Modelli Statistici

Enrico

Modello uniforme discreto

> Descrive esperimenti con numero finito di esiti equiprobabili

$$X \sim Ud(x_1, \dots, x_n)$$

Funzione di densità

$$f_X(x; x_1, \dots, x_n) = \begin{cases} \frac{1}{n} & \text{se} \quad x = x_1, \dots, x_n \\ \\ 0 & \text{altrimenti} \end{cases}$$

Inoltre:

$$E(X) = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{n+1}{2} \qquad V(X) = \sum_{i=1}^{n} \frac{(x_i - E(X))^2}{n} = \frac{n^2 - 1}{12}$$

Modello di Bernoulli

> Esperimenti nei quali si vuole solamente verificare se un evento si è verificato o meno.

(esempio n articoli, x difettosi)

$$X \sim Ber(\pi)$$

$$f_x(X) = \pi^x (1 - \pi)^{1-x}$$

con x = 0, 1

Inoltre:

$$E(X) = \pi V(X) = \pi(1 - \pi)$$

Modello binomiale

> Esperimenti che possono essere rappresentati come estrazioni con reinserimento da un'urna di composizione nota (esperimento bernoulliano)

$$X \sim Bi(n, p)$$

Con $S_X = \{0, \dots, n\}$ e n che indica il numero di prove indipendenti, p la probabilità comune di successo. Funzione di densità:

$$f_X(x;\ n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{se} \quad x \in S_X \\ 0 & \text{altrimenti} \end{cases}$$

Inoltre:

$$E(X_i) = np V(X) = np(1-p)$$

Modello Poisson

> Descrive problemi di conteggio quando non c'è una limitazione superiore per il supporto o problemi in cui tale limitazione è irrilevante.

$$X \sim P(\lambda)$$

Con $S_X = \mathbb{N}$ Funzione di densità:

$$f_X(x; \lambda) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{se} \quad x \in S_X \\ 0 & \text{altrimenti} \end{cases}$$

Si ha che:

$$E(X) = \lambda V(X) = E(X^2) - (E(X))^2 = \lambda$$

Quindi media e varianza coincidono a λ

Modello geometrico

> Descrive il tempo di attesa, espresso come il numero di replicazioni indipendenti di un esperimento bernoulliano, con probabilità di successo p per osservare per la prima volta un successo.

$$X \sim Ge(p)$$

con $S_X = \mathbb{N}^+$ Funzione di densità:

$$f_X(x;\lambda) = \begin{cases} (1-p)^{x-1}p & \text{se} \quad x \in S_X \\ 0 & \text{altrimenti} \end{cases}$$

Si ha che:

$$E(X) = \frac{1}{p} V(X) = \frac{1-p}{p^2}$$

Modello caratterizzato dall'assenza di memoria

Modello uniforme continuo

> Descrive esperimenti aleatori rappresentabili come un'estrazione casuale di un numero dall'intervallo $[a, b]$

$$X \sim U(a, b)$$

Con $S_X = [a, b]$ Funzione di densità:

$$fX(x;a,b) = \begin{cases} \frac{1}{b-a} & \text{se} \quad a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$$

Funzione di ripartizione:

$$F_X(x;a,b) = \begin{cases} 0 & \text{se} \quad x < a \\ \frac{x-a}{b-a} & \text{se} \quad a \le x \le b \\ 1 & \text{se} \quad x \ge b \end{cases}$$

Si ha che:

$$E(X) = \int_a^b x\frac{1}{b-a}\,dx = \frac{b+a}{2} V(X) = \frac{(b-a)^2}{12}$$

Modello esponenziale

> Utilizzato per rappresentare durate e tempi di vita o funzionamento, nel caso in cui sia plausibile assumere la proprietà di assenza di memoria o usura.

$$X \sim Esp(\lambda)$$

Con $S_X = [0, +\infty[$ Funzione di densità:

$$f_X(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{se} \quad x \in S_X \\ 0 & \text{altrimenti} \end{cases}$$

