

Άσκηση 1: Έστω X_1, \dots, X_n τυχαίο δείγμα μεγέθους n από την κατανομή με σ.π.π.

$$p(x, \theta) = 2\theta^2 x^{-3}, \quad x > \theta, \quad \theta > 0.$$

α) Δείξτε ότι η στατιστική συνάρτηση $T(\underline{X}) := X_{(1)} = \min_{1 \leq i \leq n} X_i$ είναι επαρκής για την παράμετρο θ και πλήρης.

β) Βρείτε Α.Ε.Ε.Δ. του $\theta^{-\kappa}$.

Λύση:

α) Το στήριγμα $S = \{x \in \mathbb{R} : p(x, \theta) > 0\} = \{\theta, +\infty\}$ εξαρτάται από το θ και για αυτό γράφουμε την σ.π.π. στην εξής μορφή:

$$p(x, \theta) = 2\theta^2 x^{-3} I_{(\theta, +\infty)}(x), \quad x \in \mathbb{R},$$

όπου

$$I_{(\theta, +\infty)}(x) := \begin{cases} 1, & x > \theta \\ 0, & x \leq \theta \end{cases}.$$

Η από κοινού σ.π.π. του δείγματος X_1, \dots, X_n γράφεται τότε:

$$p(\underline{x}, \theta) = 2^n \theta^{2n} \prod_{i=1}^n x_i^{-3} \prod_{i=1}^n I_{(\theta, +\infty)}(x_i) \quad .$$

Ορίζουμε τη συνάρτηση:

$$f(\theta, x_{(1)}) := \begin{cases} 1, & \theta \leq x_{(1)} \\ 0, & \theta > x_{(1)} \end{cases},$$

όπου $x_{(1)} = \min_{1 \leq i \leq n} x_i$. Οι συναρτήσεις K και L με τιμές $K(\underline{x}) = \prod_{i=1}^n I_{(\theta, +\infty)}(x_i)$ και

$L(\underline{x}) = f(\theta, x_{(1)})$ είναι ίσες διότι παίρνουν μόνο τις τιμές 0 και 1 και επιπλέον $K(\underline{x}) = 1$ αν και μόνο αν $L(\underline{x}) = 1$. Επομένως

$$p(\underline{x}, \theta) = \{2^n \theta^{2n} f(\theta, x_{(1)})\} \prod_{i=1}^n x_i^{-3} = G(x_{(1)}, \theta) H(\underline{x}) \quad .$$

Έτσι από το παραγοντικό κριτήριο Neyman η στατιστική συνάρτηση $T(\underline{X}) := x_{(1)} = \min_{1 \leq i \leq n} X_i$ είναι επαρκής για το θ .

Για να αποδείξουμε την πληρότητα της T , χρειάζεται να γνωρίζουμε την κατανομή της. Για την συνάρτηση κατανομής $F_X(x, \theta)$ της τ.μ. X με σ.π.π. $p(x, \theta)$, έχουμε:

$$F_X(x, \theta) = \int_{\theta}^x p(u, \theta) du = 2\theta^2 \int_{\theta}^x u^{-3} du = 1 - \left(\frac{\theta}{x}\right)^2, \quad x > \theta$$

και $F_X(x, \theta) = 0$ αν $x \leq \theta$. Συνεπώς αν $F_T(t, \theta)$ είναι η συνάρτηση κατανομής της T τότε:

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

$$F_T(t, \theta) = P(T \leq t) = 1 - P(T > t) = 1 - P(X_{(1)} > t) = 1 - P(X_i > t, i = 1, \dots, n) = 1 - \prod_{i=1}^n P(X_i > t) =$$

$$1 - [(1 - F_X(t, \theta))^n] = 1 - \left(\frac{\theta}{t}\right)^{2n}, t > \theta.$$

Παραγωγίζοντας ως προς t λαμβάνουμε την σ.π.π. της T :

$$f_T(t, \theta) = 2n\theta^{2n}t^{-2n-1}, t > \theta.$$

Έστω τώρα $E(h(T), \theta) = 0 \quad \forall \theta > 0$ και για τυχούσα μετρήσιμη πραγματική συνάρτηση $h: \mathbb{R} \rightarrow \mathbb{R}$. Έχουμε ότι:

$$E(h(T), \theta) = 0 \quad \forall \theta > 0$$

$$\Rightarrow \int_{\theta}^{+\infty} h(t) f_T(t, \theta) dt = 0 \Rightarrow 2n\theta^{2n} \int_{\theta}^{+\infty} h(t) t^{-2n-1} dt = 0 \Rightarrow \int_{\theta}^{+\infty} h(t) t^{-2n-1} dt = 0.$$

Παραγωγίζοντας την τελευταία ισότητα ως προς θ λαμβάνουμε ότι

$$-h(\theta)\theta^{-2n-1} = 0 \quad \forall \theta > 0 \Rightarrow h(\theta) = 0 \quad \forall \theta > 0,$$

δηλαδή $h(y) = 0 \quad \forall y \in (0, +\infty)$. Για το σύνολο $(-\infty, 0]$ ισχύει ότι $P(T \leq 0) = 0 \quad \forall \theta > 0$, δηλαδή το $(-\infty, 0]$ έχει μέτρο μηδέν. Άρα $h=0$ σχεδόν παντού και συνεπώς η $T(X) := X_{(1)} = \min_{1 \leq i \leq n} X_i$ είναι πλήρης.

β) Έστω $U = \psi(T)$ α.ε. του $\theta^{-\kappa}$. Τότε:

$$E(\psi(T)) = \theta^{-\kappa} \Rightarrow \int_{\theta}^{+\infty} \psi(t) f_T(t, \theta) dt = \theta^{-\kappa} \Rightarrow 2n\theta^{2n} \int_{\theta}^{+\infty} \psi(t) t^{-2n-1} dt = \theta^{-\kappa} \Rightarrow \int_{\theta}^{+\infty} \psi(t) t^{-2n-1} dt = \frac{\theta^{-\kappa}}{2n\theta^{2n}}, \quad \forall \theta > 0$$

Παραγωγίζοντας την τελευταία ισότητα ως προς θ προκύπτει:

$$-\psi(\theta)\theta^{-2n-1} = \frac{-(2n+\kappa)}{2n\theta^{\kappa+2n+1}}, \quad \forall \theta > 0 \Rightarrow \psi(\theta) = \frac{(2n+\kappa)}{2n} \theta^{-\kappa} \quad \forall \theta > 0.$$

Άρα η στατιστική συνάρτηση $U(X) = \psi(T(X)) = \frac{(2n+\kappa)}{2n} T(X)^{-\kappa} = \frac{(2n+\kappa)}{2n} X_{(1)}^{-\kappa}$ είναι

η μοναδική Α.Ε.Ε.Δ. του $\theta^{-\kappa}$, ως συνάρτηση μόνο της επαρκούς και πλήρους στατιστικής συνάρτησης T .

