CI SIAMO OCCUPATI FINORA DI STIME PUNTUALI PER $\psi(\theta)$.
PIÙ SPECIFICATAMENTE DEI CASI IN CUI ψ È IL VALORE MEDIO μ O LA VARIANZA σ^2, CHE SONO LE QUANTITÀ PIÙ UTILI QUANDO DOBBIAMO FORNIRE I PARAMETRI DI UNA CERTA DISTRIBUZIONE (ESPONENZIALE, GAUSSIANA, ECC.).

UNO STIMATORE H_n (COME AD ESEMPIO $\bar{X}_n$ PER μ, O S_n^2 PER σ^2) FORNISCE, A CAMPIONAMENTO ESEGUITO, UNA STIMA DEL VALORE DI $\psi(\theta)$ DI CUI PERÒ NON CONOSCIAMO L'ACCURATEZZA. SE LO STIMATORE È CONSISTENTE E CORRETTO, $E[H_n]$ FORNIRÀ UNA STIMA SEMPRE PIÙ ACCURATA AL CRESCERE DELLA TAGLIA n DEL CAMPIONE. TUTTAVIA, NON SAPPIAMO CON CHE VELOCITÀ LA STIMA STA MIGLIORANDO, NÉ È SEMPRE POSSIBILE AUMENTARE n.

RISULTA QUINDI NECESSARIO ESTRARRE DAL CAMPIONE ANCHE UNA STIMA DELL'ACCURATEZZA DELLA STIMA PUNTUALE.
TALE TIPO DI STIMA SI BASA SULLA COSTRUZIONE DI UN INTERVALLO, DETTO INTERVALLO DI CONFIDENZA, CHE PRESUMIBILMENTE CONTIENE IL VALORE REALE DI $\psi(\theta)$.
IN CONTRAPPOSIZIONE ALLA STIMA PUNTUALE, QUESTA VIENE CHIAMATA STIMA PER INTERVALLI DI $\psi(\theta)$.

- STIMA DELLA MEDIA DI UNA POPOLAZIONE NORMALE CON VARIANZA <u>NOTA</u>.

<u>ESEMPIO</u> NELLA PROGETTAZIONE DELL'ABITACOLO DI UN'AUTO OCCORRE TENER CONTO DEI VALORI ANTROPOMETRICI MEDI DEL GUIDATORE (CIOÈ STATURA, PESO, ECC). SUPPONIAMO CHE LA STATURA DEI GUIDATORI SEGUA UNA LEGGE NORMALE $\mathcal{N}(\mu,\sigma^2)$. VOGLIAMO STIMARE LA MEDIA μ A PARTIRE, AD ESEMPIO, DA UN CAMPIONE DI 100 PATENTATI. POSSIAMO SUPPORRE CHE LA VARIANZA σ^2 SIA UGUALE A QUELLA DELLA POPOLAZIONE ADULTA COMPLESSIVA CHE (DA RILEVAZIONI PRECEDENTI) RISULTA ESSERE $\sigma^2 = 37{,}21$.

USIAMO LA STATISTICA MEDIA CAMPIONARIA $\bar{X}_n$ PER STIMARE μ. INDICATA CON X_k, $k=1,\dots,100$, L'ALTEZZA DEI CAMPIONI, SAPPIAMO CHE $X_k \sim \mathcal{N}(\mu,\sigma^2)$ CON $\sigma^2 = 37{,}21$. DA QUESTO SEGUE CHE

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) = \mathcal{N}\left(\mu, \frac{37{,}21}{100}\right)$$

DA CUI,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - \mu}{0{,}61} \sim \mathcal{N}(0,1). \qquad (*)$$

