Στοιχεία Θεωρίας Πιθανοφάνειας

Τα παρακάτω αφορούν σε ζητήματα θεωρίας πιθανοφάνειας στα πλαίσια της παρακάτω "περιορισμένης" εκδοχής του στατιστικού προβλήματος:

1. Μία "περιορισμένη" εκδοχή του στατιστικού προβλήματος

Έστω ότι ο χώρος πιθανότητας $(\mathbb{R}, \Sigma_{\mathbb{R}}, P_{\theta_0})$ "περιγράφει" κάποιο φαινόμενο που μας ενδιαφέρει και ότι η κατανομή P_{θ_0} είναι άγνωστη και χαρακτηρίζεται από την άγνωστη τιμή της παραμέτρου θ_0. Συνεπώς το αντικείμενο της στατιστικής επαγωγής είναι η εύρεση (ή ασθενέστερα η προσέγγιση) του θ_0, η οποία είναι απλώς ένα σημείο σε κάποιο Ευκλείδιο χώρο. □

Παράδειγμα. Έστω ότι $P_{\theta_0} = N(\mu_0, 1)$ όπου το $\mu_0 = \theta_0$ άγνωστο. △

2. Παραμετρικό Στατιστικό Υπόδειγμα

Η πληροφορία που έχουμε από το 1. συνεπάγεται την επιλογή συνόλου από κατανομές $\Upsilon_{\Theta} = \{P_\theta, \theta \in \Theta\}$ στο οποίο βρίσκεται η άγνωστη κατανομή και το οποίο ονομάζεται (παραμετρικό) στατιστικό υπόδειγμα. Κάθε κατανομή εκεί χαρακτηρίζεται από τιμή της παραμέτρου θ ή οποία βρίσκεται στο σύνολο Θ το οποίο είναι υποσύνολο κάποιου Ευκλείδιου χώρου, και ονομάζεται παραμετρικός χώρος του υποδείγματος, ενώ η διάσταση του Θ (έστω d) ονομάζεται διάσταση του στατιστικού υποδείγματος. Προφανώς $\theta_0 \in \Theta$. Σημειώνεται ότι βάσει του 1, και εφόσον α) $P_{\theta_0} \in \Upsilon_{\Theta}$, και β) το Υ_{Θ} αντανακλά όλη την πληροφορία που μας είναι δεδομένη στο 1, για την άγνωστη κατανομή, τότε (τουλάχιστον στις περιπτώσεις που θα μας απασχολήσουν) η επιλογή του στατιστικού υποδείγματος είναι μονοσήμαντη. Αυτό είναι δυνατόν να μην ισχύει αν αγνοηθούν οι προϋποθέσεις α) ή/και β) κάτι το οποίο είναι δυνατόν να έχει συνέπειες σε ιδιότητες διαδικασιών στατιστικής επαγωγής. □

Παράδειγμα (συν. από παραπάνω). Υποθέτουμε ότι η άγνωστη κατανομή είναι κανονική με μοναδιαία διακύμανση. Αυτό συνεπάγεται ότι προκειμένου να ισχύουν οι α) και β) παραπάνω $\mathcal{Y}_{\Theta} = \{N(\mu,1),\ \mu\in\Theta=\mathbb{R}\}$ (οπότε $d=1$). Π.χ. αν είχαμε επιλέξει $\mathcal{Y}_{\Theta'} = \{N(\mu,v),\ \mu\in\mathbb{R},\ v>0\}$ (οπότε $\Theta' = \{(x_1,x_2),\ x_1\in\mathbb{R},\ x_2>0\}$ και $d=2$) θα αγνοούσαμε την γνωστή σε εμάς πληροφορία ότι για την P_{θ_0}, $v=1$, ενώ αν π.χ. επιλέγαμε $\mathcal{Y}_{\Theta''} = \{N(\mu,1),\ \mu<0\}$ (οπότε $\Theta'' = (-\infty,0)$, $d=1$) διατρέχουμε τον κίνδυνο ότι $P_{\theta_0}\notin \mathcal{Y}_{\Theta''}$ κάτι το οποίο θα ισχύει ανν $\mu_0\geq 0$. Στην τελευταία περίπτωση και εφόσον $\mu_0>0$ το $\mathcal{Y}_{\Theta''}$ θα ονομάζεται "κακώς εξειδικευμένο" στατιστικό υπόδειγμα (επειδή $P_{\theta_0}\notin\mathcal{Y}_{\Theta''}$). Στα παρακάτω δεν θα ασχοληθούμε καθόλου με κακώς εξειδικευμένα υποδείγματα, κάτι το οποίο γενικά αιτιολογεί τον παραπάνω χαρακτηρισμό "περιορισμένη" εκδοχή του στατιστικού προβλήματος. □

