a)$

Funzione di densità di probabilità → permette di calcolare la probabilità che X appartenga ad un determinato intervallo.

Proprietà

1. $F(x) > 0$ per qualunque x nell'intervallo ei valori ammissibili e $f(x)$ altrove
2. L'area sottesa alla funzione di densità di probabilità, $f(x)$, su tutto l'intervallo di valori ammissibili di X vale 1
3. Supponiamo di rappresentare graficamente la funzione di densità di probabilità. Siano a e b due possibili valori della variabile aleatoria, X , con $a < b$. la probabilità che X assuma valori tra a e b è l'area sottesa alla funzione di densità sull'intervallo: $P(a < X < b) = 1$
4. La funzione di ripartizione, $F(x_0)$, è l'area sottesa alla funzione di densità di probabilità, $f(x)$, fino a x_0 → $F(x_0) = 1$

Aree sottese alle funzioni di densità di probabilità di variabili aleatorie continue

Sia X una variabile aleatoria continua con funzione di densità di probabilità $f(x)$ e funzione di ripartizione $F(x)$. Valgono le seguenti proprietà:

- L'area totale sottesa alla curva $f(x)$ vale 1
- L'area sottesa alla curva $f(x)$ alla sinistra di x_0 è $F(x_0)$, dove x_0 è un qualsiasi valore assumibile dalla variabile aleatoria

Distribuzione uniforme

Si consideri ora la funzione di densità della distribuzione uniforme di probabilità nell'intervallo da 0 a 1.

La funzione di densità uniforme vale:

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

Funzione di ripartizione

$$F(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } x \in [a, b] \\ 1 & \text{se } x > b \end{cases}$$

Varianza

$$\sigma^2 = \frac{(b-a)^2}{12}$$

Valori attesi di variabili aleatorie continue

Poiché, per una variabile aleatoria continua, la probabilità associata a ogni singolo valore in un intervallo è 0 dovremo utilizzare un altro metodo.

In un esperimento con n ripetizioni indipendenti il valore atteso della variabile aleatoria continua è la media dei valori assunti quando il numero di ripetizioni tende a infinito. Viene indicato con E(X)

Media → è il valore atteso di X e si indica con media $x = E(X)$

Varianza → indicata con σ_x^2 è definita come il valore atteso del quadrato degli scarti della variabile aleatoria della sua media, $(X - \text{media } x)^2$.

$\sigma_x^2 = E[(X - \text{media } x)^2]$, ma si può calcolare anche come $\sigma_x^2 = E(X^2) - (\text{media } x)^2$

La **deviazione standard** è la radice quadrata della devianza, ma con segno positivo

Media

$$\mu = \frac{a+b}{2}$$

Varianza

$$\sigma^2 = \frac{(b-a)^2}{12}$$

(formule da utilizzare)

Trasformazioni lineari id variabili aleatorie

Sia X una variabile aleatoria e siano a e b due costanti prefissate. Definiamo la variabile Y come:

$$Y = a + bX$$

La media sarà → $a + b(\text{media } x)$

La varianza sarà → $b^2 * \text{varianza di } x$

Un caso particolare di questi risultati è la variabile aleatoria standardizzata:

$$Z = \frac{X - \text{media } x}{\text{deviazione standard } x} \rightarrow \text{ha media 0 e varianza 1.}$$

Distribuzione normale

È il tipo di distribuzione più utilizzato perché:

- Descrive bene molti fenomeni
- Ha proprietà matematiche convenienti
- Il teorema limite centrale afferma che asintoticamente (al crescere del numero di osservazioni) la distribuzione della media campionaria tende ad una normale, qualunque sia la distribuzione di probabilità delle osservazioni

Caratteristiche:

- Ha una forma campanulare
- È simmetrica
- Media, mediana e moda coincidono
- La tendenza centrale è determinata dal parametro media
- La variabilità è determinata dalla deviazione standard (la varianza va da 0 a +infinito)
- La variabile ha un campo di variazione che va da -infinito a +infinito

Per ogni coppia (media, della normale è

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}\right]}$$

varianza) la funzione di densità

Se si cambia la media la distribuzione si sposta verso sinistra o destra. Cambiando invece la deviazione standard aumenta o diminuisce la dispersione. Media e varianza sono due parametri distinti nella distribuzione normale → la varianza non dipende dalla media, come invece accade per molte distribuzioni.

Proprietà

Per ogni coppia di media e varianza la funzione di densità ha le seguenti caratteristiche:

- È positiva per ogni x reale
- L'area sottesa alla curva è 1
- È simmetrica unimodale, per cui la media coincide con mediana e moda
- Quando x tende a +- infinito la curva tende a 0 senza mai diventare 0 (l'asse delle ascisse è un asintoto della curva)
- La moda è il punto in cui la curva ha la massima altezza

Come si calcola la probabilità?

