*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 della 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 $E\left(X^2\right) = \frac{\alpha\cdot(\alpha+1)}{\vartheta^2}$, 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 $\vartheta$

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*

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

*Teorema*

Sia $X_1$ una v.c. normale di aspettativa $\mu_1$ e varianza $\sigma_1^2$. Sia $X_2$ una v.c. una v.c. normale di aspettativa $\mu_2$ e varianza $\sigma_2^2$. 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 $\sigma_1^2+\sigma_2^2$.

*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*

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

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

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

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

*Teorema*

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

Allora:

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

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

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

*La distribuzione t di "Student"*

La v.c. $X$ ha distribuzione *t di Student*, con $k > 0$ g.d.l., se la sua densità è la seguente:

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \cdot \frac{1}{\sqrt{k\pi}} \cdot \frac{1}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}}, \quad x \in R$$

*(Disegnare il grafico)*

Si può dimostrare che:

$$E(X) = 0 \quad (k > 1); \qquad Var(X) = \frac{k}{k-2} \quad (k > 2)$$

Inoltre $\mu_r$ esiste per $r < k$.

*Teorema*
Sia $Z$ una v.c. distribuita come una normale standardizzata. Sia $U$ una v.c. distribuita come una chi-quadrato con $k$ g.d.l.. Siano $Z$ e $U$ indipendenti in probabilità. Allora la v.c. $X = \frac{Z}{\sqrt{U/k}}$ si distribuisce secondo una *t di Student* con $k$ g.d.l..

*Corollario*
Sia $\left(X_1, X_2, \cdots, X_n\right)$ un campione casuale proveniente dalla v.c. normale di aspettativa $\mu$ e varianza $\sigma^2$. Allora la v.c. $T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}}$ ha distribuzione *t di Student* con $(n-1)$ g.d.l..

*Dimostrazione*
Si considerino le variabili casuali $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ e $U = \frac{(n-1)S^2}{\sigma^2}$. Poiché esse sono indipendenti,

si può applicare il teorema precedente, ottenendo la v.c.:

$$T = \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\bar{X}-\mu}{\sqrt{S^2/n}},$$

la cui distribuzione è *t di Student* con $(n-1)$ g.d.l..

*La distribuzione F di Fisher*

La v.c. $X$ ha distribuzione *F di Fisher* con $m$ e $n$ gradi di libertà se ha la seguente funzione di densità:

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}\left(\dfrac{m}{n}\right)^{\frac{m-2}{2}} \dfrac{x^{\frac{m-2}{2}}}{\left(1+\dfrac{m}{n}x\right)^{\frac{m+n}{2}}} & x>0 \\ 0 & altrove \end{cases}$$

$m$ rappresenta i gradi di libertà del numeratore; $n$ rappresenta i gradi di libertà del denominatore.

Si può dimostrare che:

- $E(X) = \dfrac{n}{n-2} \quad (n > 2)$
- $Var(X) = \dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad (n > 4)$
- $\mu_r = E(X^r)$ esiste solo per $r < \frac{n}{2}$.

*Teorema*
Siano $U$ e $V$ due v.c. chi-quadrato indipendenti con $m$ e $n$ g.d.l. rispettivamente. Allora la v.c.

$$X = \frac{U/m}{V/n}$$

è distribuita come una v.c. $F$ con $m$ e $n$ g.d.l..

*Corollario 1*
Sia $(X_1, X_2, \cdots, X_m)$ un campione casuale proveniente dalla v.c. normale di aspettativa $\mu_X$ e varianza $\sigma^2$. Sia $(Y_1, Y_2, \cdots, Y_n)$ un campione casuale proveniente dalla v.c. normale di aspettativa $\mu_Y$ e varianza $\sigma^2$. Siano

i due campioni indipendenti, cioè estratti da popolazioni differenti. Allora, la v.c.

$$X = \frac{\left. \dfrac{\sum_{i=1}^{m}\left(X_i - \overline{X}\right)^2}{\sigma^2} \middle/ (m-1) \right.}{\left. \dfrac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}{\sigma^2} \middle/ (n-1) \right.} = \frac{S_X^2}{S_Y^2}$$

ha distribuzione $F$ con $(m-1)$ e $(n-1)$ g.d.l.