IN QUESTO AMBITO IL PARAMETRO θ È $\mu \in \mathbb{R}$. DATO $\alpha \in \mathbb{R}$, DA $(*)$ SEGUE CHE $P_\mu\left\{\frac{\bar{X}_n - \mu}{0{,}61} < \alpha\right\} = \phi(\alpha)$ O ANCHE

$$P_\mu\left\{|\bar{X}_n-\mu| < 0{,}61\cdot\alpha\right\} = P_\mu\left\{-\alpha < \frac{\bar{X}_n-\mu}{0{,}61} < \alpha\right\}$$

$$= \phi(\alpha)-\phi(-\alpha) = 2\,\phi(\alpha)-1$$

RICORDANDO CHE $\phi(x)+\phi(-x)=1$

QUI COME AL SOLITO ϕ INDICA LA FUNZIONE DI RIPARTIZIONE DI $N(0,1)$. DATO $\beta\in(0,1)$, SIA $\alpha(\beta)$ TALE CHE $2\phi(\alpha)-1=\beta$. TALE α ESISTE ED È UNICO PER L'INVERTIBILITÀ DI ϕ.

POSSIAMO SCRIVERE

$$P_\mu\left\{|\bar{X}_n-\mu| < 0{,}61\cdot\alpha(\beta)\right\} = \beta$$

FISSIAMO ORA, PER ESEMPIO, $\beta=0{,}95$. UNA COMPUTAZIONE NUMERICA FORNISCE $\alpha(\beta)\simeq 1{,}96$. QUINDI ABBIAMO LA RELAZIONE

$$P_\mu\left\{|\bar{X}_n-\mu| < 1{,}1956\right\} \simeq 0{,}95.$$

QUESTO SIGNIFICA CHE <u>PRIMA DI ESEGUIRE IL CAMPIONAMENTO</u> (CIOÈ FORNIRE I VALORI $(x_1,\dots,x_{100})$ PER $(X_1,\dots,X_{100})$) VALUTIAMO PARI A $0{,}95$ LA PROBABILITÀ P_μ CHE RISULTI:

$$|\bar{X}_n-\mu| < 1{,}1956 \quad \text{CIOÈ} \quad \mu\in\left(\bar{X}_n-1{,}1956,\ \bar{X}_n+1{,}1956\right)$$

QUESTO È UN INTERVALLO ALEATORIO (I SUOI ESTREMI SONO DELLE V.A.). È DETTO <u>INTERVALLO DI CONFIDENZA</u> PER μ AL LIVELLO DEL 95%

ESEGUIAMO ORA IL CAMPIONAMENTO E SUPPONIAMO DI TROVARE TRAMITE 100 MISURAZIONI CHE

$$\bar{x}_n = \frac{1}{n} \sum_{k=1}^{n} x_k = 178{,}5 \text{ cm}$$

L'INTERVALLO DI CONFIDENZA DIVENTA UN INTERVALLO NUMERICO

$$(178{,}5 - 1{,}1956 \,,\; 178{,}5 + 1{,}1956) = (177{,}3 \,,\; 179{,}7)$$

QUESTO INTERVALLO NUMERICO SI CHIAMA INTERVALLO DI CONFIDENZA PER μ AL LIVELLO DEL 95% CALCOLATO DAL CAMPIONE.

GENERALIZZIAMO E FORMALIZZIAMO IL TUTTO

DEFINIZIONE SIA $\{X_n\}_{n \in \mathbb{N}}$ UN CAMPIONE CON DISTRIBUZIONE AVENTE DENSITÀ $f(x, \theta)$. FISSATA L'AMPIEZZA n DEL CAMPIONE, SIANO $H_n = h_n(X_1, \dots, X_n)$ E $G_n = g_n(X_1, \dots, X_n)$ DUE STATISTICHE E SIA $\psi = \psi(\theta)$ UNA FUNZIONE DEL PARAMETRO CHE VOGLIAMO STIMARE. FISSATO UN NUMERO $\beta \in (0,1)$, L'INTERVALLO ALEATORIO (H_n, G_n) È DETTO INTERVALLO DI CONFIDENZA AL LIVELLO $100\beta\%$ PER $\psi(\theta)$ SE

$$P_\theta\{H_n < \psi(\theta) < G_n\} = \beta$$

A CAMPIONAMENTO REALIZZATO, L'INTERVALLO NUMERICO

$$\left(h_m(x_1,\dots,x_m),\ g_m(x_1,\dots,x_m)\right)$$

È DETTO INTERVALLO DI CONFIDENZA AL LIVELLO $100\beta\%$ PER $\psi(\theta)$ CALCOLATO DAL CAMPIONE. GLI ESTREMI DELL'INTERVALLO SONO CHIAMATI LIMITI DI CONFIDENZA.