3. Υποθέτουμε ότι οι κατανομές που βρίσκονται στο στατιστικό υπόδειγμα έχουν συναρτήσεις πυκνότητας (αναγκαστικά) της μορφής $f(x,\theta),\ \theta\in\Theta$. Όταν η τελευταία ειδωθεί και ως συνάρτηση του θ τότε αναπαριστά το στατιστικό υπόδειγμα. Υπό κάποιες προϋποθέσεις η $E(\ln(f(x,\theta)))$ (εφόσον η εν λόγω ολοκλήρωση γίνεται ως προς την άγνωστη κατανομή P_{θ_0}) μεγιστοποιείται ολικά και μοναδικά στο θ_0. Συνεπώς το στατιστικό πρόβλημα θα μπορούσε να επιλυθεί με ακρίβεια αν η ήταν γνωστή η παραπάνω συνάρτηση και ευχερής η μεγιστοποίηση της. Η τελευταία εξαρτάται όμως συνήθως και από το θ_0 οπότε δεν μπορεί να χρησιμοποιηθεί για την επίλυση του στατιστικού προβλήματος. □

Παράδειγμα (συν. από παραπάνω). Έχουμε ότι επειδή $\mathcal{Y}_{\Theta} = \{N(\mu,1),\ \mu\in\mathbb{R}\}$, $f(x,\mu) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-\mu)^2\right)$, $f(x,\mu)>0\ \forall x,\mu\in\mathbb{R}$,

$$\ln f(x,\varphi) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}(x-\varphi)^2 = -\frac{1}{2}\ln 2\pi - \frac{1}{2}(x^2 - 2x\varphi + \varphi^2).$$

Αν $X \sim P_{\theta_0} = N(\varphi_0, 1)$, τότε $E(\ln f(X,\varphi)) = E\left[-\frac{1}{2}\ln 2\pi - \frac{1}{2}(X^2 - 2X\varphi + \varphi^2)\right] = -\frac{1}{2}\ln 2\pi - \frac{1}{2}E(X^2 - 2X\varphi + \varphi^2) \overset{(\text{γιατί;})}{=} -\frac{1}{2}\ln 2\pi - \frac{1}{2}[E(X^2) - 2\varphi E(X) + \varphi^2] = -\frac{1}{2}\ln 2\pi - \frac{1}{2}(1 + \varphi_0^2 - \varphi\varphi_0 + \varphi^2)$. Μεγιστοποιούμε την $E(\ln(f(X,\varphi)))$ ως προς φ (χρησιμοποιώντας συνθήκες 1ης και 2ης τάξης - γιατί;) οπότε έχουμε:

1ης - $\frac{dE(\ln(f(X,\varphi)))}{d\varphi} = 0 \Leftrightarrow d\left[-\frac{1}{2}\ln 2\pi - \frac{1}{2}(1 + \varphi_0^2 - 2\varphi\varphi_0 + \varphi^2)\right]/d\varphi = 0 \Rightarrow -\frac{1}{2}[-2\varphi_0 + 2\varphi] = 0 \Leftrightarrow \varphi = \varphi_0.$

2ης $\left.\frac{d^2E(\ln f(X,\varphi))}{d\varphi^2}\right|_{\varphi=\varphi_0} = -1\big|_{\varphi=\varphi_0} = -1 < 0.$

