1. IL TEOREMA DEL LIMITE CENTRALE

Il teorema del limite centrale è un risultato di grande importanza in quanto sancisce il fatto che la sovrapposizione di un gran numero di variabili aleatorie aventi media e varianza comune conduce ad una variabile con distribuzione normale (gaussiana).

Più precisamente possiamo dire che

Siano

$$\xi_1, \xi_2, ..., \xi_n$$

variabili aleatorie indipendenti aventi la stessa distribuzione di probabilità con media μ e varianza σ^2.

Siano

$$\eta_1, \eta_2, ..., \eta_n$$

le corrispondenti variabili normalizzate

$$\eta_k = \frac{\xi_k - \mu}{\sigma}$$

e consideriamo la variabile aleatoria ζ definita da

$$\zeta_n = \frac{\eta_1 + \eta_2 + ... + \eta_n}{\sqrt{n}}$$

(Poichè $E(\eta_i) = 0$ avremo che $E(\zeta_n) = nE(\eta_1 + \eta_2 + ... + \eta_n) = 0$ e inoltre, poichè $\text{Var}(\eta_i) = 1$ avremo che $\text{Var}(\eta_1 + \eta_2 + ... + \eta_n) = n$ e quindi $\text{Var}(\zeta_n) = 1$)

Si ha

$$\zeta_n = \frac{\xi_1 + \xi_2 + ... + \xi_n - n\mu}{\sqrt{n}\sigma}$$

e si può dimostrare che

$$\boxed{\lim_{n\to+\infty} \mathcal{P}(\alpha \le \zeta_n \le \beta) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-x^2/2} dx}$$

o equivalentemente che

$$\boxed{\lim_{n\to+\infty} \mathcal{P}(a \le \xi_1 + \xi_2 + ... + \xi_n \le b) = \frac{1}{\sqrt{2\pi}} \int_{\frac{a-n\mu}{\sqrt{n}\sigma}}^{\frac{b-n\mu}{\sqrt{n}\sigma}} e^{-x^2/2} dx}$$

1.1 *Un caso particolare del Teorema del Limite Centrale: il teorema di DeMoivre-Laplace*

Sia ξ_n la variabile aleatoria binomiale definita dalla somma di n variabili Bernoulliane indipendenti ξ relative ad una prova con probabilità di successo p e probabilità di insuccesso $q = 1 - p$.

$$\xi_n = \xi + \xi + \xi + ... + \xi$$

La media di ξ è $\mu = p$, la sua varianza è $\sigma = \sqrt{pq}$ per cui la media di ξ_n è np la sua varianza è $\sqrt{npq}$ mentre la sua densità di probabilità è definita, per $k - \frac{1}{2} \le x \le k + \frac{1}{2}$ da

$$\mathcal{B}_n(x) = \mathcal{P}(\xi_n = k) = \mathcal{P}(k - \frac{1}{2} \le \xi_n \le k + \frac{1}{2}) = \binom{n}{k} p^k q^{n-k}$$

e, utilizzando la Formula di Stirling,

$$\mathcal{B}_n(h) \approx\approx \frac{n^n e^{-n}\sqrt{2\pi n}}{h^h e^{-h}\sqrt{2\pi h}(n-h)^{(n-h)} e^{-(n-h)}\sqrt{2\pi(n-h)}} p^h q^{(n-h)}$$

Sia

$$\zeta_n = \frac{\xi_n - np}{\sqrt{npq}}$$

la variabile aleatoria ottenuta normalizzando ξ_n e sia $G_n(x)$ la sua PDF.

Avremo che

$$\begin{aligned}\int_{k-\frac{1}{2\sqrt{npq}}}^{k+\frac{1}{2\sqrt{npq}}} G_n(x)dx = \mathcal{P}(k - \frac{1}{2\sqrt{npq}} \le \zeta_n \le k + \frac{1}{2\sqrt{npq}}) = \\ = \mathcal{P}(k\sqrt{npq} - \frac{1}{2} \le \xi_n - np \le k\sqrt{npq} + \frac{1}{2}) = \\ = \mathcal{P}(np + k\sqrt{npq} - \frac{1}{2} \le \xi_n \le np + k\sqrt{npq} + \frac{1}{2}) = \\ = \mathcal{B}_n(np + k\sqrt{npq})\end{aligned}$$

da cui

$$\frac{1}{\sqrt{npq}} G_n(k) \approx \mathcal{B}_n(np + k\sqrt{npq})$$

e

$$G_n(k) \approx \sqrt{npq}\mathcal{B}_n(np + k\sqrt{npq})$$

Possiamo ora mostrare che

$$\sqrt{npq}\mathcal{B}_n(np + k\sqrt{npq}) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{k^2}{2}}$$

Avremo

$$\mathcal{B}_n(h) =\approx \frac{n^n e^{-n}\sqrt{2\pi n}}{h^h e^{-h}\sqrt{2\pi h}(n-h)^{(n-h)}e^{-(n-h)}\sqrt{2\pi(n-h)}}p^h q^{(n-h)} =$$
$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{h(n-h)}}\frac{n^n}{h^h(n-h)^{n-h}}p^h q^{(n-h)}$$

e definendo $\delta = k\sqrt{pq}$,

$$h = np + k\sqrt{npq} = np + \delta\sqrt{n} \quad \text{da cui} \quad n-h = n - np - k\sqrt{npq} = nq + \delta\sqrt{n}$$

si ha

$$\sqrt{npq}\mathcal{B}_n(np+\delta\sqrt{n}) \approx$$
$$= \alpha_n\beta_n = \left(\sqrt{npq}\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{(np+\delta\sqrt{n})(nq-\delta\sqrt{n})}}\right)$$
$$\left(\frac{n^n}{(np+\delta\sqrt{n})^{(np+\delta\sqrt{n})}(nq-\delta\sqrt{n})^{(nq-\delta\sqrt{n}))}}p^{(np+\delta\sqrt{n})}q^{(nq-\delta\sqrt{n})}\right)$$

Osserviamo subito che

$$\alpha_n = \sqrt{npq}\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{(np+\delta\sqrt{n})(nq-\delta\sqrt{n})}} \to \frac{1}{\sqrt{2\pi}}$$

e che

$$\beta_n = \frac{n^n}{(np+\delta\sqrt{n})^{(np+\delta\sqrt{n})}(nq-\delta\sqrt{n})^{(nq-\delta\sqrt{n})}}p^{(np+\delta\sqrt{n})}q^{(nq-\delta\sqrt{n})} =$$
$$= \frac{n^{np+\delta\sqrt{n}+nq-\delta\sqrt{n}}}{(np+\delta\sqrt{n})^{(np+\delta\sqrt{n})}(nq-\delta\sqrt{n})^{(nq-\delta\sqrt{n})}}p^{-(np+\delta\sqrt{n})}q^{(nq-\delta\sqrt{n})} =$$
$$\frac{np^{np+\delta\sqrt{n}}}{(np+\delta\sqrt{n})^{(np+\delta\sqrt{n})}}\frac{nq^{nq-\delta\sqrt{n}}}{(nq-\delta\sqrt{n})^{(nq-\delta\sqrt{n})}}$$

Si ha

$$\beta_n = \left(1+\frac{\delta}{p\sqrt{n}}\right)^{-np-\delta\sqrt{n}}\left(1-\frac{\delta}{q\sqrt{n}}\right)^{-nq+\delta\sqrt{n}} =$$
$$= e^{(-np-\delta\sqrt{n})\ln\left(1+\frac{\delta}{p\sqrt{n}}\right)+(-nq+\delta\sqrt{n})\ln\left(1-\frac{\delta}{q\sqrt{n}}\right)}$$

Ma

$$(-np-\delta\sqrt{n})\ln\left(1+\frac{\delta}{p\sqrt{n}}\right)+(-nq+\delta\sqrt{n})\ln\left(1-\frac{\delta}{q\sqrt{n}}\right)\approx$$
$$\approx(-np-\delta\sqrt{n})\left(\frac{\delta}{p\sqrt{n}}-\frac{\delta^2}{2p^2n}\right)+(-nq+\delta\sqrt{n})\left(-\frac{\delta}{q\sqrt{n}}-\frac{\delta^2}{2q^2n}\right)\approx$$
$$-\frac{np\delta}{p\sqrt{n}}+\frac{np\delta^2}{2p^2n}-\frac{\delta^2}{p}+\frac{\delta^3}{2p^2\sqrt{n}}+\frac{nq\delta}{q\sqrt{n}}+\frac{nq\delta^2}{2q^2n}-\frac{\delta^2}{q}+\frac{\delta^3}{2q^2\sqrt{n}}=$$
$$=+\frac{\delta^2}{2p}-\frac{\delta^2}{p}+\frac{\delta^3}{2p^2\sqrt{n}}+\frac{\delta^2}{2q}-\frac{\delta^2}{q}+\frac{\delta^3}{2q^2\sqrt{n}}\to$$
$$\to-\left(\frac{\delta^2}{2p}+\frac{\delta^2}{2q}\right)=$$
$$=-\frac{h^2pq}{2p}-\frac{h^2pq}{2q}=-\frac{h^2}{2}$$

Pertanto

$$\beta_n\to e^{\frac{-k^2}{2}}$$

ed infine

$$\sqrt{npq}\mathcal{B}_n(np+k\sqrt{npq})=\sqrt{npq}\mathcal{B}_n(np+\delta\sqrt{n})\to\frac{1}{2\sqrt{\pi}}e^{-\frac{k^2}{2}}$$

Quindi possiamo affermare che

$$\mathcal{P}(a\le\xi_n\le b)=\mathcal{P}(a\le\xi+\xi+\xi+....+\xi\le b)\approx$$
$$\approx\frac{1}{\sqrt{2\pi}}\int_{\frac{a-np}{\sqrt{npq}}}^{\frac{b-np}{\sqrt{npq}}}e^{-x^2/2}dx=G(\frac{b-np}{\sqrt{npq}})-G(\frac{a-np}{\sqrt{npq}})$$

dove

$$G(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-x^2/2}dx$$

è la funzione di distribuzione cumulativa Gaussiana standardizzata.

