Politecnico di Milano, Statistica INF, TEL [A-LZ], Epifani I., AA 07/08

1 Famiglia delle densità gamma

Le espressioni delle densità esponenziale di parametro θ e χ^2_1 date da

$$(\mathcal{E}(\beta)) \qquad 1/\theta \mathrm{e}^{-x/\beta}\mathbf{1}_{(0,\infty)}(x), \qquad \beta > 0$$

$$(\chi^2_1) \qquad \frac{(1/2)^{1/2}x^{1/2-1}\mathrm{e}^{-x/2}}{\sqrt{\pi}}\mathbf{1}_{(0,\infty)}(x)$$

sono casi particolari di una densità di forma:

$$f(x,a,\beta) = \frac{h(x,a,\beta)\mathbf{1}_{(0,\infty)}(x)}{\int_0^\infty h(x,a,\beta)\,dx}$$

dove

$$h(x,a,\beta) = \mathrm{e}^{-x/\beta}x^{a-1}$$

Usando il cambio di variabile $y = x/\beta$ si trova che

$$\int_0^\infty \mathrm{e}^{-x/\beta}x^{a-1}\,dx = \beta^a \int_0^\infty \mathrm{e}^{-y}y^{a-1}\,dx, \qquad a > 0$$

Def. 1 L'*integrale gamma* $\Gamma(a)$ è

$$\Gamma(a) = \int_0^\infty \mathrm{e}^{-x} x^{a-1}\, dx, \qquad a > 0$$

Siamo pronti per definire la famiglia delle densità gamma:

Def. 2 X ha *densità gamma di parametri* $a, \beta > 0$, $X \sim \Gamma(a, \beta)$, se

$$f(x, a, \beta) = \frac{(1/\beta)^a}{\Gamma(a)} \mathrm{e}^{-x/\beta} x^{a-1} \mathbf{1}_{(0,\infty)}(x)$$

Osservazione: $\mathcal{E}(\beta) = \Gamma(1, \beta)$ e $\chi_1^2 = \Gamma(1/2, 2)$

Proprietà di $\Gamma(a)$:

1. Essendo che $\mathcal{E}(\beta) = \Gamma(1, \beta)$ e $\chi_1^2 = \Gamma(1/2, 2)$, necessariamente: $\Gamma(1) = 1$ e $\Gamma(1/2) = \sqrt{\pi}$
2. $\Gamma(a+1) = \int_0^\infty \mathrm{e}^{-x} x^{a+1-1}\, dx = -\mathrm{e}^{-x} x^a |_0^\infty + \int_0^\infty \mathrm{e}^{-x} a x^{a-1}\, dx = a\Gamma(a)$
3. Se a è un naturale $n \geq 1$, abbiamo: $\Gamma(n+1) = n! \qquad \forall n \in \mathbb{N}$

Prop. 3 $X \sim \Gamma(a, \beta)$ *ha funzione generatrice dei momenti*

$$M(t) = \mathrm{E}(\mathrm{e}^{tX}) = \frac{1}{(1-\beta t)^a} \qquad \forall t < 1/\beta \tag{1}$$

Dimostrazione

$$\mathrm{E}(\mathrm{e}^{tX}) = \int_0^\infty \mathrm{e}^{tx} \frac{1/\beta^a}{\Gamma(a)} \mathrm{e}^{-x/\beta} x^{a-1} dx = \int_0^\infty \frac{1/\beta^a}{\Gamma(a)} \mathrm{e}^{-x(\frac{1}{\beta}-t)} x^{a-1} dx$$

$$= \frac{1}{(1-\beta t)^a} \int_0^\infty \frac{[(1-\beta t)/\beta]^a}{\Gamma(a)} \mathrm{e}^{-x[(1-\beta t)/\beta]} x^{a-1} dx = \frac{1}{(1-\beta t)^a}, t < \frac{1}{\beta}$$

perché, se $t < 1/\beta$ l'integranda a destra è la densità $\Gamma(a, \frac{\beta}{1-\beta t})$ ∎

$$\mathrm{E}(X) = M'(t)|_{t=0} = a\beta(1-\beta t)^{-a-1}|_{t=0} = a\beta$$

$$\mathrm{E}(X^2) = M''(t)|_{t=0} = a(a+1)\beta \times \beta(1-\beta t)^{-a-2}|_{t=0} = a(a+1)\beta^2$$

$$\mathrm{Var}(X) = a(a+1)b^2 - a^2b^2 = a\beta^2$$

Riassumiamo nella seguente proposizione alcune proprietà della famiglia gamma.

