

1 Famiglia delle densità gamma

Le espressioni delle densità esponenziale di parametro θ e χ_1^2 date da

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

$$(\chi_1^2) \quad \frac{(1/2)^{1/2} x^{1/2-1} e^{-x/2}}{\sqrt{\pi}} \mathbf{1}_{(0,\infty)}(x)$$

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

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

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

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

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

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

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

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

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. $\Gamma(a+1) = \int_0^{\infty} e^{-x} x^{a+1-1} dx = -e^{-x} x^a|_0^{\infty} + \int_0^{\infty} e^{-x} a x^{a-1} dx = a\Gamma(a)$
3. Se a è un naturale $n \geq 1$, abbiamo: $\Gamma(n+1) = n! \quad \forall n \in \mathbb{N}$

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

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

Dimostrazione

$$\begin{aligned} E(e^{tX}) &= \int_0^\infty e^{tx} \frac{1/\beta^a}{\Gamma(a)} e^{-x/\beta} x^{a-1} dx = \int_0^\infty \frac{1/\beta^a}{\Gamma(a)} e^{-x(\frac{1}{\beta} - t)} x^{a-1} dx \\ &= \frac{1}{(1 - \beta t)^a} \int_0^\infty \frac{[(1 - \beta t)/\beta]^a}{\Gamma(a)} e^{-x[(1 - \beta t)/\beta]} x^{a-1} dx = \frac{1}{(1 - \beta t)^a}, t < \frac{1}{\beta} \end{aligned}$$

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

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

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

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

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 variabili 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 variabili 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

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

$$\begin{aligned} ii) M_{X+Y}(t) &= E(e^{t(X+Y)}) \\ &= E(e^{tX} e^{tY}) = E(e^{tX}) E(e^{tY}) \quad [\text{per l'indipendenza di } X, Y] \\ &= M_X(t) M_Y(t) = \frac{1}{(1 - \beta t)^a} \times \frac{1}{(1 - \beta t)^c} = \frac{1}{(1 - \beta t)^{a+c}} \end{aligned}$$

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

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

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_n^2 .

Già sappiamo che

Prop. 6 Se $X \sim N(0, 1)$ allora $X^2 \sim \chi_1^2$.

Ora verifichiamo che

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

Dimostrazione X_1^2, \dots, X_n^2 sono i.i.d. con comune densità χ_1^2 . Dal punto *ii*) della Proposizione 4 applicato a n va $\Gamma(1/2, 2)$ e indipendenti abbiamo $\sum_{j=1}^n X_j^2 \sim \Gamma(n/2, 2) = \chi_n^2$. ■

Voi avrete le tavole della fdr χ_n^2 . 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 $2X/\theta \sim \Gamma(n, 2) = \chi_{2n}^2$ e quindi:

$F_{\Gamma(n, \theta)}(k) = F_{\chi_{2n}^2}(2k/\theta)$

2 Varianza campionaria, proprietà di non distorsione

X_1, \dots, 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

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

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

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

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

$$(3) \quad \mathbb{E}\left(\sum_{j=1}^n (X_j - \bar{X})^2\right) = (n-1)\sigma^2$$

Dimostrazione della Proposizione 8

$$\begin{aligned} \sum_{j=1}^n (X_j - \bar{X})^2 &= \sum_{j=1}^n [(X_j - \mu) - (\bar{X} - \mu)]^2 \\ &= \sum_{j=1}^n [(X_j - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_j - \mu)(\bar{X} - \mu)] \\ &= \sum_{j=1}^n (X_j - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{j=1}^n (X_j - \mu) \\ &= \sum_{j=1}^n (X_j - \mu)^2 + n(\bar{X} - \mu)^2 - 2n(\bar{X} - \mu)^2 \\ &= \sum_{j=1}^n (X_j - \mu)^2 - n(\bar{X} - \mu)^2 \\ \mathbb{E} \sum_{j=1}^n (X_j - \bar{X})^2 &= \mathbb{E} \sum_{j=1}^n (X_j - \mu)^2 - n \mathbb{E}(\bar{X} - \mu)^2 \\ &= \sum_{j=1}^n \mathbb{E}(X_j - \mu)^2 - n \text{Var}(\bar{X}) = n\sigma^2 - \sigma^2 \quad \blacksquare \end{aligned}$$

Def. 9 La statistica

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

è detta *varianza campionaria*.

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

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 $E(\bar{X}) = \mu$.

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:

$$P\left(a \leq \frac{S^2}{\sigma^2} \leq b\right) \quad \text{con } a < b \text{ prossimi a 1, oppure } P\left(\frac{S^2}{\sigma^2} \leq c\right)$$

\implies occorre la densità di S^2 . Riesco a determinarla per un campione gaussiano $N(\mu, \sigma^2)$.

3 Distribuzioni delle statistiche media e varianza campio- narie di popolazione gaussiana

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

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

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

- i) $\bar{X} \sim N(\mu, \sigma^2/n)$
- ii) le statistiche S^2 e \bar{X} sono indipendenti
- iii) $(n-1)S^2/\sigma^2 = \sum_{j=1}^n (X_j - \bar{X})^2/\sigma^2 \sim \chi_{n-1}^2$
- iv) La statistica $\frac{\bar{X} - \mu}{S/\sqrt{n}}$, 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:

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

Passo 2: Abbiamo già dimostrato la seguente decomposizione

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

da cui deriviamo che

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

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*),
- $Y \sim \chi_{n-1}^2$ (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 $Z \sim N(0, 1)$, $Y \sim \chi_n^2$ e Z e Y indipendenti.*

Allora $\frac{Z}{\sqrt{Y/n}}$ ha densità t di student con n gradi di libertà.

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

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

Dimostrazione del Lemma 11:

$$\begin{aligned}
P\left(\frac{Z}{\sqrt{Y/n}} \leq t\right) &= P\left(Z \leq t\sqrt{Y/n}\right) = \int_{A_t} f_Y(y) f_Z(z) dz dy \\
&\quad [\text{con } A_t := \{(y, z) \in (0, \infty) \times \mathbb{R} : z \leq t\sqrt{y/n}\}] \\
&= \int_0^\infty \left(\int_{-\infty}^{t\sqrt{y/n}} \varphi(z) dz\right) f_Y(y) dy = \int_0^\infty f_Y(y) \Phi\left(t\sqrt{y/n}\right) dy
\end{aligned}$$

Allora

$$\begin{aligned}
\frac{\partial}{\partial t} P\left(\frac{Z}{\sqrt{Y/n}} \leq t\right)(t) &= \int_0^\infty f_Y(y) \frac{\partial}{\partial t} \Phi(t\sqrt{y/n}) dy \\
&= \int_0^\infty f_Y(y) \varphi(t\sqrt{y/n}) \sqrt{\frac{y}{n}} dy = \int_0^\infty \frac{1}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{y}{2}} y^{\frac{n}{2}-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2 y}{2n}} \sqrt{\frac{y}{n}} dy \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \int_0^\infty \frac{\left(\frac{n+t^2}{2n}\right)^{\frac{n+1}{2}}}{\Gamma\left(\frac{n+1}{2}\right)} e^{-y \frac{n+t^2}{2n}} y^{\frac{n+1}{2}-1} dy
\end{aligned}$$

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