Επομένως το φ_0 αποτελεί το μοναδικό μεγιστοποιούν σημείο της $-\frac{1}{2}\ln 2\pi - \frac{1}{2}(1 + \varphi_0^2 - 2\varphi\varphi_0 + \varphi^2)$ η οποία όμως μας είναι εν μέρει άγνωστη αφού εξαρτάται από το άγνωστο φ_0. □

4. Αρχή της Αναλογίας

Η αρχή της αναλογίας υποδεικνύει ότι μπορούμε να αποκτήσουμε προσέγγιση του αγνώστου θ_0 μέσω της μεγιστοποιήσης κάποιας δειγματικής προσέγγισης της $E(\ln(f(x,\theta)))$. Επομένως για να προχωρήσουμε απαιτείται περαιτέρω πληροφορία για την κατασκευή αυτής της προσέγγισης. Αυτή δίνεται από την ύπαρξη του δείγματος. □

5. Στατιστικό Υπόδειγμα, Δείγμα, και Συνάρτηση Πιθανοφάνειας

Έχουμε στην διάθεση μας n iid τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$, κάθε μία από τις οποίες ακολουθεί την P_{θ_0}. Από αυτές κατασκευάζουμε τις τρεις εκδοχές της συνάρτησης πιθανοφάνειας, ως εξής:

I. (συνάρτηση πιθανοφάνειας), $L_n(\theta) := \prod_{i=1}^{n} f(X_i,\theta)$

II. (λογαριθμική συνάρτηση πιθανοφάνειας), $\ell_n(\theta) := \ln[L_n(\theta)] = \ln \prod_{i=1}^{n} f(X_i,\theta) = \sum_{i=1}^{n} \ln f(X_i,\theta)$ (εφόσον ο λογάριθμος είναι καλώς ορισμένος).

III. (μέση λογαριθμική συνάρτηση πιθανοφάνειας), $\bar{\ell}_n(\theta) := \frac{1}{n}\ell_n(\theta)$

$$= \frac{1}{n}\sum_{i=1}^{n} \ln(f(X_i,\theta)).$$

Παρατηρούμε ότι η $L_n(\theta)$ θα μπορούσε να ερμηνευθεί ως η συνάρτηση πυκνότητας (ή ακριβέστερα το δεξιό της μέρος) της από κοινού κατανομής του $(X_1, X_2, \ldots, X_n)$ αν $X_i \sim P_\theta$ $\forall i=1,\ldots,n$ στα πλαίσια της iid υπόθεσης αν αυτή υπολογισθεί στο $(X_1, X_2, \ldots, X_n)$.

Έτσι η εν λόγω συνάρτηση (και συνεπώς και οι άλλοι δύο μετασχηματισμοί της) αναπαριστά ταυτόγχρονα και την πληροφορία που έχουμε για την άγνωστη κατανομή από το στατιστικό μας υπόδειγμα, αλλά και αυτή που έχουμε για την άγνωστη κατανομή από το δείγμα. Η μέση λογαριθμική συνιστά υπό προϋποθέσεις την δειγματική προσέγγιση της $E(\ln f(x,\theta))$ παραπάνω.

Συνεπώς μπορούμε να περιμένουμε ότι η μεγιστοποίηση αυτής ως προς $\theta \in \Theta$, θα μας δώσει "δειγματική" προσέγγιση του θ_0 και συνεπώς μια διαδικασία στατιστικής εκτιμητικής. □

Παράδειγμα (συν. από παραπάνω) Έχουμε ότι $X_1, X_2, \dots, X_n \sim N(\varphi_0, 1)$ και βάσει των ορισμών:

$$\lambda_n(\varphi) = \prod_{i=1}^{n} f(X_i, \varphi) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(X_i - \varphi)^2\right)$$

$$= \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i-\varphi)^2\right)$$

$$= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i-\varphi)^2\right).$$

$$l_n(\varphi) = \ln(\lambda_n(\varphi)) = \ln\left[(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i-\varphi)^2\right)\right]$$

$$= -\frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{n}(X_i-\varphi)^2.$$

$$\bar{l}_n(\varphi) = \frac{1}{2} l_n(\varphi) = -\frac{1}{2}\ln 2\pi - \frac{1}{2n}\sum_{i=1}^{n}(X_i-\varphi)^2.$$