NELL'ESEMPIO ABBIAMO VISTO CHE SE $\theta=\mu$ E $f(x,\theta)=\mathcal{N}(\mu,\sigma^2)$ (σ^2 È NOTA), ALLORA

$$\left(\bar{X}_m-\alpha(\beta)\frac{\sigma}{\sqrt{m}},\ \bar{X}_m+\alpha(\beta)\frac{\sigma}{\sqrt{m}}\right)$$

È INTERVALLO DI CONFIDENZA AL LIVELLO $100\beta\%$ PER μ.

OSSERVAZIONE MANTENENDO L'AMPIEZZA DEL CAMPIONE, SE AUMENTIAMO LA CONFIDENZA (CIOÈ β), ALLORA CRESCERÀ ANCHE α E QUINDI L'AMPIEZZA DELL'INTERVALLO (E VICEVERSA). PER AUMENTARE LA CONFIDENZA, SENZA AUMENTARE L'AMPIEZZA DELL'INTERVALLO, DOBBIAMO AUMENTARE NECESSARIAMENTE L'AMPIEZZA m DEL CAMPIONE. SIMILMENTE, PER DIMINUIRE L'AMPIEZZA DELL'INTERVALLO, SENZA DIMINUIRE LA CONFIDENZA, DOBBIAMO AUMENTARE L'AMPIEZZA DEL CAMPIONE.

• STIMA DELLA MEDIA DI UNA POPOLAZIONE NORMALE CON VARIANZA <u>INCOGNITA</u>.

NELL'ESEMPIO PRECEDENTE ABBIAMO RICHIAMATO CHE

$$\bar{X}_n \sim N(\mu, \sigma^2/n) \quad \text{DA CUI} \quad \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

ESSENDO NOTA σ^2, GRAZIE ALLA DENSITÀ A DESTRA ABBIAMO FATTO UNA STIMA DEGLI EVENTI $\left\{ |\bar{X}_n - \mu| < \alpha \frac{\sigma}{\sqrt{n}} \right\}$, $\alpha > 0$.
LA COSA QUI <u>NON</u> FUNZIONA PERCHÉ ANCHE σ^2 È INCOGNITO. AVENDO UN CAMPIONE POSSIAMO PROCURARCI INFORMAZIONI NON SOLO SU $\bar{X}_n$ MA ANCHE SU S_n^2, CHE USEREMO PER SOPPERIRE ALLA MANCANZA DI INFORMAZIONI SU σ^2.
SE LE MISURE X_k DEL CAMPIONE SEGUONO UNA LEGGE NORMALE $N(\mu, \sigma^2)$, ALLORA

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \sim t(n-1)$$

GRAZIE ALLA DENSITÀ A DESTRA EFFETTUIAMO UNA STIMA DEGLI EVENTI

$$\left\{ |\bar{X}_n - \mu| < \alpha \frac{\sqrt{S_n^2}}{\sqrt{n}} \right\}$$

PIÙ PRECISAMENTE, SIA F LA FUNZIONE DI RIPARTIZIONE

DELLA $t(n-1)$. ALLORA, POSTO $\Theta = (\mu, \sigma^2)$, ABBIAMO

$$P_\Theta \left\{ \bar{X}_n - \alpha \frac{\sqrt{S_n^2}}{\sqrt{n}} < \mu < \bar{X}_n + \alpha \frac{\sqrt{S_n}}{\sqrt{n}} \right\} =$$

$$= P_\Theta \left\{ -\alpha < \sqrt{n} \, \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} < \alpha \right\} = F(\alpha) - F(-\alpha) = 2F(\alpha) - 1$$

DOVE L'ULTIMA UGUAGLIANZA È DOVUTA AL FATTO CHE LA DENSITÀ $t(n-1)$ È PARI.