Άσκηση 2: Έστω X_1, \dots, X_n τυχαίο δείγμα από την κατανομή $G\left(\frac{1}{\theta}, p\right)$ με σ.π.π.

$$p(x, \theta) = \frac{1}{\theta^p \Gamma(p)} x^{p-1} e^{-\frac{x}{\theta}}, x > 0, \theta > 0,$$

όπου p γνωστό.

α) Να υπολογιστεί η πληροφορία $I(\theta)$ κατά Fisher μιας παρατήρησης για το θ .

β) Να βρεθεί Α.Ε.Ε.Δ. της θ και να συγκριθεί η διασπορά της με το κάτω φράγμα της ανισότητας Cramer-Rao.

γ) Να βρεθεί Α.Ε.Ε.Δ. της $1/\theta$ και να συγκριθεί η διασπορά της με το κάτω φράγμα της ανισότητας Cramer-Rao.

Λύση:

α)

$$\begin{aligned}\ln p(x, \theta) &= -\ln \Gamma(p) - p \ln \theta - \frac{x}{\theta} + (p-1) \ln x \\ \frac{\partial}{\partial \theta} \ln p(x, \theta) &= -\frac{p}{\theta} + \frac{x}{\theta^2} \\ \frac{\partial^2}{\partial^2 \theta} \ln p(x, \theta) &= \frac{p}{\theta^2} - \frac{2x}{\theta^3}\end{aligned}$$

Άρα:

$$I(\theta) = -E \left\{ \frac{\partial^2}{\partial^2 \theta} \ln p(x, \theta) \right\} = E \left\{ \frac{2X}{\theta^3} \right\} - \frac{p}{\theta^2} = \frac{2}{\theta^3} E(X) - \frac{p}{\theta^2} = \frac{2}{\theta^3} p\theta - \frac{p}{\theta^2} = \frac{p}{\theta^2}.$$

β) Το στήριγμα $S = \{x \in \mathbb{R} : p(x, \theta) > 0\} = \{0, +\infty\}$ δεν εξαρτάται από το θ . Επιπλέον:

$$p(x, \theta) = \frac{1}{\theta^p \Gamma(p)} x^{p-1} e^{-\frac{x}{\theta}} = c(\theta) e^{Q(\theta)T(x)} h(x)$$

ανήκει δηλαδή στην Ε.Ο.Κ. και συνεπώς η $T=T(\underline{X}) := \sum_{i=1}^n T(X_i) = \sum_{i=1}^n X_i$ είναι επαρκής και πλήρης για το θ . Ακόμα έχουμε ότι:

$$E(T) = \sum_{i=1}^n E(X_i) = np\theta \Rightarrow E \left\{ \frac{T}{np} \right\} = \theta.$$

Συμπεραίνουμε λοιπόν ότι η α.ε. $T^*(\underline{X}) = \frac{\sum_{i=1}^n X_i}{np} = \bar{X}$ του θ είναι και Α.Ε.Ε.Δ. ως

συνάρτηση της επαρκούς και πλήρους T .

Επειδή οι X_1, \dots, X_n είναι ανεξάρτητες (άρα και ασυσχέτιστες) έπεται ότι:

$$V(T^*) = \frac{\sum_{i=1}^n V(X_i)}{(np)^2} = \frac{np\theta^2}{(np)^2} = \frac{\theta^2}{np}.$$

Το κάτω φράγμα Cramer-Rao δίνεται από την

$$LB = \frac{1}{nI(\theta)} = \frac{\theta^2}{np} = V(T^*).$$

γ) Ισχύει ότι αν $X_i \sim G\left(\frac{1}{\theta}, p\right) \Rightarrow \sum_{i=1}^n X_i \sim G\left(\frac{1}{\theta}, np\right)$ και άρα

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

$$E\left\{\frac{1}{T^\kappa}\right\} = \int_0^{+\infty} \frac{1}{t^\kappa} f_T(t, \theta) dt = \int_0^{+\infty} \frac{1}{t^\kappa} \frac{1}{\theta^{np} \Gamma(np)} t^{np-1} e^{-\frac{x}{\theta}} dt = \frac{1}{\theta^{np} \Gamma(np)} \int_0^{+\infty} t^{np-\kappa-1} e^{-\frac{x}{\theta}} dt \stackrel{y=t/\theta}{=} \theta \frac{\theta^{np-\kappa-1}}{\theta^{np} \Gamma(np)} \int_0^{+\infty} y^{np-\kappa-1} e^{-y} dy = \frac{\Gamma(np-\kappa)}{\theta^\kappa \Gamma(np)},$$

αφού:

$$\int_0^{+\infty} y^{np-\kappa-1} e^{-y} dy = \Gamma(np-\kappa)$$

Για $\kappa=1$ έχουμε

$$E\left\{\frac{1}{T}\right\} = \frac{\Gamma(np-1)}{\theta \Gamma(np)} = \frac{1}{\theta(np-1)},$$

αφού:

$$\Gamma(a) = (a-1)\Gamma(a-1).$$

Τελικά:

$$E\left\{\frac{np-1}{T}\right\} = \frac{1}{\theta},$$

και άρα η στατιστική συνάρτηση $\frac{np-1}{T}$ είναι Α.Ε.Ε.Δ. του $1/\theta$.

