Si consideri un campione $X_1, \dots, X_n$ di v.a. discrete con distribuzione discreta:

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

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

① $p(x) \geq 0 \ \forall x$

$$\begin{cases} p(-1) = \frac{\vartheta}{2} \geq 0 \\ p(1) = \frac{\vartheta}{2} \geq 0 \\ p(0) = 1-\vartheta \geq 0 \end{cases} \Rightarrow 0 \leq \vartheta \leq 1$$

$$\sum_x p(x) = 1 \qquad 1 = \sum_x p(x) = \frac{\vartheta}{2} + \frac{\vartheta}{2} + 1 - \vartheta = 1 \quad \forall \vartheta$$

p È MASSA $\forall \vartheta \in [0;1]$

② $$\mathbb{E}X_i = \sum x \cdot p(x) = -1 \cdot \frac{\vartheta}{2} + 1\,\frac{\vartheta}{2} = 0$$

$$\mathbb{E}X_i^2 = \frac{\vartheta}{2} + \frac{\vartheta}{2} = \vartheta$$

$$\operatorname{Var} X_i = \mathbb{E}X_i^2 - \mathbb{E}^2 X_i = \vartheta$$

INOLTRE $$\mathbb{E}|X_i| = \sum |x| \cdot p(x) = |-1| \cdot \frac{\vartheta}{2} + |1|\,\frac{\vartheta}{2} = \vartheta$$

Si considerino i seguenti due stimatori di θ

$$\Theta_n = \frac{1}{n}\sum_{i=1}^{n} |X_i| \qquad T_n = |X_n|$$

(i) Si stabilisca se Θ_n e T_n sono distorti.

$$D_{\Theta_n} = \mathbb{E}\Theta_n - \vartheta = \mathbb{E}\,\frac{1}{n}\sum^{n} |X_i| - \vartheta =$$

$$= \frac{1}{n}\sum^{n} \mathbb{E}|X_i| - \vartheta = \frac{1}{n}\sum^{n} \vartheta - \vartheta =$$

$$= \vartheta - \vartheta = 0$$

Θ_n È NON DISTORTO

$$D_{T_n} = \mathbb{E}T_n - \vartheta = \mathbb{E}|X_n| - \vartheta = \vartheta - \vartheta = 0$$

T_n È NON DISTORTO

(ii) Si calcoli l'errore quadratico medio (EQM) degli stimatori Θ_n e T_n e se ne studi il comportamento per $n \to +\infty$.

$$EQM(\Theta_n) = \operatorname{Var} \Theta_n = \operatorname{Var} \frac{1}{n} \sum^{n} |X_i| =$$

$$= \frac{1}{n^2} \operatorname{Var} \sum |X_i| = \frac{1}{n^2} \sum^{n} \operatorname{Var} |X_i| =$$

$$= \frac{1}{n} \operatorname{Var} |X_i| = \frac{1}{n} \left[\mathbb{E} |X_i|^2 - \left(\mathbb{E} |X_i| \right)^2 \right] =$$

$$= \frac{1}{n} \left[\mathbb{E} X^2 - \left(\mathbb{E} |X| \right)^2 \right] =$$

$$= \frac{1}{n} \left(\vartheta - \vartheta^2 \right) \quad \to 0 \text{ PER } n \to \infty$$

$$EQM(T_n) = \operatorname{Var} T_n = \operatorname{Var} |X_n| = \operatorname{Var} |X| =$$

$$= \vartheta - \vartheta^2 \quad \text{COSTANTE PER } n \to \infty$$

(iii) Studiando l'efficienza relativa dei due stimatori, quale vi sembra migliore?

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

Θ_n È + EFFICIENTE

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:

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

$$X_i = X \sim B(p) \quad \forall i = 1,2,3$$

$\{X_i\}$ fam. va i.i.d.

$$\mathbb{E}X_i = p \qquad Var X_i = p(1-p)$$

1. Mostrare che T è uno stimatore distorto e valutarne l'errore quadratico medio.

$$\mathbb{E}T = \mathbb{E}\left[\frac{1}{12}X_1 + 3X_2 + 3X_3\right] = \frac{1}{12}\mathbb{E}X_1 + 3\mathbb{E}X_2 + 3\mathbb{E}X_3 =$$

$$= \left(\frac{1}{12} + 3 + 3\right)p = \frac{73}{12}p \quad \Rightarrow \quad T \text{ DISTORTO}$$

$$D_T = \mathbb{E}T - p = \frac{61}{12}p \quad ; \quad D_T^2 = \left(\frac{61}{12}\right)^2 p^2$$

$$EQM(T) = Var\,T + D_T^2 = Var\left(\frac{1}{12}X_1 + 3X_2 + 3X_3\right) + D_T^2 =$$

$$= \frac{1}{144} Var X_1 + 9 Var X_2 + 9 Var X_3 + D_T^2 =$$

$$= \left(\frac{1}{144} + 18\right) p(1-p) + D_T^2$$

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.

