STATISTICA COMPLEMENTI

Si richiamano alcuni concetti.

Dicesi *variabile casuale* (v.c.) *discreta* l'insieme delle coppie di numeri reali (x_i, p_i) $i = 1, 2, \cdots, k$ (k può assumere anche il valore ∞), dove x_i indica l'i-esima *modalità* e p_i la *probabilità* ad essa associata.

Le probabilità godono delle seguenti due proprietà

- $p_i \geq 0$
- $\sum_{i=1}^{k} p_i = 1$

La v.c. X si può descrivere nella seguente forma:

$$X = \begin{cases} x_1 & x_2 & \cdots & x_k \\ p_1 & p_2 & \cdots & p_k \end{cases}$$

Una *variabile casuale continua* X è caratterizzata dalla funzione di densità di probabilità $f(x)$, che gode della seguenti proprietà:

- $f(x) \geq 0$

- continua
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

Se I è il generico intervallo, allora la densità ha il seguente significato:

$$P\{X \in I\} = \int_I f(x)dx$$

Si dice *momento dall'origine di ordine r (o r-esimo)* della variabile casuale X la seguente espressione:

$$\mu_r \equiv E\left(X^r\right) = \sum_{i=1}^{k} x_i^r p_i \qquad r = 1, 2, ...$$

(caso discreto)

$$\mu_r \equiv E\left(X^r\right) = \int_{-\infty}^{+\infty} x^r f(x)dx \qquad r = 1, 2, ...$$

(caso continuo)

Il momento di ordine 1 è l'aspettativa della variabile casuale X.

Dicesi *funzione cumulata delle probabilità (oppure, funzione di ripartizione) della variabile casuale X*, la seguente funzione:

$$F(x) = P\{X \le x\}$$

Essa gode delle seguenti proprietà:

- $F : R \to [0,1]$
- $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to +\infty} F(x) = 1$
- è funzione continua da destra: per ogni numero reale *x*, vale: $\lim_{0<h \to 0} F(x+h) = F(x)$
- è funzione monotona non decrescente: per ogni coppia di valori reali x_1, x_2 tali che $x_1 < x_2$, vale $F(x_1) \leq F(x_2)$.

Per una variabile casuale discreta essa si presenta nella forma:

$$F(x) = \sum_{i:\, x_i \leq x} p_i \, .$$

Per una variabile casuale continua essa si presenta nella forma:

$$F(x) = \int_{-\infty}^{x} f(t)dt \, .$$

La funzione generatrice dei momenti

Si consideri la seguente espressione:

$$E\left(e^{tX}\right) = \begin{cases} \int_{-\infty}^{+\infty} e^{tx} f(x)dx & \textit{caso continuo} \\ \sum_i e^{tx_i} p(x_i) & \textit{caso discreto} \end{cases}$$

Definizione.
Si dice che la v.c. X possiede la funzione generatrice dei momenti se esiste $\delta > 0$ tale che $m_X(t) < \infty$ per ogni $t \in (-\delta, \delta)$. Allora, $m_X(t)$, $t \in (-\delta, \delta)$, è detta funzione generatrice dei momenti di X.

Teorema
Sia X una v.c. tale che $E(e^{tX}) < \infty$ per ogni $t \in (-\delta, \delta)$. Allora X ha momenti di ogni ordine, e, per ogni $t \in (-\delta, \delta)$, risulta:

$$m_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu_r .$$

Inoltre

$$\mu_r = \left. \frac{d^r}{dt^r} m_X(t) \right|_{t=0} .$$

Teorema
Siano $m_X(t)$ e $m_Y(t)$ le f.g.m. delle v.c. X e Y. Si supponga che esista δ tale che per ogni $t \in (-\delta, \delta)$ valga $m_X(t) = m_Y(t)$. Allora le due v.c. hanno la stessa distribuzione, cioè $F_X(v) = F_Y(v)$, per ogni v reale.

