Statistica Matematica

Git

1 Introduzione

1.1 Funzione generatrice dei momenti

Lezione del 18/02, ultima modifica 04/03

Definizione 1 *Sia X una variabile casuale (discreta o assolutamente continua). Se esiste $t_0 > 0$ tale per cui $\mathbb{E}(e^{tX}) < +\infty$ $\forall t \in (-t_0, t_0)$, chiameremo la funzione*

$$M_X := \mathbb{E}(e^{tX})$$

funzione generatrice dei Momenti di X.

Esempi

1. $X \sim b(1, p)$ con $p \in (0, 1)$. Si ha:

$$\begin{aligned} M_X(t) = \mathbb{E}(e^{tX}) &= \sum_{x=0}^{1} e^{tx}\mathbb{P}(X = x) \\ &= \sum_{x=0}^{1} e^{tx}p^x(1-p)^{1-x} = pe^t + (1-p) \end{aligned}$$

2. $X \sim P(\lambda)$ con $\lambda > 0$. Si ha:

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{x=0}^{+\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{\lambda(e^t-1)}$$

3. $X \sim G(\alpha, \beta)$, ovvero

$$f_X(x; \alpha, \beta) := \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{1}{\beta}x}$$

con $\alpha > 0$, $\beta > 0$, $x > 0$ e

$$\Gamma(\alpha) := \int_0^{+\infty} x^{\alpha-1} e^{-x} \mathrm{d}x$$

(nota: $\alpha \in \mathbb{N} \Longrightarrow \Gamma(\alpha) = (\alpha - 1))$

Abbiamo che:

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_0^{+\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{1}{\beta}x} \mathrm{d}x$$

$$= ...[\text{sostituzione } \sigma := x(\tfrac{1}{\beta} - t)]...$$

$$= \frac{1}{(1 - \beta t)^\alpha}$$

con $t < \frac{1}{\beta}$

Momenti di una variabile casuale

Definizione 2 *Se una variabile casuale ammette FGM derivabile infinite volte in un intorno di $t = 0$ e se tutti i suoi momenti sono finiti, allora definiamo il momento di ordine s non centrato:*

$$\mu'_s := \mathbb{E}(X^s) = \frac{d^s}{dt^s} M_X(t)\,|_{t=0}$$

Il momento di di ordine s centrato in $a \in \mathbb{R}$ è:

$$\mu_s(a) := \mathbb{E}((X - a)^s)$$

Ovvero $\mu'_s = \mu_s(0)$. E' chiaro che $\mu'_1 = \mathbb{E}(X)$. Chiameremo infine momento di ordine s centrato (senza specificare altro, intenderemo centrato in μ'_1):

$$\mu_s := \mathbb{E}((X - \mu'_1)^s)$$

Teorema 1 *Vale la seguente relazione tra momenti centrati e non:*

$$\mathbb{E}((X-\mu'_1)^s) = \sum_{m=0}^{s} (-1)^m \binom{s}{m} \mu'_{s-m} (\mu'_1)^m$$

Osserviamo che $\mu_2 = \mathbb{E}((X-\mu'_1)^2) = Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mu'_2 - (\mu'_1)^2$

Teorema 2 *Date due (o più) v.c. X e Y aventi f densità / f massa f_X e f_Y e fgm $M_X(t)$ e $M_Y(t)$ rispettivamente e assunte X e Y essere indipendenti, allora si ha*

$$M_{X+Y} = M_X(t)M_Y(t)$$

Teorema 3 *Siano X e Y v.c. con funzioni di ripartizione $F_X(x)$ e $F_Y(y)$ rispettivamente. Siano $M_X(t)$ e $M_Y(t)$ le fgm di X e Y. Se $M_X(t) = M_Y(t)$ per ogni t in un intorno dell'origine, allora*

$$X \stackrel{d}{=} Y$$

Osservazione 1 *Il teorema appena visto ci dice sostanzialmente che, se esiste, la fgm caratterizza la distribuzione della corrispondente v.c.*

Esempio Siano $(X_1, ..., X_n)$ risultati della replicazione di un esponenzionale casuiale dicotomico $(X_i \sim b(1,p))$. Vogliamo trovare la distribuzione di $S_n := \sum_{i=1}^{n} X_i$. Calcoliamo quindi la sua fgm:

$$M_{S_n}(t) = \mathbb{E}(e^{tS_n}) = \mathbb{E}(e^{t\sum_{i=1}^{n} X_i})$$

$$\stackrel{TEO1}{=} \prod_{i=1}^{n} \mathbb{E}(e^{tX_i}) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} (pe^t + (1-p)) = (pe^t + (1-p))^n$$

ovvero S_n è distribuita come $b(n,p)$ per il Teorema 2.