Prop. 4 *i) Se* $X \sim \Gamma(a, \beta)$*,* $c > 0$ *e* $Y = cX$ *allora* $Y \sim \Gamma(a, c\beta)$*.*

ii) Se X*,* Y *sono va indipendenti con* $X \sim \Gamma(a, \beta)$ *e* $Y \sim \Gamma(c, \beta)$ *allora* $X + Y \sim \Gamma(a + c, \beta)$*.*

iii) Viceversa, se X*,* Y *sono va indipendenti e* $Y \sim \Gamma(c, \beta)$ *e* $X + Y \sim \Gamma(a + c, \beta)$*, con* $c > 0$*, allora* $X \sim \Gamma(a, \beta)$*.*

Dimostrazione

ii) $M_{cY}(t) = \mathrm{E}(\mathrm{e}^{tcX}) = M_X(ct) = \dfrac{1}{(1-\beta ct)^a}, \qquad t < \dfrac{1}{c\beta}$

ii) $M_{X+Y}(t) = \mathrm{E}(\mathrm{e}^{t(X+Y)})$

$$= \mathrm{E}(\mathrm{e}^{tX}\mathrm{e}^{tY}) = \mathrm{E}(\mathrm{e}^{tX})\,\mathrm{E}(\mathrm{e}^{tY}) \quad \text{[per l'indipendenza di } X, Y\text{]}$$

$$= M_X(t)M_Y(t) = \frac{1}{(1-\beta t)^a} \times \frac{1}{(1-\beta t)^c} = \frac{1}{(1-\beta t)^{a+c}}$$

iii) $M_Y(t) > 0 \, \forall t < 1/\beta$, quindi $M_{X+Y}(t) = M_X(t)M_Y(t)$ **sse**

$$M_X(t) = \frac{M_{X+Y}(t)}{M_Y(t)} = \frac{(1-\beta t)^{a+c}}{(1-\beta t)^c} = \frac{1}{(1-\beta t)^a}, \; \forall t < \frac{1}{\beta} \qquad \blacksquare$$

1.1 Legami fra va gaussiane e chiquadrato

Def. 5 Chiamiamo la densità $\Gamma(n/2, 2)$ *chiquadrato con n gradi di libertà* e la indichiamo con il simbolo χ^2_n.

Già sappiamo che

Prop. 6 *Se* $X \sim N(0,1)$ *allora* $X^2 \sim \chi^2_1$.

Ora verifichiamo che

Prop. 7 *Se* $X_1, \dots, X_n$ *sono va* i.i.d. $\sim N(0,1)$, *allora* $\sum_{j=1}^n X_j^2 \sim \chi^2_n$.

Dimostrazione $X_1^2, \dots, X_n^2$ sono i.i.d. con comune densità χ^2_1. Dal punto $ii)$ della Proposizione 4 applicato a n va $\Gamma(1/2, 2)$ e indipendenti abbiamo $\sum_{j=1}^n X_j^2 \sim \Gamma(n/2, 2) = \chi^2_n$. ∎

Voi avrete le tavole della fdr χ^2_n. Potrete usarle anche per calcolare probabilità del tipo $P(X \leq k)$, con $X \sim \Gamma(n, \theta)$. Infatti: se $X \sim \Gamma(n, \theta)$ allora $2X/\theta \sim \Gamma(n, 2) = \chi_{2n}$ e quindi: $\boxed{F_{\Gamma(n,\theta)}(k) = F_{\chi^2_{2n}}(2k/\theta)}$

2 Varianza campionaria, proprietà di non distorsione

$X_1, \ldots, X_n$ i.i.d. $\sim f$ con media μ e varianza σ^2 finite e incognite
Obiettivo: costruire uno stimatore per σ^2.
Partiamo dalla statistica

$$\sum_{j=1}^{n}(X_j - \overline{X})^2$$

$\sum_{j=1}^{n}(X_j - \overline{X})^2$ è interpretabile come indice (aleatorio) della dispersione di $X_1, \ldots, X_n$ intorno a $\overline{X}$.