Teorema

Sia $m_{X_i}(t)$ la f.g.m. della v.c. X_i, $i = 1,2,\cdots,n$.

E siano le v.c. X_i indipendenti.

Allora, $m_Y(t) = \prod_{i=1}^{n} m_{X_i}(t)$,

con $Y = \sum_{i=1}^{n} X_i$.

Alcune variabili casuali discrete

La variabile casuale binomiale

La v.c. binomiale è definita dalla seguente funzione di probabilità:

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \cdots, n \\ 0 & altrove \end{cases},$$

con $p \in (0,1)$, $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

$n! = 1 \times 2 \times \cdots \times n$ e $0! \equiv 1$.

$$\mu_1 = E(X) = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} =$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} =$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} =$$

$$= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} = np.$$

Per la determinazione della varianza si procede nel modo seguente:

$$\mu_2 - \mu_1 = \mu_2 - np = E(X(X-1)) =$$

$$= \sum_{x=0}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} =$$

$$= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x (1-p)^{n-x} =$$

$$= n(n-1)p^2 \sum_{x=1}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} (1-p)^{n-x} =$$

$$= n(n-1)p^2 \sum_{y=0}^{n-2} \frac{(n-2)!}{y!(n-2-y)!} p^y (1-p)^{n-2-y} =$$

$$= n(n-1)p^2,$$

da cui si ottiene:

$$\mu_2 = n(n-1)p^2 + np.$$

$$Var(X) = \mu_2 - \mu_1^2 = n^2 p^2 - np^2 + np - (np)^2 =$$

$= np(1-p).$

La f.g.m. della v.c. binomiale si deduce nel modo seguente:

$$m_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx} \cdot \binom{n}{x} p^x (1-p)^{n-x} =$$

$$= \sum_{x=0}^{n} \binom{n}{x} \left(pe^t\right)^x (1-p)^{n-x} = \left[(1-p) + pe^t\right]^n$$

La variabile casuale binomiale esprime *la somma di n variabili casuali indicatore indipendenti aventi la medesima probabilità di successo*.

Essa esprime, in altre parole, *il numero di successi in n prove indipendenti*.

La variabile casuale ipergeometrica

La v.c. ipergeometrica è definita dalla seguente funzione di probabilità:

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}},$$

$$Max(0, n+r-N) \leq x \leq Min(r, n).$$

$$E(X)=\sum_{x=0}^{n} x\cdot\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}=$$

$$=n\frac{r}{N}\sum_{x=1}^{n}\frac{\frac{(r-1)!}{(x-1)!(n-x)!}\frac{(N-r)!}{(n-x)!(N-r-n+x)}}{\frac{(N-1)!}{(n-1)!(N-n)!}}=$$

$$=n\frac{r}{N}\sum_{y=0}^{n-1}\frac{\frac{(r-1)!}{y!(r-1-y)!}\frac{(N-1-r+1)!}{(n-1-y)!(N-r-n+1+y)}}{\frac{(N-1)!}{(n-1)!(N-n)!}}=$$

$$=n\frac{r}{N}\sum_{y=0}^{n-1}\frac{\binom{r-1}{y}\binom{N-1-(r-1)}{n-1-y}}{\binom{N-1}{n-1}}=n\frac{r}{N}.$$

Per determinare la varianza, si procede in questo modo:

$$E\left(X\left(X-1\right)\right)=\sum_{x=0}^{n} x(x-1)\cdot\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}=$$

$$=n(n-1)\frac{r(r-1)}{N(N-1)}\sum_{x=2}^{n}\frac{\binom{r-2}{x-2}\binom{N-r}{n-x}}{\binom{N-2}{n-2}}=$$

$$=n(n-1)\frac{r(r-1)}{N(N-1)}\sum_{y=0}^{n-2}\frac{\binom{r-2}{y}\binom{N-2-r+2}{n-2-y}}{\binom{N-2}{n-2}}=$$

$$=n(n-1)\frac{r(r-1)}{N(N-1)}.$$

$$Var(X)=E\left(X\left(X-1\right)\right)+E(X)-\left(E(X)\right)^2=$$

$$=n(n-1)\frac{r(r-1)}{N(N-1)}+n\frac{r}{N}-\left(n\frac{r}{N}\right)^2=$$

$$= n\frac{r}{N}\left[(n-1)\frac{r-1}{N-1}+1-n\frac{r}{N}\right] = n\frac{r}{N}\left(1-\frac{r}{N}\right)\frac{N-n}{N-1}$$