DATO $\beta \in (0,1)$, SIA $\alpha(\beta)$ TALE CHE $2F(\alpha) - 1 = \beta$.
POSSIAMO RISCRIVERE L'EQUAZIONE SOPRA COME

$$P_\Theta \left\{ \bar{X}_n - \alpha(\beta) \frac{\sqrt{S_n^2}}{\sqrt{n}} < \mu < \bar{X}_n + \alpha(\beta) \frac{\sqrt{S_n^2}}{\sqrt{n}} \right\} = \beta$$

CIOÈ

$$\left(\bar{X}_n - \alpha(\beta) \frac{\sqrt{S_n^2}}{\sqrt{n}} , \bar{X}_n + \alpha(\beta) \frac{\sqrt{S_n^2}}{\sqrt{n}} \right)$$

È UN INTERVALLO DI CONFIDENZA PER $\mu = \varphi(\Theta)$ (CON φ PROIEZIONE SULLA PRIMA COORDINATA) AL LIVELLO DEL $100\beta\%$.

Esempio La ditta dei kiwi precedente procede ad una misurazione a campione su 50 kiwi e trova che il diametro medio del campione è di 2,04 cm, con una varianza campionaria di 0,0225 cm^2.

Supponendo che il diametro dei frutti segua una legge normale $N(\mu, \sigma^2)$, individuare un intervallo di confidenza al livello del 95% per la media calcolato dal campione.

I dati in nostro possesso sono, per $n=50$,

la media campionaria $\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} x_k = 2{,}04$

la varianza campionaria $\bar{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \bar{X}_n)^2 = 0{,}0225$

La computazione di $\alpha(\beta)$ per $\beta = 0{,}95$ fornisce $\alpha = 1{,}6765$.

Quindi

$$\bar{X}_n - \alpha(\beta) \frac{\sqrt{\bar{\sigma}_n^2}}{\sqrt{n}} = 2{,}004 \qquad \bar{X}_n + \alpha(\beta) \frac{\sqrt{\bar{\sigma}_n^2}}{\sqrt{n}} = 2{,}076$$

e l'intervallo richiesto è $(2{,}004 , 2{,}076)$

Nota operativa: l'approssimazione dell'estremo inferiore si effettua dal basso (qui il valore numerico era 2,0044...), mentre l'approssimazione dell'estremo superiore si effettua dall'alto (qui il valore numerico era 2,0756...).

- STIMA DELLA VARIANZA DI UNA POPOLAZIONE NORMALE CON MEDIA <u>NOTA</u>.

SAPPIAMO DUE COSE UTILI:

i) $H_n = \frac{1}{n} \sum_{k=1}^{n} (X_k - \mu)^2$ È UNO STIMATORE CORRETTO PER LA VARIANZA σ^2

ii) $\frac{n H_n}{\sigma^2} = \sum_{k=1}^{n} \left(\frac{X_k - \mu}{\sigma} \right)^2 \sim \chi^2(n)$

IN QUESTO CASO IL PARAMETRO θ È σ^2. IL PUNTO (ii) CI DICE CHE POSSO STIMARE LA PROBABILITÀ DI EVENTI DEL TIPO

$$\left\{ \alpha_1 < \frac{n H_n}{\sigma^2} < \alpha_2 \right\}$$

MENTRE IL PUNTO (i) CI DICE CHE LA STIMA SARÀ PRESUMIBILMENTE BUONA. DETTA F LA FUNZIONE DI RIPARTIZIONE ASSOCIATA A $\chi^2(n)$, ABBIAMO

$$P_{\sigma^2}\left\{ \frac{n H_n}{\alpha_2} < \sigma^2 < \frac{n H_n}{\alpha_1} \right\} = P_{\sigma^2}\left\{ \alpha_1 < \frac{n H_n}{\sigma^2} < \alpha_2 \right\}$$

