

4. Εκτιμητική

Σύνδεση θεωρίας πιθανοτήτων - περιγραφικής στατιστικής.

Η περιγραφική στατιστική (ΣΤΑΤΙΣΤΙΚΗ Ι) αφορά κυρίως τη μελέτη κάποιων «μεγεθών» (π.χ. μέση τιμή, διασπορά, διάμεσος, κ.ο.κ.) ενός «δείγματος» n συγκεκριμένων τιμών $x_1, x_2, \dots, x_n$ που προέρχονται από έναν πληθυσμό. Στις περισσότερες περιπτώσεις, οι τιμές αυτές μπορεί να θεωρηθεί ότι προέρχονται από n τυχαία πειράματα (π.χ. τυχαίες επιλογές n ατόμων από τον πληθυσμό). Στα πλαίσια της θεωρίας των πιθανοτήτων (ΣΤΑΤΙΣΤΙΚΗ ΙΙ, ΙΙΙ) τέτοιες τιμές συμβολίζονται με $X_1, X_2, \dots, X_n$. Ποια η σχέση μεταξύ των τ.μ. $X_1, X_2, \dots, X_n$ και των τιμών $x_1, x_2, \dots, x_n$;

Η X_i μπορεί να θεωρηθεί ως μία τυχαία μεταβλητή που εκφράζει το αποτέλεσμα του i -τυχαίου πειράματος, πριν από την εκτέλεσή του (ή ισοδύναμα, πριν μας γίνει γνωστό το αποτέλεσμα του) ενώ αντίθετα, η τιμή x_i εκφράζει το ακριβές αποτέλεσμα του πειράματος μετά την εκτέλεσή του. Η τιμή x_i καλείται και «πραγματοποίηση» του X_i . Στην ουσία, η x_i είναι ίση με το $X_i(\omega)$ όπου ω είναι το στοιχειώδες ενδεχόμενο του Ω που τελικά πραγματοποιήθηκε (μην ξεχνάμε ότι μία τ.μ. είναι μία απεικόνιση από το Ω στο $\mathbb{R}$). Στο εξής, για απλότητα θα γράφουμε $X_1, X_2, \dots, X_n$ εννοώντας τις τ.μ. (όταν είμαστε πριν την πραγματοποίηση του πειράματος) καθώς και τις τιμές $x_1, x_2, \dots, x_n$ (όταν είμαστε μετά την πραγματοποίηση του πειράματος).

Συνεπώς, η περιγραφική στατιστική μπορεί να θεωρηθεί ότι αφορά τη μελέτη των μεγεθών ενός «δείγματος» που προέρχεται από έναν (θεωρητικά άπειρο) πληθυσμό. Συνήθως όμως αυτό που μας ενδιαφέρει περισσότερο είναι τα αντίστοιχα μεγέθη του **πληθυσμού** από τον οποίο προέρχεται το δείγμα. Μία από τις κύριες επιδιώξεις της Στατιστικής συμπερασματολογίας (που αποτελεί και το μεγαλύτερο μέρος της Στατιστικής ΙΙΙ) είναι η εκτίμηση των μεγεθών ενός πληθυσμού (π.χ. μέση τιμή, διασπορά, κατανομή) με βάση ένα δείγμα $X_1, X_2, \dots, X_n$ από αυτόν.

Εκτιμητική

Κύρια επιδίωξη της εκτιμητικής αποτελεί η εκτίμηση των παραμέτρων της κατανομής κάποιου χαρακτηριστικού ενός πληθυσμού. Το πρόβλημα έχει ως εξής: υποθέτουμε ότι κάποιο συγκεκριμένο χαρακτηριστικό ενός πληθυσμού (π.χ. το ύψος των ανδρών σε μία πόλη) ακολουθεί κάποια γνωστή κατανομή $F(x; \theta)$ (π.χ. κανονική κατανομή) με άγνωστες όμως παραμέτρους $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ (π.χ. $\theta = (\mu, \sigma^2)$). Με ποιο τρόπο θα μπορούσαμε να προσδιορίσουμε (εκτιμήσουμε) τις παραμέτρους αυτές; (π.χ. να δούμε ποίο είναι το μέσο ύψος μ του πληθυσμού και ποιά η διασπορά σ^2 των υψών).

Είναι φανερό ότι, για την εκτίμηση των παραμέτρων θ της κατανομής κάποιου χαρακτηριστικού, είναι αναγκαία η γνώση της τιμής του χαρακτηριστικού τουλάχιστον σε κάποιες μονάδες του πληθυσμού. Δηλαδή, είναι απαραίτητη η εκλογή ενός «δείγματος» του πληθυσμού (ή ενός συνόλου παρατηρήσεων). Για να είναι όμως σωστή η εκτίμηση του χαρακτηριστικού θα πρέπει η εκλογή αυτή των μονάδων του δείγματος να είναι «τυχαία» από όλον τον πληθυσμό, δηλαδή κάθε μονάδα του πληθυσμού να έχει την ίδια πιθανότητα να εκλεγεί στο δείγμα. Για παράδειγμα, για την εκτίμηση του μέσου ύψους μ χρειαζόμαστε τα ύψη ενός δείγματος ανδρών που εκλέξαμε τυχαία από τον πληθυσμό. (π.χ. αν εσκεμμένα πάρουμε σαν δείγμα τους ψηλότερους άνδρες μίας περιοχής τότε προφανώς θα υπερεκτιμήσουμε το μέσο ύψος). Στην περίπτωση μίας τυχαίας εκλογής ενός δείγματος, οι τιμές του προς εξέταση χαρακτηριστικού στο δείγμα (παρατηρήσεις) μπορούν να θεωρηθούν ότι είναι τυχαίες μεταβλητές που ακολουθούν την κατανομή του χαρακτηριστικού στον πληθυσμό. Φτάνουμε λοιπόν φυσιολογικά στον επόμενο ορισμό.

Ορισμός 4.1. i) *Τυχαίο δείγμα* μεγέθους n από την κατανομή $F(x; \theta)$ (ή τη σ.π. $f(x; \theta)$) θα καλείται ένα σύνολο ανεξάρτητων και ισόνομων τ.μ. $X_1, X_2, \dots, X_n$ που ακολουθούν την κατανομή $F(x; \theta)$ (ή τη σ.π. $f(x; \theta)$).

ii) *Δειγματοληπτικός χώρος* θα καλείται το σύνολο των δυνατών τιμών του δείγματος (π.χ. αν $X_i \in \mathbf{R}$, τότε ο δειγματοληπτικός χώρος θα είναι ο $\mathbf{R}^n$).

iii) *Παραμετρικός χώρος* θα καλείται το σύνολο των επιτρεπτών τιμών της παραμέτρου θ .

Έστω λοιπόν $X_1, X_2, \dots, X_n$ ένα τυχαίο δείγμα από την $F(x; \theta)$. Η εκτίμηση των παραμέτρων $\theta = (\theta_1, \dots, \theta_k)$ γίνεται μέσω κάποιων συναρτήσεων των $X_1, X_2, \dots, X_n$. Πιο συγκεκριμένα θα έχουμε ότι:

iv) *Στατιστική (ή δειγματική) συνάρτηση* θα λέγεται κάθε συνάρτηση $T(\mathbf{X}) = T(X_1, X_2, \dots, X_n)$ των τ.μ. του δείγματος $X_1, X_2, \dots, X_n$ που δεν εξαρτάται από τις προς εκτίμηση παραμέτρους. Προφανώς, κάθε στατιστική συνάρτηση είναι και αυτή μία τ.μ. Για παράδειγμα, γνωστές στατιστικές συναρτήσεις είναι:

$$- \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (\text{δειγματικός μέσος})$$

$$- m_r = \frac{1}{n} \sum_{i=1}^n X_i^r \quad (\text{δειγματικές ροπές τάξεως } r)$$

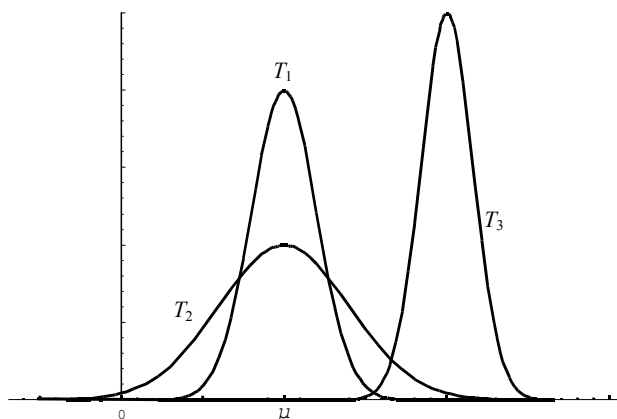
$$- S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{δειγματική διασπορά})$$

$$- R = \max\{X_1, \dots, X_n\} - \min\{X_1, \dots, X_n\} = X_{(n)} - X_{(1)} \quad (\text{δειγματικό εύρος})$$

Τέλος, v) *εκτιμήτρια συνάρτηση* μίας παραμέτρου θ θα καλείται μία στατιστική συνάρτηση $T(X_1, X_2, \dots, X_n)$ η οποία χρησιμοποιείται για την εκτίμηση της θ .

Ιδιότητες εκτιμητριών

Έστω $X_1, X_2, \dots, X_n$ ένα τυχαίο δείγμα από την κατανομή $F(x; \theta)$ και έστω $T(X_1, X_2, \dots, X_n)$ μία εκτιμήτρια συνάρτηση της παραμετρικής συνάρτησης $g(\theta)$ (π.χ. $\theta = (\mu, \sigma^2)$ και $g(\theta) = \mu$ ή μ^2 ή $\mu + 3\sigma$ κ.ο.κ). Όπως είναι φανερό, η εκτιμήτρια T είναι και αυτή μία τ.μ., δηλαδή κάθε φορά που παίρνουμε ένα διαφορετικό δείγμα $X_1, X_2, \dots, X_n$, η T θα μας δίνει διαφορετική τιμή. Για να μπορεί η T να θεωρηθεί ότι είναι «καλή» εκτιμήτρια της $g(\theta)$ θα πρέπει να έχει κάποια συγκεκριμένα χαρακτηριστικά όπως π.χ. να παίρνει τιμές «πολύ κοντά» στην $g(\theta)$ με «μεγάλη» πιθανότητα. Αυτό μπορεί να γίνει απαιτώντας η τ.μ. T να έχει μέση τιμή $g(\theta)$ ή «σχεδόν» $g(\theta)$ και να έχει πολύ μικρή διασπορά (οι τιμές της τ.μ. T να βρίσκονται «μαζεμένες» γύρω από τη μέση της τιμή). Π.χ. αν έχουμε τρεις διαφορετικές εκτιμήτριες T_1, T_2, T_3 για τη μέση τιμή μ ενός πληθυσμού:



τότε θα προτιμήσουμε να χρησιμοποιήσουμε την T_1 γιατί (ανάλογα με το δείγμα) παίρνει τιμές πολύ κοντά στο μ . (Η T_2 παίρνει και αυτή τιμές γύρω από το μ αλλά οι τιμές της μπορεί να διαφέρουν αρκετά από το μ γιατί έχει μεγάλη διασπορά. Η T_3 παίρνει τιμές που καμία σχέση δεν έχουν

με την μ). Ιδιότητες εκτιμητριών σχετικές με τη μέση τιμή και τη διασπορά τους εξετάζονται στις επόμενες δύο παραγράφους.

