Statistica Matematica A

Esercitatore: Dott. Zucca

Esercitazione # 5

Esercizio 1 *Si consideri un campione* $X_1, ..., X_n$ *di v.a. discrete con funzione di massa:*

$$p(-1) = \frac{\theta}{2},\ p(0) = 1 - \theta,\ p(1) = \frac{\theta}{2}.$$

1. *Determinare per quali* θ *la funzione* p *è una funzione di massa.*
2. *Calcolare media e varianza di* X_1*.*

Si considerino i seguenti due stimatori di θ

$$\begin{aligned} \Theta_n &= \frac{1}{n}\sum_{i=1}^{n}|X_i| \\ T_n &= |X_n| \end{aligned}$$

1. *Si stabilisca se* Θ_n *e* T_n *sono distorti.*
2. *Si calcoli l'errore quadratico medio (EQM) degli stimatori* Θ_n *e* T_n *e se ne studi il comportamento per* $n \to +\infty$*.*
3. *Studiando l'efficienza relativa dei due stimatori, quale vi sembra migliore?*

Esercizio 2 *Sia* (X_1, X_2, X_3) *un campione bernoulliano estratto da una popolazione* X*.*

Al fine di stimare la media μ *della popolazione è stato proposto il seguente stimatore:*

$$T = \frac{1}{12}X_1 + 3(X_2 + X_3)$$

1. *Mostrare che* T *è uno stimatore distorto e valutarne l'errore quadratico medio.*
2. *Si trovi la costante* c *tale che* $W = cT$ *sia uno stimatore non distorto per* μ*. Si valuti l'errore quadratico medio di* W*.*
3. *Quale dei due stimatori risulta migliore?*

Esercizio 3 *Il tempo di risposta di un calcolatore all'input di un terminale si descrive mediante una v.a. di legge esponenziale di parametro* λ. *Si intendono misurare* n *tempi di risposta* $T_1, ..., T_n$.

1. *Mostrare che* $\bar{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_i$ *è uno stimatore non distorto di* $\frac{1}{\lambda}$.
2. *Calcolare l'errore quadratico medio di* $\bar{T}_n$ *e studiarne il comportamento asintotico.*
3. *Esiste una costante* $c \in \mathbb{R}$ *tale che (in base all'efficienza relativa) per ogni* $1 \leq n \leq 6$ *lo stimatore* $W = cT_1 + (1-c)T_2$ *sia migliore di* $\bar{T}_n$*?*

Esercizio 4 X_1 *è un campione unidimensionale estratto da una Poisson di parametro* λ.

Siano $T_1(X_1) = X_1$ *e* $T_2(X_1) = 1$ *due stimatori per* λ. *Quando* T_2 *è preferito a* T_1*?*

Esercizio 5 *Il numero di elementi spuri in un litro di una soluzione prodotta in laboratorio è descritto da una variabile aleatoria* X *con media* μ *e varianza* σ^2. *La produzione giornaliera sia di* n *confezioni da un litro l'una; il controllo viene effettuato prelevando* m *confezioni a caso tra quelle prodotte. Sia* $\beta = n/m$.

1. *Si stimi, in approssimazione normale, la probabilità che la media del campione prelevato differisca dalla media vera di almeno* $\alpha\sigma$.
2. *Si stimi, in approssimazione normale, la probabilità che la media del campione prelevato differisca dalla media della produzione giornaliera di almeno* $\alpha\sigma$.
3. *Si mostri che, in approssimazione normale, la stima della media giornaliera data dalla media dei campioni prelevati migliora all'aumentare di* n *(*β *fissato). Cosa si può dire invece della stima, fatta utilizzando il campione prelevato, del numero di elementi spuri totale prodotti in una giornata?*

1 Svolgimenti

Soluzione esercizio 1

1. Dobbiamo verificare che $p(k) \geq 0$ per ogni determinazione k e che $\sum_k p(k) = 1$. Condizioni verificate per ogni $0 \leq \theta \leq 1$.
2. $\mathbb{E}[X_1] = -1 \cdot p(-1) + 0 \cdot p(0) + 1 \cdot p(1) = 0$,

   mentre $\operatorname{var}(X_1) = \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))^2\right] = \theta$

3. Siccome $X_1, ..., X_n$ sono indipendenti ed identicamente distribuiti (dal momento che, per ipotesi, sono un campione), $\mathbb{E}[\Theta_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}|X_i|\right] = \mathbb{E}[|X_i|] = 1 \cdot p(-1) + 0 \cdot p(0) + 1 \cdot p(1) = \theta$

   $\mathbb{E}[T_n] = \mathbb{E}[|X_n|] = \theta$

   Quindi Θ_n e T_n sono due stimatori non distorti.