$$= F(\alpha_2) - F(\alpha_1) \qquad \text{PER } \alpha_1, \alpha_2 > 0$$

DATO $\beta \in (0,1)$, BASTA PRENDERE α_1, α_2 TALI CHE $F(\alpha_2) = \frac{1+\beta}{2}$ E $F(\alpha_1) = \frac{1-\beta}{2}$ PER AVERE CHE $\left(\frac{n H_n}{\alpha_2}, \frac{n H_n}{\alpha_1} \right)$ È UN INTERVALLO DI CONFIDENZA AL LIVELLO DEL $100\beta\%$ PER σ^2.

- STIMA DELLA VARIANZA DI UNA POPOLAZIONE NORMALE CON MEDIA INCOGNITA

QUESTA VOLTA USIAMO IL FATTO CHE

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

IN QUESTO CASO IL PARAMETRO Θ È (μ,σ^2) E $\psi(\Theta)=\sigma^2$ (CIOÈ ψ È LA PROIEZIONE SULLA SECONDA COORDINATA).
DETTA F LA FUNZIONE DI RIPARTIZIONE ASSOCIATA A $\chi^2(n-1)$, ABBIAMO

$$P_\Theta\left\{\frac{(n-1)S_n^2}{\alpha_2} < \sigma^2 < \frac{(n-1)S_n^2}{\alpha_1}\right\} = P_\Theta\left\{\alpha_1 < \frac{(n-1)S_n^2}{\sigma^2} < \alpha_2\right\}$$

$$= F(\alpha_2) - F(\alpha_1) \qquad \text{PER } \alpha_1, \alpha_2 > 0$$

DATO $\beta \in (0,1)$, BASTA PRENDERE α_1, α_2 TALI CHE $F(\alpha_2) = \frac{1+\beta}{2}$ E $F(\alpha_1) = \frac{1-\beta}{2}$ PER AVERE CHE $\left(\frac{(n-1)S_n^2}{\alpha_2}, \frac{(n-1)S_n^2}{\alpha_1}\right)$ È UN INTERVALLO DI CONFIDENZA AL LIVELLO DEL $100\beta\%$ PER σ^2.

ESEMPIO TORNIAMO ANCORA UNA VOLTA DALLA NOSTRA DITTA DI KIWI. SU UN CAMPIONE DI 40 KIWI È STATA EFFETTUATA UNA VARIANZA CAMPIONARIA CHE HA RESTITUITO COME RISULTATO $\bar{\sigma}_n^2 = 1{,}20\ \text{cm}^2$ PER QUANTO RIGUARDA LE DIMENSIONI. SAPENDO CHE QUESTE SEGUONO UNA LEGGE NORMALE $\mathcal{N}(\mu, \sigma^2)$, INDIVIDUARE UN INTERVALLO DI CONFIDENZA DEL 98% PER LA VARIANZA CALCOLATO SUL CAMPIONE.

SIA F LA FUNZIONE DI RIPARTIZIONE DI $\chi^2(n-1)$, $n = 40$.

CALCOLIAMOCI α_1 E α_2 TALI CHE

$$F(\alpha_2) = \frac{1+\beta}{2} \qquad F(\alpha_1) = \frac{1-\beta}{2} \qquad \text{PER } \beta = 0{,}98$$

ABBIAMO $F(\alpha_2) = 0{,}99$ QUINDI $\alpha_2 = 62{,}43$

$F(\alpha_1) = 0{,}01$ QUINDI $\alpha_1 = 21{,}42$

DA CUI $\dfrac{(n-1)\bar{\sigma}_n^2}{\alpha_2} = 0{,}749 \qquad \dfrac{(n-1)\bar{\sigma}_n^2}{\alpha_1} = 2{,}185$

L'INTERVALLO DI CONFIDENZA A LIVELLO DEL 98% PER σ^2 CALCOLATO DAL CAMPIONE È $(0{,}749\,,\ 2{,}185)$.