Αμερόληπτες εκτιμήτριες

Ορισμός 4.2. Μία εκτιμήτρια συνάρτηση T της $g(\theta)$ θα καλείται *αμερόληπτη* εάν

$$E(T) = E(T(X_1, X_2, \dots, X_n)) = g(\theta) \text{ για κάθε } \theta.$$

Η T θα θεωρείται *ασυμπτωτικά αμερόληπτη* εάν

$$\lim_{n \rightarrow \infty} E(T(X_1, X_2, \dots, X_n)) = g(\theta) \text{ για κάθε } \theta.$$

Επίσης, το μέγεθος

$$b(T) = E(T) - g(\theta)$$

καλείται *μεροληψία* της εκτιμήτριας T . Προφανώς η μεροληψία μιας αμερόληπτης εκτιμήτριας είναι 0.

Είναι φανερό από τα παραπάνω ότι μας ενδιαφέρουν αμερόληπτες ή σχεδόν αμερόληπτες εκτιμήτριες γιατί σε αντίθετες περιπτώσεις είναι δυνατό να έχουμε υποεκτίμηση ή υπερεκτίμηση της ζητούμενης παραμέτρου.

Πρόταση 4.1. Έστω $X_1, X_2, \dots, X_n$ ένα τυχαίο δείγμα από μία οποιαδήποτε κατανομή $F(x; \theta)$ με μέση τιμή $\mu (= \mu(\theta))$ και διασπορά $\sigma^2 (= \sigma^2(\theta))$.

1) Η στατιστική συνάρτηση

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (\text{δειγματικός μέσος})$$

είναι αμερόληπτη εκτιμήτρια της παραμέτρου $\mu = E(X_i)$ (μέσης τιμής) της κατανομής $F(x; \theta)$ και έχει διασπορά $\text{Var}(\bar{X}) = \sigma^2/n$.

2) Η στατιστική συνάρτηση

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{δειγματική διασπορά})$$

είναι αμερόληπτη εκτιμήτρια της παραμέτρου $\sigma^2 = V(X_i)$ (διασποράς) της κατανομής $F(x; \theta)$.

Απόδειξη. 1) Ισχύει ότι

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu.$$

Επίσης,

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n}.$$

2) Παρατηρούμε ότι

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n ((X_i - \mu) - (\bar{X} - \mu))^2 \\ &= \sum_{i=1}^n ((X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)) \\ &= \sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 - 2(\bar{X} - \mu)n(\bar{X} - \mu) \\
&= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2.
\end{aligned}$$

Συνεπώς, θα είναι

$$\begin{aligned}
E(S^2) &= \frac{1}{n-1} E\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2\right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^n E((X_i - \mu)^2) - nE((\bar{X} - \mu)^2)\right).
\end{aligned}$$

Γνωρίζουμε όμως ότι

$$E((X_i - \mu)^2) = V(X_i) = \sigma^2, \quad E((\bar{X} - \mu)^2) = V(\bar{X}) = \frac{\sigma^2}{n}$$

και επομένως,

$$E(S^2) = \frac{1}{n-1} \left(n\sigma^2 - n \frac{\sigma^2}{n}\right) = \frac{1}{n-1} (n\sigma^2 - \sigma^2) = \sigma^2.$$

Άρα η S^2 είναι αμερόληπτη εκτιμήτρια της διασποράς ενός οποιουδήποτε πληθυσμού. $\square$

Άσκηση 4.1. Έστω ότι θέλουμε να εκτιμήσουμε το ποσοστό p των κατοίκων μίας μεγάλης πόλης που υποστηρίζουν μία συγκεκριμένη άποψη Α. Δείξτε ότι το ποσοστό n τυχαία επιλεγμένων ατόμων που υποστηρίζουν την Α αποτελεί μία αμερόληπτη εκτιμήτρια του p .

Λύση. Έστω $X_1, X_2, \dots, X_n$ οι τ.μ. που εκφράζουν τις απόνψεις των n τυχαία επιλεγμένων ατόμων (από την πόλη αυτή). Συγκεκριμένα, θα θεωρούμε ότι $X_i=1$ αν το i -άτομο υποστηρίζει την Α και 0 διαφορετικά. Εφόσον πρόκειται για τυχαία επιλεγμένα άτομα (κάθε άτομο του πληθυσμού έχει την ίδια πιθανότητα να επιλεγεί στο δείγμα), θα ισχύει ότι

$$P(X_i=1)=1-P(X_i=0)=p,$$

και άρα $\mu = E(X_i) = 0 \cdot P(X_i=0) + 1 \cdot P(X_i=1) = p$ (οι $X_1, X_2, \dots, X_n$ αποτελούν ένα τυχαίο δείγμα από την $B(1, p)$). Το ποσοστό των n τυχαία επιλεγμένων ατόμων που υποστηρίζουν την Α θα είναι ίσο με

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Από την Πρόταση 4.1. είδαμε όμως ότι $E(\bar{X}) = E(X_i) = p$ και επομένως το ποσοστό $\bar{X}$ των ατόμων του δείγματος που υποστηρίζουν την Α αποτελεί αμερόληπτη εκτιμήτρια του συνολικού ποσοστού των κατοίκων που υποστηρίζουν την Α.

Άσκηση 4.2. Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. από μία κατανομή με μέση μ και διασπορά σ^2 . Δείξτε ότι, ενώ η στατιστική συνάρτηση $\bar{X}$ είναι αμερόληπτη εκτιμήτρια του μ , η στατιστική συνάρτηση

$$\bar{X}^2 = \left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2$$

δεν είναι αμερόληπτη εκτιμήτρια του μ^2 . Πιο συγκεκριμένα δείξτε ότι $b(\bar{X}^2) = \sigma^2/n$.

Λύση. Θα υπολογίσουμε τη μέση τιμή της $\bar{X}^2$. Ισχύει ότι,

$$E(\bar{X}^2) = V(\bar{X}) + E(\bar{X})^2 = \frac{\sigma^2}{n} + \mu^2 \neq \mu^2$$

και άρα

$$b(\bar{X}^2) = E(\bar{X}^2) - \mu^2 = \frac{\sigma^2}{n} + \mu^2 - \mu^2 = \frac{\sigma^2}{n} \neq 0.$$

Συνεπώς η στατιστική συνάρτηση $\bar{X}^2$ δεν είναι αμερόληπτη εκτιμήτρια του μ^2 . Παρατηρούμε όμως ότι αποτελεί ασυμπτωτικά αμερόληπτη εκτιμήτρια του μ^2 διότι

$$E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2 \xrightarrow{n \rightarrow \infty} \mu^2.$$

Άσκηση 4.3. Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. από την κανονική κατανομή $N(\mu, \sigma^2)$. Αν $T = \sum_{i=1}^n X_i$ και $U = aT^2 + b$, τότε να βρεθούν τα a, b ώστε η σ.σ. U να είναι α.ε. του μ^2 .

Λύση. Ισχύει ότι (οι τ.μ. $X_1, X_2, \dots, X_n$ είναι ανεξάρτητες)

$$E(T) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = n\mu, \quad V(T) = V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) = n\sigma^2$$

και επομένως

$$E(U) = E(aT^2 + b) = aE(T^2) + b = a(V(T) + E(T)^2) + b = an\sigma^2 + an^2\mu^2 + b.$$

Για να είναι η σ.σ. U α.ε. του μ^2 θα πρέπει

$$E(U) = an\sigma^2 + an^2\mu^2 + b = \mu^2$$

για κάθε μ , και επομένως θα πρέπει

$$an^2 = 1, \quad an\sigma^2 + b = 0 \Leftrightarrow a = \frac{1}{n^2}, \quad b = -\frac{\sigma^2}{n}.$$

Άρα τελικά η σ.σ.

$$\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2 - \frac{\sigma^2}{n} = \bar{X}^2 - \frac{\sigma^2}{n}$$

είναι αμερόληπτη εκτιμήτρια του μ^2 , γεγονός που συμφωνεί και με την Άσκηση 4.2.

Άσκηση 4.4. Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. από κατανομή με μέση τιμή μ και διασπορά σ^2 . Να βρεθεί το c ώστε η σ.σ.

$$U = c \sum_{k=1}^{n-1} (X_{k+1} - X_k)^2$$

να είναι α.ε. του σ^2 .

Λύση. Ζητάμε να ισχύει ότι

$$\begin{aligned} E(U) &= c \sum_{k=1}^{n-1} E(X_{k+1} - X_k)^2 = c \sum_{k=1}^{n-1} E(X_{k+1}^2 + X_k^2 - 2X_{k+1}X_k) \\ &= c \sum_{k=1}^{n-1} (E(X_{k+1}^2) + E(X_k^2) - 2E(X_{k+1}X_k)). \end{aligned}$$

Γνωρίζουμε όμως ότι $E(X_k^2) = \text{Var}(X_k) + E(X_k)^2 = \sigma^2 + \mu^2$ για όλα τα $k=1, 2, \dots, n$. Επίσης είναι γνωστό ότι αν δύο τ.μ. X, Y είναι ανεξάρτητες τότε $E(XY) = E(X)E(Y)$. Συνεπώς,

$$E(X_{k+1}X_k) = E(X_{k+1})E(X_k) = \mu^2$$

και άρα

$$E(U) = c \sum_{k=1}^{n-1} (\sigma^2 + \mu^2 + \sigma^2 + \mu^2 - 2\mu^2) = c \sum_{k=1}^{n-1} 2\sigma^2 = c2(n-1)\sigma^2.$$

Για να είναι η U α.ε. του σ^2 αρκεί $E(U) = c2(n-1)\sigma^2 = \sigma^2$, ή ισοδύναμα,

$$c = \frac{1}{2(n-1)}.$$

Τελικά συμπεραίνουμε ότι η εκτιμήτρια

$$U = \frac{1}{2(n-1)} \sum_{k=1}^{n-1} (X_{k+1} - X_k)^2$$

είναι α.ε. του σ^2 .

Άσκηση 4.5. Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. από την $B(1, p)$ (διωνυμική με παραμέτρους 1 και p) και $U = c\bar{X}(1 - \bar{X})$. Να βρεθεί το c ώστε η σ.σ. U να είναι α.ε. του $p(1-p)$ ($=V(X_i)$).

Λύση. Ισχύει ότι

$$P(X_i = j) = \binom{1}{j} p^j (1-p)^{1-j} = p^j (1-p)^{1-j}, j=0,1$$

και συνεπώς

$$E(X_i) = 1 \cdot p = p, V(X_i) = 1 \cdot p \cdot (1-p).$$

Από την Πρόταση 4.1.(1) θα ισχύει τώρα ότι

$$E(\bar{X}) = p, E(\bar{X}^2) = V(\bar{X}) + E(\bar{X})^2 = \frac{p(1-p)}{n} + p^2 = \frac{p + (n-1)p^2}{n}.$$

Η μέση τιμή της U θα είναι ίση με

$$E(U) = cE(\bar{X} - \bar{X}^2) = c(E(\bar{X}) - E(\bar{X}^2)) = c\left(p - \frac{p + (n-1)p^2}{n}\right) = c \frac{(n-1)}{n} p(1-p)$$

και άρα θα πρέπει

$$c \frac{(n-1)}{n} p(1-p) = p(1-p) \quad \text{ή ισοδύναμα,} \quad c = \frac{n}{n-1}.$$

Τελικά, διαπιστώθηκε ότι η εκτιμήτρια $U = n\bar{X}(1 - \bar{X})/(n-1)$ είναι α.ε. του $p(1-p)$.