1.2 *Un altro caso particolare del Teorema del Limite Centrale*

Sia ξ_n la variabile aleatoria definita dalla somma di n variabili Poissoniane ξ di media λ.

$$\xi_n=\xi+\xi+\xi+....+\xi$$

La media di ξ è $\mu=\lambda$, la sua varianza è $\sigma=\sqrt{\lambda}$ per cui la media di ξ_n è $n\lambda$ la sua varianza è $\sqrt{n\lambda}$ mentre la sua densità di probabilità è definita da

$$H_n(k)=\mathcal{P}(\xi_n=k)=\frac{(n\lambda)^k}{k!}e^{-n\lambda}$$

Possiamo normalizzare la variabile aleatoria ξ_n mediante la

$$\frac{\xi_n - n\lambda}{\sqrt{n\lambda}}$$

la cui funzione densità di probabilità è data da

$$\sqrt{n\lambda} H_n(n\lambda + k\sqrt{n\lambda})$$

Possiamo mostrare che

$$\sqrt{n\lambda} H_n(n\lambda + k\sqrt{n\lambda}) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{k^2}{2}}$$

Avremo infatti che

$$\begin{aligned}
&\sqrt{n\lambda} H_n(n\lambda + k\sqrt{n\lambda}) = \\
&\quad = \sqrt{n\lambda} \frac{(n\lambda)^{n\lambda + k\sqrt{n\lambda}} e^{-n\lambda}}{(n\lambda + k\sqrt{n\lambda})^{n\lambda + k\sqrt{n\lambda}} e^{-n\lambda - k\sqrt{n\lambda}} \sqrt{2\pi(n\lambda + k\sqrt{n\lambda})}} = \\
&\qquad = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{1}{1 + \frac{k}{\sqrt{n\lambda}}}} \frac{1}{\left(1 + \frac{k}{\sqrt{n\lambda}}\right)^{n\lambda + k\sqrt{n\lambda}}} \frac{1}{e^{-k\sqrt{n\lambda}}} =
\end{aligned}$$

Si verifica subito che

$$\frac{1}{\sqrt{2\pi}} \sqrt{\frac{1}{1 + \frac{k}{\sqrt{n\lambda}}}} \to \frac{1}{\sqrt{2\pi}}$$

mentre

$$\frac{1}{\left(1 + \frac{k}{\sqrt{n\lambda}}\right)^{n\lambda + k\sqrt{n\lambda}}} \frac{1}{e^{-k\sqrt{n\lambda}}} = e^{-\epsilon_n}$$

dove

$$\epsilon_n = (n\lambda + k\sqrt{n\lambda}) \ln\left(1 + \frac{k}{\sqrt{n\lambda}}\right) - k\sqrt{n\lambda}$$

Usando lo sviluppo di Taylor del logaritmo avremo allora

$$\begin{aligned}
&\epsilon_n \approx (n\lambda + k\sqrt{n\lambda}) \left(\frac{k}{\sqrt{n\lambda}} - \frac{k^2}{2n\lambda}\right) - k\sqrt{n\lambda} = \\
&\qquad = k\sqrt{n\lambda} - \frac{k^2}{2} + k^2 - \frac{k^2}{2} - \frac{k^3}{2\sqrt{n\lambda}} - k\sqrt{n\lambda} \to \frac{k^2}{2}
\end{aligned}$$

Se ne deduce che

$$\epsilon_n \to e^{-\frac{x^2}{2}}$$

e si può concludere.

1.2.1 Approssimazione della Distribuzione Binomiale mediante la Distribuzione Normale

Il teorema di De Moivre-Laplace, qualora il numero di prove n sia grande, permette di approssimare una distribuzione binomiale mediante la distribuzione normale standardizzata.

Nella pratica l'approssimazione si usa se $np, nq \geq 5$ essendo p la probabilità di successo e $q = 1 - p$.

Ad esempio si può calcolare la probabilità che su 10 lanci di una moneta non truccata si abbiano un numero di teste ξ compreso tra 3 e 6, tenendo conto che

$$P(3 \leq \xi \leq 6) = P\left(\frac{3 - \frac{1}{2} - 5}{\sqrt{2.5}} \leq z_n \leq \frac{6 + \frac{1}{2} - 5}{\sqrt{2.5}}\right) = \\ = P(-1.58 \leq z_n \leq 0.95)$$

Se G è la CDF Gaussiana standardizzata avremo:

$$P(-1.58 \leq z_n \leq 0.95) = G(0.95) - G(1.58) = 0.8289 - 0.0571 = 0.7718$$

Possiamo confrontare il risultato con quello ottenuto direttamente mediante i valori della distribuzione cumulativa binomiale B_{10} relativa a 10 lanci; in questo modo si ottiene

$$B_{10}(6) - B_{10}(2) = 0.8281 - 0.0547 = 0.7734$$

ed osservare che l'errore di approssimazione commesso è dell'ordine di 0.0016

Seguono le istruzioni Matlab per calcolare

$G(0.95) - G(1.58)) =$ `normcdf(0.95,0,1)-normcdf(-1.58,0,1)`

e

$B_{10}(6) - B_{10}(2) =$ `cdf('bino',6,10,1/2)-cdf('bino',3,10,1/2)`

2. I Test Statistici

2.1 Le distribuzioni legate ai test statistici.

2.1.1 La distribuzione χ^2

La variabile aleatoria χ^2 a ν gradi di libertà restituisce la somma dei quadrati di ν variabili aleatorie ξ_i indipendenti, aventi distribuzione gaussiano con media 0 e varianza 1 (distribuzioni normali standardizzate).

$$\chi^2 = \xi_1^2 + \xi_2^2 + ... + \xi_\nu^2$$

Per ricavare la PDF della distribuzione χ^2 a ν gradi di libertà è opportuno procedere come segue.

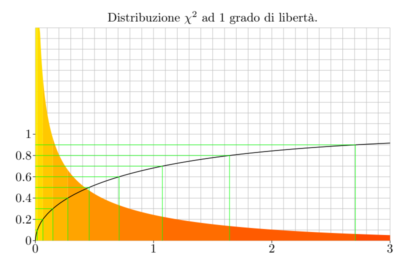

Figure 2.1: .

2.1.2 χ^2 ad 1 grado di libertà

Se ξ è una variabile aleatoria gaussiana standard e se $\eta = \xi^2$, la PDF φ di η è data da

$$\varphi(t) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{s}{2}}}{\sqrt{s}} & s \geq 0 \\ 0 & s < 0 \end{cases}$$

Infatti si ha

$$\mathcal{P}(\eta \leq \alpha) = \mathcal{P}(\xi^2 \leq \alpha) = \mathcal{P}(-\sqrt{\alpha} \leq \xi \leq \sqrt{\alpha}) =$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{\alpha}}^{\sqrt{\alpha}} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{\alpha}} e^{-\frac{t^2}{2}} dt =$$

posto $s = t^2$, per cui $ds = 2tdt$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{\alpha} \frac{e^{-\frac{s}{2}}}{2\sqrt{s}} ds = \frac{1}{\sqrt{2\pi}} \int_0^{\alpha} \frac{e^{-\frac{s}{2}}}{\sqrt{s}} ds$$

2.1.3 χ^2 a 2 gradi di libertà

Se η_1, η_2 sono variabili aleatorie $\eta_i = \xi_i^2$ con ξ_i gaussiana standard e se $\eta = \eta_1 + \eta_2$, la PDF φ di η è data da

$$\varphi(t) = \begin{cases} \frac{1}{2}e^{-\frac{s}{2}} & s \geq 0 \\ 0 & s < 0 \end{cases}$$

Infatti, se chiamiamo φ_i la PDF di η_i, si ha

$$\varphi(s) = \int_{-\infty}^{+\infty} \varphi_1(t)\varphi_2(s-t)dt =$$
$$= \int_0^s \varphi_1(t)\varphi_2(s-t)dt = \frac{1}{2\pi}\int_0^s \frac{e^{-\frac{t}{2}}e^{-\frac{s-t}{2}}}{\sqrt{t}\sqrt{s-t}}dt =$$
$$= \frac{1}{2\pi}e^{-\frac{s}{2}}\int_0^s \frac{1}{\sqrt{t}\sqrt{s-t}}dt$$

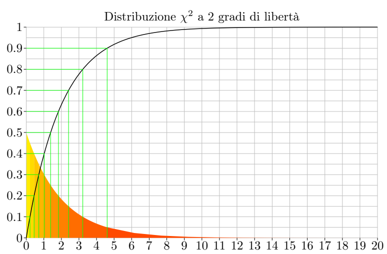

Figure 2.2: .

