χ^2_n, $t(n)$ e campioni gaussiani, stimatori di massima verosimiglianza

Titolo nota 20/11/2017

$\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ distribuzione di Pearson a n gradi di libertà o anche χ^2_n.

$$f(x) = \begin{cases} \frac{1}{\Gamma(\frac{n}{2})} \left(\frac{1}{2}\right)^{n/2} x^{n/2 - 1} \exp\left(-\frac{x}{2}\right) & x > 0 \\ 0 & x \leq 0 \end{cases}$$

X e Y indipendenti $\mathbb{P}_X = \chi^2_n$, $\mathbb{P}_Y = \chi^2_k$

$\Rightarrow \mathbb{P}_{X+Y} = \chi^2_{n+k}$

$\mathbb{P}_X = N(0,1) \Rightarrow \mathbb{P}_{X^2} = \chi^2_1$

$X_1 \dots X_n$ è un campione gaussiano standard

$\Rightarrow \sum_{i=1}^n X_i^2$ ha distribuzione χ^2_n

$X_1 \dots X_n$ è un campione gaussiano $N(\mu, \sigma^2)$

$Z_i = \frac{X_i - \mu}{\sigma}$ ha distribuzione $N(0,1)$

$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}$ ha distribuzione χ^2_n.

TEO Sia $X_1 \dots X_n$ campione gaussiano con valore atteso μ e varianza σ^2

Allora la media campionaria $\overline{X}$ e la varianza campionaria S^2 sono variabili indipendenti.

Inoltre se $Z_1 \dots Z_n$ è un campione gaussiano standard allora $\sum_{i=1}^n (Z_i - \overline{Z})^2$ ha distribuzione χ^2_{n-1}

Osservazione Se $Z_i = \frac{X_i - \mu}{\sigma}$ $\quad \overline{Z} = \frac{1}{n} \sum_{i=1}^n Z_i =$

$$= \frac{1}{n\sigma} \sum_{i=1}^n (X_i - \mu) = \frac{1}{\sigma}\left(\frac{1}{n} \sum X_i - \frac{1}{n} n\mu\right) = \frac{1}{\sigma}(\overline{X} - \mu)$$

$$\sum_{i=1}^{n} (Z_i - \bar{Z})^2 = \sum_{i=1}^{n} \left(\frac{X_i - \cancel{\mu}}{\sigma} - \frac{\bar{X} - \cancel{\mu}}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$\Rightarrow$ il Teorema dice che se $X_1 \text{ — } X_n$ è un campione gaussiano $N(\mu, \sigma^2)$ $\Rightarrow$ $\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2$

— $\Rightarrow$ —

DIM (traccia)

$n = 2$ $\quad \bar{X} = \frac{X_1 + X_2}{2} \longleftarrow$

$$S^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 = \left(X_1 - \frac{X_1 + X_2}{2} \right)^2 + \left(X_2 - \frac{X_1 + X_2}{2} \right)^2$$

$$= \frac{1}{4} (X_1 - X_2)^2 + \frac{1}{4} (X_2 - X_1)^2 = \frac{1}{2} \underline{(X_1 - X_2)^2}$$

$\bar{X}$ e S^2 sono v.a. indipendenti SSE $X_1 + X_2$ e $X_1 - X_2$ sono indipendenti

$\mathbb{P}_{X_1} = \mathbb{P}_{X_2} = N(\mu, \sigma^2) \qquad \mathbb{P}_{X_1 + X_2} = N(2\mu, 2\sigma^2)$

$\mathbb{P}_{-X_2} = N(-\mu, \sigma^2) \quad \mathbb{P}_{X_1 - X_2} = N(0, 2\sigma^2)$

$U = X_1 + X_2, \quad V = X_1 - X_2$

$(U, V) = \varphi(X_1, X_2) \qquad \varphi : (x, y) \in \mathbb{R}^2 \mapsto (x + y, x - y) \in \mathbb{R}^2$

$\psi : \mathbb{R}^2 \to \mathbb{R}$ di Borel non negativa

$$\int_{\mathbb{R}^2} \psi(u, v) \, \mathbb{P}_{U,V}(du\,dv) = \int_{\mathbb{R}^2} \psi(\varphi(x, y)) \, \mathbb{P}_{X_1, X_2}(dx\,dy)$$

$$= \int_{\mathbb{R}^2} \psi(x + y, x - y) \, \mathbb{P}_{X_1}(dx) \, \mathbb{P}_{X_2}(dx_2) =$$

$$= \int_{\mathbb{R}^2} \psi(x + y, x - y) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2} \right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y - \mu)^2}{2\sigma^2} \right) dx\,dy$$