Σύγκριση μεταξύ αμερόληπτων εκτιμητριών

Όπως έχει αναφερθεί και παραπάνω, για να μπορεί η T να θεωρηθεί ότι είναι «καλή» εκτιμήτρια της $g(\theta)$ θα πρέπει να έχει μέση τιμή $g(\theta)$ ή «σχεδόν» $g(\theta)$ και να έχει μικρή διασπορά. Την πρώτη απαίτηση την εξετάσαμε στην προηγούμενη παράγραφο όπου εισαγάγαμε την έννοια της αμεροληψίας. Δεν αρκεί όμως μόνο η αμεροληψία για να πούμε ότι μία εκτιμήτρια είναι βέλτιστη. Το γεγονός αυτό φαίνεται και από το ότι για την ίδια παραμετρική συνάρτηση $g(\theta)$ μπορούν να βρεθούν πάρα πολλές αμερόληπτες εκτιμήτριες (βλ. π.χ. Ασκήσεις 4.4-4.5). Ποια από αυτές είναι η καλύτερη; Σύμφωνα με τα όσα έχουμε πει καλύτερη θα είναι αυτή που θα παίρνει τιμές πολύ «κοντά» στην $g(\theta)$. Δηλαδή με άλλα λόγια αυτή που έχει τη μικρότερη διασπορά. Σε αντιδιαστολή με την προηγούμενη παράγραφο όπου ασχοληθήκαμε με τη μέση τιμή μιας εκτιμή-

τριας, σε αυτή την παράγραφο θα μελετήσουμε τις εκτιμήτριες ως προς τη διασπορά τους. Σχετικά λοιπόν με τη διασπορά δύο εκτιμητριών θα έχουμε τον επόμενο ορισμό:

Ορισμός 4.3. Έστω T_1, T_2 δύο αμερόληπτες εκτιμήτριες της $g(\theta)$. Η T_1 θα καλείται *αποτελεσματικότερη* της T_2 εάν ισχύει ότι

$$V(T_1) < V(T_2).$$

Επίσης, αν μία εκτιμήτρια T έχει τη μικρότερη διασπορά μεταξύ όλων των αμερόληπτων εκτιμητριών του $g(\theta)$, τότε θα καλείται *άριστη εκτιμήτρια* ή *αμερόληπτη εκτιμήτρια ελαχίστης διασποράς* (α.ε.ε.δ.) του $g(\theta)$.

Συνεπώς, τη βέλτιστη επιλογή εκτιμήτριας, προκειμένου να εκτιμήσουμε μία παραμετρική συνάρτηση $g(\theta)$, αποτελεί η επιλογή μίας άριστης εκτιμήτριας ή τουλάχιστον μίας εκτιμήτριας που ασυμπτωτικά (για μεγάλο n) είναι άριστη.

Άσκηση 4.6. Έστω $X_1, X_2, \dots, X_n$ και $Y_1, Y_2, \dots, Y_k$ δύο ανεξάρτητα μεταξύ τους τ.δ. από μία κατανομή F με άγνωστη μέση τιμή μ και διασπορά σ^2 . Ποια από τις 3 εκτιμήτριες θα χρησιμοποιούσατε για να εκτιμήσετε το μ ($n=100, k=10$):

$$i) \bar{X}, \quad ii) \bar{Y}, \quad iii) \frac{\bar{X} + \bar{Y}}{2}.$$

Μπορεί να βρεθεί καλύτερη εκτιμήτρια που να γράφεται στη μορφή $a\bar{X} + b\bar{Y}$;

Λύση. Από την Πρόταση 4.1.(1) γνωρίζουμε ότι $E(\bar{X})=\mu$, $E(\bar{Y})=\mu$ και επομένως

$$E\left(\frac{1}{2}(\bar{X} + \bar{Y})\right) = \frac{1}{2}(E(\bar{X}) + E(\bar{Y})) = \frac{1}{2}(\mu + \mu) = \mu,$$

Άρα και οι 3 εκτιμήτριες είναι αμερόληπτες. Προφανώς θα χρησιμοποιήσουμε την εκτιμήτρια με τη μικρότερη διασπορά. Και πάλι από την Πρόταση 4.1.(1) γνωρίζουμε ότι

$$V(\bar{X}) = \frac{\sigma^2}{n} = 0,01\sigma^2, \quad V(\bar{Y}) = \frac{\sigma^2}{k} = 0,1\sigma^2,$$

και άρα

$$V\left(\frac{1}{2}(\bar{X} + \bar{Y})\right) = \frac{1}{4}(V(\bar{X}) + V(\bar{Y})) = \frac{1}{4}\left(\frac{\sigma^2}{n} + \frac{\sigma^2}{k}\right) = \frac{k+n}{4nk}\sigma^2 = \frac{110}{4000}\sigma^2 = 0,0275\sigma^2.$$

Επομένως για τις συγκεκριμένες τιμές των n, k καλύτερη εκτιμήτρια είναι η $\bar{X}$. Αυτό εκ πρώτης όψεως ίσως να φαίνεται περίεργο διότι η τρίτη εκτιμήτρια βασίζεται σε δείγμα μεγέθους $n+k$ μεγαλύτερο από το n και το k των άλλων δύο εκτιμητριών. Αυτό όμως που τελικά συμβαίνει είναι ότι η $\bar{X}$ είναι αρκετά καλή εκτιμήτρια (έχει διασπορά $0,01\sigma^2$) ενώ η $\bar{Y}$ δεν είναι τόσο καλή (έχει διασπορά αρκετά μεγαλύτερη $0,1\sigma^2$). Τελικά λαμβάνοντας ως εκτιμήτρια του μ την $(\bar{X} + \bar{Y})/2$ να μην παίρνουμε πάλι μία α.ε. του μ αλλά η $(\bar{X} + \bar{Y})/2$ βασίζεται εξίσου στην εκτιμήτρια $\bar{X}$ (που είναι αρκετά καλή) και στην $\bar{Y}$ (που δεν είναι τόσο καλή) οπότε τελικά παίρνουμε μία εκτιμήτρια που είναι μέτρια. Το βέλτιστο που θα μπορούσαμε να κάνουμε θα ήταν να πάρουμε μία εκτιμήτρια που να βασίζεται περισσότερο στην $\bar{X}$ και λιγότερο στην $\bar{Y}$. Πράγματι, θα αναζητήσουμε μία αμερόληπτη εκτιμήτρια της μορφής $a\bar{X} + b\bar{Y}$ που να έχει τη μικρότερη δυνατή διασπορά. Θα πρέπει

$$E(a\bar{X} + b\bar{Y}) = \mu \Leftrightarrow aE(\bar{X}) + bE(\bar{Y}) = \mu \Leftrightarrow a\mu + b\mu = \mu \Leftrightarrow b = 1 - a.$$

Επίσης,

$$V(a\bar{X} + b\bar{Y}) = a^2V(\bar{X}) + b^2V(\bar{Y}) = a^2\frac{\sigma^2}{n} + b^2\frac{\sigma^2}{k} = \frac{a^2k + (1-a)^2n}{nk}\sigma^2.$$

Αρκεί να βρούμε το a που ελαχιστοποιεί την παράσταση $f(a) = a^2k + (1-a)^2n$. Ισχύει ότι

$$f'(a) = 2ak - 2(1-a)n = 2ak - 2n + 2an = 0 \Leftrightarrow a = \frac{n}{k+n} \quad (f''(a) = 2k + 2n > 0)$$

και επομένως η εκτιμήτρια

$$a\bar{X} + b\bar{Y} = \frac{n}{k+n}\bar{X} + \left(1 - \frac{n}{k+n}\right)\bar{Y} = \frac{n\bar{X} + k\bar{Y}}{k+n}$$

είναι αμερόληπτη και έχει τη μικρότερη διασπορά ανάμεσα στις εκτιμήτριες της μορφής $a\bar{X} + b\bar{Y}$ (που έχουν και οι 3 πρώτες εκτιμήτριες). Η διασπορά της είναι

$$V\left(\frac{1}{n+k}(n\bar{X} + k\bar{Y})\right) = \frac{n^2V(\bar{X}) + k^2V(\bar{Y})}{(n+k)^2} = \frac{\sigma^2}{n+k} = 0,0091\sigma^2.$$

Παρατηρούμε ότι η παραπάνω εκτιμήτρια είναι ο δειγματικός μέσος του μεικτού δείγματος $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_k$.

Άσκηση 4.7. Έστω $X_1, X_2, \dots, X_n$ ένα τυχαίο δείγμα (τ.δ.) από κατανομή $F(x; \theta)$ με $E(X_i) = g(\theta)$ και

$$U = U(X_1, \dots, X_n) = \sum_{i=1}^n a_i X_i.$$

Να βρεθούν οι τιμές των $a_i \in \mathbb{N}$ ώστε η U να είναι αμερόληπτη εκτιμήτρια (α.ε.) του $g(\theta)$ και να έχει ελάχιστη διασπορά.

Λύση. Για να είναι αμερόληπτη εκτιμήτρια θα πρέπει να ισχύει,

$$E(U) = g(\theta) \Leftrightarrow E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i) = g(\theta) \sum_{i=1}^n a_i = g(\theta) \Leftrightarrow \sum_{i=1}^n a_i = 1.$$

Επίσης, δεδομένου ότι οι τ.μ. είναι ανεξάρτητες θα ισχύει ότι

$$V(U) = \sum_{i=1}^n a_i^2 V(X_i) = V(X_1) \sum_{i=1}^n a_i^2.$$

Για να είναι η διασπορά της εκτιμήτριας ελάχιστη θα πρέπει να ελαχιστοποιήσουμε τη συνάρτηση

$$\sum_{i=1}^n a_i^2 = \sum_{i=1}^{n-1} a_i^2 + a_n^2 = \sum_{i=1}^{n-1} a_i^2 + \left(1 - \sum_{i=1}^{n-1} a_i\right)^2.$$

Λαμβάνοντας τις μερικές παραγώγους της ως προς a_i θα πρέπει για να έχουμε ελάχιστο να ισχύει

$$\frac{\partial}{\partial a_j} \left(\sum_{i=1}^{n-1} a_i^2 + \left(1 - \sum_{i=1}^{n-1} a_i\right)^2 \right) = 2a_j - 2\left(1 - \sum_{i=1}^{n-1} a_i\right) = 0, \quad j = 1, 2, \dots, n-1$$

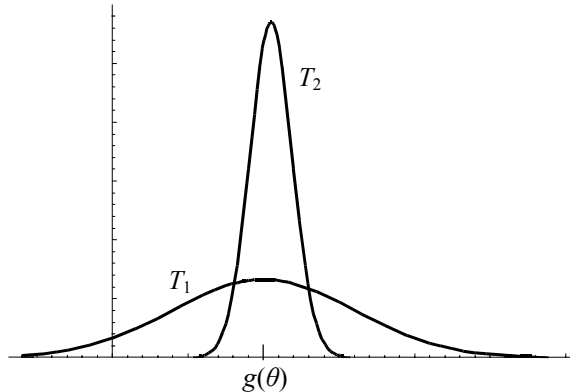
ή ισοδύναμα,

$$a_j = 1 - \sum_{i=1}^{n-1} a_i = a_n, \quad j = 1, 2, \dots, n-1.$$

Άρα όλα τα a_i θα πρέπει να είναι ίσα. Επειδή όμως αθροίζουν στη μονάδα συνάγουμε τελικά ότι για να είναι η εκτιμήτρια αμερόληπτη και με ελάχιστη διασπορά θα πρέπει $a_i = 1/n$, $i = 1, 2, \dots, n$. (Αποδεικνύεται επίσης ότι ο πίνακας Hesse είναι οριστικά θετικός).