Ακόμα:

$$V\left\{\frac{np-1}{T}\right\} = (np-1)^2 V\left\{\frac{1}{T}\right\} = (np-1)^2 \left\{E\left\{\frac{1}{T^2}\right\} - \left[E\left\{\frac{1}{T}\right\}\right]^2\right\} = (np-1)^2 \left\{\frac{\Gamma(np-2)}{\theta^2 \Gamma(np)} - \left(\frac{1}{\theta(np-1)}\right)^2\right\} = \frac{1}{\theta^2(np-2)}$$

ενώ:

$$LB = \frac{\left[\frac{d}{d\theta}\left(\frac{1}{\theta}\right)\right]^2}{nI(\theta)} = \frac{\frac{1}{\theta^4}}{n \frac{p}{\theta^2}} = \frac{1}{np\theta^2} < V\left\{\frac{np-1}{T}\right\}.$$

Άσκηση 3: Έστω X_1, \dots, X_n τυχαίο δείγμα από την κατανομή $b(N, \theta)$, $\theta \in (0, 1)$.

α) Ναδειχτεί ότι η στατιστική συνάρτηση $\frac{\bar{X}}{N}$ είναι Α.Ε.Ε.Δ. του θ .

β) Να βρεθεί η Ε.Μ.Π. του θ .

γ) Έστω το τυχαίο δείγμα X_1, \dots, X_n από την κατανομή $b(3, \theta)$. Αν το δείγμα αυτό έδωσε τις μετρήσεις $x_1=3, x_2=0, x_3=1, x_4=2, x_5=1$ να βρεθεί ο ΕΜΠ της $P(X=0)$.

Λύση:

α) Το στήριγμα $S = \{x \in \mathbb{R} : p(x, \theta) > 0\} = \{0, 1, \dots, N\}$ δεν εξαρτάται από το θ . Επιπλέον:

$$p(x, \theta) = \binom{N}{x} \theta^x (1-\theta)^{N-x} = (1-\theta)^N e^{\left(\ln \frac{\theta}{1-\theta}\right)x} \binom{N}{x} = c(\theta) e^{\varrho(\theta)T(x)} h(x)$$

ανήκει δηλαδή στην Ε.Ο.Κ. και συνεπώς η $T = T(\underline{X}) := \sum_{i=1}^n T(X_i) = \sum_{i=1}^n X_i$ είναι επαρκής και πλήρης για το θ . Ακόμα έχουμε ότι:

$$E(T) = \sum_{i=1}^n E(X_i) = nN\theta \Rightarrow E\left\{\frac{T}{nN}\right\} = \theta.$$

Συμπεραίνουμε λοιπόν ότι η α.ε. $T^*(\underline{X}) = \frac{\sum_{i=1}^n X_i}{nN} = \frac{\bar{X}}{N}$ του θ είναι και Α.Ε.Ε.Δ. ως συνάρτηση της επαρκούς και πλήρους T .

β) Έχουμε:

$$L(\theta) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{nN - \sum_{i=1}^n x_i} \prod_{i=1}^n \binom{N}{x_i},$$

$$l(\theta) = \left(\sum_{i=1}^n x_i\right) \ln \theta + \left(nN - \sum_{i=1}^n x_i\right) \ln(1-\theta) + \ln \prod_{i=1}^n \binom{N}{x_i},$$

$$\frac{d}{d\theta} l(\theta) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{nN - \sum_{i=1}^n x_i}{1-\theta} = 0 \Rightarrow \theta = \frac{\bar{x}}{N}.$$

Ακόμα:

$$\frac{d^2}{d^2\theta} l(\theta) = -\frac{nN^2}{\bar{x}} - \frac{n(N - \bar{x})}{\left(1 - \frac{\bar{x}}{N}\right)^2} < 0$$

Επομένως η $\hat{\theta} = \frac{\bar{X}}{N}$ είναι η Ε.Μ.Π. του θ .

γ) Από τις μετρήσεις $x_1=3, x_2=0, x_3=1, x_4=2, x_5=1$ παίρνουμε ότι $\hat{\theta} = \frac{\bar{x}}{N} = 0.466$. Έτσι υποθέτοντας ότι $X \sim b(3, 0.466)$ έχουμε:

$$P(X = 0) = \binom{3}{0} 0.466^0 (1 - 0.466)^{3-0} = 0.15227.$$

Άσκηση 4: Έστω X_1, \dots, X_n τυχαίο δείγμα από την κατανομή $N(\mu, \sigma^2)$.

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

Να βρεθεί η Ε.Μ.Π. του σ^2 όταν:

α) μ γνωστό.

β) μ άγνωστο.

γ) Έστω πήραμε τις εξής μετρήσεις για το X : $x_1=0.1$, $x_2=1.5$, $x_3=0.3$, $x_4=2.1$, $x_5=0.8$, $x_6=3.2$, $x_7=2.4$, $x_8=5.3$, $x_9=0.7$, $x_{10}=4.5$. Να βρεθεί ο ΕΜΠ της $P(X>2.09)$.

Λύση:

α) Έχουμε:

$$L(\sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}},$$

$$l(\sigma^2) = -n \ln \sqrt{2\pi} - n \ln \sqrt{\sigma^2} - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2},$$

$$\frac{d}{d\sigma^2} l(\sigma^2) = -\frac{n}{2\sqrt{\sigma^2}} \frac{1}{\sqrt{\sigma^2}} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2} = 0 \Rightarrow \sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}.$$

Παρατηρούμε ότι για $\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$ είναι:

$$\frac{d^2}{d^2\sigma^2} l(\sigma^2) = -\frac{n^3}{2\left\{\sum_{i=1}^n (x_i - \mu)^2\right\}^2} < 0.$$

Επομένως η $\hat{\theta} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$ είναι η Ε.Μ.Π. του σ^2 .