6. Εκτιμητής Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimator - MLE)

Ο εκτιμητής μέγιστης πιθανοφάνειας για το θ_0 $(\text{MLE}(\theta_0) = \hat{\theta}_n)$ ορίζεται ως το μεγιστοποιούν σημείο της παραπάνω συνάρτησης (η μεγιστοποίηση γίνεται δεδομένου του δείγματος ως προς το $\theta \in \Theta$) εφόσον αυτό υπάρχει και είναι μοναδικό. Εκ κατασκευής είναι μια συνάρτηση που δέχεται το δείγμα και αποδίδει μια τιμή στον παραμετρικό χώρο, δηλαδή είναι τυχαία μεταβλητή ή τυχαίο διάνυσμα (d=1 ή d>1 αντίστοιχα). Οι ιδιότητες του προσδιορίζονται από τι ς ιδιότητες της κατανομής που ακολουθεί και η τελευταία γενικά εξαρτάται από την άγνωστη P_{θ_0}.

Παρατηρώ ακόμα για το 6:

α. Ο εκτιμητής ορίζεται ως μεγιστοποιούν σημείο της $\bar{\ell}_n(\theta)$. Επειδή κάθε μορφή της πιθανοφάνειας (δηλ. οι L_n, ℓ_n, $\bar{\ell}_n$) είναι η μία γνησίως μονότονος μετασχηματισμός της άλλης (δείξτε το!), θα έχουμε ότι

$$MLE(\theta_0) := \hat{\theta}_n = \arg\max_{\theta\in\Theta} \bar{\ell}_n(\theta) = \arg\max_{\theta\in\Theta} \ell_n(\theta) = \arg\max_{\theta\in\Theta} L_n(\theta).$$

(argmax - argument that maximizes). Συνεπώς ο εκτιμητής ισοδύναμα ορίζεται ως το μεγιστοποιούν σημείο οποιασδήποτε εκδοχής της συνάρτησης πιθανοφάνειας.

β. Γενικότερα ο $MLE(\theta_0)$ ορίζεται ως το με πιθανότητα 1 μεγιστοποιούν σημείο της $\bar{\ell}_n(\theta)$ (ή της $\ell_n(\theta)$, ή της $L_n(\theta)$, ισοδύναμα), καθώς είναι δυνατόν να αγνοηθεί το ενδεχόμενο να μην υπάρχει το argmax εφόσον αυτό είναι P_{θ_0}-αμελητέο.

γ. Στα παραδείγματα που θα εξετάσουμε ο $MLE(\theta_0)$ θα ικανοποιεί συνθήκες $1^{ης}$ και $2^{ης}$ τάξης και θα είναι αναλυτικά εξαγώγιμη (κατάλληλα μετρήσιμη) συνάρτηση του $(X_1, X_2, \dots, X_n)$. Η εξαγωγή του μέσω των παραγώγων της πιθανοφάνειας ή/και η αναλυτική εξαγωγιμότητά του είναι δυνατόν να μην ισχύουν σε ελάχιστα πιο περίπλοκα παραδείγματα.

δ. Στα παρακάτω θεωρείται δεδομένη η γνώση εννοιών που αφορούν ιδιότητες εκτιμητών, όπως π.χ. η αμεροληψία, κ.ο.κ. ∎

Παράδειγμα (συν. από παραπάνω). Έχουμε ότι $\hat{\mu}_n := MLE(\mu_0) = \arg\max_{\mu\in\mathbb{R}} \bar{\ell}_n(\mu)$. Οπότε:

$1^{ης}$ τάξης $$\frac{d\bar{\ell}_n(\mu)}{d\mu} = 0 \Rightarrow d\left[-\tfrac{1}{2}\ln 2\pi - \tfrac{1}{2n}\sum_{i=1}^{n}(X_i-\mu)^2\right]\Big/ d\mu = 0$$

(γιατί;)
$$\Leftrightarrow -\frac{1}{2n}\sum_{i=1}^{n} d(X_i-\mu)^2 = 0 \Leftrightarrow -\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)(-1) = 0$$