$u = x+y \qquad x = \frac{u+v}{2}$

$v = x-y \qquad y = \frac{u-v}{2} \qquad |\det J| = \frac{1}{2}$

$$\int_{\mathbb{R}^2} f(u,v)\, \frac{1}{2}\, \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{4\sigma^2}(u-\mu)^2 - \frac{1}{4\sigma^2}v^2\right) du\,dv$$

$$= \int_{\mathbb{R}^2} f(u,v)\, \underbrace{\frac{1}{\sqrt{2\pi(\sigma\sqrt{2})^2}} \exp\left(\frac{-(u-2\mu)^2}{2(\sigma\sqrt{2})^2}\right)}_{N(2\mu,\, 2\sigma^2)} \frac{1}{\sqrt{2\pi(\sigma\sqrt{2})^2}} \exp\left(\frac{-v^2}{2(\sigma\sqrt{2})^2}\right) du\,dv$$

$N(0, 2\sigma^2)$

$\mathbb{P}_{X_1+X_2} \qquad \mathbb{P}_{X_1-X_2}$

$$(Z_1 - \bar{Z})^2 + (Z_2 - \bar{Z})^2 = \left(Z_1 - \frac{Z_1+Z_2}{2}\right)^2 + \left(Z_2 - \frac{Z_1+Z_2}{2}\right)^2$$

$$= \frac{1}{2}(Z_1 - Z_2)^2 = \left(\frac{1}{\sqrt{2}}(Z_1 - Z_2)\right)^2$$

$\mathbb{P}_{Z_1} = \mathbb{P}_{Z_2} = N(0,1) \qquad \mathbb{P}_{Z_1 - Z_2} = N(0, 2)$

$\frac{Z_1 - Z_2}{\sqrt{2}}$ ha distribuzione $N(0,1)$

$\Rightarrow$ il suo quadrato ha distribuzione χ^2_1 .

$\underline{n \geq 3}$ $\qquad \bar{X}_{n-1}$ e S^2_{n-1} le suppongo indipendenti.

$$S^2_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \Rightarrow \bar{X}_n = \bar{X}_{n-1} + \frac{1}{n}(X_n - \bar{X}_{n-1})$$

indip

$$S^2_n = \frac{1}{n-1}\left((n-2)S^2_{n-1} + \frac{n-1}{n}\underbrace{(X_n - \bar{X}_{n-1})^2}_{}\right)$$

se questo è indipendente da $\bar{X}_n$

sono a posto

$$V = X_n - \overline{X}_{n-1}$$

$$U = \overline{X}_n = \overline{X}_{n-1}\left(1 - \frac{1}{n}\right) + \frac{1}{n} X_n$$

X_n e $\overline{X}_{n-1}$ sono indipendenti

$$\mathbb{P}_{X_n} = N(\mu, \sigma^2) \qquad \mathbb{P}_{\overline{X}_{n-1}} = N\left(\mu, \frac{\sigma^2}{n-1}\right)$$

$$(U, V) = \varphi \circ (\overline{X}_{n-1}, X_n) \qquad \varphi(x,y) = \left(\frac{n-1}{n} x + \frac{1}{n} y,\ y - x\right)$$

$$\int_{\mathbb{R}^2} \psi(u,v)\, \mathbb{P}_{U,V}(du\,dv) = \int_{\mathbb{R}^2} \psi(u,v)\, f_{\overline{X}_{n-1}}(u)\, g_{X_n}(v)\, du\,dv$$

$Z_1 \dots Z_n$ campione gaussiano standard

$\sum_{i=1}^{n-1} (Z_i - \overline{Z}_{n-1})^2$ ha distribuzione χ^2_{n-2}

$$(n-1) S_n^2(Z) = \underbrace{(n-2) S^2_{n-1}(Z)}_{\sum_{i=1}^{n-1}(Z_i - \overline{Z}_{n-1})^2 \quad \chi^2_{n-2}} + \underbrace{\frac{n-1}{n}(Z_n - \overline{Z}_{n-1})^2}_{\chi^2_1}$$

$$\mathbb{P}_{Z_i} = N(0,1)$$

$$\mathbb{P}_{\overline{Z}_{n-1}} = N\left(0, \frac{1}{n-1}\right)$$

$$\mathbb{P}_{Z_n - \overline{Z}_{n-1}} = N\left(0, \frac{n}{n-1}\right)$$

$\sqrt{\frac{n-1}{n}}(Z_n - \overline{Z}_{n-1})$ ha distribuzione $N(0,1)$

$(n-1) S_n^2(Z)$ è una χ^2_{n-1}

$$\sum_{i=1}^{n} (Z_i - \overline{Z})^2$$

DISTRIBUZIONE t DI STUDENT A n GRADI DI LIBERTÀ si indica $t(n)$

È la distribuzione A.C. associata alla densità

$$\tau_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}} \quad \forall x \in \mathbb{R}$$

