DISTRIBUZIONE DELLE STATISTICHE CAMPIONARIE

__Definizione__: Per POPOLAZIONE si intende un insieme o collezione di oggetti, numeri, misure, che sono oggetto di studio. Per CAMPIONE si intende una parte della popolazione che viene selezionata per l'analisi.

__Definizione__: Sia X la variabile che descrive la probabilità di una popolazione.

Sia $(X_1, X_2, \dots, X_n)$ un insieme di variabili aleatorie indipendenti identicamente distribuite.

Un CAMPIONE è una successione di n v.a.i.i.d.

n si dice RANGO (o DIMENSIONE) del campione.

__Definizione__: Una funzione delle variabili del campione $F(X_1, \dots, X_n)$ si chiama STATISTICA.

__Definizione__: Si definisce MEDIA CAMPIONARIA la variabile aleatoria

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$$

Siano $E[X_i] = \mu$ e $Var(X_i) = \sigma^2$

- $E[\overline{X}_n] = E\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n} \cdot n E[X_i] = E[X_i] = \mu$
- $Var(\overline{X}_n) = \frac{1}{n^2} Var(X_1 + \dots + X_n) = \frac{1}{n^2} \cdot n Var(X_i) = \frac{Var(X_i)}{n} = \frac{\sigma^2}{n}$

<u>Definizione</u>: Si definisce VARIANZA CAMPIONARIA

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 = \frac{1}{n-1} \left(\sum_i X_i^2 - n \bar{X}_n^2 \right)$$

- $$E[S^2] = \frac{1}{n-1} \sum_{i=1}^{n} E[(X_i - \bar{X}_n)^2] =$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} E[X_i^2 + \bar{X}_n^2 - 2 X_i \bar{X}_n] =$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \{ E[X_i^2] + E[\bar{X}_n^2] - 2 E[X_i \bar{X}_n] \}$$

$$(n-1) E[S^2] = \sum_{i=1}^{n} E[(X_i - \bar{X}_n)^2] =$$

$$= \sum_{i=1}^{n} \{ E[X_i^2] + E[\bar{X}_n^2] - 2 E[X_i \bar{X}_n] \} =$$

$$= \sum_{i=1}^{n} \left\{ Var(X_i) + E[X_i]^2 + Var(\bar{X}_n) + E[\bar{X}_n]^2 - 2 E\left[X_i \frac{\sum_j X_j}{n} \right] \right\} =$$

$$= \sum_{i=1}^{n} \left\{ \sigma^2 + \mu^2 + \frac{\sigma^2}{n} + \mu^2 \right\} - \frac{2}{n} \sum_{i,j=1}^{n} E[X_i X_j] =$$

$$= n\sigma^2 + \sigma^2 + 2n\mu^2 - \frac{2}{n} \left\{ \sum_{i=1}^{n} E[X_i^2] + \sum_{i \neq j} E[X_i X_j] \right\} =$$

$$= (n+1)\sigma^2 + 2n\mu^2 - \frac{2}{n} \left\{ n [Var(X_i) + E[X_i]^2] + \sum_{i \neq j} E[X_i] E[X_j] \right\} =$$

$$= (n+1)\sigma^2 + 2n\mu^2 - \frac{2}{\cancel{n}} \{ \cancel{n} (\sigma^2 + \mu^2) + \cancel{n} (n-1) \mu^2 \} =$$

$$= (n+1-2)\sigma^2 + (\cancel{2n} - \cancel{2} - \cancel{2n} + \cancel{2}) \mu^2 =$$

$$= (n-1)\sigma^2$$

$$\longrightarrow E[S^2] = \sigma^2$$

TEOREMA DEL LIMITE CENTRALE

Supponiamo di avere un campione di rango n $(X_1, \dots, X_n)$ con media μ e varianza σ^2.