Posto $\sqrt{t(s-t)} = ut$ si ha

$$t(s-t) = u^2t^2 \ , \ s-t = u^2t \ , \ s = (1+u^2)t$$

$$t = \frac{s}{1+u^2} \ , \ dt = -\frac{2us}{(1+u^2)^2}du$$

per cui

$$\int \frac{1}{\sqrt{t(s-t)}}dt = \int -\frac{2us}{ut(1+u^2)^2}du =$$
$$= \int -\frac{2us}{u\frac{s}{1+u^2}(1+u^2)^2}du = -2\int \frac{1}{1+u^2}du =$$
$$= -2\arctan(u) = -2\arctan\sqrt{\frac{s-t}{t}}$$

Ne viene

$$\int_0^s \frac{1}{\sqrt{t}\sqrt{s-t}}dt = -2\arctan\sqrt{\frac{s-t}{t}}\Big|_0^s = 2\frac{\pi}{2} = \pi$$

e si può dedurre che

$$\varphi(s) = \frac{1}{2\pi}e^{-\frac{s}{2}}\pi = \frac{1}{2}e^{-\frac{s}{2}}$$

è la distribuzione di η.

2.1.4 χ^2 a ν gradi di libertà

Possiamo ora provare che la variabile aleatoria

$$\eta = \eta_1 + \eta_2 + \cdot + \eta_\nu = \xi_1^2 + \xi_2^2 + \cdot + \xi_\nu^2$$

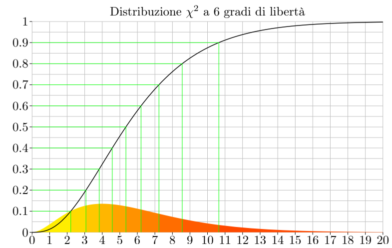

Figure 2.3: .

dove ξ_i sono variabili aleatorie gaussiane normalizzate, ha una PDF definita da

$$\varphi_\nu(u) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}}$$

Le verifiche fatte precedentemente consentono di affermare che la precedente affermazione è vera per $\nu = 1$ e per $\nu = 2$. Pertanto per verificare la medesima è sempre vera sarà sufficiente provare che, supposta vera per ν è vera anche per $\nu + 2$.

Si ha

$$\varphi_{\nu+2}(u) = \varphi_\nu(u) * \varphi_2(u) = \int_0^u \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} (u-t)^{\frac{\nu}{2}-1} e^{-\frac{u-t}{2}} \frac{1}{2} e^{-\frac{t}{2}} dt =$$

$$= \frac{1}{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{u}{2}} \int_0^u (u-t)^{\frac{\nu}{2}-1} du = \frac{1}{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{u}{2}} \left[-\frac{(u-t)^{\frac{\nu}{2}}}{\frac{\nu}{2}} \right]_0^u =$$

$$= \frac{1}{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{u}{2}} \frac{(u)^{\frac{\nu}{2}}}{\frac{\nu}{2}} = \frac{1}{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}+1\right)} u^{\frac{\nu}{2}} e^{-\frac{u}{2}}$$

Media, varianza della variabile aleatoria χ^2 sono date da

$$\mu = \nu$$
$$\sigma^2 = 2\nu$$

Infatti

$$\mu = \int_0^{+\infty} \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} u u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}} du = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{+\infty} u^{\frac{\nu}{2}} e^{-\frac{u}{2}} du =$$
$$\frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{+\infty} (2t)^{\frac{\nu}{2}} e^{-t} 2dt \frac{2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu}{2}+1\right)}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} =$$
$$\frac{2^{\frac{\nu}{2}+1}\frac{\nu}{2}\Gamma\left(\frac{\nu}{2}\right)}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} = \nu$$

mentre

$$\int_0^{+\infty} \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} u^2 u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}} du = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{+\infty} u^{\frac{\nu}{2}+1} e^{-\frac{u}{2}} du =$$
$$\frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{+\infty} (2t)^{\frac{\nu}{2}+1} e^{-t} 2dt \frac{2^{\frac{\nu}{2}+2}\Gamma\left(\frac{\nu}{2}+2\right)}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} =$$
$$\frac{2^{\frac{\nu}{2}+2}(\frac{\nu}{2}+1)\frac{\nu}{2}\Gamma\left(\frac{\nu}{2}\right)}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} = 4(\frac{\nu}{2}+1)\frac{\nu}{2}$$

e

$$\sigma^2 = 4(\frac{\nu}{2}+1)\frac{\nu}{2} - \nu^2 = 2\nu$$

Inoltre la generatrice dei momenti della variabile aleatoria χ^2 è

$$M_{\chi^2}(t) = (1-2t)^{-\nu/2}$$

2.1.5 *La distribuzione T di Student.*

È la distribuzione di una variabile aleatoria T che restituisce il rapporto

$$T = \frac{\xi}{\sqrt{\frac{\eta}{\nu}}}$$

dove ξ è una variabile aleatoria con densità di probabilità gaussiana normale (media 0 e varianza 1) ed η è una variabile aleatoria con distribuzione χ^2 a ν gradi di libertà

Per ricavare la PDF di student cominciamo con l'osservare che:

2.1.6

Se ξ è una variabile aleatoria con densità χ^2 a ν gradi di libertà, la sua PDF sarà data da

$$\varphi_\nu(u) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}}$$

Sia η la variabile aleatoria definita da

$$\eta = \sqrt{\frac{\xi}{\nu}}$$

Avremo che

$$\mathcal{P}(\eta \leq \alpha) = \mathcal{P}\left(\sqrt{\frac{\xi}{\nu}} \leq \alpha\right) = \mathcal{P}(\xi \leq \nu\alpha^2) =$$

$$= \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{\nu\alpha^2} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}} du =$$

posto $t = \sqrt{\frac{u}{\nu}}$ da cui $u = \nu t^2$ e $du = 2\nu t dt$

$$= \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{\alpha} t^{\nu-2} \nu^{\frac{\nu}{2}-1} e^{-\frac{\nu t^2}{2}} 2\nu t dt = \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{\alpha} t^{\nu-1} \nu^{\frac{\nu}{2}} e^{-\frac{\nu t^2}{2}} dt$$

e possiamo concludere che la PDF di η è data da

$$\varphi(t) = \begin{cases} \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} t^{\nu-1} \nu^{\frac{\nu}{2}} e^{-\frac{\nu t^2}{2}} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

2.1.7

Sia ora η_1 una variabile aleatoria gaussiana standard, la cui PDF è ovviamente data da

$$g_1(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

e sia $\eta_2 = \sqrt{\frac{\xi}{\nu}}$ dove ξ è una variabile aleatoria con densità χ^2 a ν gradi di libertà; per quanto detto in precedenza la PDF di η_2 è nulla prima di 0 ed è data da

$$g_2(t) = \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} t^{\nu-1} \nu^{\frac{\nu}{2}} e^{-\frac{\nu t^2}{2}}$$

per $t \geq 0$.

La variabile aleatoria T di student è definita mediante la

$$T = \frac{\eta_1}{\eta_2}$$

e possiamo ricavarne la PDF φ ricordando che si tratta del quoziente di due variabili aleatorie di cui conosciamo la densità di probabilità.

Avremo:

$$\varphi(t) = \int_0^{+\infty} x \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} x^{\nu-1} \nu^{\frac{\nu}{2}} e^{-\frac{\nu x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2 x^2}{2}} dx =$$

$$= \frac{\nu^{\frac{\nu}{2}}}{\sqrt{2\pi} 2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{+\infty} x^{\nu} e^{-\frac{x^2}{2}(\nu+t^2)} dx =$$

posto $\frac{x^2}{2}(\nu + t^2) = s$ si ha $x = \sqrt{\frac{2s}{\nu+t^2}}$ e $dx = \frac{\sqrt{2}ds}{\sqrt{\nu+t^2}2\sqrt{s}}$

$$= \frac{\nu^{\frac{\nu}{2}}}{\sqrt{2\pi} 2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \int_0^{+\infty} \frac{2^{\frac{\nu}{2}} s^{\frac{\nu}{2}}}{(\nu+t^2)^{\frac{\nu}{2}}} e^{-s} \frac{\sqrt{2}}{\sqrt{\nu+t^2}} \frac{1}{2\sqrt{s}} ds =$$

$$= \frac{\nu^{\frac{\nu}{2}} 2^{\frac{\nu}{2}-\frac{1}{2}}}{\sqrt{2\pi} 2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{(\nu+t^2)^{\frac{\nu+1}{2}}} \int_0^{+\infty} s^{\frac{\nu}{2}-\frac{1}{2}} e^{-s} ds =$$

$$= \frac{\nu^{\frac{\nu}{2}}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{(\nu)^{\frac{\nu+1}{2}} (1+\frac{t^2}{\nu})^{\frac{\nu+1}{2}}} \int_0^{+\infty} s^{\frac{\nu+1}{2}-1} e^{-s} ds =$$

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

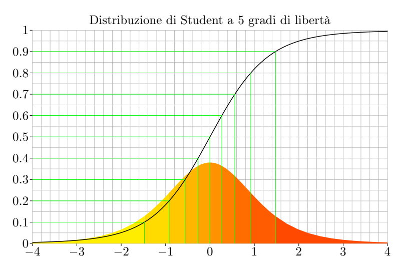

Figure 2.4: .