β) Έχουμε:

$$L(\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}},$$

$$l(\mu, \sigma^2) = -n \ln \sqrt{2\pi} - n \ln \sqrt{\sigma^2} - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2},$$

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

$$\left\{ \begin{array}{l} \frac{d}{d\mu} l(\theta) = 2 \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} = 0 \Rightarrow \mu = \bar{x} \\ \frac{d}{d\sigma^2} l(\theta) = -\frac{n}{2\sqrt{\sigma^2}} \frac{1}{\sqrt{\sigma^2}} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2} = 0 \Rightarrow \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \end{array} \right.$$

Η λύση αυτή αντιστοιχεί πράγματι σε μέγιστο αφού ο πίνακας Hesse για $\mu = \bar{x}$ και

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \text{ είναι ο:}$$

$$H = \left[\frac{\partial^2 l(\mu, \sigma^2)}{\partial \mu \partial \sigma^2} \right] = \begin{bmatrix} -\frac{n^2}{\sum_{i=1}^n (x_i - \bar{x})^2} & 0 \\ 0 & -\frac{n^3}{2 \left\{ \sum_{i=1}^n (x_i - \bar{x})^2 \right\}^2} \end{bmatrix}$$

ο οποίος είναι αρνητικά ορισμένος. Επομένως η Ε.Μ.Π. του σ^2 όταν μ άγνωστο είναι:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}.$$

γ) Από τις 10 μετρήσεις προκύπτει ότι:

$$\begin{aligned} \hat{\mu} &= \bar{x} = 2.09 \\ \hat{\sigma}^2 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = 2.855. \end{aligned}$$

Άρα $X \sim N(2.09, 2.855)$ και επομένως:

$$P(X > 2.09) = P\left(\frac{X - 2.09}{\sqrt{2.855}} > \frac{2.09 - 2.09}{\sqrt{2.855}}\right) = 1 - \Phi(0) = 0.5.$$

Άσκηση 5: Ο αριθμός $N(t)$ των εκπεμπόμενων σωματιδίων a από μια ραδιενεργό πηγή σε χρόνο t ωρών ακολουθεί την $P(\theta t)$, $\theta > 0$. Αν X_1, \dots, X_n είναι οι ενδιάμεσοι χρόνοι μεταξύ n διαδοχικών εκπομπών, να βρεθούν

- Η Ε.Μ.Π. του θ .
- Η εκτιμήτρια ροπών του θ .
- Η Ε.Μ.Π. της πιθανότητας σε χρόνο 2 ωρών να μην έχουμε εκπομπή σωματιδίου.

Λύση:

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

α) Είναι γνωστό από τις πιθανότητες ότι τα X_1, \dots, X_n είναι ανεξάρτητες και ισόνομες τ.μ. που ακολουθούν την $\text{Exp}(\theta)$. Οπότε:

$$L(\theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i},$$
$$l(\theta) = n \ln \theta - \theta \left(\sum_{i=1}^n x_i \right),$$
$$\frac{d}{d\theta} l(\theta) = \frac{n}{\theta} - \sum_{i=1}^n x_i = 0 \Rightarrow \theta = \frac{1}{\bar{x}}.$$

Ακόμα:

$$\frac{d^2}{d^2\theta} l(\theta) = -\frac{n}{\left(\frac{1}{\bar{x}}\right)^2} < 0 \text{ για } \theta = \frac{1}{\bar{x}}.$$

Επομένως η $\hat{\theta} = \frac{1}{\bar{X}}$ είναι η Ε.Μ.Π. του θ .

β) Έχουμε $\bar{x} = E(X) = 1/\theta$ και λύνοντας ως προς θ παίρνουμε ότι $\theta = \frac{1}{\bar{x}}$. Άρα η εκτιμήτρια ροπών του θ είναι η $\tilde{\theta} = \frac{1}{\bar{x}}$.

γ) Είναι $t=2$ οπότε η $N(2)$ ακολουθεί την $P(2\hat{\theta})$. Η ζητούμενη πιθανότητα είναι:

$$P(N(2) = 0) = e^{-2\hat{\theta}} \frac{(2\hat{\theta})^0}{0!} = e^{-2\hat{\theta}} = e^{-\frac{2}{\bar{x}}}.$$

Άσκηση 6: Ο χρόνος ζωής X σε ώρες μιας λυχνίας ακολουθεί κατανομή με συνάρτηση πυκνότητας

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

Αν X_1, \dots, X_n είναι οι χρόνοι ζωής n λυχνιών, τότε:

α) Να βρεθεί η Ε.Μ.Π. του θ .

β) Ποια θα ήταν η εκτίμηση σας για τον μέσο χρόνο ζωής, αν είχατε τις παρατηρήσεις: $x_1=1.2, x_2=5.7, x_3=6.01, x_4=2.3, x_5=8.4, x_6=3.22, x_7=2.8, x_8=1.51, x_9=7.2, x_{10}=4.1$;

γ) Με βάση τα δεδομένα του β) ερωτήματος και την υπόθεση περί της κατανομής του χρόνου ζωής, να εκτιμήσετε την πιθανότητα ο χρόνος ζωής μιας λυχνίας να είναι μεγαλύτερος των 10 ωρών.

δ) Εκτιμήστε μη παραμετρικά, δηλαδή χωρίς την υπόθεση περί κατανομής, την πιθανότητα στο ερώτημα γ).

Λύση:

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

α) Έχουμε:

$$L(\theta) = \frac{1}{\theta^{2n}} e^{-\frac{1}{\theta} \sum_{i=1}^n x_i} \prod_{i=1}^n x_i,$$
$$l(\theta) = -2n \ln \theta - \frac{1}{\theta} \left(\sum_{i=1}^n x_i \right) + \ln \prod_{i=1}^n x_i,$$
$$\frac{d}{d\theta} l(\theta) = -\frac{2n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i = 0 \Rightarrow \theta = \frac{\bar{x}}{2}.$$

Ακόμα για $\theta = \frac{\bar{x}}{2}$:

$$\frac{d^2}{d^2\theta} l(\theta) = -\frac{8n}{\bar{x}^2} < 0.$$

Επομένως η $\hat{\theta} = \frac{\bar{x}}{2}$ είναι η Ε.Μ.Π. του θ .

β) Ο μέσος χρόνος ζωής της λυχνίας υπολογίζεται ως εξής:

$$E(X) = \int_0^{+\infty} x \frac{1}{\theta^2} x e^{-\frac{x}{\theta}} dx = \frac{1}{\theta^2} \int_0^{+\infty} x^2 e^{-\frac{x}{\theta}} dx \stackrel{y=x/\theta}{=} \frac{\theta^2}{\theta} \int_0^{+\infty} y^2 e^{-y} dy = \theta \Gamma(3) = 2\theta,$$

εφόσον $\Gamma(3) = 2! = 2$. Άρα η Ε.Μ.Π. της $E(X)$ είναι η $2\hat{\theta} = 2 \frac{\bar{x}}{2} = \bar{x} = 4.244$.