Esercizio 1 *Ripetere il calcolo precedente supponendo $X_i \sim P(\lambda), \forall i$.*

1.2 Famiglia Esponenziale a k parametri

Una famiglia di f densità / f massa è detta essere una Famiglia Esponenziale a k parametri $\theta_1, ..., \theta_k$ se la corrispondente f densità / f massa (che è indicizzata da $\theta_1, ..., \theta_k$) può essere scritta come

$$f_X(x;\theta) = C^*(x)D^*(\theta)\{\sum_{m=1}^{k} A_m(\theta)B_m(x)\}$$

dove $C^*(x)$ è una funzione della sola x, $D^*(\theta)$ è una funzione del solo θ, $A_m(\theta)$ è una funzione del solo θ e $B_m(x)$ è una funzione della sola x.
Esempi

1. $X \sim G(\alpha, \beta) \Longrightarrow f_X(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{1}{\beta}x} \mathbb{1}_{\mathbb{R}^+}(x)$, $\alpha > 0$, $\beta > 0$
   $\mathbb{1}_{\mathbb{R}^+}$ è detto supporto della distribuzione. Quindi possiamo riscrivere $f_X(x; \alpha, \beta)$ come

   $$f_X(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \mathbb{1}_{\mathbb{R}^+}(x) exp((\alpha - 1)ln(x) - \frac{1}{\beta}x)$$

   e quindi ponendo $D^*(\alpha, \beta) := \frac{1}{\Gamma(\alpha)\beta^\alpha}$, $C^*(x) := \mathbb{1}_{\mathbb{R}^+}(x)$, $A_1(\alpha, \beta) := (\alpha - 1)$, $B_1(x) := ln(x)$, $A_2(\alpha, \beta) := -\frac{1}{\beta}$ e $B_2(x) := x$, otteniamo $G(\alpha, \beta)$ come famiglia esponenziale con $k = 2$.

2. $X \sim b(n, p) \Longrightarrow f_X(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x} \mathbb{1}_{\{0,1,...,n\}}(x)$ con $n \in \mathbb{N}$ noto. Quindi possiamo riscrivere $f_X(x; n, p)$ come

   $$f_X(x; n, p) = \binom{n}{x} \mathbb{1}_{\{0,1,...,n\}}(x)(1 - p)^n exp(ln(\frac{p}{1 - p})x)$$

   con $\frac{p}{1-p}$ detto odd ratio o parametra naturale della famiglia esponenziale.
   Quindi ponendo $D^*(p) := (1-p)^n$, $C^*(x) := \binom{n}{x} \mathbb{1}_{\{0,1,...,n\}}(x)$, $A_1(p) := ln(\frac{p}{1-p})$, $B_1(x) := x$, otteniamo $b(n, p)$ come famiglia esponenziale con $k = 1$.

Osservazione 2 *Le famiglie di esponenziali di ?...? hanno interessanti proprietà matematiche (proprietà di regolarità).*
Dal punto di vista statistico, ciò si traduce in un'interessante conseguenza: tutta l'informazione contenuta nei dati a disposizione $(X_1, ..., X_n)$ relativa alla funzione $f_X(x; \theta)$ può essere sintetizzata attraverso k quantità (funzioni di $(X_1, ..., X_n)$) che potranno essere impiegate per costruire procedure inferenziali (stima, test per la verifica di ipotesi) riguardanti il parametro θ.

Ovvero, l'appartenenza a una famiglia esponenziale permette una riduzione dei dati $(X_1, ..., X_n)$ via B_m.

1.3 Trasformazioni di variabili casuali

Lezione del 01/03, ultima modifica 03/03

Discrete

Teorema 4 *Sia X una vc con funzione di massa $f_X(x) = P(X = x)$, e sia A_X il suo supporto. Sia W=h(X) una nuova vc. Allora*

$$P(W = w) = \sum_{\{x \in A_X : h(x) = w\}} P(X = x)$$

Esempi

1. Sia $X \sim b(n, p)$ con relativa funzione di massa $f_X(x, p) = \binom{n}{p} p^x (1 - p)^{n-x} \mathbb{1}_{0,1,...,n}(x)$, n noto e $p \in (0, 1)$.