Pertanto la funzione di distribuzione della variabile aleatoria T a ν gradi di libertà è data da:

$$\boxed{\varphi(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}}$$

La media e la varianza di T risultano essere

$$\mu = 0$$
$$\sigma^2 = \frac{\nu}{\nu - 2}$$

Infatti

$$\mu = \int_{-\infty}^{+\infty} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} t \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt = 0$$

per la simmetria dell'integranda;

Inoltre

$$\sigma^2 = \int_{-\infty}^{+\infty} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} t^2 \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt =$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{+\infty} \frac{\nu t}{2} \left(-\frac{2}{\nu-1}\right) \frac{d}{dt} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu-1}{2}} dt =$$

integrando per parti

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(-\left(\frac{\nu t}{\nu-1}\right)\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu-1}{2}} \Bigg|_{-\infty}^{+\infty} + \frac{\nu}{\nu-1}\int_{-\infty}^{+\infty} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu-1}{2}} dt \right) =$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \frac{\nu}{\nu-1} \int_{-\infty}^{+\infty} \left(1+\frac{t^2}{\nu-2}\frac{\nu-2}{\nu}\right)^{-\frac{\nu-1}{2}} dt =$$

posto $s = t\sqrt{\frac{\nu-2}{\nu}}$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \frac{\nu}{\nu-1} \int_{-\infty}^{+\infty} \left(1+\frac{s^2}{\nu-2}\right)^{-\frac{\nu-1}{2}} \sqrt{\frac{\nu}{\nu-2}} ds =$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \frac{\nu}{\nu-1} \sqrt{\frac{\nu}{\nu-2}} \frac{\sqrt{(\nu-2)\pi}\Gamma\left(\frac{\nu-2}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)} =$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\nu}{\nu-1} \frac{\Gamma\left(\frac{\nu-2}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)} = \frac{\frac{\nu}{2}-\frac{1}{2}}{\frac{\nu}{2}-1} = \frac{\nu}{\nu-2}$$

2.1.8 *La distribuzione F di Fisher.*

È la distribuzione di una variabile aleatoria F che restituisce il rapporto

$$F = \frac{\eta/\mu}{\xi/\nu}$$

dove η ed ξ sono variabili aleatorie con distribuzione χ^2 a μ e ν gradi di libertà, rispettivamente.

Per ricavare la PDF di F ricordiamo che la PDF g di η/μ è data da

$$g(u) = \frac{\mu}{2^{\frac{\mu}{2}}\Gamma\left(\frac{\mu}{2}\right)} (\mu u)^{\frac{\mu}{2}-1} e^{-\frac{\mu u}{2}}$$

mentre la PDF f di ξ/ν è data da

$$f(u) = \frac{\nu}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} (\nu u)^{\frac{\nu}{2}-1} e^{-\frac{\nu u}{2}}$$

Possiamo quindi ricavare la PDF φ di $F = \frac{\eta/\mu}{\xi/\nu}$ usando quanto conosciamo sul rapporto di due variabili aleatorie.

Avremo

$$\begin{aligned}
\varphi(t) = \int_0^{+\infty} x \frac{\nu}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \nu^{\frac{\nu}{2}-1} x^{\frac{\nu}{2}-1} e^{-\frac{\nu x}{2}} \frac{\mu}{2^{\frac{\mu}{2}}\Gamma\left(\frac{\mu}{2}\right)} \mu^{\frac{\mu}{2}-1} (tx)^{\frac{\mu}{2}-1} e^{-\frac{\mu t x}{2}} dx = \\
= \frac{\nu^{\frac{\nu}{2}} \mu^{\frac{\mu}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\mu}{2}}\Gamma\left(\frac{\mu}{2}\right)} \int_0^{+\infty} x^{\frac{\nu}{2}} e^{-\frac{\nu x}{2}} t^{\frac{\mu}{2}-1} x^{\frac{\mu}{2}-1} e^{-\frac{\mu t x}{2}} dx = \\
= t^{\frac{\mu}{2}-1} C_{\nu,\mu} \int_0^{+\infty} x^{\frac{\nu+\mu}{2}-1} e^{-\frac{x}{2}(\nu+t\mu)} dx =
\end{aligned}$$

posto $\frac{x}{2}(\nu + t\mu) = s$ si ha $dx = \frac{2}{\nu+t\mu} ds$

$$\begin{aligned}
= t^{\frac{\mu}{2}-1} C_{\nu,\mu} \int_0^{+\infty} \left(\frac{2s}{\nu+t\mu}\right)^{\frac{\nu+\mu}{2}-1} e^{-s} \frac{2}{\nu+t\mu} ds = \\
= \frac{t^{\frac{\mu}{2}-1} C_{\nu,\mu} 2^{\frac{\nu+\mu}{2}}}{(\nu+t\mu)^{\frac{\nu+\mu}{2}}} \int_0^{+\infty} s^{\frac{\nu+\mu}{2}-1} e^{-s} ds = \\
= \frac{\nu^{\frac{\nu}{2}} \mu^{\frac{\mu}{2}} \Gamma\left(\frac{\nu+\mu}{2}\right) 2^{\frac{\nu+\mu}{2}}}{2^{\frac{\nu}{2}} 2^{\frac{\mu}{2}} \Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\mu}{2}\right)} \frac{t^{\frac{\mu}{2}-1}}{(\nu+t\mu)^{\frac{\nu+\mu}{2}}} = \\
= \frac{\nu^{\frac{\nu}{2}} \mu^{\frac{\mu}{2}} \Gamma\left(\frac{\nu+\mu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\mu}{2}\right)} t^{\frac{\mu}{2}-1} (\nu+t\mu)^{-\frac{\nu+\mu}{2}}
\end{aligned}$$

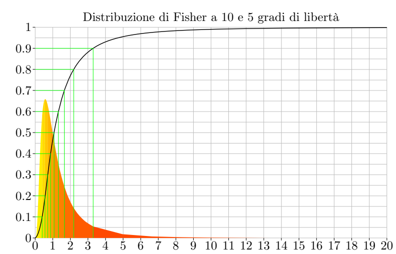

Figure 2.5: .

Pertanto la funzione di distribuzione della variabile aleatoria F è data da

$$\boxed{\varphi(t) = \begin{cases} \frac{\nu^{\frac{\nu}{2}} \mu^{\frac{\mu}{2}} \Gamma\left(\frac{\nu+\mu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\mu}{2}\right)} t^{\frac{\mu}{2}-1} (\nu+t\mu)^{-\frac{\nu+\mu}{2}} & t > 0 \\ 0 & \text{altrimenti} \end{cases}}$$

la media e varianza risultano essere

$$\begin{aligned}
\mu = \frac{\nu_2}{\nu_2 - 2} \\
\sigma^2 = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-4)(\nu_2-2)^2}
\end{aligned}$$

Infatti posto

$$C(\mu,\nu) = \frac{\nu^{\frac{\nu}{2}} \mu^{\frac{\mu}{2}} \Gamma\left(\frac{\nu+\mu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\mu}{2}\right)}$$

$$\int_0^{+\infty} t^{k+1}\varphi(t)dt = C(\mu,\nu)\int_0^{+\infty} t^{\frac{\mu}{2}+k}(\nu+t\mu)^{-\frac{\nu+\mu}{2}}dt = C(\mu,\nu)\int_0^{+\infty} t^{\frac{\mu}{2}+k}\left(\nu\left(1+t\frac{\mu}{\nu}\right)\right.$$

$$= C(\mu,\nu)\nu^{-\frac{\nu+\mu}{2}}\int_0^{+\infty} t^{\frac{\mu}{2}+k}\left(1+t\frac{\mu+2k+2}{\nu-2k-2}\ \frac{\nu-2k-2}{\mu+2k+2}\ \frac{\mu}{\nu}\right)^{-\frac{\nu+\mu}{2}} dt =$$

posto $s = \frac{\nu-2k-2}{\mu+2k+2}\frac{\mu}{\nu}t$

$$= C(\mu,\nu)\nu^{-\frac{\nu+\mu}{2}}\int_0^{+\infty} s^{\frac{\mu}{2}+k}\left(\frac{\nu}{\mu}\ \frac{\mu+2k+2}{\nu-2k-2}\right)^{\frac{\mu}{2}+k+1}\left(1+s\frac{\mu+2k+2}{\nu-2k-2}\right)^{-\frac{\nu+\mu}{2}} dt =$$

$$= C(\mu,\nu)\nu^{-\frac{\nu+\mu}{2}}\left(\frac{\nu}{\mu}\ \frac{\mu+2k+2}{\nu-2k-2}\right)^{\frac{\mu}{2}+k+1}\int_0^{+\infty} s^{\frac{\mu}{2}+k}\left(1+s\frac{\mu+2k+2}{\nu-2k-2}\right)^{-\frac{\nu+\mu}{2}} dt =$$