κρίσιμο σημείο

$\Leftrightarrow \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)=0 \Leftrightarrow \frac{1}{n}\sum_{i=1}^{n}X_i=\frac{1}{n}\sum_{i=1}^{n}\mu \Leftrightarrow \mu_n=\frac{1}{n}\sum_{i=1}^{n}X_i.$

2ης τάξης. $\left.\frac{d^2\bar{l}_n(\mu)}{d\mu^2}\right|_{\mu=\mu_n} = \left.-\frac{1}{n}\sum_{i=1}^{n}1\right|_{\mu=\mu_n} = \left.-1\right|_{\mu=\mu_n} = -1<0,$

Επομένως (γιατί;) $MLE(\mu_0)=\mu_n=\frac{1}{n}\sum_{i=1}^{n}X_i.$

Κάποιες ιδιότητες του $MLE(\mu_0)$ έχουν ως εξής:

(οι παρακάτω σχολιασμοί δίνονται ως προς την $P_{\theta_0}=N(\mu_0,1)$)

$-\ E(\mu_n) = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) \overset{(\text{γιατί;})}{=} \frac{1}{n}\sum_{i=1}^{n}E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu_0 = \mu_0$

επομένως ο εκτιμητής είναι αμερόληπτος.

$-\ Var(\mu_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) \overset{(\text{γιατί;})}{=} \frac{1}{n^2}Var\left(\sum_{i=1}^{n}X_i\right) \overset{\text{ανεξ.}}{=} \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)$

$= \frac{1}{n^2}\sum_{i=1}^{n}1 = \frac{n}{n^2} = \frac{1}{n}.$

- Από τα παραπάνω και ιδιότητα της κανονικής κατανομής έχουμε ότι $\mu_n \sim N(\mu_0, 1/n)$.

- Ένας εκτιμητής του θ_0, έστω ο θ_n, θα αναφέρεται ασθενώς συνεπής (weakly consistent) ανν, $\forall \varepsilon>0$

$$\lim_{n\to+\infty} P(|\theta_n-\theta_0|>\varepsilon)=0.$$

η παραπάνω μορφή "στοχαστικής σύγκλισης" που αναφέρεται σύγκλιση σε πιθανότητα (convergence in probability) ερμηνεύεται ως η πιθανότητα η απόσταση του θ_n από το θ_0 να είναι μεγαλύτερη του ε, συγκλίνει στο 0 καθώς $n\to+\infty$ (οπότε και αποκτούμε "όλο και περισσότερη πληροφορία" για τον εντοπισμό του θ_0)

$\forall \varepsilon > 0$ και συνεπώς καταδεικνύει κάποιας μορφής σύγκλιση του θ_n στο θ_0, καθώς $n \to +\infty$. Στην συγκεκριμένη περίπτωση, για αυθαίρετο $\varepsilon > 0$ και εφαρμογής της ανισότητας του Chebychev, έχουμε

$$P(|\varphi_n - \varphi_0| > \varepsilon) \leq \frac{E(|\varphi_n - \varphi_0|^2)}{\varepsilon^2}$$

$$= \frac{Var(\varphi_n)}{\varepsilon^2} = \frac{1}{n\varepsilon^2} \to 0 \text{ καθώς } n \to +\infty.$$

Επομένως (γιατί;) $\forall \varepsilon > 0$, $\lim_{n \to +\infty} P(|\varphi_n - \varphi_0| > \varepsilon) = 0$ και συνεπώς

ο $MLE(\varphi_0)$ είναι ασθενώς συνεπής. $\square$

Παράδειγμα 2. Έστω $X_1, X_2, \ldots, X_n$ iid τυχαίες μεταβλητές με $X_1 \sim N(0, v_0)$ όπου $v_0 > 0$ άγνωστη. Να βρεθεί ο $MLE(v_0)$.

Γνωρίζουμε ότι η P_{θ_0} είναι κανονική με μέσο 0. Επομένως $\mathcal{Y}_\Theta = \{N(0,v), v \in \Theta = (0, +\infty)\}$ (και συνεπώς $d = 1$).

