

COGNOME E NOME _____

Esercizio 1A. Sia $X_1, \dots, X_n$ un campione casuale da una popolazione binomiale di parametri incogniti (m, θ) , con funzione di massa di probabilità

$$f_X(x; \theta) = \binom{m}{x} \theta^x (1 - \theta)^{m-x}, \quad x = 0, 1, 2, \dots, m.$$

- a) Determinare lo stimatore dei momenti di θ , $\hat{\theta}_M(\mathbf{X}_n)$.
- b) Determinare distorsione, varianza ed errore quadratico medio dello stimatore $\hat{\theta}_M(\mathbf{X}_n)$.
- c) Studiare la consistenza di $\hat{\theta}_M(\mathbf{X}_n)$.
- d) Determinare lo stimatore di massima verosimiglianza di θ e confrontarne le proprietà con quelle dello stimatore dei momenti.
- e) Stabilire se esiste lo stimatore UMVUE e, in caso di risposta affermativa, determinarlo.

Svolgimento.

- a) Si osservi che:

$$E_\theta[X] = m\theta, \quad V_\theta[X] = m\theta(1 - \theta).$$

Pertanto dall'equazione $m_1(\mathbf{X}_n) = E_\theta[X]$ discende che $\bar{X}_n = m\theta$, da cui

$$\hat{\theta}_M = \frac{\bar{X}_n}{m}.$$

- b) $E_\theta[\hat{\theta}_M] = \frac{1}{m} E_\theta[\bar{X}_n] = \frac{1}{m} E_\theta[X] = \frac{m\theta}{m} = \theta, \quad \forall \theta$. Lo stimatore dei momenti è quindi non distorto ($B_\theta = 0$). Pertanto

$$MSE_\theta[\hat{\theta}_M] = V_\theta[\hat{\theta}_M] = \frac{1}{m^2} V_\theta[\bar{X}_n] = \frac{m\theta(1 - \theta)}{nm^2} = \frac{\theta(1 - \theta)}{nm}.$$

- c) Poiché

$$\lim_{n \rightarrow +\infty} MSE_\theta[\hat{\theta}_M] = \lim_{n \rightarrow +\infty} \frac{\theta(1 - \theta)}{nm} = 0$$

lo stimatore dei momenti è consistente in media quadratica.

d) Si ha che le funzioni di verosimiglianza e logverosimiglianza sono

$$L(\theta; \mathbf{x}) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{nm - \sum_{i=1}^n x_i}$$

e

$$\ell(\theta; \mathbf{x}) = \left(\sum_{i=1}^n x_i \right) \ln \theta + (nm - \sum_{i=1}^n x_i) \ln(1 - \theta).$$

Derivando e risolvendo l'equazione di logveros. si ottiene che

$$\hat{\theta}_{MV} = \frac{\bar{X}_n}{m}.$$

(Si verifica facilmente che trattasi effettivamente di un punto di massimo). Lo stimatore di MV coincide con quello dei momenti e quindi i due stimatori hanno le stesse proprietà.

e) Si può rispondere al quesito in due modi.

I modo. Poiché il modello bernoulliano è una famiglia esponenziale (fatto noto, ma comunque semplicemente verificabile), uno stimatore non distorto di θ , che sia anche funzione di una statistica sufficiente e completa, è UMVUE (teoremi di Rao-Blackwell e Lehmann-Scheffé). In questo caso $\sum_{i=1}^n X_i$ è stat. sufficiente e completa, e quindi $\bar{X}_n/m$, essendo non distorto, è UMVUE.

II modo. E' sufficiente verificare che il limite inferiore di Cramer Rao risulta pari a $\frac{\theta(1-\theta)}{nm}$ e che quindi coincide con la varianza di $\bar{X}_n/m$.