*Corollario 2*

Se $X$ ha distribuzione $F$ con $m$ e $n$ g.d.l., allora $Y = \dfrac{1}{X}$ ha distribuzione $F$ con $n$ e $m$ g.d.l..

Tale corollario è utile per la consultazione della tavole della distribuzione.

Si supponga di voler determinare il quantile $y_\alpha$. Esso è definito come $P\{Y \le y_\alpha\} = \alpha$.

Vale la seguente relazione:

$P\{Y \le y_\alpha\} = P\left\{ X = \dfrac{1}{Y} \ge \dfrac{1}{y_\alpha} \right\} = \alpha$, da cui

$$P\left\{X \le \frac{1}{y_\alpha}\right\} = 1-\alpha \;\Rightarrow\; \frac{1}{y_\alpha} = x_{(1-\alpha)} \;\Rightarrow\; y_\alpha = \frac{1}{x_{(1-\alpha)}}.$$

# Le verifiche d'ipotesi

Si considerino $n$ variabili casuali $(X_1, \cdots, X_n)$ indipendenti e aventi la medesima funzione di densità $f(x, \vartheta)$ - variabile casuale continua –, oppure la medesima funzione di probabilità $p(x, \vartheta)$ - variabile casuale discreta. Allora, si dirà che $(X_1, \cdots, X_n)$ è un campione casuale proveniente dalla funzione di densità $f(x, \vartheta)$, oppure dalla funzione di probabilità $p(x, \vartheta)$.

Per ipotesi statistica s'intende una congettura sulla forma della distribuzione dalla quale provengono i dati.

*Esempio 1*

1. l'aspettativa è pari a 3
2. la varianza è maggiore di 5
3. la distribuzione è simmetrica
4. la distribuzione è normale (gamma, Poisson, ecc.)

Nella teoria classica delle verifiche d'ipotesi, all'ipotesi statistica oggetto di verifica, detta ipotesi nulla ($H_0$), se ne deve affiancare una contraria, scelta del decisore, detta ipotesi alternativa ($H_1$).

Si dice test statistico una partizione dello spazio dei possibili risultati campionari in due sottoinsiemi disgiunti:

- la regione critica $C$, ovvero l'insieme dei risultati campionari per cui il test prescrive di rifiutare l'ipotesi nulla
- la regione di accettazione $\overline{C}$, ovvero l'insieme dei risultati campionari per cui il test prescrive di accettare l'ipotesi nulla (l'insieme complementare, o negazione, di $C$).

Si possono compiere due tipologie d'errore statistico:

- l'errore di prima specie, che consiste nel rifiutare l'ipotesi nulla quando essa è vera
- l'errore di seconda specie, che consiste nell'accettare l'ipotesi nulla quando essa è falsa.

| $D \vert V$ | $H_0$ | $H_1$ |
|---|---|---|
| $H_0$ | - | II specie |
| $H_1$ | I specie | - |

*D* è l'ipotesi scelta, *V* è l'ipotesi vera.

Data la natura statistica dell'esperimento, tali errori si possono commettere con una certa probabilità:

- $\alpha \equiv \Pr(errore\ di\ I\ specie)$

- $\beta \equiv \Pr(errore\ di\ II\ specie)$.

Un test ideale sarebbe quello che renda minimi contemporaneamente le due probabilità d'errore. Sfortunatamente, le due probabilità d'errore hanno andamenti contrapposti, come mostra il seguente

*Esempio 2*
Si voglia verificare l'ipotesi nulla che una moneta sia regolare, sulla base di $n = 5$ lanci. Pertanto, si ha un campione casuale da una variabile casuale indicatore di parametro *p* (probabilità di ottenere testa).
Sappiamo che la variabile casuale "numero di teste in 5 lanci" *X* ha distribuzione $bin(n = 5, p)$.
Se la moneta è regolare, *p* sarà pari a ½.
Occorre, come si diceva, definire un'ipotesi alternativa, supponiamo, *p* pari a 0,8.
Questo problema di verifica d'ipotesi può formalizzarsi così:

$$\begin{aligned} H_0 &: p = 0{,}5 \\ H_1 &: p = 0{,}8 \end{aligned}.$$

Si consideri il test avente la seguente regione critica:

$$C = \{\, X \geq 4 \},$$

in parole, il test prescrive di rifiutare l'ipotesi nulla se su 5 lanci almeno in 4 esce testa.