$$= \nu^{-\frac{\nu+\mu}{2}}\left(\frac{\nu}{\mu}\ \frac{\mu+2k+2}{\nu-2k-2}\right)^{\frac{\mu}{2}+k+1}\frac{C(\mu,\nu)}{C(\mu+2k+2,\nu-2k-2)} =$$

$$= \frac{\nu^{\frac{\nu}{2}}\mu^{\frac{\mu}{2}}\nu^{-\frac{\nu+\mu}{2}}\nu^{\frac{\mu}{2}+k+1}(\mu+2k+2)^{\frac{\mu}{2}+k+1}}{\mu^{\frac{\mu}{2}+k+1}(\nu-2k-2)^{\frac{\mu}{2}+k+1}(\nu-2k-2)^{\frac{\nu}{2}-k-1}(\mu+2k+2)^{\frac{\mu}{2}+k+1}}$$

$$\frac{\Gamma\left(\frac{\nu+\mu}{2}\right)\Gamma\left(\frac{\nu}{2}-k-1\right)\Gamma\left(\frac{\mu}{2}+k+1\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\mu}{2}\right)\Gamma\left(\frac{\nu+\mu}{2}\right)} = \frac{\nu^{k+1}}{\mu^{k+1}}\frac{\Gamma\left(\frac{\nu}{2}-k-1\right)\Gamma\left(\frac{\mu}{2}+k+1\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\mu}{2}\right)}$$

Siamo così in grado di calcolare i momenti rispetto all'origine di ogni ordine

Per $k = 0$ otteniamo il momento di ordine 1, Cioè la media,

$$\int_0^{+\infty} t\varphi(t)dt = \frac{\nu}{\mu}\frac{\Gamma\left(\frac{\nu}{2}-1\right)\frac{\mu}{2}\Gamma\left(\frac{\mu}{2}\right)}{\left(\frac{\nu}{2}-1\right)\Gamma\left(\frac{\nu}{2}-1\right)\Gamma\left(\frac{\mu}{2}\right)} = \frac{\nu}{\nu-2}$$

mentre per $k = 1$ possiamo calcolare la varianza come segue

$$\int_0^{+\infty} t^2\varphi(t)dt = \frac{\nu^2}{\mu^2}\frac{\Gamma\left(\frac{\nu}{2}-2\right)\frac{\mu}{2}\left(\frac{\mu}{2}+1\right)\Gamma\left(\frac{\mu}{2}\right)}{\left(\frac{\nu}{2}-1\right)\left(\frac{\nu}{2}-2\right)\Gamma\left(\frac{\nu}{2}-2\right)\Gamma\left(\frac{\mu}{2}\right)} =$$

$$= \frac{\nu^2}{\mu^2}\ \frac{\mu(\mu+2)}{(\nu-2)(\nu-4)}$$

$$\sigma^2 = \frac{\nu^2}{\mu^2}\ \frac{\mu(\mu+2)}{(\nu-2)(\nu-4)} - \left(\frac{\nu}{\nu-2}\right)^2 = \frac{2\nu^2(\mu+\nu-2)}{\mu(\nu-2)(\nu-4)}$$

2.1.9 *La Distribuzione T^2*

Il quadrato T^2 di una variabile aleatoria di Student a ν gradi di libertà è una variabile aleatoria di Fisher con $(1,\nu)$ gradi di libertà, infatti T è una variabile aleatoria che restituisce il rapporto

$$T = \frac{\xi}{\sqrt{\frac{\eta}{\nu}}}$$

dove ξ è una gaussiana normale ed η ha distribuzione χ^2 a ν gradi di libertà, per cui

$$\mathcal{P}(T^2 \leq \alpha) = \mathcal{P}\left(\frac{\xi^2}{\frac{\eta}{\nu}} \leq \alpha\right)$$

Osservando che ξ^2, in quanto quadrato di una gaussiana normale, ha una distribuzione χ^2 a un grado di libertà, possiamo concludere che T^2 è una distribuzione di Fisher a $(1, \nu)$ gradi di libertà.

2.1.10 *Test Statistici.*

Definiamo

H_0 l'affermazione che vogliamo sottoporre a test

e

H_a o H_1 una affermazione alternativa.

Va subito detto che H_a **non è necessariamente la negazione di** H_0 ed anzi, in corrispondenza di una stessa ipotesi H_0 si possono programmare diversi test semplicemente scegliendo differenti ipotesi alternative H_a ciascuna delle quali in grado di mettere in luce differenti aspetti significativi.

Un esempio classico in grado di illustrare la situazione è il meccanismo di giudizio in un sistema giuridico nel quale:

Un individuo è considerato non colpevole fino a che non è provata la sua colpevolezza oltre ogni ragionevole dubbio.

Implicitamente una tale affermazione ritiene molto più grave giudicare colpevole un non colpevole piuttosto che giudicare non colpevole un colpevole.

Nei termini prima espressi avremo

H_0 non colpevole

e

H_a colpevole oltre ogni ragionevole dubbio.

Rigettare H_0, nel caso in cui H_0 sia vera, significa giudicare colpevole un non colpevole ed è considerato più grave di di accettare H_a nel caso in cui H_a sia falsa, cioè nel caso in cui si giudichi non colpevole oltre ogni ragionevole dubbio un colpevole.

Diciamo che si commette un errore di I specie se **si rigetta** H_0 **nel caso in cui** H_0 **è vera**. (Si condanna un innocente).

Diciamo che si commette un errore di II specie se **si rigetta** H_a **nel caso in cui** H_a **è vera**. (Si assolve un colpevole, ma non oltre ogni ragionevole dubbio).

Definiamo inoltre

$P($ errore di I specie$) = \alpha$ livello di significatività del test.

$P($ errore di II specie$) = \beta$ potenza del test.

2.1.11

Riprendendiamo il semplice ma significativo esempio che abbiamo già in precedenza considerato:

progettare un test per stabilire se una moneta è truccata;

con lo scopo di sottolineare la corrispondenza tra definizioni teoriche e scelte pratiche.

Cominciamo con lo stabilire un chiaro quadro di riferimento.

Indichiamo con p la probabilità che esca Testa e con q la probabilità che esca Croce. e lanciamo la moneta $n = 100$ volte; chiamiamo T il numero di teste uscito e stabiliamo di adottare un livello di significatività $\alpha = 0.05$.

Se la moneta non è truccata T è una variabile aleatoria che ha una distribuzione binomiale (bernoulliana) di media μ e scarto quadratico σ dati da

$$\mu = np = 50 \qquad , \qquad \sigma = \sqrt{npq} = 5$$

e può essere approssimata con una variabile aleatoria normale standardizzata z definita dalla

$$z = \frac{n_t - \mu}{\sigma}$$

2.1.12

Consideriamo l'ipotesi da testare

H_0 la moneta non è truccata cioè $p = q = \frac{1}{2}$.

a fronte della'ipotesi alternativa

H_a la moneta è truccata cioè $p \neq \frac{1}{2}, q \neq \frac{1}{2}$.

In questo caso giudicheremo la moneta truccata se T è troppo grande $p > 0.5$ oppure se T è troppo piccolo $p < 0.5$ ed avremo un livello di significatività $\alpha = 0.05$ se

$$P(z < -z_a \; oppure \; z > z_a) = 0.05$$

Dovrà quindi essere

$$0.05 = P(z > z_a) + P(z < -z_a) = 2P(z > z_a) = 2(1 - P(z < z_a)) = 2(1 - F(z_a))$$

da cui

$$F(z_a) = 1 - 0.025 = 0.975 \qquad , \qquad z_a = F^{-1}(0.975) = 1.96$$

Rigetteremo cioè H_0 se $z < -z_a = -1.96$ oppure se $z > z_a = 1.96$ cioè se

$$\frac{T - \mu}{\sigma} < -1.96 \qquad \text{oppure} \qquad \frac{T - \mu}{\sigma} > 1.96$$

$$T < \mu - 1.96\sigma \approx 40 \qquad \text{oppure} \qquad T > \mu + 1.96\sigma \approx 60$$

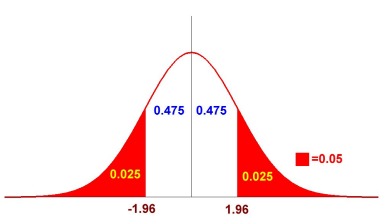

Figure 2.6: .

Riassumendo possiamo affermare che decidendo di accettare l'ipotesi H_0 che la moneta non sia truccata a fronte dell'ipotesi alternativa H_a che $p \neq 0.5$ nel caso che il numero di teste uscite sia compreso tra 40 e 60, la probabilità di commettere un errore di prima specie e cioè di rigettare l'ipotesi la moneta non è truccata, quando realmente non è truccata, è 0.05

2.1.13

Consideriamo l'ipotesi da testare

H_0 - la moneta non è truccata cioè $p = q = \frac{1}{2}$.

a fronte della'ipotesi alternativa

H_a - la moneta è truccata cioè $p > \frac{1}{2}$.

