ΣΤΑΤΙΣΤΙΚΗ Ι, ΣΕΠΤΕΜΒΡΙΟΣ 2017

Θέμα 1: Οι διοργανωτές ενός τυχερού παιχνιδιού ισχυρίζονται ότι κάθε παίχτης κερδίζει με πιθανότητα 1/5. Η Ελληνική Στατιστική Υπηρεσία σας έχει αναθέσει το ρόλο να ελέγξετε αν οι διοργανωτές δεν εξαπατούν τους παίχτες. Έχετε στη διάθεσή σας ένα τυχαίο δείγμα που προέκυψε απο τα αποτελέσματα ν παιχτών.

α): Προτείνετε έναν κατάλληλο έλεγχο υποθέσεων για να επαληθεύσετε ότι οι διοργανωτές δεν εξαπατούν τους παίχτες και βρείτε τη μορφή της ομοιόμορφα ισχυρότατης κρίσιμης περιοχής σε ε.σ.σ. α.

β): Με τη βοήθεια του Κ.Ο.Θ. καθορίστε προσεγγιστικά την παραπάνω κρίσιμη περιοχή σε ε.σ.σ. $\alpha = 0.05$. Ποιό είναι το συμπέρασμα του ελέγχου αν σε δείγμα $\nu = 400$ προέκυψαν 60 νικητές; Δίνονται οι ακόλουθες τιμές της συνάρτησης κατανομής της τυπικής κανονικής: $\Phi(1.28) = 0.9$, $\Phi(1.65) = 0.95$, $\Phi(1.96) = 0.975$.

Θέμα 2: Έστω ένα τυχαίο δείγμα $y_1, y_2, ..., y_n$ από κατανομή με αθροιστική συνάρτηση

$$F(y;\theta) = \frac{y}{\theta}, \quad \theta > 0, \quad 0 < y < \theta$$

α): Εξετάστε αν η συγκεκριμένη κατανομή ανήκει στην Ε.Ο.Κ.

β): Να βρεθεί α.ε.ε.δ. της $g(\theta) = \theta^w$ (χωρίς να κάνετε χρήση ΕΜΠ).

γ): Έστω τώρα y μια και μόνο παρατήρηση από τη συγκεκριμένη κατανομή. Να βρεθεί 95% διάστημα εμπιστοσύνης για το θ. (Υπόδειξη: Να βρείτε τη κατανομή του $\frac{y}{\theta}$.)

Θέμα 3: Έστω $(X_1, Y_1), \ldots, (X_\nu, Y_\nu)$ ένα τυχαίο δείγμα μεγέθους ν από κατανομή που προκύπτει από ένα ζεύγος τ.μ. (X, Y) με $X \sim Exp(\theta)$, $\theta > 0$, και για $x > 0$, έχουμε ότι $[Y|X = x] \sim Poisson(\mu x)$, $\mu > 0$, και άρα έχει δεσμευμένη συνάρτηση πιθανότητας $P(Y = y|X = x) = e^{-\mu x}(\mu x)^y/y!$, $y = 0, 1, \ldots$ Θέλουμε να εκτιμήσουμε τις άγνωστες παραμέτρους θ, μ. Αν $f_X(x)$ είναι η συνάρτηση πυκνότητας της $Exp(\theta)$, τότε ορίζουμε ως πυκνότητα του (X, Y), $f_{X,Y}(x, y) := f_X(x)P(Y = y|X = x)$ με στήριγμα το $(0, \infty) \times \mathbb{N}$ και δεχθείτε ότι απο κοινού πυκνότητα ανεξάρτητων ζευγών υπολογίζεται με τον ίδιο τρόπο όπως η απο κοινού συνάρτηση πιθανότητας (για διακριτές) ή πυκνότητας πιθανότητας (για απόλυτα συνεχείς) ανεξάρτητων τυχαίων μεταβλητών.

α): Με βάση τα παραπάνω ορίστε κατάλληλα και υπολογίστε τη συνάρτηση πιθανοφάνειας $L(\theta,\mu)$ $L(\theta)$ που αντιστοιχεί στο δείγμα.

β): Βρείτε την εκτιμήτρια μέγιστης πιθανοφάνειας $(\widehat{\theta}, \widehat{\mu})$ του (θ, μ).

γ): Δείξτε ότι οι $\widehat{\theta}$ και $\widehat{\mu}$ είναι συνεπείς εκτιμήτριες των αντίστοιχων παραμέτρων.