Έχουμε ότι $f(x,v) = \frac{1}{\sqrt{2\pi v}} \exp(-\frac{1}{2v} x^2)$ και συνεπώς:

$$L_n(v) := \prod_{i=1}^n f(X_i, v) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi v}} \exp(-\frac{1}{2v} X_i^2)$$

$$= (2\pi v)^{-n/2} \exp\left(-\frac{1}{2v} \sum_{i=1}^n X_i^2\right).$$

$$\ell_n(v) = \ln L_n(v) = \ln\left[(2\pi v)^{-n/2} \exp\left(-\frac{1}{2v} \sum_{i=1}^n X_i\right)\right]$$

$$= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln v - \frac{1}{2v} \sum_{i=1}^n X_i^2.$$

$$\bar{\ell}_n(v) = \frac{1}{n} \ell_n(v) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln v - \frac{1}{2vn} \sum_{i=1}^n X_i^2.$$

Επομένως $MLE(v_0) := v_n = \underset{v>0}{\arg\max}\ \bar{\ell}_n(v)$ και:

*(Παρόλο που η παραπάνω βελτιστοποίηση είναι περιορισμένη θα αποφύγουμε την χρήση ανάλογων συνθηκών τύπου Kuhn–Tucker και θα την αντιμετωπίσουμε ως απεριόριστη, κάτι το οποίο θα μας οδηγήσει στο ορθό αποτέλεσμα εξαιτίας ιδιοτήτων της $\bar{\ell}_n$ και του Θ.)

$1^{ης}$ τάξης: $\frac{d\bar{\ell}_n(v)}{dv} = 0 \Rightarrow -\frac{1}{2v} + \frac{1}{2nv^2}\sum_{i=1}^{n} X_i^2 = 0 \overset{v>0}{\iff} -\frac{v}{2} + \frac{v}{2}\frac{1}{n}\sum_{i=1}^{n} X_i^2$

$= 0 \iff \underbrace{v_n}_{\text{κρίσιμο σημείο}} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 > 0$ με πιθ. 1.

$2^{ης}$ τάξης: $\left.\frac{d^2\bar{\ell}_n(v)}{dv^2}\right|_{v=v_n} = \left.\frac{d\left(-\frac{1}{2v} + \frac{1}{2v^2}v_n\right)}{dv}\right|_{v=v_n} = \left.\left(\frac{1}{2v^2} - \frac{2}{2v^3}v_n\right)\right|_{v=v_n}$

$= \frac{1}{2v_n^2} - \frac{v_n}{v_n^3} = -\frac{1}{2v_n^2} < 0$ με πιθ. 1.

Συνεπώς (βάσει του σχολίου β) $MLE(v_0) = v_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.
(Στα παραπάνω η ολοκλήρωση γίνεται ως προς την $P_{\theta_0} = N(0, v_0)$)

– Έχουμε ότι $E(v_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) \underset{(\text{γιατί;})}{=} \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) =$

$\underset{(\text{γιατί;})}{=} \frac{1}{n}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n}\sum_{i=1}^{n} v_0 = \frac{n}{n}v_0 = v_0$, και συνεπώς έχουμε

και σε αυτή την περίπτωση αμεροληψία. (Δεν θα ασχοληθούμε με περαιτέρω ιδιότητες, π.χ. είναι σχετικά εύκολο και σε αυτή την περίπτωση να δείξουμε ότι ισχύει η ασθενής συνέπεια). $\square$

Παράδειγμα 3. Έστω $X_1, X_2, \ldots, X_n$ iid τυχαίες μεταβλητές με $X_1 \sim Exp(\lambda_0)$, όπου $\lambda_0 > 0$ άγνωστο. Να βρεθεί ο $MLE(\lambda_0)$.

