Politecnico di Milano – Facoltà di Ingegneria Informatica

Anno Accademico 2009/2010

Corso di Statistica (2L) per INF e TEL

Docente: Antonio Pievatolo; esercitazioni: Raffaele Argiento

Esercitazione del 19/03/10

Alcuni risultati sulla somma di variabili aleatorie indipendenti

1. Siano $X_1, \ldots, X_n$ variabili indipendenti ed identicamente distribuite (i.i.d.) con distribuzione Bernoulli(p). Allora
$$Y = \sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$$
2. Siano $X_1 \sim$Bin(m_1, p), $X_2 \sim$Bin(m_2, p),..., $X_n \sim$Bin(m_n, p), allora
$$Y = \sum_{i=1}^{n} X_i \sim \text{Bin}(\sum_{i=1}^{n} m_i, p)$$
3. Siano $X_1 \sim$Poisson(λ_1), $X_2 \sim$Poisson(λ_2),..., $X_n \sim$Poisson(λ_n), allora
$$Y = \sum_{i=1}^{n} X_i \sim \text{Poisson}(\sum_{i=1}^{n} \lambda_i)$$
4. Siano $X_1 \sim$gammma(α_1, β), $X_2 \sim$gamma(α_2, β),..., $X_n \sim$gamma(α_n, β), allora
$$Y = \sum_{i=1}^{n} X_i \sim \text{gamma}(\sum_{i=1}^{n} \alpha_i, \beta)$$
5. (Caso particolare di 4.) Siano $X_1, \ldots, X_n$ i.i.d con disribuione Exp(β)
$$Y = \sum_{i=1}^{n} X_i \sim \text{gamma}(n, \beta)$$
6. (Caso particolare di 4.) Siano $X_1 \sim \chi^2(m_1)$, $X_2 \sim \chi^2(m_2)$,..., $X_n \sim \chi^2(m_n)$, dove con $\chi^2(m)$ è la distribuzione chi-quadro con m gradi di libertà, allora
$$Y = \sum_{i=1}^{n} X_i \sim \chi^2(\sum_{i=1}^{n} m_i)$$

Un risultato sulla densità di variabili aleatorie funzione di variabili aleatorie

Sia X una variabile aleatoria assolutamente continua con densità $f_X(x)$, con $x \in \mathbb{R}$. Sia $g(x)$ una funzione derivabile strettamente crescente o strettamente decrescente. Allora la variabile aleatoria definita da $Y = f(X)$ è anch'essa assolutamente continua con densità:

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \left| \frac{d}{dy} g^{-1}(y) \right| \tag{1}$$

con $y \in (\alpha, \beta)$ dove $\alpha = \min\{g(-\infty), g(+\infty)\}$ e $\beta = \max\{g(-\infty), g(+\infty)\}$.

Esempio 1

Sia $X \sim$gamma(α, β) e sia $Y = aX$ con $a > 0$ qual è la distribuzione di Y?

Volendo applicare il risultato prima appena descritto si osservi che la funzione $g(x) = ax$ è derivabile e strettamente crescente in $(0, +\infty)$. La sua inversa è $g^{-1}(y) = \frac{y}{a}$ con derivata $\frac{d}{dy}g^{-1}(y) = \frac{1}{a}$. Si ha dunque che

$$f_Y(y) = \frac{1}{\Gamma(\alpha)} \frac{1}{(a\beta)^\alpha} y^{\alpha-1} \mathrm{e}^{-\frac{x}{a\beta}}.$$

In conclusione $Y \sim$gamma$(\alpha, a\beta)$

Esempio 2

Sia $X \sim$N(μ, σ^2) e $Y = aX + b$ con $a \neq 0$, determinare la densità di Y.

Si osservi che la funzione $g(x) = ax+b$ è strettamente crescente se $(a > 0)$ o strettamete decrescente se $a < 0$, e che la sua inversa è $g^{-1}(y) = \frac{y-b}{a}$ con derivata $\frac{d}{dy}g^{-1}(y) = \frac{1}{a}$. Utilizzando la (1) si ottiene

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma a} \exp\left\{ -\frac{(y - (a\mu + b))^2}{(a\sigma)^2} \right\}.$$

In conclusione $Y \sim$N$(a\mu + b, a^2\sigma^2)$.

Esercizio 1.2.2

Esercizio 1.2.2 La durata di una batteria espressa in ore è una variabile aleatoria assolutamente continua con densità $f(x, \theta)$ data da

$$f(x, \theta) = \frac{1}{2\theta\sqrt{x}} \exp\left\{ -\frac{\sqrt{x}}{\theta} \right\}, \qquad x > 0, \; \theta > 0.$$

1. Determinate la densità di $\sqrt{X}$ se $X \sim f(x, \theta)$.
2. Determinate in funzione di θ la probabilità che la batteria funzioni ancora dopo 11 ore dall'accensione.
3. Determinate la media di $X \sim f(x, \theta)$ in funzione di θ.