Θέμα 4: Έστω $X_1, X_2, ..., X_n$ τυχαίο δείγμα από την κατανομή με σ.π.π.

$$f(x;\lambda) = \frac{1}{2\lambda^3} \exp\left\{-3x - \frac{e^{-x}}{\lambda}\right\}, \qquad x \in \mathcal{R}, \quad \lambda > 0.$$

α): Να δείξετε ότι ανήκει στην Ε.Ο.Κ. και να υπολογίσετε την $E(e^{-x})$ και $V(e^{-x})$.

β): Να βρεθεί επαρκής και πλήρης στατιστική συνάρτηση για το λ.

γ): Να βρεθεί εκτιμήτρια μέγιστης πιθανοφάνειας για το λ και να δειχθεί ότι είναι αμερόληπτη και επιτυγχάνει το κάτω φράγμα των Cramer-Rao.

<u>Να απαντήσετε σε 3 από τα 4 θέματα.</u>

ΚΑΛΗ ΕΠΙΤΥΧΙΑ
ΔΙΑΡΚΕΙΑ ΕΞΕΤΑΣΗΣ: 3 ΩΡΕΣ

Λύσεις Θεμάτων Σεπτεμβρίου 17

① α) Έστω $X_i = \begin{cases} 1 & , \text{ αν ο } i\text{-παίκτης κερδίζει} \\ 0 & , \text{ αν χάνει} \end{cases}$, $1 \le i \le \nu$.

Έχουμε $X_i \sim Be(p)$, όπου p η πιθανότητα κάθε παίκτη να κερδίζει, $0 < p < 1$.

Οι διοργανωτές ισχυρίζονται ότι $p = \frac{1}{5}$.

Οι παίκτες εξαπατώνται αν $p < \frac{1}{5}$, δηλ. αν έχουν μικρότερη πιθανότητα επιτυχίας. Ένας κατάλληλος έλεγχος υποθέσεων είναι H_0 : "$p = \frac{1}{5}$" έναντι H_1 : "$p < \frac{1}{5}$"

Με τη βοήθεια του κριτηρίου Neyman-Pearson (λόγο πιθανοφάνειας) καταλήγουμε σε κρίσιμη περιοχή $\left\{ \sum_{i=1}^{\nu} X_i \le c \right\}$ (ως ενδεχόμενο) το οποίο θα μπορούσε και κάποιος να το "μαντέψει" από την αρχή.

Τυπικά, γράφουμε $C_a = \left\{ x = (x_1, \dots, x_\nu) \in \mathbb{R}^\nu : \sum_{i=1}^{\nu} x_i \le c_a \right\}$ (ή γνήσια μικρότερο) για την ομοιόμορφα ισχυρότατη κ.π. μεγέθους a.

Το C_a επιλέγεται έτσι ώστε $P_{\frac{1}{5}}\Big(\underbrace{\sum_{i=1}^{\nu} X_i \le c_a}_{\substack{\text{περιοχή απόρριψης της } H_0 \\ (\text{ενδεχόμενο})}} \Big) = a$

β) Με τη βοήθεια του Κ.Ο.Θ. $[(X_n)$ α.ι.τ.μ. $\sim Be(p)]$

$$\sqrt{n}\left(\overline{X}_n - p\right) \xrightarrow[n\to\infty]{d} N\left(0,\ p(1-p)\right)$$

ή

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{d} N(0,1),$$

απ' όπου παίρνουμε τη γνωστή προσέγγιση για μεγάλο n,

$$\sum_{i=1}^{n} X_i \sim Bin(np,\ np(1-p)) \approx N\left(np,\ np(1-p)\right),$$

και άρα κάτω από την H_0: "$p = \frac{1}{5}$", έχουμε

$$\sum_{i=1}^{n} X_i \sim N\left(\frac{n}{5},\ \frac{4n}{25}\right), \text{ άρα}$$

$$P_{\frac{1}{5}}\left(\sum_{i=1}^{n} X_i \le c\right) = 0.05 \iff P_{\frac{1}{5}}\left(\frac{\sum_{i=1}^{n} X_i - \frac{n}{5}}{\sqrt{\frac{4n}{25}}} \le \frac{5}{2}\cdot\frac{c - \frac{n}{5}}{\sqrt{n}}\right) = 0.05$$

ασυμπτ. προσέγγ.

$$\iff \Phi\left(\frac{5}{2}\,\frac{c-\frac{n}{5}}{\sqrt{n}}\right) = 0.05 \overset{\phi(-x) = 1-\phi(x)}{\iff}$$

$$\phi\left(\frac{5}{2}\cdot\frac{\frac{n}{5} - c}{\sqrt{n}}\right) = 0.95 = \phi(1.65) \iff$$

$$\frac{5}{2}\,\frac{\frac{n}{5} - c}{\sqrt{n}} = 1.65 \iff c_{0.05} = \frac{n}{5} - 0.66\sqrt{n}$$

Άρα το ενδεχόμενο απόρριψης είναι $\left\{\sum_{i=1}^{n} X_i \le \frac{n}{5} - 0.66\sqrt{n}\right\}$.