Esercizio 2A. Un prodotto viene commercializzato con due confezioni diverse, A e B. Si assume che, in un periodo di riferimento, il ricavo delle vendite (in migliaia di euro) del prodotto con confezione j ($j = A, B$) sia una v.a. normale di parametri (θ_j, σ^2) . Si considerano due campioni casuali e indipendenti tra loro di ricavi relativi a $n = 9$ punti vendita. Indichiamo con

$$X_1^A, \dots, X_i^A, \dots, X_n^A \quad \text{e} \quad X_1^B, \dots, X_i^B, \dots, X_n^B$$

i due campioni casuali considerati. I valori osservati dei ricavi nei due campioni sono riportati nella seguente tabella.

Punto Vendita	x_i^A	x_i^B
1	1.72	1.17
2	0.50	1.73
3	1.01	1.42
4	1.14	2.20
5	1.13	1.21
6	1.55	1.11
7	2.32	0.84
8	0.71	1.51
9	0.94	1.75

- Determinare un intervallo di confidenza al 95% per il parametro incognito $\Delta = \theta_A - \theta_B$, assumendo che $\sigma^2 = 0.16$. Sulla base dell'intervallo che si ottiene con i dati assegnati, è lecito supporre che il diverso confezionamento non determina una differenza statisticamente significativa nei ricavi?
- Si sottoponga a verifica il seguente sistema di ipotesi, assumendo un livello di significatività pari a $\alpha = 0.05$:

$$H_0 : \Delta = 0$$

$$H_1 : \Delta \neq 0$$

e commentare l'esito del test.

Svolgimento.

- Per i noti risultati sulle statistiche medie campionarie per popolazioni normali, si ha che

$$\bar{X}_{n_A}^A - \bar{X}_{n_B}^B \sim N\left(\theta_A - \theta_B, 2\frac{\sigma^2}{n}\right).$$

Si ottiene così la seguente quantità pivotale :

$$\frac{\bar{X}_{n_A}^A - \bar{X}_{n_B}^B - \Delta}{\sigma\sqrt{\frac{2}{n}}} \sim N(0, 1).$$

L'intervallo di confidenza per Δ è quindi:

$$[\bar{X}_{n_A}^A - \bar{X}_{n_B}^B - z_{1-\frac{\alpha}{2}}\sigma\sqrt{\frac{2}{n}}, \bar{X}_{n_A}^A - \bar{X}_{n_B}^B + z_{1-\frac{\alpha}{2}}\sigma\sqrt{\frac{2}{n}}],$$

dove $z_{1-\frac{\alpha}{2}}$ indica il percentile di livello $1 - \frac{\alpha}{2}$ della v.a. $N(0, 1)$. Assumendo $1 - \alpha = 0.95$ si ha $z_{0.975} = 1.96$. Poiché $\bar{x}_{n_A}^A = 1.22$ e $\bar{x}_{n_B}^B = 1.44$, si ottiene che l'intervallo osservato risulta essere $[-0.59, 0.15]$. Dal momento che tale intervallo include lo zero (valore che indica l'uguaglianza dei due parametri θ_A e θ_B), sulla base dei dati disponibili non possiamo concludere che il confezionamento determini una differenza statisticamente significativa nei ricavi.

b) Si può rispondere al quesito in due modi.

I modo. L'intervallo di confidenza osservato contiene tutti i valori che corrispondono a ipotesi nulle accettate dal test bilaterale di livello α , condotto con i dati a disposizione. In altre parole, se l'intervallo di confidenza di livello $1 - \alpha$ osservato contiene lo zero, l'ipotesi nulla $\Delta = 0$ contro l'alternativa $\Delta \neq 0$ viene accettata nel test di livello α condotto con gli stessi dati che si sono usati per determinare l'intervallo. (IMPORTANTE: notare la corrispondenza tra livello di confidenza $(1 - \alpha)$ e livello del test (α)).

II modo. Il valore osservato della statistica test per il confronto tra un'ipotesi nulla semplice e un'ipotesi alternativa bilaterale è

$$\frac{(x_{n_A}^A - \bar{x}_{n_B}^B) - 0}{\sigma \sqrt{\frac{2}{n}}} = \frac{0.22}{\sqrt{(0.16 \times 2)/9}} = 1.17 < 1.96 = z_{1-\frac{\alpha}{2}}.$$

Si accetta pertanto H_0 .