- Dato che non esiste una formula per la funzione di ripartizione è necessario utilizzare una approssimazione numerica
- Utilizziamo le tavole
- Poi affrontiamo il problema per una normale generica e poi ci basiamo sulla standardizzazione per riformulare il problema in termini di normale standard.

Normale standard

È il membro con media 0 e varianza 1 e funge da rappresentante della famiglia

La tavola della Normale standard data nel libro (Tavola 1 dell'Appendice) fornisce i valori della funzione di ripartizione della distribuzione normale ottenuti tramite approssimazione numerica. Per un dato valore a di Z , la tavola fornisce $f(a)$.

Standardizzazione

Data una qualunque v.a. X , si definisce standardizzata la v.a. Z

$$Z = \frac{X - \mu_X}{\sigma_X}$$

Come calcolare $P(a < X < b)$ quando X ha distribuzione normale:

- Disegna la curva normale per il problema in termini di x
- Traduci i valori di x in valore di z (standardizzazione)
- Usa la tavola della funzione di ripartizione della normale standard

Problemi diretti e inversi

- Problema diretto → dato un valore di z determinare la probabilità cumulata
- Problema inverso → dato un valore p della probabilità cumulata, determinare il valore Z_p corrispondente

Passi del problema inverso

- Trovare il valore di Z corrispondente alla probabilità data
- Convertire nelle unità di X usando l'inversa della standardizzazione cioè:

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

valori anomali

In una distribuzione normale, un valore viene considerato anomalo se p fuori dall'intervallo (media $x \pm k \cdot \text{deviazione standard } x$), dove k solitamente è presa $= 2$. Un valore non è anomalo in senso assoluto, ma relativamente ad una certa distribuzione. Questo criterio non ha senso se la distribuzione è molto diversa dalla normale. Questo metodo per valutare se i risultati sono attendibili o anomali è chiamato **regola empirica**.

- La distribuzione normale permette di sfruttare una serie di utili proprietà
- Nella maggior parte dei casi, quando la variabile in esame è continua la normale è un modello adeguato, cioè descrive in modo sufficientemente accurato la distribuzione di probabilità.
- Vi sono casi in cui la normale è un modello del tutto inadeguato e quindi usare la normale porta a risultati inattendibili
- Valutare l'ipotesi di normalità significa confrontare la distribuzione osservata con la distribuzione normale
- Per confrontare la distribuzione osservata con la normale sono:
 - Grafici
 - Calcolo delle misure di sintesi
 - Verifica della regola empirica
- La distribuzione normale può essere inadeguata per vari motivi. Due motivi frequenti sono:
 - Asimmetria → i dati hanno una natura fortemente asimmetrica, in questo caso la distribuzione osservata ha media e mediana molto diverse
 - Code pesanti → i dati possono presentare valori estremi
- La distribuzione normale ha come supporto l'intero asse dei numeri reali quindi non assegna probabilità nulla nemmeno a numeri negativi
- Se abbiamo poche osservazioni è difficile stabilire se un certo tipo di distribuzione sia o meno adeguato al nostro caso.

Approssimare la binomiale con la normale

N prove indipendenti, ogni prova ha probabilità di successo pari a p

Valore atteso → $E(X) = \text{media} = np$

Varianza → $np(1-p)$

Funzione di ripartizione →

$$P(X \leq a) \approx P(Y \leq a) = P\left(Z \leq \frac{a - np}{\sqrt{np(1-p)}}\right)$$

Somma di variabili aleatorie

- La media della loro somma è la somma delle loro medie
- Se la covarianza fra ogni coppia di queste variabili aleatorie è 0, allora la varianza della loro somma è la somma delle loro varianze.
- Se le covarianze fra le coppie di variabili non sono 0, la varianza della loro somma è:

$$\text{Var}(X_1 + X_2 + \dots + X_k) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2 + 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k \text{Cov}(X_i, X_j)$$

differenza tra variabili aleatorie

- La media della loro differenza è la differenza fra le loro medie

- Se la covarianza tra X e Y è 0, allora la varianza della loro differenza è
- Se la covarianza tra X e Y non è 0, allora la varianza della loro differenza è

$$\text{Var}(X - Y) = \sigma_X^2 + \sigma_Y^2$$

$$\text{Var}(X - Y) = \sigma_X^2 + \sigma_Y^2 - 2\text{Cov}(X, Y)$$