Prop. 8 *Se* $X_1, \ldots, X_n$ i.i.d. $\sim f$ *e* $\sigma^2 < \infty$ *allora:*

$$\sum_{j=1}^{n}(X_j - \overline{X})^2 = \sum_{j=1}^{n}(X_j - \mu)^2 - n(\overline{X} - \mu)^2 \tag{2}$$

$$\mathrm{E}\Big(\sum_{j=1}^{n}(X_j - \overline{X})^2\Big) = (n-1)\sigma^2 \tag{3}$$

Dimostrazione della Proposizione 8

$$\sum_{j=1}^{n}(X_j - \overline{X})^2 = \sum_{j=1}^{n}[(X_j - \mu) - (\overline{X} - \mu)]^2$$

$$= \sum_{j=1}^{n}[(X_j - \mu)^2 + (\overline{X} - \mu)^2 - 2(X_j - \mu)(\overline{X} - \mu)]$$

$$= \sum_{j=1}^{n}(X_j - \mu)^2 + n(\overline{X} - \mu)^2 - 2(\overline{X} - \mu)\sum_{j=1}^{n}(X_j - \mu)$$

$$= \sum_{j=1}^{n}(X_j - \mu)^2 + n(\overline{X} - \mu)^2 - 2n(\overline{X} - \mu)^2$$

$$= \sum_{j=1}^{n}(X_j - \mu)^2 - n(\overline{X} - \mu)^2$$

$$\mathrm{E}\sum_{j=1}^{n}(X_j - \overline{X})^2 = \mathrm{E}\sum_{j=1}^{n}(X_j - \mu)^2 - n\,\mathrm{E}(\overline{X} - \mu)^2$$

$$= \sum_{j=1}^{n}\mathrm{E}(X_j - \mu)^2 - n\,\mathrm{Var}(\bar{X}) = n\sigma^2 - \sigma^2 \quad \blacksquare$$

Def. 9 La statistica

$$S^2 := \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2$$

è detta *varianza campionaria*.

S^2 è uno stimatore di σ^2 per cui $\mathrm{E}(S^2) = \sigma^2$,

cioè S^2 è uno stimatore *non distorto* (o *corretto*) di σ^2.

Considerato che la media di S^2 vale σ^2, decido di stimare σ^2 con S^2.

Notate che $\bar{X}$ è uno stimatore non distorto di μ, infatti $\mathrm{E}(\bar{X}) = \mu$.

Se S^2 è un "buon" stimatore per σ^2, allora S^2/σ^2 dovrebbe essere prossimo a 1 con probabilità "grande". Quindi, per dare una misura dell'errore di approssimazione, è ragionevole calcolare:
$P\left(a \le \frac{S^2}{\sigma^2} \le b\right)$ con $a < b$ prossimi a 1, oppure $P\left(\frac{S^2}{\sigma^2} \le c\right)$
$\Longrightarrow$ occorre la densità di S^2. Riesco a determinarla per un campione gaussiano $N(\mu, \sigma^2)$.

3 Distribuzioni delle statistiche media e varianza campionarie di popolazione gaussiana

Prop. 10 *Sia* $X_1, \ldots, X_n$ *un campione casuale dalla f.d.r.* $N(\mu, \sigma^2)$ *e*

$$\bar{X} = \frac{\sum_{j=1}^{n} X_j}{n}, \qquad S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2, \qquad n \geq 2$$

Allora, per ogni $\mu \in \mathbb{R}$ *e per ogni* $\sigma^2 > 0$

i) $\overline{X} \sim N(\mu, \sigma^2/n)$

ii) le statistiche S^2 *e* $\bar{X}$ *sono indipendenti*

iii) $(n-1)S^2/\sigma^2 = \sum_{j=1}^{n}(X_j - \bar{X})^2/\sigma^2 \sim \chi^2_{n-1}$

iv) La statistica $\dfrac{\bar{X} - \mu}{S/\sqrt{n}}$*, con* $S = \sqrt{S^2}$*, ha densità* t *di student con* $n-1$ *gradi di libertà, cioè la sua densità è:*

$$f_{n-1}(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}$$

Dimostrazione della Proposizione 10

$i)$ Già dimostrato.

$ii)$ NON dimostriamo questo punto sull'indipendenza; la dimostrazione richiederebbe un lungo richiamo sui vettori gaussiani che ci risparmiamo.