Esercizio 3A. Sia $X_1, \dots, X_n$ un campione casuale di dimensione $n = 3$ proveniente da una popolazione di Poisson di parametro incognito θ , con distribuzione di probabilità:

$$f_X(x; \theta) = e^{-\theta} \frac{\theta^x}{x!}, \quad x = 0, 1, 2, \dots \quad \theta > 0.$$

Si consideri il sistema di ipotesi:

$$H_0 : \theta = \theta_0 = 2 \quad H_1 : \theta = \theta_1 = 1.$$

- a) Verificare che la regione di rifiuto del test basato sul rapporto delle verosimiglianze risulta essere

$$R = \{\mathbf{x}_n : x_1 + x_2 + x_3 \leq k\}. \quad (1)$$

- b) Calcolare la probabilità di errore di I specie, α , e la potenza del test, $1 - \beta$, che si ottengono ponendo $k = 1$.

[**Sugg.** Ricordare che se $X_1, \dots, X_n$ sono n v.a. i.i.d. distribuite secondo la legge di Poisson di parametro θ , la loro somma è una v.a. di Poisson di parametro $n\theta$.]

Svolgimento.

- a) La funzione di verosimiglianza per il modello di Poisson è

$$L(\theta; \mathbf{x}_n) = e^{-n\theta} \theta^{\sum_{i=1}^n x_i},$$

da cui si ottiene il rapporto delle verosimiglianze

$$\lambda(\mathbf{x}_n) = e^{-n(\theta_0 - \theta_1)} \left(\frac{\theta_0}{\theta_1} \right)^{\sum_{i=1}^n x_i}.$$

Per il Lemma di Neyman e Pearson, osservando che λ è funzione crescente di $\sum_{i=1}^n x_i$ (dal momento che $\theta_0 = 2 > 1 = \theta_1$), si ha che

$$R = \{\mathbf{x}_n : \lambda(\mathbf{x}_n) \leq k'\} = \{\mathbf{x}_n : \bar{x}_n \leq k\}.$$

Per $n = 3$ si ottiene quanto cercato.

- b) Osserviamo innanzitutto che $Y_3 = X_1 + X_2 + X_3 \sim \text{Poisson}(3\theta)$. Per definizione, $\alpha = \Pr\{R; \theta = \theta_0\}$ e $1 - \beta = \Pr\{R; \theta = \theta_1\}$. Assumendo $k = 1$, si ha

$$\alpha = \Pr\{Y_3 \leq 1; \theta = 2\} = \Pr\{Y_3 = 0; \theta = 2\} + \Pr\{Y_3 = 1; \theta = 2\} = e^{-6} + 6e^{-6} = 0.017,$$

$$1 - \beta = \Pr\{Y_3 \leq 1; \theta = 1\} = \Pr\{Y_3 = 0; \theta = 1\} + \Pr\{Y_3 = 1; \theta = 1\} = e^{-3} + 3e^{-3} = 0.2.$$

In questo problema la distribuzione della statistica test è nota, e non è quindi necessario ricorrere alla distribuzione asintotica che, per n sufficientemente elevato, risulta $N(n\theta, n\theta)$. Si noti inoltre che, in questo caso, essendo $n = 3$ molto piccolo, i valori che si ottengono utilizzando l'approssimazione normale sono piuttosto imprecisi. Infatti

$$\alpha \approx \Phi\left(\frac{1 - n\theta_0}{\sqrt{n\theta_0}}\right) = \Phi(-5/\sqrt{6}) = 0.0207,$$

$$1 - \beta \approx \Phi\left(\frac{1 - n\theta_1}{\sqrt{n\theta_1}}\right) = \Phi(-2/\sqrt{3}) = 0.124.$$

Esercizio 4A. Si consideri un campione casuale di dimensione n proveniente da una popolazione bernoulliana di parametro θ incognito. Siamo interessati a effettuare inferenza sul logaritmo naturale della varianza di tale popolazione, ovvero sul parametro $g(\theta) = \ln[\theta(1 - \theta)]$.