γ) Έχουμε τα εξής:

$$P(X > 10) = \int_{10}^{+\infty} \frac{1}{\theta^2} x e^{-\frac{x}{\theta}} dx = \frac{1}{\theta^2} \int_{10}^{+\infty} x e^{-\frac{x}{\theta}} dx = \int_{10}^{+\infty} y e^{-y} dy = \left(\frac{10}{\theta} + 1 \right) e^{-\frac{10}{\theta}}.$$

Για $\theta = \hat{\theta} = \frac{\bar{x}}{2} = 2.122$ έχουμε:

$$\hat{P}(X > 10) = \left(\frac{10}{\hat{\theta}} + 1 \right) e^{-\frac{10}{\hat{\theta}}} = 0.051.$$

δ) Με σκεπτικό ανάλογο αυτού της άσκησης 11, η μη παραμετρική εκτίμηση γίνεται

βάσει της διωνυμικής κατανομής $\hat{P}(X > 10) = \frac{0}{10} = 0$.

Άσκηση 7: Έστω X_1, \dots, X_n τυχαίο δείγμα από την κατανομή $U(\theta_1, \theta_2)$. Να βρεθούν:

α) Ένα επαρκές στατιστικό για το (θ_1, θ_2) .

β) Η Ε.Μ.Π. του (θ_1, θ_2) .

γ) Η εκτιμήτρια ροπών του (θ_1, θ_2) .

δ) Χρησιμοποιώντας τις παρακάτω 20 παρατηρήσεις

76 68 52 73 59 45 67 44 84 61 57 70 89 41 75 63 49 80 53 86

να εκτιμηθούν οι παράμετροι (θ_1, θ_2) με την μέθοδο των ροπών και την μέθοδο μεγίστης πιθανοφάνειας.

Λύση:

α) Η σ.π.π. της $U(\theta_1, \theta_2)$ είναι:

$$p(x, \theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & x \in (\theta_1, \theta_2) \\ 0, & x \notin (\theta_1, \theta_2) \end{cases} = \frac{1}{\theta_2 - \theta_1} I_{(\theta_1, \theta_2)}(x), \quad x \in \mathbb{R},$$

$$\text{όπου } I_{(\theta_1, \theta_2)}(x) := \begin{cases} 1, & x \in (\theta_1, \theta_2) \\ 0, & x \notin (\theta_1, \theta_2) \end{cases}.$$

Οπότε η από κοινού σ.π.π. του δείγματος είναι:

$$p(\underline{x}, \theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n} \prod_{i=1}^n I_{(\theta_1, \theta_2)}(x_i), \quad \underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Ορίζουμε την συνάρτηση:

$$f(a, b) := \begin{cases} 1, & a \leq b \\ 0, & a > b \end{cases}.$$

Τότε $f(\theta_1, x_{(1)}) f(x_{(n)}, \theta_2) = \prod_{i=1}^n I_{(\theta_1, \theta_2)}(x_i)$ για κάθε $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, με $x_{(1)} = \min x_i$ και $x_{(n)} = \max x_i$ οπότε θα έχουμε

$$p(\underline{x}, \theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n} f(\theta_1, x_{(1)}) f(x_{(n)}, \theta_2) = G(T(\underline{x}), \theta_1, \theta_2) H(\underline{x}),$$

όπου $G(T(\underline{x}), \theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n} f(\theta_1, x_{(1)}) f(x_{(n)}, \theta_2)$ και $H(\underline{x}) = 1$. Από το παραγοντικό

κριτήριο Neyman, συμπεραίνουμε ότι η $\underline{T}(\underline{x}) = (T_1(\underline{x}), T_2(\underline{x})) = (X_{(1)}, X_{(n)})$ είναι επαρκής εκτιμήτρια του (θ_1, θ_2) .

β) Η συνάρτηση πιθανοφάνειας είναι:

$$L(\theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n} \prod_{i=1}^n I_{(\theta_1, \theta_2)}(x_i) = \frac{1}{(\theta_2 - \theta_1)^n} f(\theta_1, x_{(1)}) f(x_{(n)}, \theta_2),$$

η οποία δεν παραγωγίζεται παντού ως προς θ_1 και θ_2 . Έτσι για να μεγιστοποιηθεί θα πρέπει να ελαχιστοποιηθεί η διαφορά $\theta_2 - \theta_1$ και ταυτόχρονα το γινόμενο $f(\theta_1, x_{(1)}) f(x_{(n)}, \theta_2)$ να πάρει την μέγιστη τιμή του που είναι 1. Όμως $f(\theta_1, x_{(1)}) f(x_{(n)}, \theta_2) = 1$ αν-ν $\theta_1 \leq x_{(1)}$ και $\theta_2 \geq x_{(n)}$. Συνεπώς η Ε.Μ.Π. του (θ_1, θ_2) είναι η $(X_{(1)}, X_{(n)})$.

γ) Θεωρούμε το σύστημα

$$\frac{1}{n} \sum_{i=1}^n x_i = E(X)$$

$$\frac{1}{n} \sum_{i=1}^n x_i^2 = E(X^2)$$

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

Επειδή $E(X) = \frac{\theta_1 + \theta_2}{2}$ και $E(X^2) = V(X) + [E(X)]^2 = \frac{(\theta_1 - \theta_2)^2}{12} + \left(\frac{\theta_1 + \theta_2}{2}\right)^2$ η λύση του παραπάνω συστήματος δίνει την εκτιμήτρια ροπών του (θ_1, θ_2) που είναι:

$$\left(\bar{X} - \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}, \bar{X} + \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \right).$$

δ) Οι εκτιμήσεις των θ_1 και θ_2 με την μέθοδο μεγίστης πιθανοφάνειας είναι:

$$\hat{\theta}_1(x) = x_{(1)} = 41$$

$$\hat{\theta}_2(x) = x_{(n)} = 89$$

ενώ οι εκτιμήσεις των θ_1 και θ_2 με την μέθοδο των ροπών είναι:

$$\tilde{\theta}_1(x) = \bar{x} - \sqrt{\frac{3}{20} \sum_{i=1}^{20} (x_i - \bar{x})^2} = 40.017$$

$$\tilde{\theta}_2(x) = \bar{x} + \sqrt{\frac{3}{20} \sum_{i=1}^{20} (x_i - \bar{x})^2} = 89.183$$

