Ασυμπτωτική Στατιστική – Ασκήσεις 2

2.1. Έστω $X_1, X_2, \ldots, X_n$ τυχαίο δείγμα με $\mathbb{E}|X_1|^{2k} < \infty$ για κάποιον ακέραιο $k \geq 2$ και $\mathbb{E}(X_1) = \mu$. Να βρείτε την οριακή κατανομή της ακολουθίας των διδιάστατων τυχαίων διανυσμάτων

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ M_{k,n} - \mu_k \end{pmatrix}$$

όπου $\overline{X}_n$ η ακολουθία των δειγματικών μέσων,

$$M_{k,n} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^k$$

η ακολουθία των δειγματικών κεντρικών ροπών και $\mu_k = \mathbb{E}(X_1 - \mu)^k$ η πληθυσμιακή κεντρική ροπή τάξης k.

2.2. Έστω $X_1, X_2, \ldots, X_n$ τυχαίο δείγμα με $\mathbb{E}|X_1|^{2k} < \infty$ για κάποιον ακέραιο $k \geq 3$ και $\mathbb{E}(X_1) = \mu$. Με το συμβολισμό της προηγούμενης άσκησης να βρείτε την οριακή κατανομή της ακολουθίας

$$\sqrt{n}\begin{pmatrix} M_{s,n} - \mu_s \\ M_{k,n} - \mu_k \end{pmatrix},$$

όπου s ακέραιος με $2 \leq s < k$.

2.3. Έστω $X_1, X_2, \ldots, X_n$ τυχαίο δείγμα με $\mathbb{E}|X_1|^{2k} < \infty$ για κάποιον ακέραιο $k \geq 3$ και $\mathbb{E}(X_1) = \mu$. Με το συμβολισμό της προηγούμενης άσκησης να βρείτε την οριακή κατανομή της

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ M_{s,n} - \mu_s \\ M_{k,n} - \mu_k \end{pmatrix},$$

όπου s ακέραιος με $2 \leq s < k$.

2.4. [Γενική περίπτωση από την οποία συνάγονται οι ασκήσεις 1.16, 2.1, 2.2 και 2.3]. Έστω $X_1, X_2, \ldots, X_n$ τυχαίο δείγμα με $\mathbb{E}|X_1|^{2k} < \infty$ για κάποιον ακέραιο $k \geq 2$ και $\mathbb{E}(X_1) = \mu$ $\mathrm{Var}(X_1) = \sigma^2 = \mu_2$. Να δείξετε ότι

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ M_{2,n} - \sigma^2 \\ M_{3,n} - \mu_3 \\ \vdots \\ M_{k,n} - \mu_k \end{pmatrix} \xrightarrow{\mathrm{d}} N_k(\mathbf{0}, \mathbf{\Sigma}),$$

όπου $\mathbf{\Sigma}$ ο πίνακας διασποράς του τυχαίου διανύσματος

$$\boldsymbol{Y} := \begin{pmatrix} X_1 \\ (X_1 - \mu)^2 \\ (X_1 - \mu)^3 - 3\sigma^2(X_1 - \mu) \\ \vdots \\ (X_1 - \mu)^k - k\mu_{k-1}(X_1 - \mu) \end{pmatrix}.$$

2.5. Ο συντελεστής λοξότητας (skewness) μίας κατανομής (ή της αντίστοιχης τυχαίας μεταβλητής X) είναι ο αριθμός $\beta := \mathbb{E}[(X-\mu)^3]/\sigma^3 = \mu_3/\mu_2^{3/2}$, ο οποίος είναι καλά ορισμένος όταν $\sigma^2 > 0$ και $\mathbb{E}|X|^3 < \infty$. Ο συντελεστής λοξότητας είναι ένας δείκτης της ασυμμετρίας της κατανομής του πληθυσμού.

Θεωρούμε ένα τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ με $\mathbb{E}|X_1|^6 < \infty$ και κατασκευάζουμε τον αντίστοιχο δειγματικό συντελεστή λοξότητας

$$\widehat{\beta}_n := \frac{M_{3,n}}{M_{2,n}^{3/2}}.$$

Να αποδείξετε ότι η ακολουθία $\widehat{\beta}_n$ είναι ασυμπτωτικά κανονική εκτιμήτρια του β.
[**Υπόδειξη:** Μπορείτε να χρησιμοποιήσετε την άσκηση 2.2 με $s=2, k=3$.]

2.6. Ο συντελεστής κυρτότητας (kurtosis) μίας κατανομής (ή της αντίστοιχης τυχαίας μεταβλητής X) είναι ο αριθμός $\gamma := \mathbb{E}[(X-\mu)^4]/\sigma^4 = \mu_4/\mu_2^2$, ο οποίος είναι καλά ορισμένος όταν $\sigma^2 > 0$ και $\mathbb{E}|X|^4 < \infty$. Ο συντελεστής κυρτότητας είναι ένας δείκτης της κυρτότητας της κατανομής του πληθυσμού (στις κανονικές κατανομές είναι $\gamma = 3$, οπότε $\gamma > 3$ και $\gamma < 3$ αντιστοιχεί σε λεπτόκυρτες και πλατύκυρτες κατανομές, αντίστοιχα).