   Considero quindi $W = n - X$. Come si distribuisce W?

$$P(W = w) = P(X = n - w) = \binom{n}{n-w} p^{n-w} (1 - p)^w \mathbb{1}_{0,1,...,n}(w)$$

2. Sia X una vc tale che $f_X(x) = P(X = x) = \left(\frac{1}{2}\right)^x \mathbb{1}_{\mathbb{N}}(x)$, $W = X^3$.

$$P(W = w) = P(X^3 = w) = P(X = \sqrt[3]{w}) = \left(\frac{1}{2}\right)^{\sqrt[3]{w}} \mathbb{1}_{1,8,27,...}(w)$$

Assolutamente continue

Teorema 5 *Sia X una variabile casuale (ass continua) con funzione di densità $f_X(x)$ e sia $W = h(X)$, ove h è una funzione monotona. Supponiamo inoltre che $f_X(x)$ sia continua sul supporto di X e che $h^{-1}(w)$ abbia derivata continua sul supporto di W. Allora*

$$f_W(w) = f_X(h^{-1}(w)) \left| \frac{d}{dw} h^{-1}(w) \right| \mathbb{1}_{A_W}(w)$$

Esempio (Standardizzazione di una vc normale) Sia $X \sim N(m, s^2)$. Considero $W = h(X) = \frac{X-m}{s}$. Allora, dato che $h^{-1}(w) = sw + m$, che ha derivata continua su tutto $\mathbb{R}$,

$$f_W(w) = f_X(sw + m)|s| = \frac{e^{\frac{-w^2}{2}}}{\sqrt{2\pi}}$$

Teorema 6 *Se $W = h(X)$ ove h è monotona a tratti (un numero finito k) e valgono le condizioni del teorema precedente (su ogni tratto), allora*

$$f_W(w) = \sum_{n=1}^{k} f_X(h_n^{-1}(w)) \left| \frac{d}{dw} h_n^{-1}(w) \right| \mathbb{1}_{A_W}(w)$$

Esempio (Chi-quadro): Sia $X \sim N(0,1)$ e $W = h(X) = X^2$. h è monotona sui tratti $A_0 = 0$, $A_1 = (-\infty, 0)$, $A_2 = (0, +\infty)$. Trovo inoltre $h_1^{-1}(w) = -\sqrt{w} \in A_1 \forall w \geq 0$, mentre $h_2^{-1}(w) = \sqrt{w} \in A_2 \forall w \geq 0$. $\frac{d}{dw} h_1^{-1}(w) = -\frac{1}{2\sqrt{w}}$, $\frac{d}{dw} h_2^{-1}(w) = \frac{1}{2\sqrt{w}}$ sono entrambe continue su $\mathbb{R}_+$.

$$f_W(w) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(-\sqrt{w})^2}{2}} \left| \frac{1}{2\sqrt{w}} \right| + \frac{1}{\sqrt{2\pi}} e^{\frac{-(\sqrt{w})^2}{2}} \left| \frac{1}{2\sqrt{w}} \right|$$

$$= \frac{1}{\sqrt{2\pi w}} e^{\frac{-w}{2}} \mathbb{1}_{\mathbb{R}+}(w) = \frac{1}{2^{1/2}\Gamma(1/2)} w^{\frac{1}{2}-1} e^{\frac{-w}{2}}$$

Si riconosce che $W \sim \mathcal{G}(\alpha = 1/2, \beta = 2)$ e si chiama Chi quadrato con $\nu = 1$ gradi di libertà. In generale, una vc Chi Quadro con $\nu = n$ gradi di libertà è $W = \sum_{i=1}^{n} X_i^2$, ove $X_1, X_2, ..., X_n$ sono vc iid N(0,1). Per il teorema sulla FGM di una somma di vc iid si trova immediatamente che $W \sim \mathcal{G}(\alpha = n \cdot 1/2, \beta = 2)$.

1.4 Convergenze

Convergenza in probabilità

Definizione 3 *Sia $\{X_n\}_{n \in \mathbb{N}}$ una successione di variabili casuali e sia X un'altra variabile casuale, tutte definite sullo stesso spazio campionario. Diciamo che X_n converge in probabilità a X (scriviamo $X_n \xrightarrow{p} X$) se $\forall \varepsilon > 0$*

$$\lim_{n \to \infty} P(|X_n - X| \leq \varepsilon) = 0$$

Osservazione 3 *Se $X_n \xrightarrow{p} X$ diciamo che la "massa" della differenza $|X_n - X|$ converge a 0. Inoltre, quando scriviamo $X_n \xrightarrow{p} X$, stiamo sottintendendo tutta la parte iniziale della definizione precendete, cioè il "sia $\{X_n\}_{n \in \mathbb{N}}$ una successione di variabili casuali...".*

Teorema 7 *Alcuni risultati utili:*

1. *Supponiamo che $X_n \xrightarrow{p} X$ e $Y_n \xrightarrow{p} Y$. Allora $X_n + Y_n \xrightarrow{p} X + Y$*