Άσκηση 8: Έστω $X \sim G(\theta_1, \theta_2)$ και X_1, \dots, X_n τυχαίο δείγμα. Έστω $T = \sum_{i=1}^n X_i$ και

$R = \sum_{i=1}^n \ln X_i$. Δείξτε ότι η στατιστική συνάρτηση $\tilde{T} = (T, R)$ είναι επαρκής για την παράμετρο $\underline{\theta} = (\theta_1, \theta_2)$. Είναι η $\tilde{T}^* = \left(\frac{1}{n}T, \frac{1}{n}R\right)$ επαρκής για την $\underline{\theta} = (\theta_1, \theta_2)$;

Λύση:

$$f(x, \theta_1, \theta_2) = \frac{1}{\Gamma(\theta_2)\theta_1^{\theta_2}} x^{\theta_2-1} e^{-\frac{1}{\theta_1}x} = \frac{1}{\Gamma(\theta_2)\theta_1^{\theta_2}} e^{(\theta_2-1)\ln x - \frac{1}{\theta_1}x} 1(x) = C(\theta_1, \theta_2) e^{\mathcal{Q}_1(\theta_1, \theta_2)T_1(x) + \mathcal{Q}_2(\theta_1, \theta_2)T_2(x)} h(x)$$

Άρα η κατανομή ανήκει στη διπαραμετρική εκθετική οικογένεια κατανομών και συνεπώς η στατιστική συνάρτηση $\tilde{T} = (T, R) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n \ln X_i\right)$ είναι επαρκής για την παράμετρο $\underline{\theta} = (\theta_1, \theta_2)$.

Από γνωστό πόρισμα η $\tilde{T}^* = \left(\frac{1}{n}T, \frac{1}{n}R\right)$ είναι επίσης επαρκής για την $\underline{\theta} = (\theta_1, \theta_2)$.

Άσκηση 9: Έστω η διακριτή τ.μ. $X \sim f(x, \theta) = \theta^2(x+1)(1-\theta)^x$, $x = 0, 1, \dots$ όπου $\theta \in \Omega = (0,1)$. Να βρεθεί η Ε.Μ.Π. της θ με την βοήθεια τ.δ. X_1, \dots, X_n . Στην συνέχεια να βρεθεί η Ε.Μ.Π. της $\alpha(\theta) = \frac{2(1-\theta)}{\theta}$ και να εξετάσετε αν είναι αμερόληπτη.

Λύση:

Έχουμε:

$$L(\theta) = \theta^{2n} \left[\prod_{i=1}^n (x_i - 1) \right] (1-\theta)^{\sum_{i=1}^n x_i},$$

$$l(\theta) = 2n \ln \theta + \left(\sum_{i=1}^n x_i \right) \ln(1-\theta) + \ln \prod_{i=1}^n (x_i + 1),$$

$$\frac{d}{d\theta} l(\theta) = \frac{2n}{\theta} - \frac{\sum_{i=1}^n x_i}{1-\theta} = 0 \Rightarrow \theta = \frac{2}{2 + \bar{x}}.$$

Ακόμα για $\theta = \frac{2}{2 + \bar{x}}$:

$$\frac{d^2}{d^2\theta} l(\theta) = -\frac{2n}{\left(\frac{2}{2 + \bar{x}}\right)^2} - \frac{\sum_{i=1}^n x_i}{\left(\frac{2}{2 + \bar{x}}\right)^2} < 0.$$

Επομένως η $\hat{\theta} = \frac{2}{2 + \bar{x}}$ είναι η Ε.Μ.Π. του θ .

Από γνωστό πόρισμα η Ε.Μ.Π. της $\alpha(\theta) = \frac{2(1-\theta)}{\theta}$ είναι η $\alpha(\hat{\theta}) = \frac{2(1-\hat{\theta})}{\hat{\theta}} = \bar{x}$.

Θα εξετάσουμε τώρα αν η $T = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ είναι α.ε. της $\alpha(\theta)$.

Έχουμε $E(T) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} nE(X) = E(X)$. Όμως:

$$E(X) = \sum_{x=1}^{\infty} x\theta^2(x+1)(1-\theta)^x = \theta^2 \sum_{x=1}^{\infty} (x^2 + x)(1-\theta)^x = \frac{2(1-\theta)}{\theta} = \alpha(\theta)$$

αφού $\sum_{x=1}^{\infty} x^k a^x = \frac{a(a+1)\dots(a+k-1)}{(1-a)^{k+1}}$. Άρα η T είναι α.ε. της $\alpha(\theta)$.

Άσκηση 10: Έστω $X \sim f(x, \theta) = a\theta x^{a-1} e^{-\theta x^a}$, $x > 0$, $\theta > 0$ άγνωστη παράμετρος και $a > 0$ γνωστή παράμετρος. Έστω τ.δ. X_1, \dots, X_n .

α) Να βρείτε την Ε.Μ.Π. της θ .

β) Να εξετάσετε αν η Ε.Μ.Π. είναι αμερόληπτη.

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

γ) Με βάση το αποτέλεσμα του β) να υποδείξετε μια Α.Ε.Ε.Δ. της θ .

δ) Να βρείτε το κατώτερο φράγμα Cramer-Rao και να το συγκρίνετε με τη διασπορά της Α.Ε.Ε.Δ. που βρήκατε στο γ).

Λύση:

α) Έχουμε:

$$L(\theta) = a^n \theta^n \left[\prod_{i=1}^n x_i^{a-1} \right] e^{-\theta \sum_{i=1}^n x_i^a},$$

$$l(\theta) = n \ln a + n \ln \theta - \theta \left(\sum_{i=1}^n x_i^a \right) + \ln \prod_{i=1}^n x_i^{a-1},$$

$$\frac{d}{d\theta} l(\theta) = -\frac{n}{\theta} - \sum_{i=1}^n x_i^a = 0 \Rightarrow \theta = \frac{n}{\sum_{i=1}^n x_i^a}.$$

Ακόμα:

$$\frac{d^2}{d^2\theta} l(\theta) = -\frac{n}{\theta^2} < 0.$$

Επομένως η $\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i^a}$ είναι η Ε.Μ.Π. του θ .