$iii)$ Passo 1: $X_1, \dots, X_n$ i.i.d. $\sim N(\mu, \sigma^2)$ implica $\frac{X_1-\mu}{\sigma}, \dots, \frac{X_n-\mu}{\sigma}$ i.i.d. $\sim N(0,1)$. Poi

$$\frac{X_1 - \mu}{\sigma} \sim N(0,1) \quad \text{implica} \quad \frac{(X_1-\mu)^2}{\sigma^2} \sim \chi_1^2$$

Allora

$$\sum_{j=1}^{n} (X_j - \mu)^2 / \sigma^2 \sim \chi_n^2$$

Analogamente:

$$\sqrt{n}\frac{\bar{X} - \mu}{\sigma} \sim N(0,1) \quad \text{implica} \quad n\frac{(\overline{X}-\mu)^2}{\sigma^2} \sim \chi_1^2$$

Passo 2: Abbiamo già dimostrato la seguente decomposizione

$$\sum_{j=1}^{n}(X_j - \bar{X})^2 = \sum_{j=1}^{n}(X_j - \mu)^2 - n(\overline{X} - \mu)^2$$

da cui deriviamo che

$$(n-1)S^2/\sigma^2 + (n/\sigma^2)(\overline{X} - \mu)^2 = \sum_{j=1}^{n}(X_j - \mu)^2/\sigma^2$$

Ora, osservate che

- Le due v.a. a sinistra sono indipendenti,
- la seconda ha densità $\chi_1^2 = \Gamma(1/2, 2)$,
- e la loro somma ha densità $\chi_n^2 = \Gamma(n/2, 2)$.

Dalle proprietà della famiglia delle densità gamma deduciamo che l'altro addendo $(n-1)S^2/\sigma^2$ ha densità $\Gamma(n/2 - 1/2, 2) = \chi_{n-1}^2$.

iv) Siano $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$ e $Y = \frac{S^2(n-1)}{\sigma^2}$. Allora

- Z, Y sono indipendenti (è il punto *ii*),
- $\bar{X} \sim N(\mu, \sigma^2/n)$ e quindi $Z \sim N(0,1)$ (è il punto *i*),
- $Y \sim \chi^2_{n-1}$ (è il punto *iii*)

Osserviamo poi che

- $\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{Z}{\sqrt{Y/(n-1)}}$

Applicando il seguente Lemma 11, la tesi segue. ∎

Lemma 11 *Siano $Z \sim N(0,1)$, $Y \sim \chi^2_n$ e Z e Y indipendenti. Allora $\dfrac{Z}{\sqrt{Y/n}}$ ha densità t di student con n gradi di libertà.*

La densità t di student con n gradi di libertà è data

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}(1 + \frac{t^2}{n})^{-\frac{n+1}{2}} \qquad x \in \mathbb{R}$$

Dimostrazione del Lemma 11:

$$P\Big(\frac{Z}{\sqrt{Y/n}} \le t\Big) = P\left(Z \le t\sqrt{Y/n}\right) = \int_{A_t} f_Y(y) f_Z(z)\, dz\, dy$$

$$[\text{con } A_t := \{(y,z) \in (0,\infty) \times \mathbb{R} :\ z \le t\sqrt{y/n}\}]$$

$$= \int_0^\infty \Big(\int_{-\infty}^{t\sqrt{y/n}} \varphi(z)\, dz\Big) f_Y(y)\, dy = \int_0^\infty f_Y(y)\Phi\left(t\sqrt{y/n}\right)\, dy$$

Allora

$$\frac{\partial}{\partial t} P\Big(\frac{Z}{\sqrt{Y/n}} \le t\Big)(t) = \int_0^\infty f_Y(y) \frac{\partial}{\partial t}\Phi\big(t\sqrt{y/n}\big)\, dy$$

$$= \int_0^\infty f_Y(y)\varphi\big(t\sqrt{y/n}\big)\sqrt{\frac{y}{n}}\, dy = \int_0^\infty \frac{\frac{1}{2^{n/2}}}{\Gamma\left(\frac{n}{2}\right)} \mathrm{e}^{-\frac{y}{2}} y^{\frac{n}{2}-1} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2 y}{2n}} \sqrt{\frac{y}{n}}\, dy$$

$$= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)} \big(1 + \frac{t^2}{n}\big)^{-\frac{n+1}{2}} \int_0^\infty \frac{\left(\frac{n+t^2}{2n}\right)^{\frac{n+1}{2}}}{\Gamma\left(\frac{n+1}{2}\right)} \mathrm{e}^{-y\frac{n+t^2}{2n}} y^{\frac{n+1}{2}-1}\, dy$$

Ultimo integrale vale 1 perché l'integranda è la densità $\Gamma(\frac{n+1}{2}, \frac{2n}{n+t^2})$ ∎