Αποτελεσματικότητα οποιωνδήποτε εκτιμητριών

Στην προηγούμενη παράγραφο εξετάσαμε την αποτελεσματικότητα εκτιμητριών που είναι αμερόληπτες. Συγκεκριμένα, αναφέραμε ότι, αν έχουμε ένα σύνολο $T_1, T_2, \dots, T_k$ από αμερόληπτες εκτιμήτριες μίας παραμέτρου $g(\theta)$ ($E(T_i) = g(\theta)$) τότε ως βέλτιστη για την εκτίμηση της $g(\theta)$ θα είναι η εκτιμήτρια με τη μικρότερη διασπορά. Τι γίνεται όμως στην περίπτωση που έχουμε ένα σύνολο από εκτιμήτριες του $g(\theta)$ που δεν είναι όλες αμερόληπτες; Ποια θα είναι η βέλτιστη; Έχουμε ήδη αναφέρει ότι συνήθως προτιμούμε αμερόληπτες εκτιμήτριες. Υπάρχουν όμως περιπτώσεις όπου μία μη αμερόληπτη εκτιμήτρια είναι καλύτερη από μία αμερόληπτη. Για παράδειγμα αν, για την εκτίμηση ενός $g(\theta)$, έχουμε τις εκτιμήτριες T_1, T_2 με σ.π.π. που δίνονται στο παρακάτω γράφημα τότε παρατηρούμε ότι η T_2 δεν είναι αμερόληπτη (δεν έχει μέση τιμή $g(\theta)$) αλλά παίρνει τιμές «κοντά» στο $g(\theta)$ με μεγάλη πιθανότητα σε αντίθεση με την T_1 που είναι αμερόληπτη αλλά μπορεί να πάρει τιμές πολύ μακριά από το $g(\theta)$ (έχει μεγάλη διασπορά).



Επομένως, παρότι δεν είναι αμερόληπτη, καλύτερη εκτιμήτρια σε αυτή την περίπτωση είναι η T_2 . Αν όμως η T_2 ήταν περισσότερο μεροληπτική; Ποια θα ήταν καλύτερη;

Για να βρούμε ένα κριτήριο αξιολόγησης εκτιμητριών ενός $g(\theta)$, ανεξάρτητα από το αν είναι αμερόληπτες ή όχι, παρατηρούμε ότι στην περίπτωση α.ε. του $g(\theta)$ εξετάζαμε την

$$V(T) = E((T - E(T))^2) = E((T - g(\theta))^2)$$

($E(T) = g(\theta)$ λόγω αμεροληψίας). Η ποσότητα αυτή είναι στην ουσία η μέση τιμή του τετραγώνου του σφάλματος της T από την εκτιμώμενη παράμετρο. Παρατηρούμε όμως ότι την ίδια ακριβώς ποσότητα μπορούμε να χρησιμοποιήσουμε και γενικότερα για την αξιολόγηση οποιωνδήποτε εκτιμητριών ενός θ . Ειδικότερα θα έχουμε τον επόμενο ορισμό.

Ορισμός 4.4. Έστω μία εκτιμήτρια $T = T(X_1, X_2, \dots, X_n)$ μίας παραμέτρου $g(\theta)$. Η ποσότητα

$$mse(T) = E((T - g(\theta))^2)$$

καλείται μέσο τετραγωνικό σφάλμα της T από την $g(\theta)$.

Επομένως, θα θεωρούμε ότι μεταξύ δύο εκτιμητριών μίας παραμέτρου, βέλτιστη είναι αυτή που έχει το μικρότερο μέσο τετραγωνικό σφάλμα. Παρατηρούμε ότι το μέσο τετραγωνικό σφάλμα γράφεται στη μορφή

$$\begin{aligned} mse(T) &= E((T - g(\theta))^2) = E((T - E(T)) + (E(T) - g(\theta))^2) \\ &= E((T - E(T))^2 + (E(T) - g(\theta))^2 + 2(T - E(T))(E(T) - g(\theta))) \\ &= E((T - E(T))^2) + (E(T) - g(\theta))^2 + 2E(T - E(T))(E(T) - g(\theta))) \\ &= V(T) + b(T)^2. \end{aligned}$$

Αν ένας εκτιμητής είναι αμερόληπτος τότε προφανώς το μέσο τετραγωνικό του σφάλμα είναι ίσο με τη διασπορά του. Με τη χρήση τώρα του μ.τ.σ. μπορούμε να δώσουμε ένα γενικότερο από τον 4.2 ορισμό για την αποτελεσματικότητα δύο εκτιμητριών.

Ορισμός 4.5. Έστω T_1, T_2 δύο εκτιμήτριες της $g(\theta)$. Η T_1 θα καλείται *αποτελεσματικότερη* της T_2 εάν ισχύει ότι

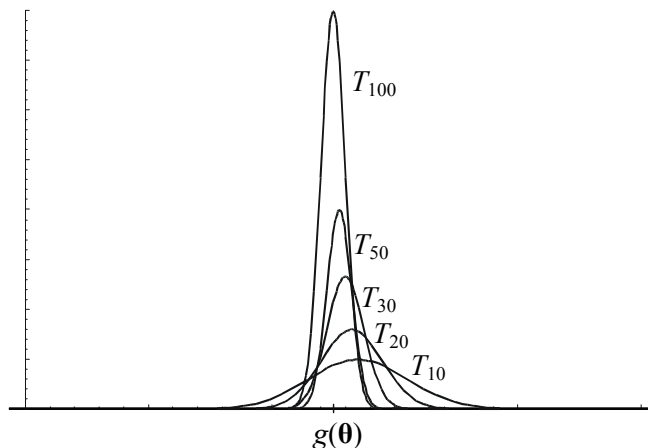
$$mse(T_1) < mse(T_2).$$

Αν οι εκτιμήτριες T_1, T_2 είναι αμερόληπτες τότε ο παραπάνω ορισμός είναι ισοδύναμος με τον Ορισμό 4.2. που αφορούσε αμερόληπτες εκτιμήτριες. Τέλος, η *σχετική αποτελεσματικότητα* μίας εκτιμήτριας T_1 σε σχέση με την T_2 ορίζεται από το πηλίκο:

$$\frac{mse(T_2)}{mse(T_1)}.$$

Συνέπεια

Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. από μία κατανομή $F(x; \theta)$ και έστω $T_n = T_n(X_1, X_2, \dots, X_n)$ μία στατιστική συνάρτηση που χρησιμοποιείται για την εκτίμηση μίας παραμετρικής συνάρτησης $g(\theta)$. Όπως έχουμε ήδη αναφέρει, η T_n θα αποτελεί ικανοποιητική εκτιμήτρια του $g(\theta)$ αν παίρνει τιμές «κοντά» στην $g(\theta)$ με μεγάλη πιθανότητα. Επιπλέον, είναι φυσικό να απαιτούμε η T_n να παίρνει τιμές «πολύ κοντά» στην $g(\theta)$ όσο το n μεγαλώνει. Αν δηλαδή πάρουμε ένα πολύ μεγάλο δείγμα (θεωρητικά άπειρο) είναι λογικό να απαιτήσουμε η εκτιμήτρια να μας δίνει ακριβώς την $g(\theta)$. Με άλλα λόγια, μία εκτιμήτρια T_n θα θεωρείται «καλή» και αν, αυξάνοντας το μέγεθος του δείγματος, γίνεται ακριβέστερη ως προς την εκτίμηση του $g(\theta)$. Οι εκτιμήτριες που έχουν αυτή την ιδιότητα θα καλούνται *συνεπείς*. Για παράδειγμα η σ.π.π. μίας συνεπούς εκτιμήτριας T_n της $g(\theta)$ θα έχει τη μορφή που δίνεται στο επόμενο σχήμα.



Πιο αυστηρά δίνεται ο επόμενος ορισμός.

Ορισμός 4.6. Μία εκτιμήτρια $T_n = T_n(X_1, X_2, \dots, X_n)$ μιας παραμέτρου $g(\theta)$ θα καλείται *συνεπής* αν ισχύει ότι

$$P(|T_n - g(\theta)| < \varepsilon) \xrightarrow{n \rightarrow \infty} 1 \text{ για κάθε } \varepsilon > 0.$$

Σύμφωνα με τον παραπάνω ορισμό μία εκτιμήτρια T_n θα είναι συνεπής αν, λαμβάνοντας ένα πολύ μεγάλο δείγμα (ασυμπτωτικά), παίρνει τιμές σε μία οσοδήποτε μικρή περιοχή $(g(\theta) - \varepsilon, g(\theta) + \varepsilon)$ γύρω από το $g(\theta)$ με πιθανότητα 1. Επειδή όμως ο έλεγχος της παραπάνω συνθήκης δεν είναι εύκολος συνήθως χρησιμοποιούμε την επόμενη πρόταση.

Πρόταση 4.2. Μία εκτιμήτρια $T_n = T_n(X_1, X_2, \dots, X_n)$ μιας παραμέτρου $g(\theta)$ θα είναι συνεπής αν ισχύουν οι παρακάτω συνθήκες

$$\text{i) } E(T_n) \xrightarrow{n \rightarrow \infty} g(\theta) \quad (\text{δηλαδή } b(T_n) \xrightarrow{n \rightarrow \infty} 0) \quad \text{ii) } V(T_n) \xrightarrow{n \rightarrow \infty} 0,$$

ή ισοδύναμα αν $mse(T_n) = V(T_n) + b(T_n)^2 \xrightarrow{n \rightarrow \infty} 0$.

Απόδειξη. Από την ανισότητα Markov για την τ.μ. $(T_n - g(\theta))^2$ και για $\alpha = \varepsilon^2$ θα ισχύει ότι

$$P((T_n - g(\theta)) \geq \varepsilon^2) \leq \frac{E[(T_n - g(\theta))^2]}{\varepsilon^2} \quad \text{για κάθε } \varepsilon > 0,$$

ή ισοδύναμα,

$$P(|T_n - g(\theta)| \geq \varepsilon) \leq \frac{mse(T_n)}{\varepsilon^2} \quad \text{για κάθε } \varepsilon > 0,$$

Επειδή όμως $mse(T_n) = V(T_n) + b(T_n)^2 \xrightarrow{n \rightarrow \infty} 0$, θα ισχύει ότι

$$P(|T_n - g(\theta)| \geq \varepsilon) \rightarrow 0 \quad \text{για κάθε } \varepsilon > 0, \text{ ή ισοδύναμα } P(|T_n - g(\theta)| < \varepsilon) \xrightarrow{n \rightarrow \infty} 1 \quad \text{για κάθε } \varepsilon > 0$$

και άρα η T_n είναι συνεπής εκτιμήτρια του $g(\theta)$.

Άσκηση 4.8. Δείξτε ότι ο δειγματικός μέσος $\bar{X}$ παρατηρήσεων που προέρχονται από μία κατανομή με μέση τιμή μ και διασπορά $\sigma^2 < \infty$ είναι συνεπής εκτιμήτρια του μ .

Λύση. Έχουμε ήδη αποδείξει παραπάνω ότι

$$E(\bar{X}) = \mu, \quad V(\bar{X}) = \frac{\sigma^2}{n}$$

και επομένως (αν ως T_n θεωρήσουμε την $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$),

$$mse(\bar{X}) = V(\bar{X}) + b(\bar{X})^2 = \frac{\sigma^2}{n} + (E(\bar{X}) - \mu)^2 = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0.$$

Η κατανομή χι-τετράγωνο (χ^2 , chi-square)

Είδαμε παραπάνω ότι η στατιστική συνάρτηση

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

αποτελεί αμερόληπτη εκτιμήτρια της διασποράς σ^2 ενός πληθυσμού. Για να εξετάσουμε όμως αν είναι αποτελεσματικότερη κάποιας άλλης εκτιμήτριας ή π.χ. αποτελεί συνεπή εκτιμήτρια του σ^2 θα πρέπει να γνωρίζουμε τη διασπορά της. Επειδή συνηθέστερη κατανομή είναι η κανονική, στη συνέχεια θα εξετάσουμε την κατανομή της S^2 όταν το δείγμα προέρχεται από μία $N(\mu, \sigma^2)$. Αρχικά θα χρειαστεί να ορίσουμε την κατανομή Γάμμα η οποία αποτελεί γενίκευση της εκθετικής κατανομής.