β) Έστω $U = \frac{n}{\sum_{i=1}^n x_i^a}$. Θα βρούμε την $E(U)$. Προηγουμένως θα βρούμε την κατανομή

της x^a και εν συνεχεία της $\sum_{i=1}^n x_i^a$. Επειδή η συνάρτηση $y = g(x) = x^a$ είναι “1-1” και συνεχώς παραγωγίσιμη στο $S = \{x \in \mathbb{R}^+ : f(x, \theta) > 0\} = \{0, +\infty\}$ με αντίστροφη $x = g^{-1}(y) = y^{1/a}$ θα έχουμε για την σ.π.π. της:

$$f_Y(y) = f(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| = a\theta y^{\frac{a-1}{a}} e^{-\theta y} \frac{1}{a} y^{\frac{1-a}{a}} = \theta e^{-\theta y}, \quad y > 0. \text{ Άρα η } Y = X^a \sim G(\theta, p=1)$$

και συνεπώς η τ.μ. $Z = \sum_{i=1}^n X_i^a \sim G(\theta, p=n)$.

Επειδή τώρα η στατιστική συνάρτηση $U = \frac{n}{\sum_{i=1}^n x_i^a} = \frac{n}{Z} = \phi(Z)$ θα έχουμε:

$$E(U) = E[\phi(Z)] = \int_0^{+\infty} \phi(z) f_Z(z) dz = \int_0^{+\infty} \frac{n}{z} \frac{\theta^p}{\Gamma(p)} z^{p-1} e^{-\theta z} dz = \frac{n\theta^p}{\Gamma(p)} \int_0^{+\infty} z^{p-2} e^{-\theta z} dz \stackrel{\theta z = \omega}{=} =$$

$$\frac{n\theta^p}{\Gamma(p)\theta^{p-1}} \int_0^{+\infty} \omega^{p-2} e^{-\omega} d\omega \stackrel{p=n}{=} \frac{n\theta^n}{\Gamma(n)} \Gamma(n-1) = \frac{n}{n-1} \theta$$

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

και συνεπώς η στατιστική συνάρτηση $U = \frac{n}{\sum_{i=1}^n x_i^a}$ δεν είναι α.ε. της θ .

γ) Θέτουμε $T^* = \frac{n-1}{\sum_{i=1}^n x_i^a}$, δηλαδή $T^* = \frac{n-1}{n}U$. Τότε $E(T^*) = \theta$, δηλαδή η T^* είναι α.ε.

της θ . Εξ άλλου $f(x, \theta) = a\theta x^{a-1} e^{-\theta x^a} 1(x) = C(\theta)e^{Q(\theta)T(x)}h(x)$, δηλαδή η δοθείσα κατανομή ανήκει στην εκθετική οικογένεια κατανομών και συνεπώς η στατιστική συνάρτηση $Z = \sum_{i=1}^n T(X_i) = \sum_{i=1}^n x_i^a$ είναι επαρκής και πλήρης για την θ . Άρα η αμερόληπτη T^* ως συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης Z είναι Α.Ε.Ε.Δ.

δ) Έχουμε:

$$\frac{\partial}{\partial \theta} \ln f(x, \theta) = \frac{\partial}{\partial \theta} (\ln a + \ln \theta - \theta x^a + (a-1) \ln x) = \frac{1}{\theta} - x^a.$$

Άρα

$$I(\theta) = E \left\{ \frac{\partial}{\partial \theta} \ln f(x, \theta) \right\}^2 = E \left\{ \frac{1}{\theta} - X^a \right\}^2 = E \left\{ \frac{1}{\theta^2} - \frac{2}{\theta} X^a + X^{2a} \right\} = \frac{1}{\theta^2} - \frac{2}{\theta} E \{ X^a \} + E \{ X^{2a} \}$$

Όπως είδαμε όμως στο β) η τ.μ $X^a \sim G(\theta, 1)$ και άρα $E \{ X^a \} = 1/\theta$, ενώ $E \{ X^{2a} \} = V(X^a) + [E \{ X^a \}]^2 = \frac{1}{\theta^2} + \frac{1}{\theta^2} = \frac{2}{\theta^2}$. Τελικά $I(\theta) = \frac{1}{\theta^2}$ και συνεπώς το κατώτατο φράγμα Cramer-Rao είναι $LB = \frac{1}{nI(\theta)} = \frac{\theta^2}{n}$.

Εξ άλλου:

$$V(T^*) = E[(T^*)^2] - [E(T^*)]^2. \text{ Βρήκαμε ήδη ότι } E(T^*) = \theta \text{ ενώ:}$$

$$E[(T^*)^2] = E \left[\frac{n-1}{Z} \right]^2 = \int_0^{+\infty} \left(\frac{n-1}{z} \right)^2 f_Z(z) dz = \int_0^{+\infty} \left(\frac{n-1}{z} \right)^2 \frac{\theta^n}{\Gamma(n)} z^{n-1} e^{-\theta z} dz =$$

$$\frac{(n-1)^2 \theta^2}{\Gamma(n)} \int_0^{+\infty} \omega^{(n-2)-1} e^{-\omega} d\omega = \frac{(n-1)}{(n-2)} \theta^2.$$

$$\text{Συνεπώς } V(T^*) = \frac{\theta^2}{(n-2)} > LB = \frac{\theta^2}{n}.$$

Άσκηση 11: Το πλάτος ενός παλμού είναι τ.μ. $X \sim N(\mu, 4)$. Στην έξοδο του μηχανήματος μπορούμε να παρατηρήσουμε μόνο αν το X υπερβαίνει την τιμή 40 ή όχι. Αν το X σε 100 παρατηρήσεις υπερέβη την τιμή αυτή 80 φορές, ποια η εκτίμηση με την μέθοδο μέγιστης πιθανοφάνειας της παραμέτρου μ ;

Λύση:

Έστω $Y = \begin{cases} 1, & \text{αν } X > 40 \\ 0, & \text{αν } X \leq 40 \end{cases}$. Σύμφωνα με την εκφώνηση μια μή παραμετρική εκτίμηση

