Esercizi di Statistica per Biotecnologie

Francesca Pizzorni Ferrarese

Esercitazione III - Inferenza I

Es. 1
Dato un campione di tre elementi estratti con campionamento bernoulliano da una popolazione di media μ e varianza σ^2, scegliere quale tra i seguenti due stimatori è preferibile:

$$T_1 = (X_1 + 2X_2 + 2X_3)/5$$

$$T_2 = (X_1 + 2X_2 + X_3)/4$$

Es. 2
Dato un campione di 5 misurazioni del diametro di una sferetta in cm

6.33 6.37 6.36 6.32 6.37

determinare una stima corretta ed efficiente del valore atteso e della varianza della popolazione.

Es. 3
Si vuole stimare il peso del contenuto di una bustina di un medicinale granulare in mg prodotta da una linea di produzione completamente automatica di una azienda farmaceutica. A tale scopo, si sono estratte 10 bustine e, una volta aperte, se ne è misurato il peso.

170, 180, 172, 171, 183, 181, 175, 178, 185, 184

Determinare:
a) una stima puntuale del valore atteso e della varianza.
b) una stima per intervallo con un livello di confidenza del 95% del valore atteso
c) una stima per intervallo con un livello di confidenza del 90% del valore atteso

Es. 4
Si vuole stimare il valore atteso relativo ai battiti cardiaci al minuto per una certa popolazione. La media campionaria di battiti al minuto per un campione di 49 soggetti è risultata uguale a 90. La popolazione è distribuita in modo normale con uno scarto quadratico medio pari a 10.
Determinare:
b) una stima per intervallo con un livello di confidenza del 99% del valore atteso
c) una stima per intervallo con un livello di confidenza del 95% del valore atteso

Es. 5
Data una v.c $X \sim N(\mu;\sigma^2)$ relativa alla sopravvivenza di un organismo unicellulare a contatto con una determinata sostanza tossica, misurata in secondi, si sono ottenute le seguenti realizzazioni,

140 136 150 144 148 152 138 141 143 151

Determinare una stima per intervallo al 95% per la varianza σ^2.

ESERCITAZIONE 3 10/01/2011

(1) Estrazione bernoulliana: le unità statistiche vengono estratte una alla volta e dopo l'estrazione sono nuovamente estraibili. Il campionamento bernoulliano garantisce che in ogni estrazione un'osservazione ha la stessa probabilità di verificarsi.

Il campione (X_1, X_2, X_3) essendo estratto con riposizione, ha componenti i.i.d. $\Rightarrow X_i \sim P$

$$\Rightarrow E[X_i] = \underset{E[P]}{\mu} \qquad Var[X_i] = \underset{Var[P]}{\sigma^2}, \quad i = 1, 2, 3.$$

Analizziamo la correttezza degli stimatori T_1 e T_2 utilizzando le proprietà del valore atteso. (Correttezza: il valore atteso dello stimatore è il parametro da stimare $E[\Theta] = \theta$)

$$E[T_1] = E\left[\frac{X_1 + 2X_2 + 2X_3}{5}\right] = \frac{1}{5} \cdot E[X_1 + 2X_2 + 2X_3] =$$

$$= \frac{1}{5}\left[E[X_1] + 2E[X_2] + 2E[X_3]\right] = \frac{1}{5}\left[\mu + 2\mu + 2\mu\right] = \mu$$

$$E[T_2] = E\left[\frac{X_1 + 2X_2 + X_3}{4}\right] = \frac{1}{4} E\left[X_1 + 2X_2 + X_3\right] =$$

$$= \frac{1}{4}\left[E[X_1] + 2E[X_2] + E[X_3]\right] = \frac{1}{4}\left[\mu + 2\mu + \mu\right] = \mu$$

Entrambi gli stimatori T_1 e T_2 risultano corretti.
Per determinare il più efficiente è necessario calcolare la loro varianza (Efficienza: lo stimatore possiede la varianza minima).