Funzione di ripartizione:

$$F_X(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & \text{se} \quad x > 0 \\ 0 & \text{se} \quad x \le 0 \end{cases}$$

Si ha che:

$$E(X) = \frac{1}{\lambda}, \qquad V(X) = \frac{1}{\lambda^2}$$

Modello normale o Gaussiano

> il più importante. Usato in vari contesti, in particolare per descrivere presenza di caratteri antropometrici (struttura e peso) e per approssimare diverse distribuzioni di probabilità discrete e continue.

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

Con $S_X = \mathbb{R}$. La funzione di densità:

$$f_X(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Il parametro μ è sia la moda che la mediana. Inoltre:

$$E(X) = \mu V(X) = \sigma^2$$

Se $\mu = 0$ e $\sigma^2 = 1$ si ottiene la **distribuzione normale standard**: $N(0,1)$.

Modello χ^2

$$Y \sim \chi^2(n)$$

è una variabile casuale continua con supporto $S_Y = [0, +\infty[$ e:

$$E(Y) = nV(Y) = 2n$$

Per $n \to +\infty$ la distribuzione di probabilità tende alla distribuzione normale Funzione di densità:

$$f(x) = \frac{1}{2^{\frac{g}{2}} \int_0^{+\infty} e^{-x} x^{(\frac{g}{2})-1}\, dx} x^{\frac{g}{2}-1} e^{-\frac{x}{2}}$$

Modello t-student

$$T \sim t(n)$$

Variabile casuale continua con supporto $S_t = \mathbb{R}$ e:

$$E(T) = 0 V(T) = \frac{n}{n-2}$$

Anche questo tende alla dist. normale per n grande. La funzione di densità ha formula:

$$f(t) = \frac{\Gamma\left(\frac{g+1}{2}\right)}{\sqrt{\pi \cdot g}\ \Gamma\left(\frac{g}{2}\right)}(1 + \frac{t^2}{g})^{-\frac{g+1}{2}}$$

Dove $\Gamma\left(\frac{g}{2}\right)$ indica l'integrale:

$$\Gamma\left(\frac{g}{2}\right) = \int_0^{+\infty} e^{-x} x^{\frac{g}{2}-1}\, dx$$

Modello F-fisher

$$F \sim F(n, m)$$

Supporto $S_F = [0, +\infty[$ e:

$$E(F) = \frac{m}{m-2}$$

La funzione di densità ha formula:

$$f(x) = \frac{\nu_1^{\nu_1/2}\ \nu_2^{\nu_2/2}\ \Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{\frac{\nu_1}{2}-1}}{(\nu_1 x + \nu_2)^{(\nu_1+\nu_2)/2}}$$

Con parametri ν_1, ν_2 positivi.

Teorema del limite centrale

> Data una successione di V.C. X_i indipendenti con media μ e varianza σ^2, allora la somma e media campionaria standardizzate coincidono.

$$Z_n = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \to Z \sim N(0,1) Z_n = \frac{(\bar{X}_n - \mu)\sqrt{n}}{\sigma} \to Z \sim N(0,1)$$

Le distribuzioni gaussiane approssimate si possono usare anche per le V.C. **non gaussiane** Per n **elevato** si usano le distribuzioni approssimate:

$$\bar{X}_n \sim \left(\mu, \frac{\sigma^2}{n}\right) \qquad S_n \sim (n\mu, n\sigma^2)$$

Inoltre, per n elevato, vale che:

$$P(a < \bar{X}_n \leq b) \doteq \Phi\left(\frac{b-\mu}{(\sigma/\sqrt{n})}\right) - \Phi\left(\frac{a-\mu}{(\sigma/\sqrt{n})}\right) P(a < S_n \leq b) \doteq \Phi\left(\frac{b-n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a-n\mu}{\sigma\sqrt{n}}\right)$$