- Για $n = 400$, έχουμε $c_{0.05} = 80 - 13.2 = \underline{66.8}$.

Όμως αν οι νικητές είναι 60, τότε $\sum_{i=1}^{400} X_i = 60 < 66.8$, και άρα απορρίπτεται η μηδενική υπόθεση δίκαιου παιχνιδιού, έναντι της εξαπάτησης των παικτών, σε ε.σ.σ. $a = 0.05$.

② α) Παρατηρούμε ότι το στήριγμα εξαρτάται από την άγνωστη παράμετρο θ, και άρα δεν αποτελεί Ε.Ο.Κ.

β) Η $F(y;\theta)$ είναι η συνάρτ. κατανομής της $\mathcal{U}(0,\theta)$ (ομοιόμορφης κατανομής), και η λύση έχει γίνει στις παραδόσεις των μαθημάτων. Αρκεί να παρατηρήσουμε ότι η $T = X_{(n)} = \max\{X_1, \ldots, X_n\}$ είναι επαρκής & πλήρης σ.σ. και με γνωστή μέθοδο καταλήξαμε ότι η $\boxed{\frac{n+w}{n} T^w = h(T)}$ είναι α.ε.ε.δ. του θ^w.

γ) $Y \sim \mathcal{U}(0,\theta) \Rightarrow \frac{Y}{\theta} \sim \mathcal{U}(0,1)$.

- Αναζητούμε c_1 και c_2 : $\boxed{P\left(c_1 < \frac{Y}{\theta} < c_2\right) = 0.95.}$

Όμως $P\left(c_1 < \frac{Y}{\theta} < c_2\right) \overset{\frac{Y}{\theta} \sim \mathcal{U}(0,1)}{=} c_2 - c_1$. για $(0 < c_1 < c_2 < 1)$

Άρα για οποιαδήποτε επιλογή c_1, c_2 : $c_2 - c_1 = 0.95$ με π.χ. $c_1 = 0.025$, $c_2 = 0.975$

έχουμε $c_1 < \frac{Y}{\theta} < c_2 \Leftrightarrow \frac{1}{c_2} < \frac{\theta}{Y} < \frac{1}{c_1} \Leftrightarrow$

$\theta \in \left(\frac{Y}{c_2}, \frac{Y}{c_1}\right)$ που είναι και το ζητούμενο

95% διάστημα εμπιστοσύνης.

③ α) Έχουμε $f_{X,Y}(x,y;\theta,\mu) = \theta e^{-\theta x} e^{-\mu x}(\mu x)^y \cdot \frac{1}{y!}$

$(x>0,\ y=0,1,2,\dots)$

Άρα

$$\boxed{\underset{\theta,\mu>0}{L(\theta,\mu)} = \prod_{i=1}^{\nu} f_{X_i,Y_i}(x_i,y_i;\theta,\mu) = \theta^{\nu} e^{-\theta\cdot\sum_{i=1}^{\nu} x_i} e^{-\mu\sum_{i=1}^{\nu}x_i} \mu^{\sum_{i=1}^{\nu} y_i} \prod_{i=1}^{\nu} x_i^{y_i} \prod_{i=1}^{\nu}\frac{1}{y_i!}}$$

β)

$$\Rightarrow \ell(\theta,\mu) = \log L(\theta,\mu) = \underbrace{\nu\log\theta - \theta\Big(\sum_{i=1}^{\nu} x_i\Big)}_{g(\theta)} + \underbrace{-\mu\Big(\sum_{i=1}^{\nu} x_i\Big) + \Big(\sum_{i=1}^{\nu} y_i\Big)\log\mu}_{h(\mu)} + \log\prod_i(\cdot)$$

Είναι φανερό ότι για να μεγιστοποιηθεί η $\ell(\theta,\mu)$, αρκεί να μεγιστοποιηθούν η $g(\theta)$ ως προς θ, και η $h(\mu)$ ως προς μ ανεξάρτητα.

$$\left.\begin{matrix} g'(\theta)=0 \\ h'(\mu)=0 \end{matrix}\right| \Longleftrightarrow \left.\begin{matrix} \frac{\nu}{\theta} - \sum_i x_i = 0 \\ -\sum_i x_i + \frac{\sum_i y_i}{\mu} = 0 \end{matrix}\right| \Longleftrightarrow \begin{matrix} \theta^* = \frac{\nu}{\sum_i x_i} \\ \mu^* = \frac{\sum_i y_i}{\sum_i x_i} \end{matrix}$$