$$Var[T_1] = Var\left[\frac{X_1 + 2X_2 + 2X_3}{5}\right] = \left(\frac{1}{5}\right)^2 Var\left[X_1 + 2X_2 + 2X_3\right] =$$

$$= \frac{1}{25}\left[Var[X_1] + 2^2 Var[X_2] + 2^2 Var[X_3]\right] =$$

$$= \frac{1}{25}\left[\sigma^2 + 4\sigma^2 + 4\sigma^2\right] = \frac{9}{25}\sigma^2 = 0,36\,\sigma^2$$

$$Var[T_2] = Var\left[\frac{X_1 + 2X_2 + X_3}{4}\right] = \left(\frac{1}{4}\right)^2 Var\left[X_1 + 2X_2 + X_3\right] =$$

$$= \left(\frac{1}{4}\right)^2\left[Var[X_1] + 2^2 Var[X_2] + Var[X_3]\right] =$$

$$= \frac{1}{16}\left[\sigma^2 + 4\sigma^2 + \sigma^2\right] = \frac{6}{16}\sigma^2 = 0,375\,\sigma^2$$

Sebbene non sia nota la già σ^2 confrontando i valori della varianza degli stimatori T_1 e T_2, dal momento che

$$Var[T_1] = 0,36\,\sigma^2 < 0,375\,\sigma^2 = Var[T_2]$$

possiamo concludere che lo stimatore più efficiente sia T_1.

(2) La stima corretta ed efficiente per il valore atteso della popolazione è la media campionaria

Ipotesi:
P: diametro sfera
campionamento bernoulliano

$$E[P] : \bar{x} = \frac{6,33+6,37+6,36+6,32+6,37}{5} = 6,35\text{ cm.}$$

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$Var[P]: s^2 = \left(\frac{\sum_i^n x_i^2}{n} - \bar{x}^2\right)\frac{n}{n-1} \overset{i.i.d.}{=} \frac{\sum_i^n P^2 - E[P]^2}{n-1} =$$

$$= \left(\frac{6,33^2+6,37^2+6,36^2+6,32^2+6,32^2}{5} - 6,35^2\right)\cdot\frac{5}{4}$$

$$= (40,32294 - 40,3225)\cdot\frac{5}{4} = 5,5\cdot 10^{-4}\text{ cm}^2$$

Anche per la varianza la stima corretta ed efficiente è la varianza campionaria.

(3) a) stima puntuale: si stima un solo valore per il parametro ignoto.

La stima corretta ed efficiente per il valore atteso della popolazione è la media campionaria

Ipotesi:
P: peso contenuto bustina in mg.
campionamento bernoull.