Sia
$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \equiv S_n$$

Allora
$$\lim_{n \to +\infty} S_n \sim N(0,1)$$

Qualunque sia la distribuzione di X_i.

Ciò implica che per $n \gg 1$ vale l'approssimazione
$$P(S_n < x) \approx \Phi(x)$$

DIM.:
$$S_n = \frac{\frac{X_1 + \dots + X_n}{n} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$$

VARIABILE CAMPIONARIA STANDARDIZZATA

$E[S_n] = 0 \qquad Var(S_n) = 1$

Sia $Y_K = \frac{X_K - \mu}{\sigma}$ una v.a.i.i.d., $K = 1, \dots, n$

$E[Y_K] = 0 \qquad Var(Y_K) = 1$

$$S_n = \frac{Y_1 + \dots + Y_n}{\sqrt{n}}$$

$$\phi_{S_n}(t) = \phi_{\frac{Y_1 + \dots + Y_n}{\sqrt{n}}}(t) = \phi_{\frac{Y_1}{\sqrt{n}} + \dots + \frac{Y_n}{\sqrt{n}}}(t) =$$

$$= \phi_{\frac{Y_1}{\sqrt{n}}}(t)\, \phi_{\frac{Y_2}{\sqrt{n}}}(t) \cdot \ldots \cdot \phi_{\frac{Y_n}{\sqrt{n}}}(t) = \left[\phi_{\frac{Y_1}{\sqrt{n}}}(t)\right]^n$$

Scanned by TapScanner

$$\phi_{\alpha X}(t) = \phi_X(\alpha t) \longrightarrow \phi_{\frac{Y_1}{\sqrt{n}}}(t) = \phi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)$$

per $n \longrightarrow \infty$: $\frac{t}{\sqrt{n}} \rightarrow 0$

per $n \gg 1$: $\phi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)$ sviluppo con MacLaurin

$$\phi_{Y_1}\left(\frac{t}{\sqrt{n}}\right) = \phi_{Y_1}(0) + \frac{t}{\sqrt{n}}\phi'_{Y_1}(0) + \frac{t^2}{2n}\phi''_{Y_1}(0) + o\left(\frac{t^2}{n}\right) =$$

$$= 1 + \frac{t}{\sqrt{n}} \cdot 0 + \frac{t^2}{2n} E[Y_1^2] + o\left(\frac{t^2}{n}\right) =$$

$$= 1 + \frac{t^2}{2n}\left[\underbrace{Var(Y_1)}_{1} + \overbrace{E[Y_1]^2}^{0}\right] + o\left(\frac{t^2}{n}\right) =$$

$$= 1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)$$

$$\longrightarrow \phi_{S_n}(t) = \left[\phi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n$$

$$\lim_{n \to \infty} \phi_{S_n}(t) = \lim_{n \to +\infty}\left[1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n =$$

$$= e^{\frac{t^2}{2}} = \Phi_Z(t)$$

$$X \sim N(\mu, \sigma^2) \qquad \phi(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Nel nostro caso $E[S_n] = 0$ e $Var(S_n) = 1$

$$\longrightarrow S_n \sim N(0,1) \quad \text{e} \quad \phi_{S_n} = e^{\frac{t^2}{2}}$$

Questo dimostra il teorema.

DISTRIBUZIONE χ^2

Supponiamo $X = Z_1^2 + \dots + Z_n^2$, $Z_i \sim N(0,1)$.