$$p = \mathbb{E}W = \mathbb{E}cT = c\,\mathbb{E}T = c\frac{73}{12}p \Rightarrow c = \frac{12}{73}$$

$$EQM(W) = Var\,W = Var\,cT = c^2 Var\,T =$$

$$= \frac{\cancel{12^2}}{73^2}\left(\frac{1}{\cancel{144}} + 18\right)p(1-p) = \frac{18 \cdot 144 + 1}{73^2}p(1-p)$$

3. Quale dei due stimatori risulta migliore?

$$\frac{EQM(W)}{EQM(T)} = \frac{c^2 Var\,T}{Var\,T + D_T^2} \overset{\text{MAGGIORAZIONE}}{<} \frac{Var\,T}{Var\,T + D_T^2} < 1$$

$\Rightarrow$ W È + EFF DI T

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* $T_1, ..., T_n$.

$\mathbb{E}T_i = \frac{1}{\lambda}$; $\mathbb{V}ar\, T = \frac{1}{\lambda^2}$

1. *Mostrare che* $\bar{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_i$ *è uno stimatore non distorto di* $\frac{1}{\lambda}$.

DEVE ESSERE $\mathbb{E}\bar{T}_n = \frac{1}{\lambda}$ INFATTI

$$\mathbb{E}\bar{T}_n = \mathbb{E}\,\frac{1}{n}\sum_{i=1}^{n} T_i = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}T_i = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda} = \frac{1}{\lambda}$$

2. *Calcolare l'errore quadratico medio di* $\bar{T}_n$ *e studiarne il comportamento asintotico.*

POICHÈ $\bar{T}_n$ È CORRETTO ALLORA $EQM(\bar{T}_n) = \mathbb{V}ar\,\bar{T}_n =$

$$= \mathbb{V}ar\left(\frac{1}{n}\sum_{i=1}^{n} T_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}ar\,T_i = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\lambda^2} = \frac{1}{n\lambda^2}$$

QUINDI $\lim_{n\to\infty} EQM(\bar{T}_n) = 0$

3. *Esiste una costante* $c \in \mathbb{R}$ *tale che (in base all'efficienza relativa) per ogni* $1 \le n \le 6$ *lo stimatore* $W = cT_1 + (1-c)\,T_2$ *sia migliore di* $\bar{T}_n$?

W È NON DISTORTO. INFATTI $\mathbb{E}W = \mathbb{E}[cT_1 + (1-c)T_2] =$

$$= c\,\mathbb{E}T_1 + (1-c)\mathbb{E}T_2 = \frac{c}{\lambda} + \frac{1-c}{\lambda} = \frac{1}{\lambda}.$$

QUINDI $EQM(W) = \mathbb{V}ar\,W = \mathbb{V}ar[cT_1 + (1-c)T_2] =$

$$= c^2\,\mathbb{V}ar\,T_1 + (1-c)^2\,\mathbb{V}ar\,T_2 = (2c^2-2c+1)\frac{1}{\lambda^2}$$

POICHÈ $\dfrac{EQM(W)}{EQM(\bar{T}_n)} = \dfrac{(2c^2-2c-1)\,{}^1/_{\lambda^2}}{\frac{1}{n}\cdot\frac{1}{\lambda^2}} = n(2c^2-2c+1)$

LA RICHIESTA DIVENTA $\exists\, c \in \mathbb{R} : n(2c^2-2c+1) < 1 \quad \forall\, 1 \le n \le 6$?

NOTIAMO CHE $2c^2-2c+1 = c^2+c^2-2c+1 = c^2+(c-1)^2 > 0 \quad \forall c \in \mathbb{R}$

E CHE $2c^2-2c+1 \geq \frac{1}{2} \;\forall\, c \in \mathbb{R}$ DUNQUE PER $n \geq 3$

$n(2c^2-2c+1) > 1 \;\; \forall c \in \mathbb{R}$.

QUINDI PER $1 \le n \le 6$ W NON È MAI + EFFICIENTE 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?