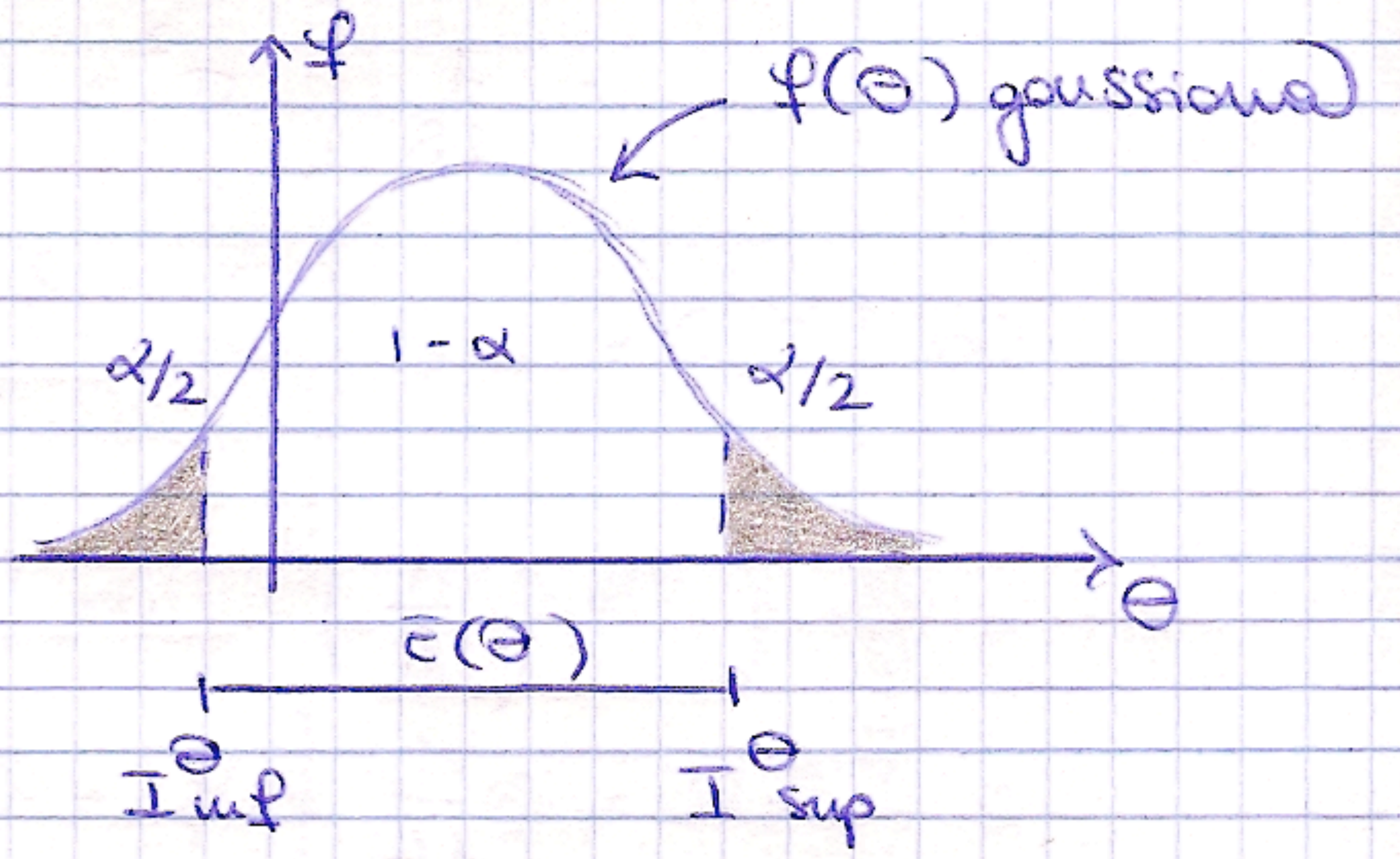
$$E[P]: \bar{x} = \frac{170+180+172+\ldots+184}{10} = 177,9\text{ mg}$$

b) stima per intervallo: si stima un intervallo in cui si è fiduciosi ricada il parametro ignoto.
Dato trovare un intervallo I^θ che abbia probabilità $1-\alpha$ di contenere la stima $\hat{\theta}$ e bipartisca la probabilità α nelle code.

$$P(I^\theta_{inf} \le \bar{X} \le I^\theta_{sup}) = 1-\alpha$$

per n "grande"

$$\bar{X} \sim N\left(E[P]; \frac{Var[P]}{n}\right)$$



Standardizzo $$P\left(-z_{\alpha/2} \le \frac{\bar{X}-E[P]}{\sqrt{\frac{Var[P]}{n}}} \le z_{\alpha/2}\right) = 1-\alpha$$

ottengo

$$P\left(\bar{X} - z_{\alpha/2}\sqrt{\frac{Var[P]}{n}} \le E[P] \le \bar{X} + z_{\alpha/2}\sqrt{\frac{Var[P]}{n}}\right) =$$

$$= 1-\alpha$$

$\Rightarrow$ ottengo l'intervallo $I = \left[\bar{x} - z_{\alpha/2}\sqrt{\frac{Var[P]}{n}} \ ; \ \bar{x} + z_{\alpha/2}\sqrt{\frac{Var[P]}{n}}\right]$ ← varianza nota

varianza non nota $\rightarrow$ la stimo usando s^2

$$I = \left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \ ; \ \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