La v.c. ipergeometrica esprime *il numero di successi in n estrazioni senza reimmissione o in blocco*.

La v.c. ipergeometrica è ben approssimata dalla v.c. binomiale se $n < N/10$. In formula:

$$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \cong \binom{n}{x}\left(\frac{r}{N}\right)^{x}\left(1-\frac{r}{N}\right)^{n-x}.$$

Esempio

$N = 52,\ r = 13,\ n = 5.$

x	Binomiale	Ipergeometrica
0	0,2373	0,2215
1	0,3955	0,4114
2	0,2637	0,2743
3	0,0879	0,0815
4	0,0146	0,0107
5	0,0010	0,0005

La variabile casuale di Poisson

La v.c. di Poisson ha la seguente funzione di probabilità:

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0, 1, 2, \ldots \ ; \lambda > 0$$

$$E(X) = \sum_{x=0}^{+\infty} x \cdot \frac{e^{-\lambda}\lambda^{x}}{x!} = \lambda \sum_{x=1}^{+\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda \sum_{y=0}^{+\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} =$$

$$= \lambda$$

$$E(X(X-1)) = \sum_{x=0}^{+\infty} x(x-1) \cdot \frac{e^{-\lambda}\lambda^{x}}{x!} =$$

$$\lambda^2 \sum_{x=2}^{+\infty} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!} = \lambda^2 \sum_{y=0}^{+\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} = \lambda^2$$

$$Var(X) = E(X(X-1)) + E(X) - [E(X)]^2 =$$

$$= \lambda^2 + \lambda - \lambda^2 = \lambda.$$

$$m_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{+\infty} e^{tx} \frac{e^{-\lambda}\lambda^{x}}{x!} = e^{-\lambda} \sum_{x=0}^{+\infty} \frac{\left(e^{t}\lambda\right)^{x}}{x!} =$$

$$= e^{-\lambda} \cdot e^{\lambda e^{t}} = e^{\lambda\left(e^{t}-1\right)},$$

poichè la sommatoria rappresenta l’espansione in serie di Taylor della funzione esponenziale:

$$e^z = \sum_{x=0}^{+\infty} \frac{z^x}{x!}.$$

Teorema

Sia X_1 una v.c. di Poisson di parametro λ_1 e Sia X_2 una v.c. di Poisson di parametro λ_2. Siano X_1 e X_2 indipendenti in probabilità. Allora la v.c. $X = X_1 + X_2$ è una v.c. di Poisson di parametro $(\lambda_1 + \lambda_2)$.

Dimostrazione

$$m_X(t) = m_{X_1}(t) \cdot m_{X_2}(t) = e^{\lambda_1 \left(e^t - 1\right)} \cdot e^{\lambda_2 \left(e^t - 1\right)} =$$
$$= e^{(\lambda_1 + \lambda_2)\left(e^t - 1\right)}.$$

Riconoscendo in tale espressione quella della f.g.m. di una v.c. di Poisson di parametro $(\lambda_1 + \lambda_2)$, *in forza del teorema dell'unicità della f.g.m.*, si può concludere che X è una v.c. di Poisson di parametro $(\lambda_1 + \lambda_2)$.