Ορισμός 4.7. Η συνεχής κατανομή με σ.π.π.

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0 \quad \text{όπου} \quad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

($f(x) = 0$ για $x \leq 0$) καλείται κατανομή Γάμμα με παραμέτρους $\lambda > 0$, $\alpha > 0$ (συμβολίζεται και με $G(\alpha, \lambda)$).

Η συνάρτηση $\Gamma: [0, \infty) \rightarrow \mathbb{R}$ καλείται *συνάρτηση γάμμα*. Χρησιμοποιώντας ολοκλήρωση κατά παράγοντες αποδεικνύεται ότι

$$\Gamma(\alpha+1) = \alpha \Gamma(\alpha), \quad \alpha > 0.$$

Επίσης εύκολα βλέπουμε ότι $\Gamma(1)=1$. Επομένως, $\Gamma(2) = 1\Gamma(1) = 1$, $\Gamma(3) = 2\Gamma(2) = 2$, $\Gamma(4) = 3\Gamma(3) = 3 \cdot 2$, κ.ο.κ, $\Gamma(\alpha)=(\alpha-1)!$ για $\alpha = 1,2,\dots$. Η συνάρτηση γάμμα μπορεί να θεωρηθεί ως μία επέκταση της συνάρτησης παραγοντικό (που είναι ορισμένη στο $\{0,1,2,\dots\}$) σε όλο το $[0,\infty)$.

Αν $X \sim G(\alpha,\lambda)$ τότε αποδεικνύεται ότι

$$E(X) = \frac{\alpha}{\lambda}, \quad V(X) = \frac{\alpha}{\lambda^2}.$$

Είναι εύκολο να δούμε ότι η εκθετική κατανομή είναι μία κατανομή $G(\lambda,1)$. Επίσης, αποδεικνύεται ότι αν $X_1, X_2, \dots, X_k$ είναι ανεξάρτητες τ.μ. με $X_i \sim G(\alpha_i, \lambda)$, $i=1,2,\dots,k$ τότε η τ.μ.

$$X_1 + X_2 + \dots + X_k \sim G(\alpha_1 + \alpha_2 + \dots + \alpha_k, \lambda).$$

Μία υποπερίπτωση της κατανομής Γάμμα, η $G(n/2, 1/2)$ για $n = 1,2,\dots$, παρουσιάζει ιδιαίτερο ενδιαφέρον στη Στατιστική. Αυτό προκύπτει από τις επόμενες προτάσεις.

Πρόταση 4.3. Αν $X \sim N(0,1)$ τότε η κατανομή της τ.μ. $Y = X^2$ ακολουθεί $G(1/2, 1/2)$ κατανομή.

Απόδειξη. Η συνάρτηση πυκνότητας πιθανότητας της τ.μ. Y για $x > 0$ θα είναι

$$\begin{aligned} f_Y(x) &= \frac{d}{dx} F_Y(x) = \frac{d}{dx} P(Y \leq x) = \frac{d}{dx} P(X^2 \leq x) = \frac{d}{dx} P(-\sqrt{x} \leq X \leq \sqrt{x}) = \frac{d}{dx} (\Phi(\sqrt{x}) - \Phi(-\sqrt{x})) \\ &= \frac{d}{dx} (2\Phi(\sqrt{x}) - 1) = 2\phi(\sqrt{x}) \frac{1}{2\sqrt{x}} = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-\frac{x}{2}}, \quad x > 0 \end{aligned}$$

ενώ $f_Y(x) = 0$ για $x \leq 0$ (διότι σε αυτή την περίπτωση $P(X^2 \leq x) = 0$). Η σ.π.π. της Y επαληθεύεται εύκολα ότι είναι η σ.π.π. της $G(1/2, 1/2)$ (ισχύει ότι $\Gamma(1/2) = \sqrt{\pi}$).

Πρόταση 4.4. Αν $X_1, X_2, \dots, X_n \sim N(0,1)$ ανεξάρτητες τ.μ. τότε η κατανομή της τ.μ. $\sum_{i=1}^n X_i^2$ ακολουθεί $G(n/2, 1/2)$ κατανομή.

Απόδειξη. Οι τ.μ. X_i^2 , σύμφωνα με την προηγούμενη πρόταση, ακολουθούν κατανομή $G(1/2, 1/2)$. Επομένως το άθροισμά τους, σύμφωνα με την παραπάνω ιδιότητα της κατανομής Γάμμα, θα ακολουθεί κατανομή $G(n/2, 1/2)$.

Η κατανομή $G(n/2, 1/2)$, δηλαδή η κατανομή του αθροίσματος των τετραγώνων τυπικών κανονικών, έχει ιδιαίτερη σημασία στη στατιστική και συναντάται σε αρκετές εφαρμογές. Για το λόγο αυτό της έχει δοθεί μία ιδιαίτερη ονομασία. Ειδικότερα έχουμε τον επόμενο ορισμό.

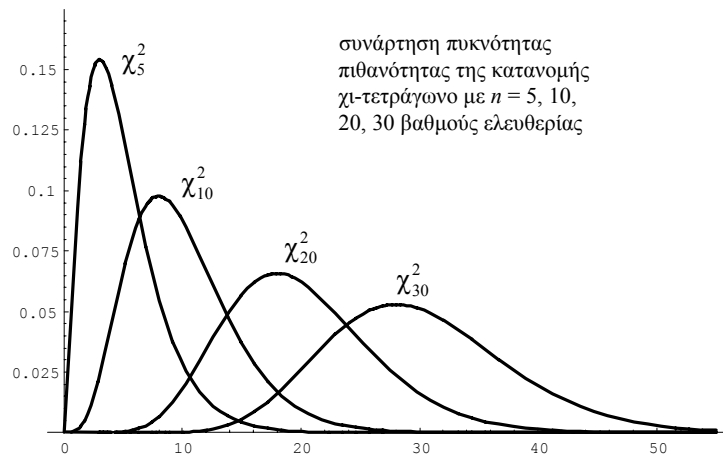
Ορισμός 4.8. Η κατανομή του αθροίσματος των τετραγώνων n ανεξάρτητων τυπικών κανονικών η οποία είναι $G(n/2, 1/2)$, θα καλείται κατανομή χ^2 -τετράγωνο με n βαθμούς ελευθερίας (*chi-square* ή χ^2 , $G(n/2, 1/2) \equiv \chi_n^2$).

Όπως είναι φανερό, οι βαθμοί ελευθερίας (β.ε.) n είναι η παράμετρος της κατανομής χ_n^2 . Αποδεικνύεται στην επόμενη πρόταση ότι οι β.ε. μιας χ_n^2 είναι και η μέση της τιμή.

Πρόταση 4.5. Αν $X \sim \chi_n^2$ τότε,

$$E(X) = n, \quad V(X) = 2n.$$

Απόδειξη. Γνωρίζουμε ότι η κατανομή χ_n^2 είναι στην ουσία μία $G(\alpha=n/2, \lambda=1/2)$ και επομένως θα έχει μέση τιμή $\alpha/\lambda=n$, και διασπορά $\alpha/\lambda^2=2n$.



Ιδιαίτερα χρήσιμη για την στατιστική είναι η επόμενη πρόταση σύμφωνα με την οποία η κατανομή της δειγματικής διασποράς S^2 από κανονικό πληθυσμό ακολουθεί (κατάλληλα τροποποιημένη) κατανομή χι-τετράγωνο.

Πρόταση 4.6. Αν $X_1, X_2, \dots, X_n$ τ.δ. από την $N(\mu, \sigma^2)$ και

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

η δειγματική διασπορά, τότε

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi_{n-1}^2.$$

Άσκηση 4.9. Αν $X_1, X_2, \dots, X_n$ τ.δ. από την $N(\mu, \sigma^2)$, εξετάσετε ως προς i) την αμεροληψία, ii) τη σχετική αποτελεσματικότητα και iii) τη συνέπεια, τις εκτιμήτριες του σ^2 :

$$S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad \text{και} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

iv) Σε ποιά περίπτωση συστήνετε τη χρήση της S_0^2 και σε ποια τη χρήση της S^2 ;

Λύση. i) Από την Πρόταση 4.1.β γνωρίζουμε ότι η S^2 αποτελεί αμερόληπτη εκτιμήτρια του σ^2 ανεξάρτητα από την κατανομή από την οποία προέρχονται οι παρατηρήσεις. Το ίδιο είναι εύκολο να δείξουμε και για την S_0^2 . Πράγματι,

$$E(S_0^2) = \frac{1}{n} \sum_{i=1}^n E((X_i - \mu)^2) = \frac{1}{n} \sum_{i=1}^n V(X_i) = \frac{1}{n} n \sigma^2 = \sigma^2.$$

Επομένως και οι δύο σ.σ. είναι αμερόληπτες εκτιμήτριες του σ^2 .

ii) Θα βρούμε τις διασπορές των εκτιμητριών S_0^2 και S^2 . Για την S_0^2 παρατηρούμε ότι η τ.μ. $n S_0^2 / \sigma^2$ γράφεται στη μορφή

$$\frac{n S_0^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^n Z_i^2,$$

όπου $Z_1, Z_2, \dots, Z_n$ είναι ανεξάρτητες $N(0,1)$ τ.μ. Επομένως, από την Πρόταση 4.4. προκύπτει ότι η τ.μ. $n S_0^2 / \sigma^2$ ακολουθεί $G(n/2, 1/2) = \chi_n^2$ κατανομή και άρα θα έχει διασπορά $2n$. Δηλαδή,

$$V\left(\frac{n S_0^2}{\sigma^2}\right) = 2n \Rightarrow \frac{n^2}{\sigma^4} V(S_0^2) = 2n \Rightarrow V(S_0^2) = \frac{2\sigma^4}{n}.$$

Για την S^2 τώρα, από την Πρόταση 4.6. παρατηρούμε ότι η τ.μ.

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$$

ακολουθεί κατανομή χ_{n-1}^2 και επομένως από την Πρόταση 4.5. θα έχει διασπορά $2(n-1)$. Δηλαδή,

$$Var\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1) \quad \text{και επομένως} \quad V(S^2) = \frac{2\sigma^4}{n-1}.$$

Από τα παραπάνω βλέπουμε ότι η εκτιμήτρια S_0^2 είναι αποτελεσματικότερη από την S^2 με σχετική αποτελεσματικότητα

$$\frac{V(S_0^2)}{V(S^2)} = 1 - \frac{1}{n},$$

iii) Στο ερώτημα (i) αποδείχθηκε ότι οι δύο εκτιμήτριες είναι α.ε. του σ^2 με διασπορές

$$V(S^2) = \frac{2\sigma^4}{n-1} \xrightarrow{n \rightarrow \infty} 0 \quad \text{και} \quad V(S_0^2) = \frac{2\sigma^4}{n} \xrightarrow{n \rightarrow \infty} 0$$

και επομένως, από την Πρόταση 4.2. και οι δύο εκτιμήτριες είναι συνεπείς.

iv) Η εκτιμήτρια S_0^2 είναι αποτελεσματικότερη από την S^2 αν και η σχετική αποτελεσματικότητα της S^2 σε σχέση με την S_0^2 είναι σχεδόν 1 για μεγάλα n (οι δύο εκτιμήτριες έχουν πρακτικά την ίδια απόδοση για σχετικά μεγάλα δείγματα). Από την άλλη όμως παρατηρούμε ότι η χρήση της εκτιμήτριας S_0^2 είναι δυνατή μόνο όταν είναι γνωστό το μ , ενώ η S^2 δεν βασίζεται στην ακριβή τιμή του μ . Επομένως, για μικρά δείγματα με γνωστό μ , συνίσταται η χρήση της S_0^2 ενώ σε όλες τις άλλες περιπτώσεις είναι προτιμότερο να χρησιμοποιούμε την S^2 .