τ_n è una funzione pari $\Rightarrow$ se T è una v.a con distribuzione $t(n)$ $\mathbb{P}_T = t(n)$

$$F_T(t) + F_T(-t) = 1$$

Si dimostra che $E[T]$ esiste finito $\Rightarrow E[T] = 0$

e $Var[T] = \begin{cases} \frac{n-2}{n} & n \geq 3 \\ +\infty & n = 1,2 \end{cases}$

TEOREMA Sia Z v.a gaussiana standard $\mathbb{P}_Z = N(0,1)$
Sia Y v.a. con distribuzione χ^2_n : $\mathbb{P}_Y = \chi^2_n$
e supponiamo che Z e Y siano indipendenti.
Allora la v.a. $T = \frac{Z\sqrt{n}}{\sqrt{Y}}$ ha distribuzione $t(n)$

DIM $T = \varphi_0(Y, Z)$ $\quad \varphi(y,z) = \begin{cases} \frac{z\sqrt{n}}{\sqrt{y}} & y > 0 \\ 0 & y \leq 0 \end{cases}$

$\psi : \mathbb{R} \to \mathbb{R}$ funzione di Borel nonnegativa

$$\int_{\mathbb{R}} \psi(t)\, \mathbb{P}_T(dt) = \int_{\mathbb{R}^2} \psi(\varphi(y,z))\, \mathbb{P}_{Y,Z}(dy\,dz)$$

$$= \int_{\mathbb{R}^2} \psi(\varphi(y,z))\, \underbrace{f_Y(y)}_{\text{è la } \chi^2_n \Rightarrow f_Y(y) = 0 \text{ se } y \leq 0}\, g_Z(z)\, dz$$

$$= \int_{y>0,\, z\in\mathbb{R}} \psi\left(\frac{z\sqrt{n}}{\sqrt{y}}\right) \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{2}\right)^{n/2} y^{n/2-1} \exp\left(-\frac{y}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dy\, dz$$

$$t = \frac{z\sqrt{n}}{\sqrt{y}} \qquad z \to -\infty \quad t \to -\infty \qquad t \in \mathbb{R}$$
$$z \to +\infty \quad t \to +\infty$$

$$\int_{t\in\mathbb{R}} \varphi(t) \frac{1}{\Gamma\left(\frac{n}{2}\right)\sqrt{2n\pi}} \left(\frac{1}{2}\right)^{n/2} \left(\int_0^{+\infty} y^{\frac{n-1}{2}} \exp\left(-\frac{y}{2}\left(1+\frac{t^2}{n}\right)\right) dy \right) dt$$

$$u = \frac{y}{2}\left(1+\frac{t^2}{n}\right)$$

$$\int_{\mathbb{R}} \varphi(t) \frac{1}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \left(1+\frac{t^2}{n}\right)^{-\frac{n-1}{2}} \left(\int_0^{+\infty} u^{\overbrace{n/2 - 1/2}^{\text{"} \frac{n+1}{2} - 1}} \exp(-u)\, du \right) dt$$

$$\Gamma\left(\frac{n+1}{2}\right)$$

$$\mathbb{P}_T = h(t)\, dt \qquad h(t) = \tau_n(t)$$

COROLLARIO $X_1 \ldots X_n$ campione gaussiano di valore atteso μ e varianza σ^2

Pongo $T = \dfrac{(\overline{X}-\mu)\sqrt{n}}{S}$

Allora $\mathbb{P}_T = t(n-1)$ $\qquad \mathbb{P}_{\overline{X}} = N\left(\mu, \frac{\sigma^2}{n}\right)$

DIM $T = \dfrac{(\overline{X}-\mu)\sqrt{n}}{S} = \qquad \mathbb{P}_{(\overline{X}-\mu)\frac{\sqrt{n}}{\sigma}} = N(0,1)$

$$= \frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \frac{\sigma}{S} = \underbrace{\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}}_{N(0,1)} \cdot \frac{\sqrt{n-1}}{\sqrt{\underbrace{\frac{S^2}{\sigma^2}(n-1)}_{\chi^2_{n-1}}}}$$

ha distribuzione $t(n-1)$

$X_1 — X_n$ campione statistico

Se θ è un parametro che caratterizza la distribuzione del campione e $Y = f(X_1 — X_n)$ è una statistica del campione il cui scopo è stimare θ, si dice che Y è uno stimatore di θ.