Per tale test, calcoliamo le probabilità d'errore statistico:

- $\alpha = P\{X \geq 4 | p = 0,5\} = \sum_{x=4}^{5} \binom{5}{x} \cdot 0,5^x \cdot 0,5^{5-x} =$

$$= \binom{5}{4} \cdot 0,5^5 + \binom{5}{5} \cdot 0,5^5 = 6 \cdot 0,5^5 = 0,1875$$

- $\beta = P\{X < 4 | p = 0,8\} = 1 - P\{X \geq 4 | p = 0,8\} =$

$$= 1 - \sum_{x=4}^{5} \binom{5}{x} \cdot 0,8^x \cdot 0,2^{5-x} = 1 - \binom{5}{4} \cdot 0,8^4 \cdot 0,2^1 - \binom{5}{5} \cdot 0,8^5 =$$
$$= 1 - 0,4096 - 0,32768 = 0,26272$$

Consideriamo ora un secondo test, avente per regione critica:

$C_1 = \{ X \geq 5\}$,

e calcoliamone i due tipi di errore:

- $\alpha_1 = P\{X \geq 5 | p = 0,5\} = \binom{5}{5} \cdot 0,5^5 \cdot 0,5^{5-5} = 0,03125$
- $\beta_1 = P\{X < 5 | p = 0,8\} = 1 - P\{X = 5 | p = 0,8\} =$

$$=1-\binom{5}{5}\cdot 0{,}8^5\cdot 0{,}2^{5-5}=0{,}67232.$$

Si nota che $\alpha > \alpha_1$ e $\beta < \beta_1$, cioè che al crescere di $\alpha$, $\beta$ diminuisce.

Non essendo possibile minimizzare i due tipi d'errore, la teoria classica delle verifiche d'ipotesi, fissa la probabilità dell'errore di prima specie $\alpha$, e cerca il test che, a parità di $\alpha$, rende minima la probabilità d'errore di seconda specie $\beta$ (test più potente).

Un'ipotesi statistica si dice <u>semplice</u>, se specifica completamente la distribuzione da cui provengono i dati. Altrimenti, essa si dice <u>composta</u>.

Le ipotesi dell'*esempio 2* sono entrambe semplici. Nell'*esempio 1*:

- l'ipotesi 1. è semplice se si campiona dalla distribuzione normale con varianza nota, composta se la varianza non è nota
- l'ipotesi 2. è composta.

Il lemma di Neyman-Pearson fornisce la regione critica del test più potente, qualora si considerino due ipotesi semplici. Generalizzando opportunamente la teoria, è possibile ricavare regioni critiche di test che possiedono buone proprietà statistiche.

Per il campionamento da distribuzione normale, ecco le regioni critiche dei test sui parametri, per vari problemi di verifica d'ipotesi:

*Verifiche d'ipotesi su* $\mu$

• Caso 1 $\sigma^2$ nota.

$H_0 : \mu \leq \mu_0$

$H_1 : \mu > \mu_0$

Si rifiuta $H_0$ se $\dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} > z_{1-\alpha}$, o, equivalentemente, se $\bar{X} > \mu_0 + z_{1-\alpha} \dfrac{\sigma}{\sqrt{n}}$.

$H_0 : \mu \geq \mu_0$

$H_1 : \mu < \mu_0$

Si rifiuta $H_0$ se $\dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} < -z_{1-\alpha}$, o, equivalentemente, se $\bar{X} < \mu_0 - z_{1-\alpha} \dfrac{\sigma}{\sqrt{n}}$.

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

Si rifiuta $H_0$ se $\dfrac{\left|\overline{X} - \mu_0\right|}{\sigma / \sqrt{n}} > z_{1-\frac{\alpha}{2}}$, o, equivalentemente, se

$$\overline{X} \notin \left[ \mu_0 - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \mu_0 + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

• Caso 2 $\sigma^2$ non nota.