In questo caso giudicheremo la moneta truccata se T è troppo grande ed se stabiliamo un livello di significatività di $\alpha = 0.05$ possiamo calcolare

$$P(z > z_a) = 0.05 = 1 - F(z_a)$$

da cui

$$z_a = F^{-1}(0.95) = 2.571$$

Rigetteremo cioè H_0 se se $z > z_a = 2.571$ da cui

$$\frac{T - \mu}{\sigma} > 2.571$$

$$T > \mu + 2.571\sigma \approx 63$$

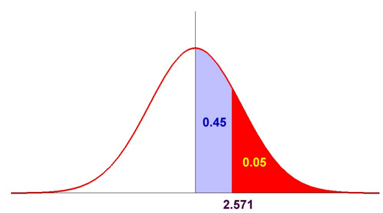

Figure 2.7: .

Riassumendo possiamo affermare che decidendo di accettare l'ipotesi H_0 che la moneta non sia truccata a fronte dell'ipotesi alternativa H_a che $p > 0.5$ nel caso che il numero di teste uscite sia maggiore di 63, la probabilità di commettere un errore di prima specie e cioè di rigettare l'ipotesi la moneta non è truccata, quando realmente non è truccata, è 0.05

2.1.14

Con riferimento ai due esempi precedenti osserviamo che se otteniamo un numero di teste $T = 62$, a parità di livello di significatività, nel primo caso rigettiamo l'ipotesi che la moneta sia truccata mentre nel secondo caso la accettiamo; ciò in conseguenza alla diversa formulazione dell'ipotesi H_a che deve essere scelta in modo da esprimere le esigenze del problema.

Se ad esempio il test è condotto allo scopo si stabilire se è equo giocare a Testa o Croce con quella moneta e si ha intenzione di puntare su Testa è chiaro che il secondo test meglio si adatta alla situazione.

2.2 *Il Test χ^2*

2.2.1 *Il Test χ^2 CASO 2 eventi*

Si consideri un esperimento in cui sono possibili 2 uscite

$$A_1 \quad , \quad A_2$$

di cui si ipotizzano le probabilità

$$p_1 \quad , \quad p_2$$

Si supponga di eseguire n volte l'esperimento e di ottenere delle frequenze di accadimento

$$x_1 \quad , \quad x_2$$

relative alla variabile aleatoria $\xi = (\xi_1, \xi_2$ he ha una distribuzione binomiale

Uno stimatore delle differenza tra frequenze osservate e frequenze ipotizzate si può definire come

$$\chi^2 = \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} \tag{2.1}$$

Chiaramente x_1 ed x_2 non sono indipendenti e si ha

$$\begin{aligned} \chi^2 = \frac{(x_1 - np_1)^2}{np_2} + \frac{(x_2 - np_2)^2}{np_2} &= \frac{(x_1 - np_1)^2}{np_1} p_2 + \frac{(x_2 - np_2)^2}{np_2} p_1 = \\ = \frac{(x_1 - np_1)^2 p_2 + ((n - x_1) - n(1 - p_1))^2 p_1}{np_1 p_2} &= \\ = \frac{(x_1 - np_1)^2 p_2 + (-x_1 + p_1))^2 p_1}{np_1 p_2} = \frac{(x_1 - np_1)^2 (p_2 + p_1)}{np_1 p_2} &= \\ = \left(\frac{x_1 - np_1}{\sqrt{np_1 p_2}} \right)^2 \end{aligned}$$

Dal momento che
$\frac{x_1 - np_1}{\sqrt{np_1 p_2}}$ tende ad una distribuzione normnale standard per il teorema del limite centrale, si ha che $\left(\frac{x_1 - np_1}{\sqrt{np_1 p_2}} \right)^2$ si può approssimare mediante una ditribuzione con densità χ^2 ad un grado di libertà.

2.2.2 *Il Test χ^2 in generale*

Si consideri un esperimento in cui sono possibili k uscite

$$A_j \qquad j = 1..k$$

di cui si ipotizzano le probabilità

$$p_j \qquad j = 1..k$$

Si supponga di eseguire n volte l'esperimento e di ottenere delle frequenze di accadimento

$$x_j \qquad j = 1..k$$

relative alla variabile aleatoria $\xi = (\xi_1, \xi_2, \ldots, \xi_k)$ che ha una distribuzione multinomiale

Uno stimatore delle differenza tra frequenze osservate e frequenze ipotizzate si può definire come

$$\chi^2 = \sum_{j=1}^{k} \frac{(x_j - np_j)^2}{np_j} \tag{2.2}$$

tenendo conto che $n_j = np_j$ è il valor medio di ξ_j

Consideriamo ora la variabile aleatoria η che ha k componenti indipendenti con distribuzione di Poisson di media $\lambda_j = np_j$ condizionate da

$$\eta_1 + \eta_2 + \cdots + \eta_k = n$$

Abbiamo visto in precedenza che η ha la stessa distribuzione di ξ e quindi possiamo utilizzare il luogo della 2.2 la

$$\chi^2 = \sum_{j=1}^{k} \frac{(y_j - \lambda_j)^2}{\lambda_j} = \sum_{j=1}^{k} \left(\frac{y_j - \lambda_j}{\sqrt{\lambda_j}} \right)^2 \tag{2.3}$$

Se ora facciamo $n \to +\infty$, per il teorema del limite centrale, possiamo affermare che

$$\frac{y_j - \lambda_j}{\sqrt{\lambda_j}} \to \zeta_j$$

dove ζ_j ha distribuzione normale standard, pertanto avremo che

$$\chi^2 = \sum_{j=1}^{k} \zeta_j^2 = \zeta_1^2 + \zeta_2^2 + \cdots + \zeta_k^2$$

è la somma di k variabili aleatorie normali standard condizionate da

$$\begin{aligned} 0 = \eta_1 + \eta_2 + \cdots + \eta_k - n = \\ \sqrt{n} \left(\sqrt{p_1} \frac{\eta_1 - np_1}{\sqrt{np_1}} + \sqrt{p_2} \frac{\eta_2 - np_2}{\sqrt{np_2}} + \cdots + \sqrt{p_k} \frac{\eta_k - np_k}{\sqrt{np_k}} \right) = \\ \sqrt{n} \left(\sqrt{p_1} \frac{\eta_1 - \lambda_1}{\sqrt{\lambda_1}} + \sqrt{p_2} \frac{\eta_2 - \lambda_2}{\sqrt{\lambda_2}} + \cdots + \sqrt{p_k} \frac{\eta_k - \lambda_k}{\sqrt{\lambda_k}} \right) \end{aligned}$$

da cui

$$0 = \left(\sqrt{p_1} \frac{\eta_1 - \lambda_1}{\sqrt{\lambda_1}} + \sqrt{p_2} \frac{\eta_2 - \lambda_2}{\sqrt{\lambda_2}} + \cdots + \sqrt{p_k} \frac{\eta_k - \lambda_k}{\sqrt{\lambda_k}} \right)$$

ed, al limite,

$$0 = \sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k$$

Sia φ la PDF della variabile aleatoria ottenuta; avremo che

$$\varphi(x) = \mathcal{P}(\zeta_1^2 + \zeta_2^2 + \cdots + \zeta_k^2 \leq x | \sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k = 0) =$$
$$\lim_{\epsilon \to 0} \mathcal{P}(\zeta_1^2 + \zeta_2^2 + \cdots + \zeta_k^2 \leq x || \sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k| < \epsilon) =$$

Dal momento che ζ_j è una variabile normale standard, la sua PDF è

$$g_j(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

per cui

$$\mathcal{P}(\zeta_1^2 + \zeta_2^2 + \cdots + \zeta_k^2 \leq x || \sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k| < \epsilon) =$$
$$\frac{\frac{1}{\sqrt{2\pi}^k} \int_V e^{-\frac{u_1^2 + u_2^2 + \cdots + u_k^2}{2}} du_1\, du_2\, \ldots, du_k}{\mathcal{P}(|\sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k| < \epsilon)}$$

essendo V la parte della sfera n-dimensionale compresa tra i piani

$$\sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k = \pm\epsilon$$

Poichè l'integranda dipende solo dalla distanza dall'origine e $|\sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k| < \epsilon$ rappresenta la parte di spazio delimitata da due piani paralleli ed equidistanti da un piano per l'origine, il calcolo non dipende da come è inclinato il piano e possiamo quindi sostituire la condizione data con la

$$|\sqrt{p_k}\zeta_k| < \epsilon$$

Ne segue che

$$\frac{1}{\sqrt{2\pi}^k} \int_V e^{-\frac{u_1^2 + u_2^2 + \cdots + u_k^2}{2}} du_1\, du_2\, \ldots, du_k \approx$$
$$\approx \frac{1}{\sqrt{2\pi}^k} 2\epsilon \int_{V_{k-1}} e^{-\frac{u_1^2 + u_2^2 + \cdots + u_{k-1}^2}{2}} du_1\, du_2\, \ldots, du_{k-1} = \frac{1}{\sqrt{2\pi}} 2\epsilon \chi_{k-1}^2(x)$$

essendo V_{k-1} la sfera $(n-1)$-dimensionale di raggio x e, conseguentemente, χ_{k-1}^2 la CDF di una variabile aleatoria di tipo χ^2 a $k-1$ gradi di libertà

D'altro canto $\sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k$ è una variabile aleatoria normale standard in quanto combinazione lineare di variabili aleatorie normali standard per cui la sua PDF è

$$g(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

e quindi

$$\mathcal{P}(|\sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k| < \epsilon) = \frac{1}{\sqrt{2\pi}} \int_{-\epsilon}^{+\epsilon} e^{-\frac{u^2}{2}} = approx \frac{1}{\sqrt{2\pi}} 2\epsilon$$

Ne segue che

$$\mathcal{P}(\zeta_1^2 + \zeta_2^2 + \cdots + \zeta_k^2 \leq x || \sqrt{p_1}\zeta_1 + \sqrt{p_2}\zeta_2 + \cdots + \sqrt{p_k}\zeta_k| < \epsilon) = \frac{\frac{1}{\sqrt{2\pi}} 2\epsilon \chi_{k-1}^2(x)}{\frac{1}{\sqrt{2\pi}} 2\epsilon} = \chi_{k-1}^2(x)$$

(supponendo che le frequenze teoriche possano essere stimate senza dover stimare statisticamente i parametri della popolazione).