Μέθοδοι εκτιμήσεως.

Στις προηγούμενες παραγράφους εξετάσαμε τα χαρακτηριστικά (π.χ. αμεροληψία, αποτελεσματικότητα, συνέπεια, κ.τ.λ.) αρκετών εκτιμητριών παραμέτρων γνωστών κατανομών. Σε όλες αυτές τις περιπτώσεις έπρεπε πρώτα να γνωρίζουμε τη μορφή της εκτιμήτριας και στη συνέχεια να εξετάσουμε αν είναι π.χ. συνεπής. Στις περισσότερες περιπτώσεις χρησιμοποιήσαμε το δειγματικό μέσο και τη δειγματική διασπορά μιας και αυτές είναι οι πιο γνωστές στατιστικές συναρτήσεις. Οι εκτιμήτριες αυτές αποδείχθηκαν αρκετά ικανοποιητικές στις περιπτώσεις που επιθυμούμε να εκτιμήσουμε μέσες τιμές ή διασπορές κανονικών πληθυσμών. Τι γίνεται όμως στη γενική περίπτωση που δίνεται κάποια κατανομή και θέλουμε να εκτιμήσουμε κάποια συνάρτηση των παραμέτρων της; Στη συνέχεια θα παρουσιάσουμε δύο μεθόδους για την εύρεση εκτιμητριών για οποιαδήποτε συνάρτηση των παραμέτρων. Η πρώτη είναι η (αρχαιότερη) μέθοδος των ροπών ενώ η δεύτερη και σημαντικότερη είναι η μέθοδος της μέγιστης πιθανοφάνειας.

α. Μέθοδος των ροπών

Έστω ένα τ.δ. $X_1, X_2, \dots, X_n$ από μία κατανομή με συνάρτηση σ.κ. $F(x; \theta = (\theta_1, \theta_2, \dots, \theta_k))$ για την οποία υποθέτουμε ότι υπάρχουν οι ροπές έως k τάξης ($E(X_i^r)$: ροπή r -τάξης της τ.μ. X_i). Έστω επίσης ότι επιθυμούμε να εκτιμήσουμε τα $\theta_1, \theta_2, \dots, \theta_k$. Οι k -πρώτες ροπές της κατανομής θα είναι συναρτήσεις των παραμέτρων $\theta_1, \theta_2, \dots, \theta_k$, δηλαδή θα είναι

$$E(X^r) = \mu'_r(\theta_1, \theta_2, \dots, \theta_k), \quad \text{για } r=1, 2, \dots, k.$$

(όπου για ευκολία συμβολίζουμε με $E(X^r)$ την κοινή ροπή r τάξης των $X_1, X_2, \dots, X_n$, δηλ. $E(X^r) = E(X_i^r)$, $i=1, 2, \dots, n$). Σύμφωνα με τη μέθοδο των ροπών, κατασκευάζουμε το ακόλουθο σύστημα που αποτελείται από k εξισώσεις και k αγνώστους $\theta_1, \theta_2, \dots, \theta_k$:

$$m_r = \mu'_r(\theta_1, \theta_2, \dots, \theta_k), \quad r=1, 2, \dots, k$$

όπου $m_r = \frac{1}{n} \sum_{i=1}^n X_i^r$ είναι η δειγματική ροπή r -τάξης. Δηλαδή στην ουσία κατασκευάζουμε το σύστημα

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i &= E(X), \\ \frac{1}{n} \sum_{i=1}^n X_i^2 &= E(X^2), \\ &\dots\dots\dots \\ \frac{1}{n} \sum_{i=1}^n X_i^k &= E(X^k). \end{aligned}$$

Εξισώνουμε δηλαδή τις k πρώτες δειγματικές ροπές m_r με τις k πρώτες πληθυσμιακές ροπές $E(X^r)$, $r=1, 2, \dots, k$. Οι λύσεις αυτού του συστήματος ως προς $\theta_1, \theta_2, \dots, \theta_k$ καλούνται εκτιμήτριες ροπών των $\theta_1, \theta_2, \dots, \theta_k$ και συμβολίζονται με $\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_k$. Σε αρκετές περιπτώσεις η μέθοδος αυτή προσφέρει εκτιμήτριες με καλές ιδιότητες (π.χ. α.ε.ε.δ.) αλλά υπάρχουν και αρκετές περιπτώσεις όπου δεν μπορεί να εφαρμοστεί ή δεν οδηγεί σε καλές εκτιμήτριες.

Πριν περάσουμε σε κάποιες εφαρμογές είναι χρήσιμο να δούμε που βασίζεται η παραπάνω μέθοδος. Συγκεκριμένα έχουμε την επόμενη πρόταση.

Πρόταση 4.7. Η δειγματική ροπή

$$m_r = \frac{1}{n} \sum_{i=1}^n X_i^r$$

τάξεως r είναι συνεπής εκτιμήτρια της ροπής $\mu'_r = E(X_i^r)$ τάξεως r ενός πληθυσμού (υποθ. ότι $V(X_i^k) < \infty$).

Απόδειξη. Έστω $X_1, X_2, \dots, X_n$ ένα τ.δ. από μία κατανομή με συνάρτηση σ.κ. $F(x; \theta)$. Ισχύει ότι

$$E(m_r) = \frac{1}{n} \sum_{i=1}^n E(X_i^r) = \frac{1}{n} n \mu'_r = \mu'_r$$

και

$$V(m_r) = \frac{1}{n^2} \sum_{i=1}^n V(X_i^r) = \frac{1}{n^2} n V(X_1^r) = \frac{V(X_1^r)}{n} \xrightarrow{n \rightarrow \infty} 0$$

και άρα η m_r είναι συνεπής εκτιμήτρια του μ'_r .

Από την παραπάνω πρόταση προκύπτει ότι, για πολύ μεγάλα δείγματα, πρακτικά θα είναι $m_r \approx \mu'_r$. Επομένως, για μεγάλο n θα ισχύει ότι,

$$m_r = \mu'_r(\theta_1, \theta_2, \dots, \theta_k), \quad r=1, 2, \dots, k,$$

από όπου λύνοντας παίρνουμε τις εκτιμήτριες ροπών.

Άσκηση. 4.10. Δεδομένου ενός τ.δ. $X_1, X_2, \dots, X_n$ να βρεθούν οι εκτιμήτριες ροπών των παραμέτρων: i) θ αν $X_i \sim$ εκθετική με παράμετρο θ , ii) θ αν $X_i \sim$ ομοιόμορφη στο $(0, \theta)$, iii) λ αν $X_i \sim$ Poisson με παράμετρο λ .

Λύση. i) Έχουμε μία άγνωστη παράμετρο $\theta_1 = \theta$ και επομένως θα πάρουμε μία εξίσωση με έναν άγνωστο ($k=1$):

$$m_1 = \mu'_1(\theta_1) \Leftrightarrow m_1 = E(X_i) \Leftrightarrow \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{\theta}$$

και άρα

$$\tilde{\theta} = \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^{-1} = \frac{1}{\bar{X}}.$$

Άρα εάν θέλουμε να εκτιμήσουμε την παράμετρο θ ενός πληθυσμού που ακολουθεί εκθετική κατανομή, τότε λαμβάνοντας ένα δείγμα $X_1, X_2, \dots, X_n$ θα έχουμε ότι $\tilde{\theta} = 1/\bar{X}$.

ii) Έχουμε και πάλι μία άγνωστη παράμετρο $\theta_1 = \theta$ και επομένως θα πάρουμε μία εξίσωση με έναν άγνωστο ($k=1$):

$$m_1 = \mu'_1(\theta_1) \Leftrightarrow m_1 = E(X_i) \Leftrightarrow \frac{1}{n} \sum_{i=1}^n X_i = \frac{\theta}{2},$$

από όπου προκύπτει ότι

$$\tilde{\theta} = \frac{2}{n} \sum_{i=1}^n X_i = 2\bar{X}.$$

Όμοια, εάν θέλουμε να εκτιμήσουμε την παράμετρο θ ενός πληθυσμού που ακολουθεί ομοιόμορφη κατανομή στο $(0, \theta)$, τότε μπορούμε να πάρουμε ότι $\tilde{\theta} = 2\bar{X}$.

iii) Εδώ $\theta_1 = \lambda$ και επομένως θα πάρουμε και πάλι μία εξίσωση με έναν άγνωστο ($k=1$):

$$m_1 = \mu'_1(\theta_1) \Leftrightarrow m_1 = E(X_i) \Leftrightarrow \frac{1}{n} \sum_{i=1}^n X_i = \lambda$$

και τελικά

$$\tilde{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

Άσκηση. 4.11. Δεδομένου ενός τ.δ. $X_1, X_2, \dots, X_n$ από $N(\mu, \sigma^2)$ να βρεθούν οι εκτιμήτριες ροπών των παραμέτρων μ, σ^2 .

Λύση. Έχουμε δύο άγνωστες παραμέτρους $\theta_1 = \mu, \theta_2 = \sigma^2$ και επομένως θα πάρουμε δύο εξισώσεις με δύο αγνώστους ($k=2$):

$$\begin{cases} m_1 = \mu'_1(\theta_1, \theta_2) \\ m_2 = \mu'_2(\theta_1, \theta_2) \end{cases} \Leftrightarrow \begin{cases} m_1 = E(X_i) \\ m_2 = E(X_i^2) \end{cases} \Leftrightarrow \begin{cases} \frac{1}{n} \sum_{i=1}^n X_i = \mu \\ \frac{1}{n} \sum_{i=1}^n X_i^2 = \sigma^2 + \mu^2 \end{cases}$$

και άρα λύνοντας ως προς μ, σ^2 προκύπτουν οι εκτιμήτριες ροπών

$$\tilde{\mu} = \bar{X}, \quad \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} S^2.$$

Επομένως η εκτιμήτρια ροπών του μ ενός κανονικού πληθυσμού είναι και πάλι ο δειγματικός μέσος ενώ η αντίστοιχη εκτιμήτρια του σ^2 είναι η $(n-1)S^2/n$ που είναι συνεπής αλλά δεν είναι αμερόληπτη εκτιμήτρια (είναι όμως ασυμπτωτικά αμερόληπτη).

Άσκηση 4.12. Έστω ένα τ.δ. $X_1, X_2, \dots, X_n$ από την ομοιόμορφη στο (a, b) κατανομή. Να βρεθούν οι εκτιμήτριες ροπών των παραμέτρων a, b .