Οι λύσεις αυτές αντιστοιχούν μάλιστα σε μέγιστα των $g(\theta)$ και $h(\mu)$ αντίστοιχα, αφού είναι μοναδικά στάσιμα σημεία και

$g''(\theta) = -\frac{\nu}{\theta^2} < 0,\ \forall\theta>0$, και

$h''(\mu) = -\frac{\sum_i y_i}{\mu^2} < 0,\ \forall\mu>0$ (εκτός εάν ⊛ $y_1 = y_2 = \dots = y_\nu = 0$ [με θετική πιθανότητα αυτό συμβαίνει])

Άρα $\hat\theta = \frac{\nu}{\sum_{i=1}^{\nu} X_i}$ και $\hat\mu = \frac{\sum_i Y_i}{\sum_i X_i}$

⊛. Σε αυτήν την περίπτωση για $\mu\in(0,+\infty)$ δεν υπάρχει η εμπ, αφού $h(\mu) = -\mu\big(\sum_i x_i\big) \searrow$ και χρειάζεται επέκταση του παραμετρικού χώρου, να συμπεριλαμβάνει το 0, ταυτίζοντας Poisson(0) την εκφυλ. τ.μ. που είναι μηδέν.

γ) • Απο τον Α.Ν.Μ.Α , έχουμε για την α.ι.τ.μ. (X_n) με $E|X_1| = E(X_1) = \frac{1}{\theta} < +\infty$, $\forall \theta > 0$, ότι

$$\overline{X}_n \xrightarrow[n\to\infty]{P} E(X_1) = \frac{1}{\theta} \xRightarrow[\substack{(\text{συνεχούς} \\ \text{απεικόνισης})}]{\text{Θ.Σ.Α.}} \hat{\theta} = \frac{1}{\overline{X}_n} \xrightarrow{P} \theta ,$$

για $g(x) = \frac{1}{x}$, $x > 0$.

και άρα είναι συνεπής εκτιμήτρια του θ.

• Για το μ έχουμε, $\hat{\mu} = \dfrac{\sum_{i=1}^{\nu} Y_i}{\sum_{i=1}^{\nu} X_i} = \dfrac{\overline{Y}_n}{\overline{X}_n}$.

<u>α' τρόπος</u> (πιο δύσκολος)

Αναζητούμε την κατανομή των Y_i , δηλ. της Y.

$$P(Y=y) = \int f_{X,Y}(x,y)\,dx = \int_0^{+\infty} \theta e^{-\theta x} e^{-\mu x} \frac{(\mu x)^y}{y!}\,dx$$

$$= \theta \cdot \frac{\mu^y}{y!} \int_0^{+\infty} e^{-(\theta+\mu)x} x^y\,dx \overset{\Gamma(y+1)=y!}{=} \frac{\theta}{(\theta+\mu)^{y+1}} \cdot \mu^y \underbrace{\int_0^{+\infty} \underbrace{\frac{(\theta+\mu)^{y+1}}{\Gamma(y+1)} x^{y+1-1} e^{-(\theta+\mu)x}}_{\text{σ.π.π. } G(y+1,\,\theta+\mu)}\,dx}_{=1}$$

$$= \frac{\theta}{\theta+\mu}\left(\frac{\mu}{\theta+\mu}\right)^y \quad , \quad y = 0, 1, 2, \ldots$$

$$\Rightarrow Y \sim Geo\left(\frac{\theta}{\theta+\mu}\right) \text{ , στο } \underline{\mathbb{N}} \quad \left[E(Geo(p)) = \frac{1-p}{p} \right]$$

$\Rightarrow$ απο Α.Ν.Μ.Α. ότι $\overline{Y}_n \xrightarrow{P} E(Y) = \dfrac{\frac{\mu}{\theta+\mu}}{\frac{\theta}{\theta+\mu}} = \dfrac{\mu}{\theta}$

για $(Y_n)_{n\geq 1}$

$$\Rightarrow \hat{\mu} = \frac{\overline{Y}_n}{\overline{X}_n} \xrightarrow{P} \frac{\frac{\mu}{\theta}}{\frac{1}{\theta}} = \mu \Rightarrow \text{συνεπής εκτιμ. του } \mu.$$

<u>β' τρόπος</u> (πιο απλός!) → <u>χωρίς εύρεση κατανομής!</u>

$$E(Y) = E[\underbrace{E(Y|X)}] = E[\mu X] = \mu E(X) = \frac{\mu}{\theta}$$

$[Y|X=x] \sim Poisson(\mu x)$, και συνεχίζουμε όπως πριν με Α.Ν.Μ.Α.