La distribuzione di Poisson è indicata per probabilizzare fenomeni che ragionevolmente soddisfino le seguenti condizioni:

- La probabilità che nell'intervallo (piccolo) di ampiezza h si abbia $X = 1$ è approssimativamente proporzionale ad h

$$P\{X = 1\} = vh + o(h)$$

- La probabilità che nell'intervallo (piccolo) di ampiezza *h* si abbia $X \geq 2$ è trascurabilr relativamente alla probabilità di avere $X = 1$

$$\frac{P\{X \geq 2\}}{P\{X = 1\}} = o(h)$$

- Il numero di presenze del fenomeno in intervalli temporali incompatibili sono indipendenti in probabilità.

Il numero $v > 0$ indica la probabilità di avere un accadimento in un intervallo unitario.

Teorema

Se le tre condizioni enunciate si verificano, si dimostra che il numero di accadimenti *X* di un fenomeno in un intervallo di tempo di ampiezza *t* segue la distribuzione di Poisson con parametro $\lambda = v \cdot t$.

Alcune variabili casuali continue

La variabile casuale gamma

La v.c. Y è distribuita secondo una gamma di parametri $\alpha > 0$ e $\vartheta > 0$ se la funzione di densità è data da:

$$f(y) = \begin{cases} \dfrac{\vartheta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\vartheta y} & y > 0 \\ 0 & altrove \end{cases},$$

dove $\Gamma(\alpha) \equiv \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$.

Proprietà dlla funzione gamma:

- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$,
- $\Gamma(n) = (n-1)!$.

$$E(Y^r) = \int_0^{+\infty} y^r \frac{\vartheta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\vartheta y} dy =$$

$$= \frac{\vartheta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(r+\alpha)}{\vartheta^{r+\alpha}} \int_0^{+\infty} \frac{\vartheta^{r+\alpha}}{\Gamma(r+\alpha)} y^{r+\alpha-1} e^{-\vartheta y} dy =$$

$$= \frac{1}{\vartheta^r} \frac{(r+\alpha-1)\cdot(r+\alpha-2)\cdot\ldots\cdot\alpha\cdot\Gamma(\alpha)}{\Gamma(\alpha)} =$$

$$= \frac{1}{\vartheta^r}(r+\alpha-1)\cdot(r+\alpha-2)\cdot\ldots\cdot\alpha.$$

Per r = 1, si ottiene $E(X)=\frac{\alpha}{\vartheta}$. Per r = 2, si ottiene $E(X^2)=\frac{\alpha\cdot(\alpha+1)}{\vartheta^2}$, da cui:

$$Var(X)=\frac{\alpha\cdot(\alpha+1)}{\vartheta^2}-\frac{\alpha^2}{\vartheta^2}=\frac{\alpha}{\vartheta^2}.$$

$$m_X(t)=\int_0^{+\infty} e^{ty}\,\frac{\vartheta^\alpha}{\Gamma(\alpha)}\,y^{\alpha-1}e^{-\vartheta y}dy=$$

$$=\frac{\vartheta^\alpha}{(\vartheta-t)^\alpha}\int_0^{+\infty}\frac{(\vartheta-t)^\alpha}{\Gamma(\alpha)}\,y^{\alpha-1}e^{-y(\vartheta-t)}dy=\left(\frac{\vartheta}{\vartheta-t}\right)^\alpha,$$

per $t<\vartheta$.

Casi particolari:

a) $\alpha=1$, v.c. esponenziale di parametro ϑ

b) $\alpha=\frac{k}{2},\vartheta=\frac{1}{2}$ (k intero), v.c. chi-quadrato con k gradi di libertà (g.d.l.).

(esplicitare i calcoli, e disegnare le densità; dare la definizione di quantile, e mostrare l'uso delle tavole.)