4. L'errore quadratico medio è definito come

   $EQM(\Theta_n) = \mathbb{E}\left[(\Theta_n - \theta)^2\right] = \text{var}(\Theta_n) + (distorsione)^2 =$

   siccome Θ_n è non distorto e $X_1, ..., X_n$ sono indipendenti ed identicamente distribuiti

   $= \text{var}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i|\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{var}(|X_i|) = \frac{1}{n}\text{var}(|X_i|) = \frac{1}{n}\left(\mathbb{E}\left[X_i^2\right] - \theta^2\right) = \frac{\theta(1-\theta)}{n} \to 0$ per $n \to +\infty$

   Analogamente: $EQM(T_n) = \theta(1-\theta)$

5. Studiando l'efficienza relativa dei due stimatori, abbiamo che:

$$\frac{EQM(\Theta_n)}{EQM(T_n)} = \frac{1}{n} < 1 \text{ per } n > 1,$$

   quindi Θ_n è migliore di T_n.

Soluzione esercizio 2

Osserviamo subito che lo stimatore della media μ, $\sum_{i=1}^{n}\alpha_i X_i$ è distorto se e solo se $\sum_{i=1}^{n}\alpha_i \neq 1$ e

$$EQM\left(\sum_{i=1}^{n}\alpha_i X_i - \mu\right) = \left(\sum_{i=1}^{n}\alpha_i^2\right)\text{var}(X_1) + \left(\sum_{i=1}^{n}\alpha_i - 1\right)^2.$$

1. Siccome $\mathbb{E}[T] = \frac{1}{12}\mathbb{E}(X_1) + 3\mathbb{E}(X_2) + 3\mathbb{E}(X_3) = \frac{73}{12}\mathbb{E}(X_i) = \frac{73}{12}\mu \neq \mu$, quindi lo stimatore T è distorto.

   $EQM(T) = \mathbb{E}\left[(T-\mu)^2\right] = \text{var}(T) + (distorsione)^2 = \text{var}(T) + \left(\frac{61}{12}\right)^2\mu^2 =$

   $= \left(18 + \frac{1}{144}\right)\mu(1-\mu) + \left(\frac{61}{12}\right)^2\mu^2$

2. siccome $\mathbb{E}(T) = \frac{73}{12}\mu$ allora $W = cT$ è non distorto se e solo se $c = \frac{12}{73}$.

   $EQM(W) = \text{var}(W) = \left(\frac{12}{73}\right)^2\text{var}(T) = \left(\frac{12}{73}\right)^2\left(18 + \frac{1}{144}\right)\mu(1-\mu)$

3. Siccome $\frac{EQM(W)}{EQM(T)} < 1$, allora W è migliore.

Soluzione esercizio 3

1. $\mathbb{E}\left[\bar{T}_n\right] = \mathbb{E}\left[T_i\right] = \frac{1}{\lambda}$, siccome la media di un'esponenziale di parametro λ è $1/\lambda$. Quindi lo stimatore $\bar{T}_n$ è uno stimatore non distorto di $\frac{1}{\lambda}$.

2. $EQM\left(\bar{T}_n\right) = \mathrm{var}\left(\bar{T}_n\right) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{var}\left(T_i\right) = \frac{1}{n}\mathrm{var}\left(T_i\right) = \frac{1}{n\lambda^2} \to 0$ per $n \to +\infty$.

3. Sia $W = cT_1 + (1-c)\,T_2$ con c costante reale.

   Allora

$$\begin{aligned}
\mathbb{E}\left[W\right] &= c\mathbb{E}\left[T_1\right] + (1-c)\,\mathbb{E}\left[T_2\right] = \mathbb{E}\left[T_1\right] = \frac{1}{\lambda} \\
EQM\left(W\right) &= \mathbb{E}\left[\left(W - \frac{1}{\lambda}\right)^2\right] = \mathrm{var}\left(W\right) = c^2\mathrm{var}\left(T_1\right) + (1-c)^2\,\mathrm{var}\left(T_2\right) = \\
&= \frac{1}{\lambda^2}\left(c^2 + 1 + c^2 - 2c\right) = \frac{2c^2 - 2c + 1}{\lambda^2}.
\end{aligned}$$