2. *Supponiamo che* $X_n \xrightarrow{p} X$ *e sia a una costante. Allora* $aX_n \xrightarrow{p} aX$
3. *Supponiamo che* $X_n \xrightarrow{p} a$ *costante, e sia g una funzione reale continua in a. Allora* $g(X_n) \xrightarrow{p} g(a)$
4. *(Corollario di 3.) Se* $X_n \xrightarrow{p} a$, *allora* $X_n^2 \xrightarrow{p} a^2$, $\frac{1}{X_n} \xrightarrow{p} \frac{1}{a}$ *(se* $a \neq 0$*),* $\sqrt{X_n} \xrightarrow{p} a$ *(*$a \geq 0$*).*
5. $X_n \xrightarrow{p} X$ *e* $Y_n \xrightarrow{p} Y$ *allora* $X_nY_n \xrightarrow{p} XY$

Convergenza in distribuzione

Definizione 4 *Sia* $\{X_n\}_{n\in\mathbb{N}}$ *una successione di variabili casuali e sia X un'altra variabile casuale, tutte definite sullo stesso spazio campionario.*

Siano F_{X_n} *e* F_X *le relative funzioni di ripartizione (di distribuzione, sinonimo). Sia* $C(F_X)$ *l'insieme dei punti ove* F_X *è continua. Diciamo che* X_n *converge in distribuzione (o in legge) a X (scriviamo* $X_n \xrightarrow{d} X$*) se*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \forall x \in C(F_X)$$

Esempio

Teorema 8 *Se* $X_n \xrightarrow{p} X$ *allora* $X_n \xrightarrow{d} X$.

Osservazione 4 *Il contrario in generale non vale, tranne nel caso in cui X è una vc degenere (cioè costante).*

Teorema 9 *Supponiamo che* $X_n \xrightarrow{d} X$ *e sia g una funzione continua sul supporto di X. Allora* $g(X_n) \xrightarrow{d} g(X)$

Teorema 10 (Slutsky) *Supponiamo che* $X_n \xrightarrow{d} X$, $A_n \xrightarrow{p} a$ *costante e* $B_n \xrightarrow{p} b$ *costante. Allora* $A_n + B_nX_n \xrightarrow{d} a + bX$

1.5 Teoria asintotica

Lezione del 04/03, ultima modifica 05/03

Teorema 11 *(*Δ*-method) Sia* $\{X_n\}_{n\in\mathbb{N}}$ *una successione di vc tale che* $\sqrt{n}(X_n - \vartheta) \xrightarrow{d} N(0, \sigma^2)$. *Supponiamo che una funzione g(X) sia derivabile in* ϑ *e che* $g'(\vartheta) \neq 0$. *Allora*

$$\sqrt{n}(g(X_n) - g(\vartheta)) \xrightarrow{d} N(0, \sigma^2(g'(\vartheta))^2)$$

Esempi

1. Considero
$$Y_n = \frac{\chi_n^2 - n}{\sqrt{2n}} = \sqrt{n}\left(\frac{\chi_n^2}{\sqrt{2}n} - \frac{1}{\sqrt{2}}\right)$$
ove χ_n^2 è la chiquadro con n gradi di libertà. Ricordiamo che $\mathbb{E}(\chi_n^2) = n$ e che $Var(\chi_n^2) = 2n$ (discende dal fatto che $\chi_n^2 \sim \mathcal{G}(\alpha = n/2, \beta = 2)$. Siccome $Y_n \xrightarrow{d} N(0,1)$, scrivendo Y_n nella forma $Y_n = \sqrt{n}\left(\frac{\chi_n^2}{\sqrt{2}n} - \frac{1}{\sqrt{2}}\right)$ riconosciamo che la prima parte delle ipotesi del Δ-method sono soddisfatte. Considero quindi $g(t) = \sqrt{t}$, che è derivabile in $\vartheta = 1/\sqrt{2}$, $g'(t) = \frac{1}{2\sqrt{t}}|_{\vartheta=1/\sqrt{2}} = 2^{-3/4}$. Allora
$$\sqrt{n}(g\left(\frac{\chi_n^2}{\sqrt{2}n}\right) - g(\vartheta)) = \sqrt{n}\left(\sqrt{\frac{\chi_n^2}{\sqrt{2}n}} - \sqrt{\frac{1}{\sqrt{2}}}\right) \xrightarrow{d} N(0, 1^2 \cdot 2^{-3/2})$$

Teorema 12 *(Teorema centrale del limite) Siano $X_1, ...X_n$ vc iid dotate di media μ e varianza finita σ^2. Allora*
$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\cdot\sigma} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