La v.c. normale

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad x\in R,\ \mu\in R,\ \sigma>0$$

$$m_X(t)=\int_{-\infty}^{+\infty} e^{tx}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx=$$

$$=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{x^2+\mu^2+2x\mu-2xt\sigma^2+\sigma^4t^2+2\mu t\sigma^2-\sigma^4t^2-2\mu t\sigma^2}{2\sigma^2}}dx$$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left[x-\left(m+t\sigma^2\right)\right]^2}dx\ e^{\frac{1}{2}\sigma^2t^2+\mu t}=$$

$$=e^{\frac{1}{2}\sigma^2t^2+\mu t}.$$

Verificare che:

- $\frac{d}{dt}\left(e^{\frac{1}{2}\sigma^2t^2+\mu t}\right)\Big|_{t=0}=\mu$
- $\frac{d^2}{dt^2}\left(e^{\frac{1}{2}\sigma^2t^2+\mu t}\right)\Big|_{t=0}=\sigma^2+\mu^2.$

(Conseguenze)

Teorema

Sia X_1 una v.c. normale di aspettativa μ_1 e varianza σ_1^2. Sia X_2 una v.c. una v.c. normale di aspettativa μ_2 e varianza σ_2^2. Siano X_1 e X_2 indipendenti in probabilità. Allora la v.c. $X = X_1 + X_2$ è una v.c. normale di aspettativa $\mu_1 + \mu_2$ e varianza $\sigma_1^2 + \sigma_2^2$.

Dimostrazione

$$m_X(t) = m_{X_1}(t) \cdot m_{X_2}(t) = e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2} \cdot e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2} =$$

$$= e^{(\mu_1 + \mu_2)t + \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right)t^2}.$$

Riconoscendo in tale espressione quella della f.g.m. di una v.c. normale di aspettativa $\mu_1 + \mu_2$ e varianza $\sigma_1^2 + \sigma_2^2$, *in forza del teorema dell'unicità della f.g.m.*, si può concludere che la v.c. X è normale di aspettativa $\mu_1 + \mu_2$ e varianza $\sigma_1^2 + \sigma_2^2$.

(Conseguenze).

Teorema (sulla f.g.m.)

Sia X dotata di f.g.m.. Allora, posto $Y = a + bX$, esiste $m_Y(t) = e^{ta} \cdot m_X(bt)$.

Dimostrazione

$$m_Y(t)=E\left(e^{tY}\right)=E\left(e^{t(a+bX)}\right)=e^{ta}\cdot E\left(e^{btX}\right)=$$

$$=e^{ta}\cdot m_X(bt).$$

(Standardizzazione)

Teorema

Sia Z una v.c. normale standardizzata. Allora la v.c. $Y=Z^2$ è una v.c. chi-quadrato con 1 g.d.l..

Dimostrazione

$$F_Y(y)=P\{Y\le y\}=P\left\{Z^2\le y\right\}=P\left\{-\sqrt{y}\le Z\le+\sqrt{y}\right\}=$$

$$=\int_{-\sqrt{y}}^{+\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=2\int_{0}^{+\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=$$

$$=2\int_0^y\frac{1}{\sqrt{2}}\frac{1}{\sqrt{\pi}}\frac{1}{2\sqrt{v}}e^{-\frac{v}{2}}dv=\int_0^y\left(\frac{1}{2}\right)^{\frac{1}{2}}\frac{1}{\Gamma\left(\frac{1}{2}\right)}v^{\left(1-\frac{1}{2}\right)}e^{-\frac{v}{2}}dv,$$

che è la funzione cumulata della probabilità di una v.c. chi-quadrato con 1 g.d.l..

Teorema

Siano $(X_1,X_2,\cdots,X_n)$ n variabili casuali normali di aspettativa μ e varianza σ^2 e indipendenti in probabilità.

Allora:

a) La v.c. $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (media campionaria) si distribuisce come una normale di aspettativa μ e varianza $\frac{\sigma^2}{n}$.

b) Le variabili casuali $\bar{X}$ e $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ sono indipendenti

c) La v.c. $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i - \bar{X})^2}{\sigma^2}$ (varianza campionaria corretta) si distribuisce come una chi-quadrato con $(n-1)$ g.d.l..

(Si dimostrano le parti a) e c)).