Poichè i dati sono discreti e la variabile χ^2 è continua può essere opportuno apportare una correzione allo stimatore usando

$$\bar{\chi}^2 = \sum_{j=1}^{k} \frac{(|x_j - np_j| - .5)^2}{np_j}$$

$\bar{\chi}^2$ si chiama correzione di Yates.

Gli stimatori introdotti possono essere usati in test che tendano a stabilire se le frequenze teoriche p_j siano in accordo con i risultati ottenuti x_j.

2.2.3 *Test sulle Medie*

2.2.4

Siano $\xi_1, \xi_2, .., \xi_n$ variabili aleatorie costituenti un campione di grandezza n. (Con ciò si intende che ξ_k è una variabile aleatoria che si ottiene estraendo un elemento dalla popolazione).

Definiamo una nuova variabile aleatoria $\bar{\xi}$, che chiamiamo media campionaria, mediante la

$$\bar{\xi} = \frac{\xi_1 + \xi_2 + .. + \xi_n}{n}$$

La distribuzione di probabilità di $\bar{\xi}$ è detta distribuzione campionaria della media.

Si può verificare che, se la popolazione ha media μ e varianza σ^2 allora la media e la varianza di $\bar{\xi}$ sono date da

$$\mu_{\bar{\xi}} = \mu \qquad , \qquad \sigma^2_{\bar{\xi}} = \frac{\sigma^2}{n} \qquad , \qquad \left(\sigma_{\bar{\xi}} = \frac{\sigma^2}{\sqrt{n}}\right)$$

Si può inoltre dimostrare che la variabile aleatoria

$$\frac{\bar{\xi} - \mu}{\sigma/\sqrt{n}}$$

ha, per n grande, una distribuzione normale standard.

$\frac{\bar{\xi}-\mu}{\sigma/\sqrt{n}}$ si usa come stimatore della media.

2.2.5 Test sulle Varianze

2.2.6

Siano $\xi_1, \xi_2, .., \xi_n$ variabili aleatorie costituenti un campione di grandezza n. (Con ciò si intende che ξ_k è una variabile aleatoria che si ottiene estraendo un elemento dalla popolazione).

Definiamo una nuova variabile aleatoria s^2, che chiamiamo varianza campionaria, mediante la

$$s^2 = \frac{(\xi_1 - \bar{\xi})^2 + (\xi_2 - \bar{\xi})^2 + .. + (\xi_n - \bar{\xi})^2}{n}$$

La distribuzione di probabilità di s^2 è detta distribuzione campionaria della varianza.

Si può verificare che, se la popolazione ha media μ e varianza σ^2 allora la media di s^2 è data da

$$\mu_{s^2} = \frac{n-1}{n}\sigma^2$$

Inoltre se la popolazione è distribuita normalmente, la variabile aleatoria definita da

$$\frac{ns^2}{\sigma^2}$$

ha una distribuzione di tipo χ^2 con $n-1$ gradi di libertà.

$\frac{ns^2}{\sigma^2}$ si usa come stimatore della varianza.

2.3 Stima di parametri

2.3.1 Popolazioni

Diciamo che è assegnata una popolazione se è assegnato un insieme $\mathcal{U}$ ed una variabile aleatoria ξ definita su $\mathcal{U}$.

Se ad esempio siamo interessati a stimare il diametro di sfere di acciaio per cuscinetti, possiamo considerare la popolazione i cui elementi sono le sfere prodotte e la variabile aleatoria ξ che associa ad ogni sfera s_k il suo diametro $\xi(k) = d_k$. La variabile ξ che restituisce il diametro di ciascuna sfera descrive la popolazione che stiamo esaminando. ξ potrebbe ad esempio avere una funzione di distribuzione, ad esempio gaussiana di media μ e varianza σ^2.

2.3.2 *Campioni*

Data una popolazione definita dalla variabile aleatoria ξ sullo spazio $\mathcal{U}$ diciamo che è assegnato un campione di taglia n se sono assegnate n variabili aleatorie $x_1, x_2, ..., x_n$ sullo spazio $\mathcal{U}$ Indichiamo con x la n-pla delle variabili aleatorie $x_1, x_2, ..., x_n$ che costituiscono il campione.

Ad esempio possiamo considerare x_k come la variabile aleatoria che restituisce il diametro di una k-esima sferetta scelta in $\mathcal{U}$.

2.3.3 *Stimatore*

Diciamo che è assegnato uno stimatore, o riassunto campionario, se è assegnata una funzione $\phi = \phi(x_1, x_2, ..., x_n)$.

Possiamo ad esempio considerare uno stimatore del diametro delle sferette considerando la media dei diametri delle n sferette che abbiamo estratto per costruire il campione.

In tal caso

$$\phi(x_1, x_2, ..., x_n) = \frac{1}{n}(x_1 + x_2 + ... + x_n)$$

Uno stimatore è a sua volta una variabile aleatoria.

2.4 *Risultati sulle distribuzioni campionarie*

Si consideri una popolazione individuata da una variabile aleatoria ξ su un insieme $\mathcal{U}$ di media μ e varianza σ ed un campione $x = (x_1, x_2, ..., x_n)$ di taglia n.

Sia

$$\mu = E(\xi)$$

la media della popolazione e

$$\bar{x}_n = \frac{1}{n}(x_1 + x_2 + ... + x_n)$$

la media campionaria

2.4.1 *La media campionaria*

Sia $\mu_{\bar{x}_n}$ la media della variabile aleatoria $\bar{x}_n$, cioè il valor medio della media campionaria; si ha

- $$\mu_{\bar{x}_n} = E(\bar{x}_n) = \mu$$

  infatti

  $$E(\bar{x}_n) = \frac{1}{n}\left(E(x_1) + E(x_2) + \cdots E(x_n)\right) = \frac{n\mu}{n} = \mu$$

- $$\sigma^2_{\bar{x}_n} = E((\bar{x}_n - \mu_{\bar{x}_n})^2) = E((\bar{x}_n - \mu)^2) = \frac{\sigma^2}{n}$$

  infatti, tenendo conto che le variabili x_i sono indipendenti,

  $$\begin{aligned} &E((\bar{x}_n - \mu_{\bar{x}_n})^2) = E((\bar{x}_n - \mu)^2) = \\ &= \frac{E((x_1-\mu)^2) + E((x_2-\mu)^2) + \cdots E((x_n-\mu)^2)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

2.4.2 *Distribuzione della media campionaria*

$\bar{x}_n$ è distribuito normalmente con media μ e varianza σ^2/n. Infatti, poichè ξ è distribuita normalmente con media μ e varianza σ^2 allora anche ciascuna delle x_i segue la stessa distribuzione e, dal momento che la somma di distribuzioni gaussiane è ancora gaussiana, anche $\bar{x}_n$ ha una PDF Gaussiana la cui media e varianza sono, in accordo con il punto precedente, μ e $\frac{\sigma^2}{n}$, rispettivamente.

Inoltre, per il teorema del limite centrale, si ottiene che

$$\frac{\bar{x}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

è asintoticamente ($n \geq 30$) distribuito normalmente con media 1 e varianza 0.