In questo esercizio siamo nel caso di varianza non nota $\Rightarrow$

$$s^2 = \left(\frac{170^2 + 180^2 + 172^2 + \dots + 184^2}{10} - 177,9^2\right)\cdot\frac{10}{9}$$

$$= (31676,5 - 31648,41)\cdot\frac{10}{9} = 28,09\cdot\frac{10}{9} = 31,21 \ mg^2$$

$s = 5,59$ mg

Nel nostro caso $\alpha = 0,05$ per cui $\frac{\alpha}{2} = 0,025 \rightarrow z_{0,025} = 1,96$

di conseguenza la stima richiesta risulta

$$I = \left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \ ; \ \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right] = \left[177,9 - 1,96\cdot\frac{5,59}{\sqrt{10}} \ ; \ 177,9 + 1,96\cdot\frac{5,59}{\sqrt{10}}\right] = [174,91 \ ; \ 180,89]$$

c) con procedimento analogo è possibile ottenere l'intervallo di confidenza al 90%.

$\alpha = 0,1$ per cui $\frac{\alpha}{2} = 0,05 \rightarrow z_{0,05} = 1,64(5)$

$$I = \left[177,9 - 1,64\cdot\frac{5,59}{\sqrt{10}} \ ; \ 177,9 + 1,64\cdot\frac{5,59}{\sqrt{10}}\right] =$$

$$= [175 \ ; \ 180,8]$$

④ In questo caso la varianza è nota $Var[P]=100$

$\bar{x} = 90$ media campionaria

a) $\alpha = 0,01$ $\alpha/2 = 0,005$ $z_{0,005} = 2,57(6)$

La stima intervallare risulta

$$I = \left[\bar{x} - z_{\alpha/2}\sqrt{\frac{Var[P]}{n}} \,;\, \bar{x} + z_{\alpha/2}\sqrt{\frac{Var[P]}{n}}\right] =$$

$$= \left[90 - 2,57 \cdot \frac{10}{7} \,;\, 90 + 2,57 \cdot \frac{10}{7}\right] =$$

$$= \left[86,33 \,;\, 93,67\right]$$

b) $\alpha = 0,05$ $\alpha/2 = 0,025$ $z_{0,025} = 1,96$

$$I = \left[90 - 1,96 \cdot \frac{10}{7} \,;\, 90 + 1,96 \cdot \frac{10}{7}\right] =$$