Soluzione

1. Si osservi che la funzione $g(x) = \sqrt{x}$ è derivabile e crescente su $(0, \infty)$. Inoltre $g^{-1}(y) = y^2$ e $\frac{d}{dy}g^{-1}(y) = 2y$. Applicando allora la formula (1) si ha che

$$f_Y(y) = \frac{1}{2\theta\sqrt{y^2}} \exp\left\{ -\frac{\sqrt{y^2}}{\theta} \right\} 2y = \frac{1}{\theta} \exp\left\{ -\frac{y}{\theta} \right\}, \quad y > 0.$$

   Si ha dunque che $Y \sim$Exp(θ)

   attenzione. questa funzione di ripartizione sembra avere l'integrale errato!!

2. Se $Y \sim$Exp(θ) allora la sua funzione di ripartizione sarà $F_Y(y) = 1 - \exp\left\{\frac{y}{\theta}\right\}$. Quindi

$$\mathbb{P}(X > 11) = \mathbb{P}(Y > \sqrt{11}) = 1 - F_Y(\sqrt{11}) = \exp\left\{ \frac{\sqrt{11}}{\theta} \right\}$$

   1 - (1 - EXP(y/theta)) =
   1 - 1 + EXP(y(theta)

3. $\mathbb{E}(X) = \mathbb{E}(Y^2) = 2\theta^2$.

∎

Esercizio 1.3.1

Sia $X_1, \ldots, X_n$ un campione casuale da $\mathrm{N}(4.2, 4)$ e sia $\bar{X}_n$ la media campionaria.

1. Calcolate $P(|\bar{X}_n - 4.2| \leq 0.3)$, con $n = 4$ e $n = 25$ e confrontate i risultati;

2. Per quali valori di n $P(|\bar{X}_n - 4.2| \leq 0.3) \geq 0.8$.

perché: var media campionaria è Sigma^2/n. l'altra la traslo in zero.

Soluzione Si osservi che $\bar{X}_n - 4.2$ ha distribuzione $\mathrm{N}(0, \frac{4}{n})$, allora sia $Z \sim \mathrm{N}(0,1)$

1. $\mathbb{P}(-0.3 \leq \bar{X}_n - 4.2 \leq 0.3) = \mathbb{P}(-0.3\frac{\sqrt{n}}{2} \leq Z \leq 0.3\frac{\sqrt{n}}{2}) = 2\phi\left(0.3\frac{\sqrt{n}}{2}\right) - 1$ Quindi se $n = 4$ si ottiene $P(|\bar{X}_4 - 4.2| \leq 0.3) = 2\phi(0.3) - 1 = 0.2358$ moltiplico per la deviazione standard della media campionaria
se $n = 25$ si ottiene $P(|\bar{X}_4 - 4.2| \leq 0.3) = 2\phi(0.75) - 1 = 0.5467$
Osserviamo come, al crescere di n, cresce la probabilità che la media campionaria disti meno di 0.3 dalla sua media.

2. $P(|\bar{X}_n - 4.2| \leq 0.3) \geq 0.8 \Rightarrow 2\phi\left(0.15\sqrt{n} - 1 \geq 0.8 \Rightarrow \phi(-.15\sqrt{n}) \geq 0.9 \Rightarrow 0.15\sqrt{n} \geq \phi^{-1}(0.9).\right)$
Utilizzando quindi le tavole della distribuzione normale si ottiene $n \geq 72.99$ ovvero, dato che n è un intero, $n \geq 73$.

∎

Esercizio 1.3.3

Sia $Z_1, \ldots, Z_n$ un campione casuale estratto dalla popolazione di densità gaussiana standard e sia Sia $S_n = Z_1^2 + \ldots + Z_n^2$. In teoria delle probabilità una distribuzione (chi quadrato o chi quadro) è una distribuzione di probabilità che descrive la somma dei quadrati di alcune variabili aleatorie indipendenti aventi distribuzione normale standard.

1. Calcolate media e varianza di S_n;

2. Sia $n = 18$. Calcolate $\mathbb{P}(S_n \leq 9.39)$.

3. Sia $n = 300$. Calcolate $\mathbb{P}(S_n \leq 312.98)$.

Soluzione S_n è somma di n variabili aleatore indipendenti con distribuzione chi quadro con un grado di libertà. Dunque $S_n \sim \chi^2(n) \stackrel{d}{=} \mathrm{gamma}(\frac{n}{2}, 2)$.