Esempi/Applicazioni

1. $X \sim b(n,p)$, $X \overset{a}{\sim} N(np, np(1-p))$
2. $X_1, ..., X_n$ vc $P(\lambda = 1)$. Considero $Y_n = \sum X_i$. Dato che $Y_n \overset{a}{\sim} N(n\lambda, n\lambda)$, $\frac{Y_n}{n} \overset{a}{\sim} N(1, 1/n)$
3. Considerato $W_n = \sqrt{n}(Y_n/n - 1) = \frac{Y_n/n-1}{1/\sqrt{n}} = \frac{\bar{Y_n} - \mathbb{E}(\bar{Y_n})}{\sqrt{Var(Y_n/n)}}$

Teorema 13 *Sia $\{X_n\}$ una succ di vc iid con FGM $M_{X_n}(t)$ definita e $< \infty$ per $t \in (-h, h) \forall n$, e sia X un'altra vc con FGM $M_X(t)$ definita e $< \infty$ per $t \in (-h_1, h_1), h_1 \leq h$. Se*
$$\lim_{n\to+\infty} M_{X_n}(t) = M_X(t) \forall |t| \leq h_1$$
allora $X_n \xrightarrow{d} X$.

Applicazione

Sia $X_n \sim b(n,p)$. Ricordiamo che $X_n = \sum X_i$ ove $X_i \sim b(1,p)$, ed inoltre $\mu = \mathbb{E}(X) = np$. Siccome $M_{X_n}(t) = \mathbb{E}(e^{tX_n}) = [(1-p) + pe^t]^n = [1 + \frac{\mu}{n}(e^t - 1)]^n$,
$$M_{X_n}(t) \overset{n\to\infty}{\longrightarrow} e^{\mu(e^t - 1)}$$
che è la FGM di una Poisson di parametro μ.

2 Approccio alla Statistica Matematica

2.1 Introduzione

Definizione 5 *(Campione Casuale) Il vettore casuale* $(X_1, ..., X_n)$ *si dice Campione Casuale relativamente ad una vc* $X \sim F_X(x, \vartheta)$ *se i suoi elementi sono vc i.i.d.*

Osservazione Il fatto che le vc siano i.i.d. implica che

$$F_{X_1,...,X_n}(X_1, ..., X_n) = \prod_{i=1}^{n} F_{X_i}(X_i)$$

e

$$f_{X_1,...,X_n}(X_1, ..., X_n) = \prod_{i=1}^{n} f_{X_i}(X_i)$$

Definizione 6 *(Statistica) Sia* $(X_1, ..., X_n)$ *un campione casuale da una distribuzione associata alla vc* X*, e sia* Ω *lo spazio campionario di* $(X_1, ..., X_n)$ *Ogni funzione*

$$T(X_1, ..., X_n) : \Omega \longrightarrow \mathbb{R}^k$$

che NON dipende da parametri incogniti è detta Statistica.

Osservazioni Le cose scritte tra virgolette "" sono concetti e/o definizioni non ancora introdotti, che vengono usati per dare un'idea intuitiva di quello che si andrà a vedere, cose che poi durante il corso verranno trattate con rigore.

1. Una statistica T è una "caratteristica numerica" del campione: si presta a *sintetizzare* l'informazione su ϑ contenuta nel campione.

2. $\sum_{i=1}^n X_i$ e $\sum_{i=1}^n X_i^2$ sono entrambe statistiche: sono alla base di due "stimatori" molto importanti:

   Media Campionaria: $\bar{X}_n = \frac{1}{n} \sum X_i$

   Varianza Campionaria: $S_n^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2 = \frac{1}{n-1} \sum X_i^2 - \frac{n}{n-1} \bar{X}_n^2$

3. Ogni statistica è una vc: ha quindi una distribuzione, che dipende dal parametro.

   Esempio Considero $\bar{X} = \frac{1}{n} \sum X_i$ ove $X_i \sim N(\mu, \sigma^2)$. Allora $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. Da questo si potrà dedurre la "bontà" di $\bar{X}$ come "stimatore" di μ.

4. Tra tutti i modi di sintetizzare l'informazione contenuta in $(X_1, ..., X_n)$ relativamente a ϑ, siamo interessati a quelli che NON tralasciano informazioni o quote di informazioni rilevanti per il parametro.

5. In relazione ad uno stimatore potremmo essere interessati ad alcune proprietà, in particolare a queste due:

   · Accuratezza [concetto legato alla media dello stimatore] ($Non\ distorsione$)

   · Precisione [concetto legato alla varianza dello stimatore] ($efficienza$ o $consistenza$)