Λύση. Έχουμε δύο άγνωστες παραμέτρους $\theta_1=a$, $\theta_2=b$ και επομένως θα πάρουμε δύο εξισώσεις με δύο αγνώστους ($k=2$):

$$\begin{cases} m_1 = \mu'_1(\theta_1, \theta_2) \\ m_2 = \mu'_2(\theta_1, \theta_2) \end{cases} \Leftrightarrow \begin{cases} m_1 = E(X_i) \\ m_2 = E(X_i^2) = \text{Var}(X_i) + E(X_i)^2 \end{cases} \Leftrightarrow \begin{cases} \frac{1}{n} \sum_{i=1}^n X_i = \frac{a+b}{2} \\ \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2 \end{cases}$$

και ισοδύναμα,

$$\begin{cases} \bar{X} = \frac{a+b}{2} \\ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{(b-a)^2}{12} \end{cases} \Leftrightarrow \begin{cases} 2\bar{X} = a+b \\ \sqrt{\frac{12}{n} \sum_{i=1}^n (X_i - \bar{X})^2} = b-a \end{cases}$$

Λύνοντας ως προς a , b προκύπτουν οι εκτιμήτριες ροπών

$$\tilde{a} = \bar{X} - \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}, \quad \tilde{b} = \bar{X} + \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

Άρα εάν θέλουμε να εκτιμήσουμε τις παραμέτρους a , b ενός πληθυσμού που ακολουθεί την ομοιόμορφη στο (a, b) κατανομή, τότε λαμβάνοντας ένα δείγμα $X_1, X_2, \dots, X_n$ μπορούμε να χρησιμοποιήσουμε τις παραπάνω εκτιμήτριες $\tilde{a}$, $\tilde{b}$.

β. Μέθοδος μέγιστης πιθανοφάνειας.

Η μέθοδος αυτή είναι αρκετά ισχυρή διότι, με μία σχετικά εύκολη διαδικασία, προσφέρει εκτιμήτριες με πολύ καλές ιδιότητες. Όπως και στην προηγούμενη παράγραφο, θεωρούμε αρχικά ένα τ.δ. $X_1, X_2, \dots, X_n$ από μία κατανομή με συνάρτηση σ.π. ή σ.π.π. $f(x; \theta)$. Έστω ότι επιθυμούμε να εκτιμήσουμε το θ ή οποιαδήποτε συνάρτηση του, $g(\theta)$. Η από κοινού σ.π.π ή σ.π. των $X_1, X_2, \dots, X_n$ θεωρούμενη ως συνάρτηση του θ , δηλαδή η

$$L_X(\theta) = f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta),$$

θα καλείται συνάρτηση πιθανοφάνειας του δείγματος. Η εκτιμήτρια μέγιστης πιθανοφάνειας δίνεται στον επόμενο ορισμό.

Ορισμός 4.8. Μία εκτιμήτρια $\hat{\theta}$ θα καλείται *εκτιμήτρια μέγιστης πιθανοφάνειας* (ε.μ.π. ή MLE) της παραμέτρου (ή των παραμέτρων) θ αν ισχύει ότι

$$L_X(\hat{\theta}) = \sup_{\theta \in \Theta} L_X(\theta) = \sup_{\theta \in \Theta} f(X_1, X_2, \dots, X_n; \theta).$$

Άρα η ε.μ.π. $\hat{\theta}$ είναι στην ουσία η τιμή της παραμέτρου θ που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας $L(\theta)$. Συνήθως, αντί να αναζητούμε το σημείο μεγίστου της $L(\theta)$, είναι πιο εύκολο να αναζητούμε το σημείο μεγίστου του $\ln L(\theta)$ (έχουν το ίδιο σημείο μεγίστου διότι η συνάρτηση $\ln$ είναι αύξουσα). Προφανώς, αν $k=1$ ($\theta = \theta$) και η συνάρτηση $\ln L(\theta)$ παραγωγίζεται σε ολόκληρο τον παραμετρικό χώρο Θ (και Θ ανοικτό διάστημα) μπορούμε να βρούμε το σημείο μεγίστου μέσα από τη λύση της εξίσωσης $(\ln L(\theta))' = 0$, ελέγχοντας παράλληλα ότι $(\ln L(\hat{\theta}))'' < 0$.

Άσκηση 4.13. Έστω $X_1, X_2, \dots, X_n$ τ.δ. από κατανομή Poisson με μέση τιμή λ . Βρείτε την ε.μ.π. του λ .

Λύση. Είναι γνωστό ότι η σ.π. της κατανομής Poisson (λ) είναι

$$f(x;\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x=0,1,\dots, \quad \lambda>0$$

και άρα ο λογάριθμος της συνάρτησης πιθανοφάνειας του δείγματος θα δίνεται από τον τύπο

$$\begin{aligned} \ln L(\lambda) &= \ln \prod_{i=1}^n f(x_i; \lambda) = \sum_{i=1}^n \ln f(x_i; \lambda) = \sum_{i=1}^n \ln \left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right) \\ &= \sum_{i=1}^n (\ln e^{-\lambda} + \ln \lambda^{x_i} - \ln x_i!) = \sum_{i=1}^n (-\lambda + x_i \ln \lambda - \ln x_i!) = -n\lambda + \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i!. \end{aligned}$$

Παρατηρούμε ότι η παραπάνω συνάρτηση του λ είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο $(0, \infty)$. Επομένως, τα πιθανά λ για τα οποία μεγιστοποιείται η $\ln L(\lambda)$ θα είναι ρίζες της εξίσωσης $(\ln L(\lambda))' = 0$. Ειδικότερα θα έχουμε ότι

$$\frac{\partial}{\partial \lambda} (\ln L(\lambda)) = \frac{\partial}{\partial \lambda} \left(-n\lambda + \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i! \right) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i = 0$$

και επομένως

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X},$$

η οποία συμπίπτει με την εκτιμήτρια ροπών του λ . Απομένει φυσικά να επαληθεύσουμε ότι η παραπάνω τιμή αποτελεί μέγιστο. Πράγματι, ισχύει ότι

$$\frac{\partial^2}{\partial \lambda^2} (\ln L(\hat{\lambda})) = -\frac{1}{(\hat{\lambda})^2} \sum_{i=1}^n x_i = -\frac{1}{\bar{X}^2} \sum_{i=1}^n x_i = -\frac{n\bar{X}}{\bar{X}^2} = -\frac{n}{\bar{X}} < 0.$$

Αξίζει να παρατηρήσουμε ότι οι ε.μ.π. έχουν ασυμπτωτικά πολύ καλές ιδιότητες. Πιο συγκεκριμένα αν $\hat{\theta}_n$ είναι ε.μ.π. του θ από δείγμα μεγέθους n , αποδεικνύεται ότι, κάτω από κατάλληλες συνθήκες ομαλότητας, η $\hat{\theta}_n$ ακολουθεί ασυμπτωτικά κανονική κατανομή. Επίσης, $E(\hat{\theta}_n) \rightarrow \theta$ και $V(\hat{\theta}_n) \rightarrow 0$ και επομένως η $\hat{\theta}_n$ είναι συνεπής εκτιμήτρια του θ . Αποδεικνύεται επίσης ότι ασυμπτωτικά ($n \rightarrow \infty$), η $\hat{\theta}_n$ είναι άριστη εκτιμήτρια του θ . Άρα, η βέλτιστη εκτιμήτρια μιάς παραμέτρου θ όταν έχουμε ένα σχετικά μεγάλο δείγμα, θα είναι η ε.μ.π. του θ .

Άσκηση. 4.14. Δεδομένου ενός τ.δ. $X_1, X_2, \dots, X_n$ από $N(\mu, \sigma^2)$ να βρεθούν: i) η ε.μ.π. του μ αν σ^2 γνωστό, ii) η ε.μ.π. του σ^2 αν μ γνωστό, iii) οι ε.μ.π. των παραμέτρων μ, σ^2 (και τα δύο είναι άγνωστα).

Λύση. i) Είναι γνωστό ότι η σ.π. της $N(\mu, \sigma^2)$ κατανομής είναι

$$f(x; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in R$$

και άρα ο λογάριθμος της συνάρτησης πιθανοφάνειας του δείγματος θα δίνεται από τον τύπο

$$\begin{aligned} \ln L(\mu) &= \ln \prod_{i=1}^n f(x_i; \mu) = \sum_{i=1}^n \ln f(x_i; \mu) = \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) \\ &= -n \ln \sqrt{2\pi\sigma^2} - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}. \end{aligned}$$

Η παραπάνω συνάρτηση του μ είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο $(-\infty, \infty)$. Επομένως, τα πιθανά μ για τα οποία μεγιστοποιείται η $\ln L(\mu)$ θα είναι ρίζες της εξίσωσης $(\ln L(\mu))' = 0$. Ειδικότερα,

$$\frac{\partial}{\partial \mu}(\ln L(\mu)) = \frac{\partial}{\partial \mu} \left(-n \ln \sqrt{2\pi\sigma^2} - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \right) = \sum_{i=1}^n \frac{2(x_i - \mu)}{2\sigma^2} = \frac{1}{\sigma^2} (\sum_{i=1}^n x_i - n\mu) = 0$$

και άρα

$$\hat{\mu} = \bar{X},$$

η οποία συμπίπτει με την εκτιμήτρια ροπών του μ . Η παραπάνω τιμή αποτελεί μέγιστο διότι

$$\frac{\partial^2}{\partial \mu^2}(\ln L(\mu)) = \frac{\partial}{\partial \mu} \left(\frac{1}{\sigma^2} \sum_{i=1}^n x_i - \frac{n}{\sigma^2} \mu \right) = -\frac{n}{\sigma^2} < 0.$$

ii) Ο λογάριθμος της συνάρτησης πιθανοφάνειας του δείγματος θα ισούται (θέτουμε $\theta = \sigma^2$)

$$\begin{aligned} \ln L(\theta) &= \ln \prod_{i=1}^n f(x_i; \theta) = \sum_{i=1}^n \ln f(x_i; \theta) = \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i - \mu)^2}{2\theta}} \right) \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \theta - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\theta}. \end{aligned}$$

Η παραπάνω συνάρτηση του θ είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο $(0, \infty)$. Επομένως, όμοια με παραπάνω, θα έχουμε ότι

$$\frac{\partial}{\partial \theta}(\ln L(\theta)) = \frac{\partial}{\partial \theta} \left(-\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \theta - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\theta} \right) = -\frac{n}{2\theta} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\theta^2} = 0$$

και άρα

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2,$$

η οποία, όπως έχουμε υποθέσει, προϋποθέτει τη γνώση του μ . Η παραπάνω τιμή αποτελεί μέγιστο διότι

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2}(\ln L(\hat{\theta})) &= \frac{\partial}{\partial \theta} \left(-\frac{n}{2\theta} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\theta^2} \right)_{\theta=\hat{\theta}} = \frac{n}{2(\hat{\theta})^2} - \frac{1}{(\hat{\theta})^3} \sum_{i=1}^n (x_i - \mu)^2 = \\ &= \frac{n}{2(\hat{\theta})^2} - \frac{n}{(\hat{\theta})^3} \hat{\theta} = \frac{n}{2(\hat{\theta})^2} - \frac{n}{(\hat{\theta})^2} = -\frac{n}{2(\hat{\theta})^2} < 0. \end{aligned}$$

iii) Ο λογάριθμος της συνάρτησης πιθανοφάνειας του δείγματος θα ισούται (θέτουμε $\theta = \sigma^2$)

$$\begin{aligned} \ln L(\mu, \theta) &= \ln \prod_{i=1}^n f(x_i; \mu, \theta) = \sum_{i=1}^n \ln f(x_i; \mu, \theta) = \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i - \mu)^2}{2\theta}} \right) \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \theta - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\theta}. \end{aligned}$$

Η παραπάνω συνάρτηση των θ, μ είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο $(0, \infty) \times (-\infty, \infty)$. Επομένως, για να βρούμε το μέγιστο εξισώνουμε τις μερικές παραγώγους με το 0. Συγκεκριμένα θα έχουμε ότι

$$\frac{\partial}{\partial \mu}(\ln L(\mu, \theta)) = \frac{\partial}{\partial \mu} \left(-\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \theta - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\theta} \right) = \sum_{i=1}^n \frac{(x_i - \mu)}{\theta} = 0$$

$$\frac{\partial}{\partial \theta}(\ln L(\mu, \theta)) = \frac{\partial}{\partial \theta} \left(-\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \theta - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\theta} \right) = -\frac{n}{2\theta} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\theta^2} = 0$$

και άρα λύνοντας ως προς μ , θ προκύπτει ότι

$$\begin{cases} \hat{\mu} = \bar{X} \\ \hat{\theta} = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} S^2. \end{cases}$$

Παρατηρούμε ότι συμπίπτουν με τις εκτιμήτριες ροπών.