για το $\hat{p} = \hat{P}(X > 40) = \frac{80}{100}$. Όμως :

$$p = P(X > 40) = P\left(\frac{X - \mu}{2} > \frac{40 - \mu}{2}\right) = P\left(Z > \frac{40 - \mu}{2}\right) = 1 - \Phi\left(\frac{40 - \mu}{2}\right) = \Phi\left(\frac{\mu - 40}{2}\right).$$

$$\text{Συνεπώς : } \hat{\Phi}\left(\frac{\mu - 40}{2}\right) = \hat{p} = 0.8 \stackrel{\Phi^{-1}}{\Rightarrow} \Phi\left(\frac{\hat{\mu} - 40}{2}\right) = 0.8 \stackrel{\Phi(0.85)=0.8}{\Rightarrow} \frac{\hat{\mu} - 40}{2} = 0.85 \Rightarrow \hat{\mu} = 41.7.$$

Άσκηση 12: Ένα ερώτημα σχετικό με την μελέτη του μηχανισμού ήχο-εντόπισης νυχτερίδων είναι η απόσταση νυχτερίδας και εντόμου όταν η νυχτερίδα πρώτο-αντιληφθεί το έντομο. Λόγω τεχνικών δυσκολιών στην μέτρηση της απόστασης, μόνο $n=11$ παρατηρήσεις λήφθηκαν, από τις οποίες έχουμε $\bar{X} = 48.36$ και $S^2 = 327.05$.

α) Τι υποθέσεις πρέπει να κάνουμε για να μπορέσουμε να κατασκευάσουμε ένα διάστημα εμπιστοσύνης (Δ.Ε.);

β) Ποιο το βασικό αποτέλεσμα που χρησιμοποιείται για την κατασκευή Δ.Ε. στην πάνω περίπτωση;

γ) Κατασκευάστε ένα 95% Δ.Ε. για το μ .

Λύση:

α) Πρέπει να υποθέσουμε ότι η μεταβλητή “απόσταση” έχει την κανονική κατανομή.

$$\beta) \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}.$$

$$\text{Απόδειξη: } \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{(\bar{X} - \mu)}{\frac{\sigma/\sqrt{n}}{\frac{s}{\sigma}}} = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} = Z \sim N(0, 1) \text{ και}$$

$$\frac{(n-1)s^2}{\sigma^2} = Y \sim X_{n-1}^2, \text{ με } Z, Y \text{ ανεξάρτητες τ.μ. Άρα } \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{Z}{\sqrt{\frac{Y}{n-1}}} \sim t_{n-1} \text{ εξ ορισμού.}$$

$$\gamma) \bar{x} \pm t_{10}(0.025) \frac{s}{\sqrt{11}} = \bar{x} \pm 2.228 \frac{s}{\sqrt{11}}.$$

Άσκηση 13: Οι ερευνητές της προηγούμενης άσκησης θέλουν με το πείραμα τους να στηρίξουν την υποψία τους ότι $\mu < 55$.

α) Διατυπώστε την μηδενική και την εναλλακτική υπόθεση.

β) Έστω ότι πιστεύουν ότι η πληθυσμιακή διασπορά είναι $\sigma^2 = 18^2 = 308$. Ποιο το έλεγχο-στατιστικό και ποια η p-τιμή αν $\bar{X} = 48.36$ και $n=11$;

γ) Απορρίπτεται η μηδενική υπόθεση σε επίπεδο σημαντικότητας $\alpha = 0.1$;

Λύση:

α) $H_0 : \mu=55 \quad H_1: \mu<55$

β) $Z = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} = -1.2$

p-τιμή = $\Phi(Z) = 0.1151$

γ) Όχι.

Άσκηση 14:

α) Δείξτε ότι ο πληθυσμιακός μέσος μ ικανοποιεί την σχέση:

$$E(X - \mu)^2 = \min_a E(X - a)^2 \quad (\text{εδώ } \mu = E(X)).$$

Δηλαδή αν προβλέψουμε την μεταβλητή X με το μ το μέσο τετραγωνικό σφάλμα είναι το μικρότερο δυνατό.

β) Έστω ότι ο μέσος μ εκτιμάται με τον δειγματικό μέσο τυχαίου δείγματος X_1, \dots, X_n , και έστω X_{n+1} μία άλλη μεταβλητή ισόνομη και ανεξάρτητη των X_1, \dots, X_n . Ποια η κατανομή

του $\frac{\bar{X} - X_{n+1}}{s\sqrt{1 + \frac{1}{n}}}$;

γ) Με χρήση του προηγούμενου αποτελέσματος κατασκευάστε ένα $(1-\alpha)100\%$ διάστημα εμπιστοσύνης για το X_{n+1} (λέγεται διάστημα πρόβλεψης).

Λύση:

α)

$$E(X - a)^2 = E[(X - \mu) + (\mu - a)]^2 =$$

$$E(X - \mu)^2 + 2(\mu - a)E(X - \mu) + (\mu - a)^2 = E(X - \mu)^2 + (\mu - a)^2 \geq E(X - \mu)^2.$$

β)

$$\frac{\bar{X} - X_{n+1}}{s\sqrt{1 + \frac{1}{n}}} = \frac{\frac{\bar{X} - X_{n+1}}{\sigma\sqrt{1 + \frac{1}{n}}}}{\frac{s\sqrt{1 + \frac{1}{n}}}{\sigma}} = \frac{\frac{\bar{X} - X_{n+1}}{\sigma\sqrt{1 + \frac{1}{n}}}}{\frac{s}{\sigma}}.$$

Δ. Φουσκάκης – Ασκήσεις Στατιστικής

Αλλά $\bar{X} - X_{n+1} \sim N\left(0, \sigma^2 + \frac{\sigma^2}{n}\right)$ και συνεπώς $\frac{\bar{X} - X_{n+1}}{\sigma\sqrt{1+\frac{1}{n}}} = Z \sim N(0,1)$. Επίσης

$\frac{(n-1)s^2}{\sigma^2} = Y \sim \chi_{n-1}^2$, με Z, Y ανεξάρτητες τ.μ. Άρα $\frac{\bar{X} - X_{n+1}}{s\sqrt{1+\frac{1}{n}}} = \frac{Z}{\sqrt{\frac{Y}{n-1}}} \sim t_{n-1}$.

γ) $\bar{X} \pm t_{n-1}(a/2)s\sqrt{1+\frac{1}{n}}$.