Allora $X \sim \chi^2_n$

(CHI - QUADRO ad n GRADI DI LIBERTA')

$$P(X \geq \chi^2_{\alpha,n}) = \alpha$$

DISTRIBUZIONE DI STUDENT

Sia $Z \sim N(0,1)$, $C_n \sim \chi_n^2$.

$$T_n = \frac{Z}{\sqrt{C_n / n}}$$

LEGGE DI STUDENT $\left(\frac{\bar{X}_n - \mu}{S / \sqrt{n}} \right)$

($t_n(\alpha)$ o $t_\alpha(n)$ $\longrightarrow$ quantili)

CORREZIONE DI CONTINUITA'

Distribuzione qualsiasi $\xrightarrow[n \gg 1]{} N(0,1)$

(se la distribuzione è discreta $\longrightarrow$ continua)

DISTRIBUZIONE DELLE VARIABILI CAMPIONARIE

DATI RAGGRUPPATI

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_i m_i f_i$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 = \frac{1}{n-1} \sum_i (\underbrace{m_i}_{\text{VALORE CENTRALE DEI GRUPPO}} - \bar{X}_n)^2 \underbrace{f_i}_{\text{FREQUENZA}}$$

$$E[\bar{X}_n] = \mu \qquad E[s^2] = \sigma^2$$

$$Var(\bar{X}_n) = \frac{\sigma^2}{n}$$

Se $(X_1, \dots, X_n)$ sono normali,

allora $\bar{X}_n$ è normale con $\mu, \frac{\sigma^2}{n}$

Se $(X_1, \dots, X_n)$ non sono normali,

allora $\bar{X}_n \xrightarrow[n \to \infty]{} N(\mu, \frac{\sigma^2}{n})$

Ovviamente $\dfrac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$ $\begin{cases} \sim N(0,1) \\ \xrightarrow[n \to \infty]{} N(0,1) \end{cases}$

Considerazioni:

1) $\bar{X}_n$ e S^2 sono indipendenti

2) $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$

<u>DIM.</u>: 2) $(n-1)\frac{S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X}_n)^2}{\sigma^2} =$

$$= \sum_{i=1}^{n} \frac{(X_i - \mu + \mu - \bar{X}_n)^2}{\sigma^2} =$$

$$= \sum_{i=1}^{n} \frac{(X_i - \mu)^2 + (\bar{X}_n - \mu)^2 - 2(X_i - \mu)(\bar{X}_n - \mu)}{\sigma^2} =$$

$$= \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 + n\left(\frac{\bar{X}_n - \mu}{\sigma}\right)^2 - 2\,\frac{\bar{X}_n - \mu}{\sigma} \overbrace{\sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}}^{n\,\frac{\bar{X}_n - \mu}{\sigma}} =$$

$$= \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 + \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2 - 2\,\frac{\bar{X}_n - \mu}{\sigma} \cdot n\,\frac{\bar{X}_n - \mu}{\sigma} =$$

$$= \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 + \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2 - 2\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2 =$$

$$= \sum_{i=1}^{n} \underbrace{\left(\frac{X_i - \mu}{\sigma}\right)^2}_{Z_i^2} - \underbrace{\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2}_{Z^2} =$$

$$= \sum_{i=1}^{n} Z_i^2 - Z^2 \sim \chi^2_{n-1}$$

$$(n-1)\frac{S^2}{\sigma^2} = \sum_{i=1}^{n} \underbrace{\left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2}_{\downarrow} \sim \chi^2_{n-1}$$

INVECE $\left(\frac{X_i - \mu}{\sigma}\right) \sim Z$

Se $(X_1, \dots, X_n)$ sono v. a. di un campione normale con μ, σ^2 $(X_i \sim N(\mu, \sigma^2))$,

allora $T_n = \dfrac{\bar{X}_n - \mu}{S / \sqrt{n}} \sim t_{n-1}$

STUDENT CON n-1 GRADI DI LIBERTA'

<u>DIM.</u>: $T_n = \dfrac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \cdot \dfrac{\sigma}{S} \dfrac{(n-1)}{(n-1)} =$

$$= \frac{(\bar{X}_n - \mu) / (\sigma / \sqrt{n})}{\sqrt{\dfrac{S^2 (n-1)}{\sigma^2 (n-1)}}} \sim \frac{Z}{\sqrt{\chi^2_{n-1} / (n-1)}} \sim t_{n-1}$$