   Lo stimatore W sarebbe migliore di $\bar{T}_n$ in base all'efficienza relativa se e solo se $\frac{EQM(W)}{EQM\left(\bar{T}_n\right)} < 1$. Tuttavia

$$\frac{EQM\left(W\right)}{EQM\left(\bar{T}_n\right)} = n\left(2c^2 - 2c + 1\right) \leq 6\left(2c^2 - 2c + 1\right)$$

   per ogni $1 \leq n \leq 6$ siccome $2c^2 - 2c + 1 > 0$. Di conseguenza, se esiste una costante $c \in \mathbb{R}$ tale che per ogni $1 \leq n \leq 6$ lo stimatore $W = cT_1 + (1-c)\,T_2$ sia migliore di $\bar{T}_n$ allora tale c dovrebbe soddisfare

$$\begin{aligned}
6\left(2c^2 - 2c + 1\right) &< 1 \\
12c^2 - 12c + 5 &< 0 \text{ mai verificata,}
\end{aligned}$$

   quindi NON esiste nessuna costante c che soddisfa la richiesta del problema.

Soluzione esercizio 4

Se X_1 un campione unidimensionale estratto da una Poisson di parametro λ, allora lo stimatore $T_1\left(X_1\right) = X_1$ è non distorto, mentre $T_2\left(X_1\right) = 1$ è distorto per $\lambda \neq 1$. Infatti:

$$\begin{aligned}
\mathbb{E}\left[T_1\right] &= \mathbb{E}\left[X_1\right] = \lambda \\
\mathbb{E}\left[T_2\right] &= 1.
\end{aligned}$$

Calcolando gli errori quadratici medi:

$$\begin{aligned}
EQM\left(T_1\right) &= \mathbb{E}\left[\left(T_1 - \lambda\right)^2\right] = \mathrm{var}\left(X_1\right) = \lambda \\
EQM\left(T_2\right) &= \mathbb{E}\left[\left(T_2 - \lambda\right)^2\right] = (1-\lambda)^2 = 1 + \lambda^2 - 2\lambda.
\end{aligned}$$

Quindi T_2 è preferito a T_1 se e solo se

$$\begin{aligned}\frac{EQM\,(T_2)}{EQM\,(T_1)} &<1\\ 1+\lambda^2-2\lambda &<\lambda\\ 0,382 \approx \frac{3-\sqrt{5}}{2} &<\lambda < \frac{3+\sqrt{5}}{2} \approx 2.618.\end{aligned}$$

Soluzione esercizio 5

Sia $\{X_i\}_{i=1}^n$ la produzione giornaliera (sono variabili i.i.d.) e sia $\{X_i\}_{i=1}^m$ il campione prelevato; la variabile aleatoria che controlla la bontà della nostra stima è

$$Z := \frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{m}\sum_{i=1}^n X_i \equiv (1-\beta)\left(\frac{1}{n-m}\sum_{i=m+1}^n X_i - \frac{1}{m}\sum_{i=1}^m X_i\right).$$

In approssimazione normale

$$\begin{aligned}\frac{1}{n-m}\sum_{i=m+1}^n X_i &\approx \mathcal{N}(\mu, \sigma^2/(n-m))\\ \frac{1}{m}\sum_{i=1}^m X_i &\approx \mathcal{N}(\mu, \sigma^2/m)\end{aligned}$$

e sono indipendenti, per cui $Z \approx \mathcal{N}(0, (1/\beta-1)\sigma^2/n)$.

1. In approssimazione normale, standardizzando lo stimatore,

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^m X_i - \mu\right| \geq \alpha\sigma\right) = 2\left(1-\Phi\left(\alpha\sqrt{m}\right)\right). \tag{1}$$

2. Come nel punto precedente

$$\mathbb{P}\left(|Z| \geq \alpha\sigma\right) = 2\left(1-\Phi\left(\frac{\alpha\sqrt{n}}{\sqrt{1/\beta-1}}\right)\right). \tag{2}$$

3. Dall'equazione (2) si ha che

$$2\left(1-\Phi\left(\frac{\alpha\sqrt{n}}{\sqrt{1/\beta-1}}\right)\right) \downarrow 0, \qquad n\to+\infty.$$

La differenza tra la stima del numero totale di particelle spurie prodotte ed il numero effettivamente prodotto è rappresentato da $Z_1 = nZ$ pertanto

$$\mathbb{P}\left(|Z_1| \geq \alpha\sigma\right) = 2\left(1-\Phi\left(\frac{\alpha}{\sqrt{n(1/\beta-1)}}\right)\right)$$

e quindi

$$2\left(1-\Phi\left(\frac{\alpha}{\sqrt{n(1/\beta-1)}}\right)\right)\uparrow 1, \qquad n\to+\infty.$$