Supponiamo che la distribuzione del campione sia dotata di densità che dipenderà del parametro $g(x|\theta)$

La densità congiunta di $X_1 — X_n$ è
$\prod_{i=1}^{n} g(x_i|\theta)$ e si indica $f(x_1 — x_n|\theta)$

$$\mathbb{P}\left(|X_i - x_i| < \frac{\delta}{2}\right) = \mathbb{P}\left(x_i - \frac{\delta}{2} < X_i < x_i + \frac{\delta}{2}\right) = \int_{x_i - \frac{\delta}{2}}^{x_i + \delta/2} g(y|\theta)\,dy$$

$$\simeq g(x_i|\theta)\,\delta$$

$$\mathbb{P}\left(|X_1 - x_1| < \frac{\delta}{2}, |X_2 - x_2| < \frac{\delta}{2}, \dots |X_n - x_n| < \frac{\delta}{2}\right)$$

$$= \mathbb{P}\left(|X_1 - x_1| < \frac{\delta}{2}\right)\mathbb{P}\left(|X_2 - x_2| < \frac{\delta}{2}\right) \dots \mathbb{P}\left(|X_n - x_n| < \frac{\delta}{2}\right)$$

$$\simeq \prod_{i=1}^{n}\left(g(x_i|\theta)\,\delta\right) = \delta^n \prod_{i=1}^{n} g(x_i|\theta) = \delta^n f(x_1 — x_n|\theta)$$

Supposti noti $x_1 — x_n$ cerco $\hat{\theta}$ che massimizza

$h(\theta) = f(x_1 — x_n|\theta) = \prod_{i=1}^{n} g(x_i|\theta)$

$\frac{dh}{d\theta}(\theta) = 0$

$h(\theta)$ è massimo quando $\log h(\theta)$ è massimo

$$\log h(\theta) = \log\left(\prod_{i=1}^{n} g(x_i, \theta)\right) = \sum_{i=1}^{n} \log g(x_i | \theta)$$

$$\frac{d}{dh} \log h(\theta) = \sum_{i=1}^{n} \frac{1}{g(x_i|\theta)} \frac{\partial g(x_i|\theta)}{\partial \theta}$$

DISTRIBUZIONE DI BERNOULLI

$$\mathbb{P}(X=1) = p \Rightarrow \mathbb{P}(X=0) = 1-p$$

$$g(x_i | p) = \begin{cases} p & x_i = 1 \\ 1-p & x_i = 0 \end{cases}$$

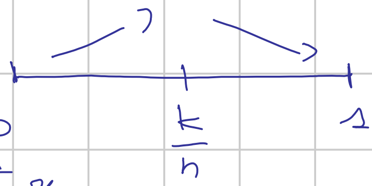
$$f(x_1 \ldots x_n | p) = \prod_{i=1}^{n} g(x_i | p) = p^k (1-p)^{n-k}$$

$$k := \sum_{i=1}^{n} x_i$$

$$\frac{df}{dp} = k p^{k-1} (1-p)^{n-k} + p^k (n-k)(1-p)^{n-k-1}(-1)$$

$$= p^{k-1}(1-p)^{n-k-1}\left(k(1-\cancel{p}) - p(n-\cancel{k})\right) \gtrless 0$$

$$k - np \gtrless 0 \qquad p \lessgtr \frac{k}{n}$$



$\Rightarrow$ ho il max quando $p = \frac{k}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$

$\Rightarrow$ lo stimatore è $\frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$

DISTRIBUZIONE DI POISSON

$$g(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots$$

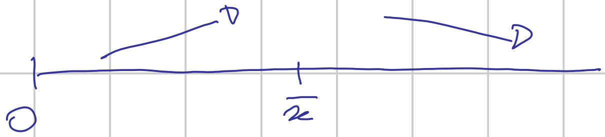
$$f(x_1 \ldots x_n|\lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}$$

$$\log f(x_1 \ldots x_n|\lambda) = \log \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} =$$

$$= \sum_{i=1}^{n} \log\left(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right) = \sum_{i=1}^{n}\left(-\lambda + x_i \log\lambda - \log(x_i!)\right)$$

$$= -n\lambda + (\log\lambda)\left(\sum_{i=1}^{n} x_i\right) - \sum_{i=1}^{n}\log(x_i!)$$

$$\frac{d}{d\lambda}\log f = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i \gtreqless 0 \qquad \lambda \lesseqgtr \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$



La media campionaria $\bar{X}$ è uno stimatore di massima verosimiglianza per il parametro λ di una distribuzione di Poisson.