1. $\mathbb{E}(S_n) = \frac{n}{2} \times 2 = n$, mentre $\mathrm{Var}(S_n) = \frac{n}{2} \times 2^2 = 2n$

2. Dalle tavole della distribuzione chi quadro $\mathbb{P}(S_{18} \leq 9.39) = 0.05$.

3. La distribuzione chi quadro con n gradi di liberà non è tabulata per valori di $n > 60$. In questo caso si può applicare il teorema del limite centrale ottenedo che approssimativamente

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n}{\sqrt{2n}} \stackrel{approx}{\sim} \mathrm{N}(0,1).$$

Dunque $\mathbb{P}(S_n \leq 312.98) = \mathbb{P}(\frac{S_n - n}{\sqrt{2n}} \leq \frac{312.98 - n}{\sqrt{2n}}) \simeq \mathbb{P}(Z \leq \frac{312.98 - 300}{\sqrt{600}}) = \mathbb{P}(Z \leq 0.53) = 0.7019$. Dove $Z \sim \mathrm{N}(0,1)$.

∎

Esercizio 1.3.4

Abbiamo estratto il campione casuale $X_1, \ldots, X_n$ dalla popolazione di densità esponenziale di parametro θ: $f(x,\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \mathbb{1}_{(0,\infty)}(x), \ \theta > 0$

1. Determinate la densità della media campionaria $\bar{X}$.

2. Determinate la densit'a della variabile aleatoria $\frac{2n\bar{X}}{\theta}$

3. Sia $\alpha = 5\%,\ n = 3$ e $\theta = 2$: determinate k tale che $\mathbb{P}(\bar{X} \leq k) = \alpha$
4. Sia $n = 3$ e $\theta = 1.49$: calcolate $\mathbb{P}(\bar{X} > \frac{1.64}{3})$
5. Sia $n = 35$ e $\theta 1.49$: calcolate $\mathbb{P}(\bar{X} > \frac{1.64}{3})$
6. Determinate k dipendente da θ e da n tale che $\mathbb{P}(\sum_{j=1}^n X_j > k) = 0.9$

Soluzione

1. Si osservi che $X_1 + \cdots + X_n \sim$gamma(n, θ) e quindi $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \sim$gamma$(n, \frac{\theta}{n})$.
2. Si ha che $\frac{2n\bar{X}}{\theta} \sim$gamma$(n, 2) \stackrel{d}{=} \chi^2(2n)$ no no. dovrebbe essere al quadrato! non viene fuori una chi quadro
3. $\mathbb{P}(\bar{X} \leq k) = \alpha \Rightarrow \mathbb{P}(\frac{2n\bar{X}}{\theta} \leq \frac{2nk}{\theta}) \Rightarrow \frac{2nk}{\theta} = q_{\chi^2_{2n}}(\alpha)$ dove con $q_{\chi^2_n}$ si è indicata le funzione prcentile di una chi quadro con n gradi di libertà. Con l'ausilio delle tavole si ottiene $k = \frac{1.64}{3}$
4. $\mathbb{P}(\bar{X} > \frac{1.64}{3}) = \mathbb{P}(\frac{2n\bar{X}}{\theta} > \frac{2n}{\theta}\frac{1.64}{3}) = \mathbb{P}(\chi^2_6 > 2.2013) = 0.9$ Abbiamo utilizzato il simbolo χ^2_6 si per una v.a. che per la sua distribuzione.
5. $\mathbb{P}(\bar{X} > \frac{1.64}{3}) = \mathbb{P}(\frac{2n\bar{X}}{\theta} > \frac{2n}{\theta}\frac{1.64}{3}) = \mathbb{P}(\chi^2_{70} > 25.6823) \simeq \mathbb{P}(Z > \frac{25.6823-70}{\sqrt{140}} = \mathbb{P}\left(Z > -375\right) = 0.9999 \simeq 1$ Nelle ultime uguaglianze si è usato il Teorema del limite centrale, e Z al solito ha distribuzione $\mathrm{N}(0, 1)$.
6. $\mathbb{P}\left(\sum_{j=1}^n X_i > k\right) = 0.9 \Rightarrow \mathbb{P}\left(\frac{2}{\theta}\sum_{j=1}^n X_i \leq \frac{2}{\theta}k\right) = 0.1 \Rightarrow \mathbb{P}\left(\chi^2_{2n} < \frac{2}{\theta}k\right) = 0.1$. Qindi $\frac{2}{\theta}k = q_{\chi^2_{2n}}(0.1) \Rightarrow k = \frac{\theta}{2}q_{\chi^2_{2n}}(0.1)$. Dove con $q_{\chi^2_n}(p)$ si è indicata la funzione quantile di una chi quadro con due gradi di libertà.

∎