- Determinare lo stimatore di massima verosimiglianza di $g(\theta)$, $\widehat{g(\theta)}$.
- Determinare la varianza asintotica di $\widehat{g(\theta)}$ e uno stimatore di tale quantità.
- Determinare l'approssimazione normale della distribuzione campionaria dello stimatore di massima verosimiglianza di $g(\theta)$ e l'espressione generica di un intervallo di confidenza approssimato per $g(\theta)$.
- Supponendo di avere osservato in un campione di $n = 100$ osservazioni un valore della media campionaria pari a 0.6, determinare un intervallo di confidenza al 95% per $g(\theta)$.

Svolgimento.

- Per la proprietà di equivarianza degli stimatori di MV si ha

$$\widehat{g(\theta)} = g(\hat{\theta}_{MV}) = \ln \bar{X}_n(1 - \bar{X}_n).$$

- Per il *Delta Method* (le cui condizioni di applicabilità sono qui soddisfatte) si ha che $V_\theta[\widehat{g(\theta)}] \approx [g'(\theta)]^2 I_n(\theta)^{-1}$, dove, nel problema qui considerato, $g'(\theta) = \frac{d}{d\theta} \ln \theta(1 - \theta) = \frac{1-2\theta}{\theta(1-\theta)}$ e $I_n(\theta) = \frac{n}{\theta(1-\theta)}$ è l'informazione attesa di Fisher per il modello bernoulliano. Si ha quindi che

$$V_\theta[\ln \bar{X}_n(1 - \bar{X}_n)] \approx \frac{(1 - 2\theta)^2}{n\theta(1 - \theta)}.$$

Inoltre, ricordando che l'informazione osservata risulta $I_n^{oss}(\mathbf{x}_n) = -\frac{d}{d\theta} \ln L(\theta; \mathbf{x}_n)|_{\theta=\hat{\theta}_{MV}} = \frac{n}{\bar{x}_n(1-\bar{x}_n)}$, si ha

$$V_\theta[\widehat{g(\theta)}] \approx [g'(\hat{\theta}_{MV})]^2 [I_n^{oss}(\mathbf{x}_n)]^{-1} = \frac{(1 - 2\bar{X}_n)^2}{n\bar{X}_n(1 - \bar{X}_n)}.$$

- Per il *Delta Method* si ha che $\widehat{g(\theta)}$ ha distribuzione approssimativamente

$$N(g(\theta), [g'(\theta)]^2 I_n(\theta)^{-1},)$$

da cui si ottiene l'intervallo di confidenza di livello $1 - \alpha$ approssimato

$$\widehat{g(\theta)} \pm z_{1-\frac{\alpha}{2}} \sqrt{[g'(\hat{\theta}_{MV})]^2 [I_n^{oss}(\mathbf{x}_n)]^{-1}}$$

In questo esercizio si ha quindi che

$$\ln \bar{X}_n(1 - \bar{X}_n) \sim N\left(\ln \theta(1 - \theta), \frac{|1 - 2\theta|}{\sqrt{n\theta(1 - \theta)}}\right)$$

e che l'intervallo approssimato è

$$\ln \bar{X}_n(1 - \bar{X}_n) \pm z_{1-\frac{\alpha}{2}} \frac{|1 - 2\bar{x}_n|}{\sqrt{n\bar{x}_n(1 - \bar{x}_n)}}.$$

Per i dati a disposizione si ottiene quindi:

$$\ln(0.6 * 0.4) \pm 1.96 \frac{0.2}{\sqrt{100 * 0.6 * 0.4}},$$

da cui si ottiene l'intervallo $[-1.51, -1.35]$.