$$= \left[87,2 \,;\, 92,8\right]$$

⑤ S^2 = stima puntuale di $Var[P]$

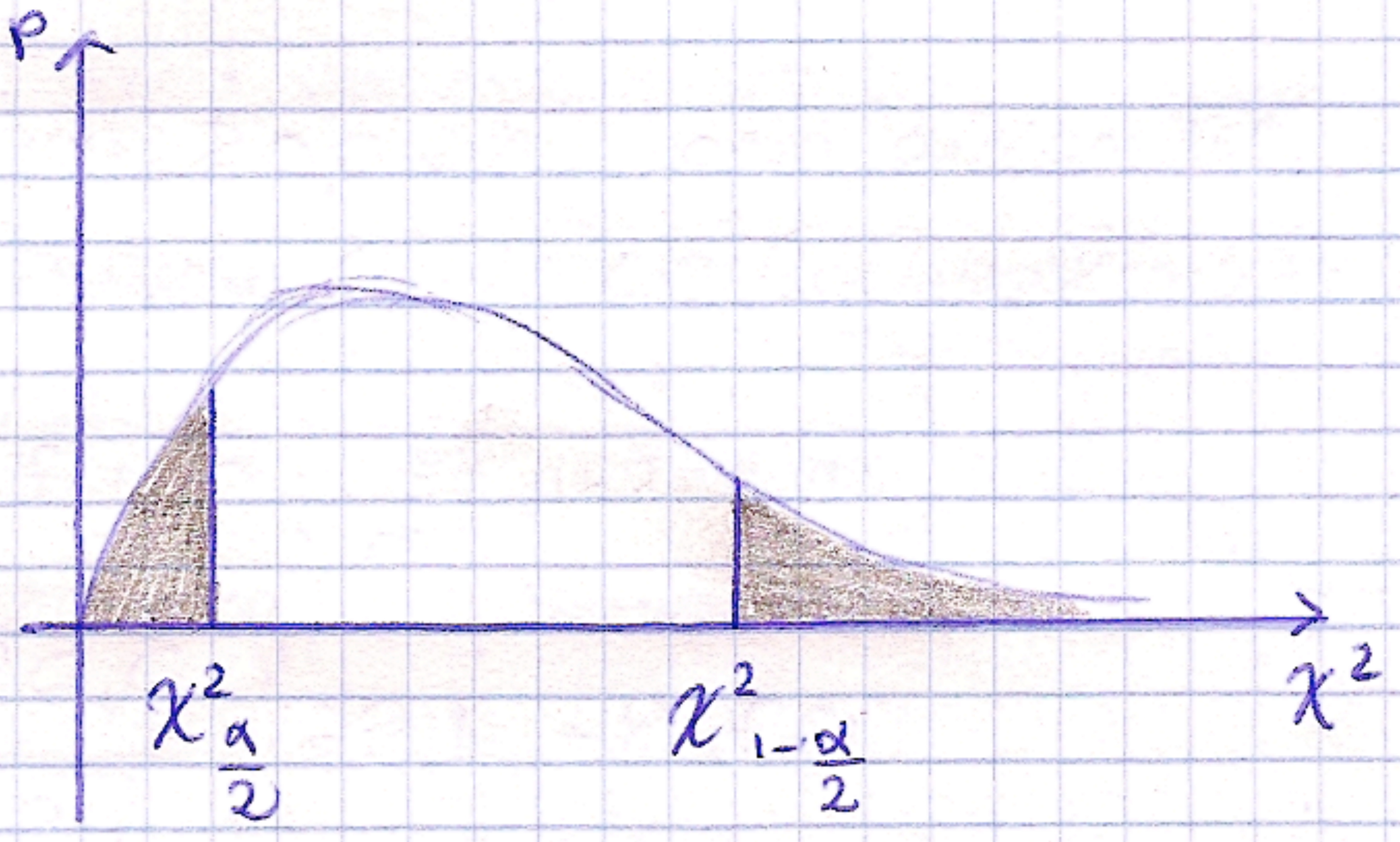
Per n grande e P gaussiana $S^2 \sim \frac{Var[P]}{n-1}\chi^2(n-1)$

Analogamente a prima

$P(I_{inf}^{\Theta} \le S^2 \le I_{sup}^{\Theta}) = 1-\alpha$

Mi riconduco ad una distribuzione nota

$$\frac{S^2(n-1)}{Var[P]} \sim \chi^2(n-1)$$



$$\Rightarrow P\left(\chi^2_{\frac{\alpha}{2}}(n-1) \le \frac{n-1}{Var[P]} S^2 \le \chi^2_{1-\frac{\alpha}{2}}(n-1)\right) = 1-\alpha$$

ottengo la stima

$$I = \left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \,;\, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right]$$

Media campionaria $\bar{x} = \frac{140+136+\dots+151}{10} = 144,3$ s

$$S^2 = \left(\frac{140^2+136^2+\dots+151^2}{10} - 144,3^2\right) \cdot \frac{10}{9} =$$

$$= (20851,5 - 20822,49) \cdot \frac{10}{9} = 32,2\bar{3} \text{ s}^2$$

Calcoli chi-quadrato. $\alpha = 0,05$ $\alpha/2 = 0,025$

$$\chi^2_{0,025}(9) = 19,02 \qquad \chi^2_{0,975}(9) = 2,7$$

$1-\alpha/2$

$$\Rightarrow \text{stima } I = \left[\frac{9 \cdot 32,23}{19,02} ; \frac{9 \cdot 32,23}{2,7} \right] =$$

$$= \left[15,25 ; 107,43 \right]$$