④ α) $f(x;\lambda) = \frac{1}{2\lambda^3} e^{-3x - \frac{e^{-x}}{\lambda}}$, $x \in \mathbb{R}$, $\lambda > 0$.

- έχουμε ότι το $S_f = \{x \in \mathbb{R} : f(x;\lambda) > 0\} = \mathbb{R}$ ανεξ. του λ.
- $f(x;\lambda) = \underbrace{\frac{1}{2\lambda^3}}_{= b(\lambda)} \underbrace{e^{-\frac{1}{\lambda} \cdot e^{-x}}}_{e^{\eta(\lambda) T(x)}} \cdot \underbrace{e^{-3x}}_{h(x)}$, άρα αποτελεί Ε.Ο.Κ.

Σε κανονική μορφή $A(\eta) = -\log(b(\lambda(\eta))) =$

$$-\log\left(\frac{1}{2\lambda^3}\right) \overset{\eta = -\frac{1}{\lambda}}{=} -\log\left(\frac{-1}{2}\cdot\eta^3\right) = \log 2 - 3\log(-\eta),\ \underline{\underline{\eta < 0}}$$

Άρα $E\left(e^{-X}\right) = E[T(X)] = A'(\eta) = -\frac{3}{\eta} = \boxed{3\lambda}$

$Var\left(e^{-X}\right) = Var[T(X)] = A''(\eta) = \frac{3}{\eta^2} = \boxed{3\lambda^2}$

β) Απο γνωστή Πρόταση για Ε.Ο.Κ. έχουμε ότι επαρκής και πλήρης σ.σ. για το λ, θα είναι $\boxed{T^* = \sum_{i=1}^{n} T(X_i) = \sum_{i=1}^{n} e^{-X_i}}$

γ) $L(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n} \frac{1}{2\lambda^3} e^{-\frac{1}{\lambda} e^{-x_i}} \cdot e^{-3x_i} =$

$(2\lambda^3)^{-n} \cdot e^{-\frac{1}{\lambda}\sum_i e^{-x_i}} \cdot e^{-3\sum_i x_i} \implies$

$\ell(\lambda) = -n\log(2\lambda^3) - \frac{1}{\lambda}\sum_i e^{-x_i} - 3\sum_i x_i$, $\lambda > 0$.

$\ell'(\lambda) = 0 \iff -\frac{3n}{\lambda} + \frac{1}{\lambda^2}\sum_i e^{-x_i} = 0 \iff \lambda^* = \frac{\sum_{i=1}^{n} e^{-x_i}}{3n}$.

Επιπλέον $\ell''(\lambda) = \frac{3n}{\lambda^2} - 2\cdot\frac{1}{\lambda^3}\sum_i e^{-x_i} = \frac{1}{\lambda^2}\left(3n - \frac{2}{\lambda}\sum_i e^{-x_i}\right)$

Έχουμε $\ell''(\lambda^*) = \frac{1}{(\lambda^*)^2}(3n - 6n) = -\frac{3n}{(\lambda^*)^2} < 0 \Rightarrow \boxed{\hat{\lambda} = \frac{\sum_{i=1}^{n} e^{-X_i}}{3n}}$

Έχουμε $E(\hat{\lambda}) = \frac{1}{3n} \sum_{i=1}^{n} E(e^{-X_i}) = \frac{\cancel{3n}\lambda}{\cancel{3n}} = \lambda, \ \forall \lambda > 0.$

$\Rightarrow$ $\underline{\hat{\lambda} \text{ α.ε. του } \lambda.}$

$$Var(\hat{\lambda}) = \frac{1}{9n^2} \sum_{i=1}^{n} Var(e^{-X_i}) = \frac{3\cancel{n}\lambda^2}{9n^{\cancel{2}}} = \boxed{\frac{\lambda^2}{3n}}$$

Όμως κ.φ.-C.R. $= \frac{1}{I(\lambda)}$, όπου

$$I(\lambda) = Var(\ell'(\lambda)) = Var\left(-\frac{3n}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^{n} e^{-X_i}\right)$$

$$= \frac{1}{\lambda^4} n \cdot (3\lambda^2) = \frac{3n}{\lambda^2} \Rightarrow \boxed{\text{κ.φ.-CR} = \frac{\lambda^2}{3n}}$$

Άρα η $\hat{\lambda}$ πράγματι επιτυγχάνει το κ.φ.-C.R.