2.4.3 *La varianza campionaria*

Chiamiamo varianza campionaria la variabile aleatoria definita da

$$S_n^2 = \frac{(x_1 - \bar{x}_n)^2 + (x_2 - \bar{x}_n)^2 + \ldots.. + (x_n - \bar{x}_n)^2}{n}$$

Si ha che

$$E(S_n^2) = \frac{n-1}{n}\sigma^2$$

infatti possiamo facilmente calcolare che

$$\begin{aligned} (x_i - \bar{x}_n)^2 &= (x_i - \mu + \mu - \bar{x}_n)^2 = \\ &= (x_i - \mu)^2 - 2(x_i - \mu)(\bar{x}_n - \mu) + (\bar{x}_n - \mu)^2 \end{aligned}$$

perciò

$$\sum_{i=1}^{n}(x_i-\bar{x}_n)^2 = \sum_{i=1}^{n}(x_i-\mu)^2 - 2(\bar{x}_n-\mu)\sum_{i=1}^{n}(x_i-\mu) + n(\bar{x}_n-\mu)^2 =$$
$$= \sum_{i=1}^{n}(x_i-\mu)^2 - 2n(\bar{x}_n-\mu)^2 + n(\bar{x}_n-\mu)^2 = \sum_{i=1}^{n}(x_i-\mu)^2 - n(\bar{x}_n-\mu)^2 \tag{2.4}$$

e ne segue

$$E(S_n^2) =$$
$$= \frac{1}{n}E\left(\sum_{i=1}^{n}(x_i-\bar{x}_n)^2\right) = \frac{1}{n}\left(E\left(\sum_{i=1}^{n}(x_i-\mu)^2\right) - nE(\bar{x}_n-\mu)^2)\right) =$$
$$= \frac{1}{n}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \frac{n-1}{n}\sigma^2$$

Definiamo inoltre

$$\hat{S}_n^2 = \frac{(x_1-\bar{x}_n)^2 + (x_2-\bar{x}_n)^2 + ... + (x_n-\bar{x}_n)^2}{n-1}$$

$\hat{S}_n^2$ è una definizione alternativa di varianza campionaria per la quale si ha

$$nS_n^2 = (n-1)\hat{S}_n^2 = \sum_{i=1}^{n}(x_i-\bar{x}_n)^2$$

2.4.4 *Indipendenza di media e varianza campionaria*

Possiamo dimostrare che, nel caso in cui le variabili aleatorie x_i siano Gaussiane di media μ e varianza σ^2, la media campionaria $\bar{x}_n$ e la varianza campionaria $\hat{S}_n^2$ sono tra loro indipendenti.

Infatti, supponiamo, senza perdere di generalità , che $\mu = 0$ e $\sigma = 1$, avremo che

$$\hat{S}_n^2 = \frac{1}{n-1}\left((x_1-\bar{x}_n)^2 + \sum_{i=2}^{n}(x_i-\bar{x}_n)^2\right) =$$

essendo $\sum_{i=1}^{n}(x_i-\bar{x}_n) = 0$

$$\frac{1}{n-1}\left(\left(\sum_{i=2}^{n}(x_i-\bar{x}_n)\right)^2 + \sum_{i=2}^{n}(x_i-\bar{x}_n)^2\right) =$$

Ora sia $\varphi(t) = \varphi(t_1, t_2, \ldots, t_n)$ la PDF congiunta delle variabili aleatorie indipendenti gaussiane $(x_1, x_2, \ldots, x_n)$

$$\varphi(t) = \frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\sum_{i=1}^{n}t_i^2}$$

e, posto $\bar{t} = \frac{1}{n}\sum t_i$, consideriamo la trasformazione di coordinate definita da:

$$\begin{cases} s_1 = \bar{t} \\ s_2 = t_2 - \bar{t} \\ \dots \\ s_n = t_n - \bar{t} \end{cases}$$

La trasformazione è lineare ed il suo Jacobiano è $\frac{1}{n}$ inoltre si ha $t_i = s_i + s_1$ per $i = 2, \dots, n$ e

$$t_1 = \sum_{i=1}^{n} t_i - \sum_{i=2}^{n} t_i = n\bar{t} - \sum_{i=2}^{n} t_i = \bar{t} - \sum_{i=2}^{n} t_i + (n-1)\bar{t} = \\ = \bar{t} - \sum_{i=2}^{n}(t_i - \bar{t}) = \bar{t} - \sum_{i=2}^{n} s_i = s_1 - \sum_{i=2}^{n} s_i$$

per cui possiamo esprimere φ in funzione di s mediante la

$$\varphi(s) = \frac{n}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\left((s_1 - \sum_{i=2}^{n} s_i)^2 + \sum_{i=2}^{n}(s_i + s_1)^2\right)}$$

e dal momento che

$$\left(s_1 - \sum_{i=2}^{n} s_i\right)^2 + \sum_{i=2}^{n}(s_i + s_1)^2 = \\ = s_1^2 + \left(\sum_{i=2}^{n} s_i\right)^2 - 2s_1\sum_{i=2}^{n} s_i + \sum_{i=2}^{n} s_i^2 + \sum_{i=2}^{n} s_1^2 + 2s_1\sum_{i=2}^{n} s_i = \\ = ns_1^2 + \left(\sum_{i=2}^{n} s_i\right)^2 + \sum_{i=2}^{n} s_i^2$$

si ha

$$\varphi(s) = \frac{n}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}ns_1^2} e^{-\frac{1}{2}\left(\left(\sum_{i=2}^{n} s_i\right)^2 + \sum_{i=2}^{n} s_i^2\right)}$$

Quindi, se $(y_1, y_2, \dots, y_n)$ è la variabile aleatoria ottenuta da $(x_1, x_2, \dots, x_n)$ mediante la trasformazione lineare indicata, si ha che y_1 è indipendente da $y_2, \dots, y_n$ e $\bar{x}$ è indipendente da $\hat{S}^2$

È un risultato notevole che la variabile aleatoria definita da $\hat{S}_n^2$ ha una distribuzione di tipo χ_{n-1}^2, cioè ha una distribuzione di tipo χ^2 ad $(n-1)$ gradi di libertà. Infatti, dal momento che

$$\bar{x}_{n+1} = \frac{x_{n+1} + n\bar{x}_n}{n+1} = \bar{x}_n + \frac{1}{n+1}(x_{n+1} - \bar{x}_n)$$

si ha

$$n\hat{S}^2_{n+1} = \sum_{i=1}^{n+1}(x_i - \bar{x}_{n+1})^2 = \sum_{i=1}^{n+1}\left((x_i - \bar{x}_n) - \frac{1}{n+1}(x_{n+1} - \bar{x}_n)\right)^2 =$$

$$= \sum_{i=1}^{n+1}\left((x_i - \bar{x}_n)^2 - 2\frac{(x_{n+1} - \bar{x}_n)(x_i - \bar{x}_n)}{n+1} + \frac{(x_{n+1} - \bar{x}_n)^2}{(n+1)^2}\right) =$$

$$= \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + (x_{n+1} - \bar{x}_n)^2 - 2\frac{1}{n+1}(x_{n+1} - \bar{x}_n)^2 + \frac{n+1}{(n+1)^2}(x_{n+1} - \bar{x}_n)^2 =$$

$$= (n-1)\hat{S}^2_n + \frac{n}{n+1}(x_{n+1} - \bar{x}_n)^2$$

Ora

$$\hat{S}^2_2 = \frac{1}{2}(x_1 - x_2)^2$$

ha una distribuzione χ^2_1 in quanto quadrato di una gaussiana;

inoltre la formula precedente consente di verificare che $\hat{S}^2_{n+1}$ ha una distribuzione χ^2_n se $\hat{S}^2_n$ ha una distribuzione χ^2_{n-1} ricordando che $\frac{n}{n+1}(x_{n+1} - \bar{x}_n)^2$ è gaussiana.

Pertanto per il principio di induzione $\hat{S}^2_n$ ha una distribuzione di tipo χ^2_{n-1} per ogni $n \in \mathbb{N}$

Possiamo quindi affermare che la variabile aleatoria

$$\xi_2 = \sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2$$

ha una distribuzione di tipo χ^2 ad $n-1$ gradi di libertà.

2.4.5 *T-Test*

Abbiamo quindi verificato che

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

è una variabile aleatoria gaussiana standard (media 0 e varianza 1).

$$\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2 = \frac{n}{\sigma^2}S^2 = \frac{n-1}{\sigma^2}\hat{S}^2$$

è una variabile aleatoria con distribuzione di tipo χ^2 ad $n-1$ gradi di libertà.

Quindi

$$\sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2}{n-1}} = \sqrt{\frac{nS^2}{\sigma(n-1)}} = \sqrt{\frac{\hat{S}^2}{\sigma}}$$

è la radice di una variabile aleatoria di tipo χ^2 ad $n-1$ gradi di libertà divisa per i suoi gradi di libertà e pertanto

$$\frac{\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\hat{S}^2}{\sigma^2}}} = \sqrt{n}\frac{\bar{x}-\mu}{\hat{S}} = \sqrt{n-1}\frac{\bar{x}-\mu}{S}$$

ha una distribuzione di Student ad $n-1$ gradi di libertà.

2.4.6 *F-Test*

2.4.7

Se ξ_1, ξ_2 sono due popolazioni normalmente distribuite con varianza σ_1^2, σ_2^2 e se X_1, X_2 sono due campioni di taglia n_1 ed n_2 rispettivamente estratti dalle due popolazioni, le variabili aleatorie

$$\frac{n_i}{\sigma_i^2}S_i^2 = \frac{n_i - 1}{\sigma_i^2}\hat{S}_i^2$$

hanno distribuzioni di tipo χ^2 ad $n_i - 1$ gradi di libertà. Poichè

$$\frac{\frac{n_1}{(n_1-1)\sigma_1^2}S_1^2}{\frac{n_2}{(n_2-1)\sigma_2^2}S_2^2} = \frac{\frac{1}{\sigma_1^2}\hat{S}_1^2}{\frac{1}{\sigma_2^2}\hat{S}_2^2}$$

possiamo affermare che

$$\frac{1}{\sigma_2^2}\sigma_1^2\frac{1}{\hat{S}_1^2}\hat{S}_2^2$$

ha una distribuzione di probabilità di Fisher con $n_1 - 1$, $n_2 - 1$ gradi di libertà.