$H_0 : \mu \leq \mu_0$

$H_1 : \mu > \mu_0$

Si rifiuta $H_0$ se

$$\frac{\overline{X} - \mu_0}{S / \sqrt{n}} > t_{1-\alpha}(n-1),$$

o, equivalentemente, se

$$\overline{X} > \mu_0 + t_{1-\alpha}(n-1) \frac{S}{\sqrt{n}}.$$

$t_{1-\alpha}(n-1)$ indica il quantile di ordine $(1-\alpha)$ della distribuzione $t$ con $(n-1)$ g.d.l..

$H_0 : \mu \geq \mu_0$

$H_1 : \mu < \mu_0$

Si rifiuta $H_0$ se $\dfrac{\bar{X} - \mu_0}{t / \sqrt{n}} < -t_{1-\alpha}(n-1)$, o, equivalentemente, se $\bar{X} < \mu_0 - t_{1-\alpha}(n-1)\dfrac{S}{\sqrt{n}}$.

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

Si rifiuta $H_0$ se

$\dfrac{\left|\bar{X} - \mu_0\right|}{S / \sqrt{n}} > t_{1-\frac{\alpha}{2}}(n-1)$,

o, equivalentemente, se

$\bar{X} \notin \left[\mu_0 - t_{1-\frac{\alpha}{2}}(n-1)\dfrac{S}{\sqrt{n}} ; \mu_0 + t_{1-\frac{\alpha}{2}}(n-1)\dfrac{S}{\sqrt{n}}\right]$.

*Verifiche d'ipotesi su* $\sigma^2$.

- Caso 1 $\mu$ nota

$H_0 : \sigma^2 \leq \sigma_0^2$

$H_1 : \sigma^2 > \sigma_0^2$

Si rifiuta $H_0$ se vale

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2} > \chi^2_{1-\alpha}(n),$$

dove $\chi^2_{1-\alpha}(n)$ indica il quantile di ordine $(1-\alpha)$ della distribuzione chi-quadrato con $n$ g.d.l..

$H_0 : \sigma^2 \geq \sigma_0^2$

$H_1 : \sigma^2 < \sigma_0^2$

Si rifiuta $H_0$ se vale

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2} < \chi^2_{\alpha}(n).$$

$H_0 : \sigma^2 = \sigma_0^2$

$H_1 : \sigma^2 \neq \sigma_0^2$

Si rifiuta $H_0$ se vale

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2} \notin \left[\chi^2_{\frac{\alpha}{2}}(n);\chi^2_{1-\frac{\alpha}{2}}(n)\right].$$

- Caso 2 $\mu$ non nota

$H_0 : \sigma^2 \le \sigma_0^2$

$H_1 : \sigma^2 > \sigma_0^2$

Si rifiuta $H_0$ se vale

$$\frac{\sum_{i=1}^{n} \left( X_i - \overline{X} \right)^2}{\sigma_0^2} > \chi^2_{1-\alpha}(n-1),$$

dove $\chi^2_{1-\alpha}(n-1)$ indica il quantile di ordine $(1-\alpha)$ della distribuzione chi-quadrato con $(n-1)$ g.d.l..

$H_0 : \sigma^2 \ge \sigma_0^2$

$H_1 : \sigma^2 < \sigma_0^2$

Si rifiuta $H_0$ se vale

$$\frac{\sum_{i=1}^{n} \left( X_i - \overline{X} \right)^2}{\sigma_0^2} < \chi^2_{\alpha}(n-1).$$

$H_0 : \sigma^2 = \sigma_0^2$

$H_1 : \sigma^2 \neq \sigma_0^2$

Si rifiuta $H_0$ se vale

$$\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{\sigma_0^2}\notin\left[\chi^2_{\frac{\alpha}{2}}(n-1);\chi^2_{1-\frac{\alpha}{2}}(n-1)\right].$$

*Definizione*

Dicesi *p-value* il valore minimo (massimo) della probabilità dell'errore di prima specie per il quale si rifiuta (si accetta) l'ipotesi nulla $H_0$ .