Γνωρίζουμε ότι η P_{θ_0} είναι κάποια συνεχική κατανομή. Επομένως $\mathcal{Y}_\Theta = \{ Exp(\lambda), \lambda \in \Theta = (0, +\infty) \}$ (οπότε $d=1$). Έχουμε ότι

$$f(x,\lambda) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}.$$ Οι συναρτήσεις πυκνότητας

στο συγκεκριμένο υπόδειγμα έχουν και μηδενικό μέρος. Επειδή $P_{\theta_0}(X_i = 0) = 0 \;\; \forall i = 1, \ldots, n$ (γιατί;) σε τέτοιες περιπτώσεις για την κατασκευή της πιθανοφάνειας χρησιμοποιούμε το αυστηρά θετικό μέρος (**). Έχουμε λοιπόν ότι:

$$L_n(\lambda) = \prod_{i=1}^n f(X_i, \lambda) 1_{X_i \geq 0} = \prod_{i=1}^n \lambda \exp(-\lambda X_i) =$$

$$= \lambda^n \exp\left(-\lambda \sum_{i=1}^n X_i\right).$$

$$\ell_n(\lambda) = \ln[L_n(\lambda)]$$

$$= \ln\left[\lambda^n \exp\left(-\lambda \sum_{i=1}^n X_i\right)\right]$$

$$= n \ln\lambda - \lambda \sum_{i=1}^n X_i.$$

$$\bar{\ell}_n(\lambda) = \frac{1}{n} \ell_n(\lambda) = \ln\lambda - \lambda \frac{1}{n} \sum_{i=1}^n X_i$$

$$= \ln\lambda - \lambda \mu_n.$$

Επομένως $MLE(\lambda_0) = \underset{\lambda > 0}{\operatorname{argmax}}\, \bar{\ell}_n(\lambda)$, και:

(και σε αυτό το παράδειγμα ισχύει το (*) από προηγουμένως)

$1^{ης}$ τάξης: $\dfrac{d\bar{\ell}_n(\lambda)}{d\lambda} = 0 \Rightarrow \dfrac{1}{\lambda} - \psi_n = 0 \Leftrightarrow$

$P_{\theta_0}(\psi_n > 0) = 1$
(γιατί;)

$$\lambda_n = \psi_n^{-1} = \frac{n}{\sum_{i=1}^{n} X_i} \text{ με πιθ. } 1.$$

κρίσιμο σημείο

$2^{ης}$ τάξης: $\left.\dfrac{d^2\bar{\ell}_n(\lambda)}{d\lambda^2}\right|_{\lambda=\lambda_n} = \left.\dfrac{d\left(\frac{1}{\lambda} - \psi_n\right)}{d\lambda}\right|_{\lambda=\lambda_n} = \left.-\dfrac{1}{\lambda^2}\right|_{\lambda=\lambda_n}$

$$= -\frac{1}{\lambda_n^2} < 0 \text{ με πιθ. } 1.$$

Συνεπώς (και βάσει του παραπάνω σχολίου b) $MLE(\lambda_0) = $
$= \lambda_n = \psi_n^{-1} = \dfrac{n}{\sum_{i=1}^{n} X_i}$.

- Είναι δυνατόν να δειχθεί ότι $E(\lambda_n) = E\left(\dfrac{n}{\sum_{i=1}^{n} X_i}\right)$

$$= nE\left(\frac{1}{\sum_{i=1}^{n} X_i}\right) \neq \frac{n}{E\left(\sum_{i=1}^{n} X_i\right)}$$ (η ολοκλήρωση γίνεται

ως προς την $P_{\theta_0} = \mathcal{E}xp(\lambda_0)$) $\underset{\text{(γιατί;)}}{=} \dfrac{n}{\sum_{i=1}^{n} E(X_i)} = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{\lambda_0}}$

$= \dfrac{n}{n}\dfrac{1}{1/\lambda_0} = \lambda_0$. Επομένως ο $MLE(\lambda_0)$ είναι

μεροληπτικός (γενικά και παρά τα προηγούμενα παραδείγματα, η αμεροληψία είναι μια "σχετικά σπάνια" ιδιότητα για διάφορους λόγους), ενώ δεν είναι δύσκολο να δείξουμε ότι και αυτός ο εκτιμητής είναι ασθενώς συνεπής. □

Οι σημειώσεις βρίσκονται σε στάδιο διαρκούς διόρθωσης και δεν υποκαθιστούν τις διαλέξεις. Παρακαλώ αναφέρτε όποιο λάθος στο stelios@aueb.gr ή στο e-class του μαθήματος.