Άσκηση. 4.15. Δεδομένου ενός τ.δ. $X_1, X_2, \dots, X_n$ να βρεθούν οι ε.μ.π. των παραμέτρων: i) θ αν $X_i \sim$ ομοιόμορφη στο $(0, \theta)$. ii) θ αν $X_i \sim$ εκθετική με παράμετρο θ .

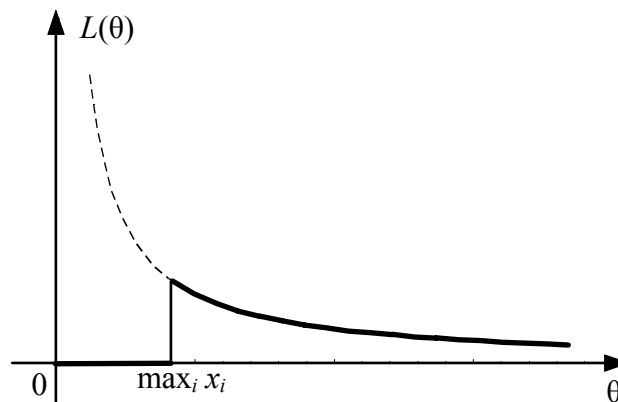
Λύση. i) Η σ.π. της ομοιόμορφης στο $(0, \theta)$ κατανομής είναι

$$f(x; \theta) = \begin{cases} \frac{1}{\theta}, & x \in (0, \theta) \\ 0, & x > \theta \end{cases}$$

και άρα η συνάρτηση πιθανοφάνειας του δείγματος θα είναι

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \begin{cases} \frac{1}{\theta^n}, & \text{αν } x_1, x_2, \dots, x_n \in (0, \theta) \\ 0 & \text{αν } \exists x_i > \theta \end{cases} = \frac{1}{\theta^n} I_x(\theta),$$

όπου $I_x(\theta) = 1$ αν $x_1, \dots, x_n < \theta \Leftrightarrow \theta > \max_i x_i$ και $I_x(\theta) = 0$ αν $\theta \leq \max_i x_i$. Η παραπάνω συνάρτηση του θ δεν είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο $(0, \infty)$. Πράγματι, δεν είναι παραγωγίσιμη στο σημείο $\theta = \max_i x_i$. Επομένως δεν μπορούμε να ακολουθήσουμε τη γνωστή διαδικασία παραγωγίζοντας και εξισώνοντας με το 0. Εναλλακτικά, θα βρούμε το θ για το οποίο μεγιστοποιείται η $L(\theta)$ από το γράφημά της. Η γραφική παράσταση της συνάρτησης πιθανοφάνειας $L(\theta)$ θα είναι της μορφής:



Πράγματι, η συνάρτηση $h(\theta) = \frac{1}{\theta^n} I_x(\theta)$ είναι ίση με 0 αν $\theta \leq \max_i x_i$ ενώ για $\theta > \max_i x_i$ είναι φθίνουσα. Επομένως το supremum των τιμών της συνάρτησης λαμβάνεται στο $\max_i x_i$. Από τον ορισμό της ε.μ.π. συμπεραίνουμε τελικά ότι

$$\hat{\theta} = \max \{X_1, X_2, \dots, X_n\}.$$

Η εκτιμήτρια αυτή είναι καλύτερη από την εκτιμήτρια ροπών $\tilde{\theta} = 2\bar{X}$ (αποδεικνύεται ότι έχει μικρότερο μέσο τετραγωνικό σφάλμα, ενώ είναι ασυμπτωτικά αμερόληπτη).

ii) Η σ.π. της εκθετικής κατανομής με παράμετρο θ είναι

$$f(x; \theta) = \theta e^{-\theta x}, \quad x > 0.$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας του δείγματος θα δίνεται από τον τύπο

$$\ln L(\theta) = \ln \prod_{i=1}^n f(x_i; \theta) = \sum_{i=1}^n \ln f(x_i; \theta) = \sum_{i=1}^n \ln(\theta e^{-\theta x_i}) = \sum_{i=1}^n (\ln \theta - \theta x_i) = n \ln \theta - \theta \sum_{i=1}^n x_i.$$

Η παραπάνω συνάρτηση του θ είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο $(0, \infty)$. Επομένως, τα πιθανά θ για τα οποία μεγιστοποιείται η $\ln L(\theta)$ θα είναι ρίζες της $(\ln L(\theta))' = 0$. Ειδικότερα,

$$\frac{\partial}{\partial \theta} (\ln L(\theta)) = \frac{\partial}{\partial \theta} \left(n \ln \theta - \theta \sum_{i=1}^n x_i \right) = \frac{n}{\theta} - \sum_{i=1}^n x_i = 0$$

και άρα

$$\hat{\theta} = \frac{1}{\bar{X}}.$$

η οποία συμπίπτει με την εκτιμήτρια ροπών του θ . Η παραπάνω τιμή αποτελεί μέγιστο διότι

$$\frac{\partial^2}{\partial \theta^2} (\ln L(\theta)) = \frac{\partial}{\partial \theta} \left(\frac{n}{\theta} - \sum_{i=1}^n x_i \right) = -\frac{n}{\theta^2} < 0.$$

Άσκηση 4.16. Έστω ένα τ.δ. $X_1, X_2, \dots, X_n$ από τη διωνυμική (n, p) κατανομή. Να βρεθεί η ε.μ.π. της παραμέτρου p .

Λύση. Η σ.π. της διωνυμικής (n, p) κατανομής είναι

$$f(x, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n.$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας του δείγματος δίνεται από τον τύπο

$$\begin{aligned} \ln L(p) &= \ln \prod_{i=1}^n f(x_i; p) = \sum_{i=1}^n \ln f(x_i; p) = \sum_{i=1}^n \ln \left(\binom{n}{x_i} p^{x_i} (1-p)^{n-x_i} \right) \\ &= \sum_{i=1}^n \ln \binom{n}{x_i} + \sum_{i=1}^n x_i \ln p + \sum_{i=1}^n (n-x_i) \ln(1-p). \end{aligned}$$

Η παραπάνω συνάρτηση του p είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο $(0, 1)$. Επομένως, θα έχουμε

$$\begin{aligned} \frac{\partial}{\partial p} (\ln L(p)) &= \frac{\partial}{\partial p} \left(\sum_{i=1}^n \ln \binom{n}{x_i} + \sum_{i=1}^n x_i \ln p + \sum_{i=1}^n (n-x_i) \ln(1-p) \right) \\ &= \frac{1}{p} \sum_{i=1}^n x_i - \frac{1}{1-p} \sum_{i=1}^n (n-x_i) = \frac{n}{p} \bar{x} - \frac{n}{1-p} (n - \bar{x}) = 0 \end{aligned}$$

και άρα, λύνοντας ως προς p θα έχουμε ότι

$$\hat{p} = \frac{\bar{X}}{v}.$$

Η παραπάνω τιμή αποτελεί μέγιστο διότι

$$\frac{\partial^2}{\partial p^2}(\ln L(p)) = \frac{\partial}{\partial p} \left(\frac{n}{p} \bar{x} - \frac{n}{1-p} (v - \bar{x}) \right)_{p=\hat{p}} = \frac{-nv}{p(1-p)} < 0.$$

Πρόταση 4.8. (αναλλοίωτου της ε.μ.π.) Αν g είναι μία 1-1 συνάρτηση τότε η ε.μ.π. της π.σ. $g(\theta)$ είναι η $g(\hat{\theta})$ (δηλαδή, $\hat{g}(\theta) = g(\hat{\theta})$).

Άσκηση 4.17. Δεδομένου ενός τ.δ. $X_1, X_2, \dots, X_n$ από $N(\mu, \sigma^2)$ να βρεθεί η ε.μ.π. του σ .

Λύση. Από την Άσκηση 4.14 έχουμε βρει ότι οι ε.μ.π. των μ, σ^2 από κανονική κατανομή είναι

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} S^2.$$

Αρα, χρησιμοποιώντας και την Πρόταση 4.8.(i), θα έχουμε ότι η ε.μ.π. του $g(\sigma^2) = \sigma = \sqrt{\sigma^2}$ θα είναι (η $g(\theta) = \sqrt{\theta}$ είναι 1-1 για $\theta > 0$)

$$g(\hat{\sigma}^2) = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

Άσκηση 4.18. Έστω ότι η ημερήσια ζήτηση ενός προϊόντος από ένα πολυκατάστημα ακολουθεί κατανομή Poisson (λ) (και είναι ανεξάρτητη από τις άλλες ημέρες). Για λόγους συντήρησης, το πολυκατάστημα δεν θα διαθέτει το προϊόν αυτό για k ημέρες. Χρησιμοποιώντας στοιχεία που υπάρχουν για τη ζήτηση του προϊόντος τις n τελευταίες ημέρες, να εκτιμήσετε την πιθανότητα να ζητηθεί το προϊόν αυτό κατά τη διάρκεια της συντήρησης (χρησιμ. ε.μ.π.).

Λύση. Έστω $X_1, X_2, \dots, X_n$ η ζήτηση του προϊόντος τις n τελευταίες ημέρες. Σύμφωνα με την εκφώνηση οι τ.μ. $X_1, X_2, \dots, X_n$ είναι ένα τ.δ. από την Poisson με μέση τιμή λ . Έστω $Y_1, Y_2, \dots, Y_k$ οι τ.μ. που εκφράζουν τη ζήτηση τις k ημέρες της συντήρησης. Ζητείται η πιθανότητα

$$P(Y_1 > 0 \text{ ή } Y_2 > 0 \text{ ή } \dots \text{ ή } Y_k > 0) = 1 - P(Y_1 = 0, Y_2 = 0, Y_k = 0) = 1 - P(Y_1 = 0)P(Y_2 = 0)P(Y_k = 0).$$

Επειδή η ημερήσια ζήτηση ακολουθεί κατανομή Poisson η παραπάνω πιθανότητα θα είναι ίση με

$$1 - e^{-\lambda} \cdot e^{-\lambda} \dots e^{-\lambda} = 1 - e^{-k\lambda}$$

Αρα, δεδομένου ενός τ.δ. $X_1, X_2, \dots, X_n$ από την Poisson (λ), ζητείται η ε.μ.π. της

$$g(\lambda) = 1 - e^{-k\lambda}.$$

Γνωρίζουμε από προηγούμενη άσκηση ότι η ε.μ.π. του λ είναι το $\bar{X}$ και επομένως, σύμφωνα και με την Πρόταση 4.8.(i), η ε.μ.π. της πιθανότητας να ζητηθεί το προϊόν αυτό κατά τη διάρκεια της συντήρησης είναι

$$\hat{g}(\lambda) = 1 - e^{-k\bar{X}}.$$