$\mathbb{E}T_1 = \mathbb{E}X_1 = \lambda \Rightarrow T_1$ NON DISTORTO

$\mathbb{E}T_2 = \mathbb{E}1 = 1 \Rightarrow T_2$ DISTORTO E $D_{T_2} = 1-\lambda$

QUINDI $EQM(T_1) = \mathbb{V}ar\, T_1 = \mathbb{V}ar\, X_1 = \lambda$

MENTRE $EQM(T_2) = \mathbb{V}ar\, T_2 + D^2_{T_2} = \mathbb{V}ar\, 1 + (1-\lambda)^2 = (1-\lambda)^2$

E L'EFFICIENZA RELATIVA È

$$\frac{EQM(T_1)}{EQM(T_2)} = \frac{\lambda}{(1-\lambda)^2} < 1 \quad \text{DA CUI}$$

$$\lambda < (1-\lambda)^2 \Rightarrow \lambda^2 - 2\lambda + 1 - \lambda > 0$$

T_1 È + EFF. DI $T_2 \iff \lambda^2 - 3\lambda + 1 > 0$

OVVERO SE E SOLO SE $\lambda < \lambda_1 \vee \lambda > \lambda_2$ DOVE

$$\lambda_{1,2} = \frac{3 \pm \sqrt{9-4}}{2} = \frac{3 \pm \sqrt{5}}{2} \quad \text{CON } \lambda_1 < \lambda_2$$

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 $\beta =$* m/n

1. *Si stimi, in approssimazione normale, la probabilità che la media del campione prelevato differisca dalla media vera di almeno $\alpha\sigma$.*

$$\mathbb{P}\left[|\bar{X}_m-\mu| \geq \alpha\sigma\right] = \mathbb{P}\left[\left|\frac{\bar{X}_m-\mu}{\sigma/\sqrt{m}}\right| \geq \frac{\alpha\sigma}{\sigma/\sqrt{m}}\right] =$$

$$\mathbb{P}\left[|Z| \geq \sigma\sqrt{m}\right] = 2\left(1-\Phi(\sigma\sqrt{m})\right)$$

2. *Si stimi, in approssimazione normale, la probabilità che la media del campione prelevato differisca dalla media della produzione giornaliera di almeno $\alpha\sigma$.*

$$\mathbb{P}\left[|\bar{X}_n-\bar{X}_m| \geq \alpha\sigma\right] = \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{i=1}^{m} X_i\right| \geq \alpha\sigma\right] =$$

OSS.:

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

$$= \frac{n-m}{n}\,\frac{1}{n-m}\sum_{i=m+1}^{n} X_i - \frac{n-m}{n}\,\frac{1}{m}\sum_{i=1}^{m} X_i =$$

$$= \frac{n-m}{n}\left(\bar{X}_{n-m} - \bar{X}_m\right) = (1-\beta)\left(\bar{X}_{n-m}-\bar{X}_m\right)$$

2. $$\bar{X}_{n-m} \sim N\left(\mu, \frac{\sigma^2}{n-m}\right) \; ; \; \bar{X}_m \sim N\left(\mu, \frac{\sigma^2}{m}\right)$$

3. $\bar{X}_{n-m}$ E $\bar{X}_m$ SONO INDIPENDENTI (NON HANNO ALCUN ADDENDO X_i IN COMUNE) QUINDI

$$\bar{X}_{n-m} - \bar{X}_m \sim N\left(\mu-\mu, \frac{\sigma^2}{n-m} + \frac{\sigma^2}{m}\right)$$

QUINDI $$(1-\beta)\left(\bar{X}_{n-m}-\bar{X}_m\right) \sim N\left(0, \frac{(1-\beta)\sigma^2}{m}\right)$$

$$= \mathbb{P}\left[|Z| \geq \frac{\alpha\sigma}{\sqrt{\frac{1-\beta}{m}}\,\sigma}\right] = \mathbb{P}\left[|Z| \geq \alpha\sqrt{\frac{m}{1-\beta}}\right] = 2\left[1-\Phi\left(\alpha\sqrt{\frac{m}{1-\beta}}\right)\right]$$

3. *Si mostri che, in approssimazione normale, la stima della media giornaliera data dalla media dei campioni prelevati migliora all'aumentare di* m *(β fissato). Cosa si può dire invece della stima, fatta utilizzando il campione prelevato, del numero di elementi spuri totale prodotti in una giornata?*

$$\lim_{m \to \infty} 2\left[1 - \Phi\left(\alpha\sqrt{\frac{m}{1-\beta}}\right)\right] = 0$$