Θεωρούμε ένα τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ με $\mathbb{E}|X_1|^8 < \infty$ και κατασκευάζουμε τον αντίστοιχο δειγματικό συντελεστή κυρτότητας

$$\widehat{\gamma}_n := \frac{M_{4,n}}{M_{2,n}^2}.$$

Να αποδείξετε ότι η ακολουθία $\widehat{\gamma}_n$ είναι ασυμπτωτικά κανονική εκτιμήτρια του γ.
[**Υπόδειξη:** Μπορείτε να χρησιμοποιήσετε την άσκηση 2.2 με $s=2, k=4$.]

2.7. Με τον συμβολισμό των προηγούμενων δύο ασκήσεων, να βρείτε την οριακή κατανομή της της ακολουθίας τυχαίων διανυσμάτων

$$\sqrt{n}\begin{pmatrix} \widehat{\beta}_n - \beta \\ \widehat{\gamma}_n - \gamma \end{pmatrix},$$

ενός τυχαίου δείγματος $X_1, X_2, \ldots, X_n$ από κατανομή με θετική διασπορά και πεπερασμένη όγδοη ροπή.

2.8. Ο αριθμός

$$\kappa := \frac{\sigma}{\mu} = \frac{\sqrt{\operatorname{Var}(X)}}{\mathbb{E}(X)}, \quad \mu \neq 0,$$

ονομάζεται συντελεστής μεταβλητότητας (coefficient of variation) της κατανομής (ή της αντίστοιχης τυχαίας μεταβλητής X) και είναι καλά ορισμένος όταν $\mu \neq 0$ και $\mathbb{E}|X|^2 < \infty$. Συνήθως χρησιμοποιείται για θετικά δεδομένα/θετικές τυχαίες μεταβλητές.

Θεωρούμε ένα τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ με $\mathbb{E}|X_1|^4 < \infty$ και κατασκευάζουμε τον αντίστοιχο δειγματικό συντελεστή μεταβλητότητας

$$\widehat{\kappa}_n := \frac{M_{2,n}^{1/2}}{\overline{X}_n}.$$

Να αποδείξετε ότι η ακολουθία $\widehat{\kappa}_n$ είναι ασυμπτωτικά κανονική εκτιμήτρια του κ.
[**Υπόδειξη:** Μπορείτε να χρησιμοποιήσετε την άσκηση 2.1 με $k = 2$.]

2.9. Ασυμπτωτική Κανονικότητα του Δειγματικού Συντελεστής Συσχέτισης. Έστω $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ ένα τυχαίο δείγμα από διδιάστατη κατανομή με $\mathbb{E}(X_1^2 + Y_1^2) < \infty$, $\mathbb{E}(X_1) = \mu_1$, $\mathbb{E}(Y_1) = \mu_2$, $\mathrm{Var}(X_1) = \sigma_1^2 > 0$, $\mathrm{Var}(Y_1) = \sigma_2^2 > 0$ και $\mathrm{Cov}(X_1, Y_1) = \rho\sigma_1\sigma_2$. Ο δειγματικός συντελεστής συσχέτισης ορίζεται ως

$$\widehat{\rho}_n := \frac{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\left[\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2\right]^{1/2}\left[\frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2\right]^{1/2}}.$$

Δείξτε αρχικά τα εξής:

(α1) Για κάθε c_1, $\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \sum_{i=1}^n (X_i - c_1)^2 - n(\overline{X}_n - c_1)^2$.

(α2) Για κάθε c_2, $\sum_{i=1}^n (Y_i - \overline{Y}_n)^2 = \sum_{i=1}^n (Y_i - c_2)^2 - n(\overline{Y}_n - c_2)^2$.

(α3) Για κάθε c_1, c_2, $\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n) = \sum_{i=1}^n (X_i - c_1)(Y_i - c_2) - n(\overline{X}_n - c_1)(\overline{Y}_n - c_2)$.

Στη συνέχεια, για $k, s \in \{0, 1, \ldots\}$ ορίζουμε τις τυποποιημένες κεντρικές ροπές της διδιάστατης κατανομής των (X_i, Y_i) ως εξής:

$$\mu_{k,s} := \mathbb{E}\left\{\left(\frac{X_1 - \mu_1}{\sigma_1}\right)^k \left(\frac{Y_1 - \mu_2}{\sigma_2}\right)^s\right\}.$$

Επομένως, $\mu_{0,1} = \mu_{1,0} = 0$, $\mu_{2,0} = \mu_{0,2} = 1$, $\mu_{1,1} = \rho$. Αντίστοιχα, ορίζονται οι αντίστοιχες 'ψευδοδειγματικές' ποσότητες

$$m_{k,s;n} := \frac{1}{n}\sum_{i=1}^n \left\{\left(\frac{X_i - \mu_1}{\sigma_1}\right)^k \left(\frac{Y_i - \mu_2}{\sigma_2}\right)^s\right\},$$

οι οποίες βέβαια δεν είναι στατιστικές συναρτήσεις διότι εξαρτώνται από τις άγνωστες μέσες τιμές και διασπορές των X_i και Y_i, είναι όμως αμερόληπτες: Αν $\mathbb{E}\{|X_1|^k |Y_1|^s\} < \infty$ τότε $\mathbb{E}[m_{k,s;n}] = \mu_{k,s}$.

Εκφράζοντας τον $\widehat{\rho}_n$ συναρτήσει των $\mu_{k,s;n}$ να αποδείξετε ότι αν $\mathbb{E}(X_1^4) < \infty$ και $\mathbb{E}(Y_1^4) < \infty$ τότε

$$\sqrt{n}(\widehat{\rho}_n - \rho) \xrightarrow{\mathrm{d}} N(0, \tau^2),$$

όπου η οριακή διασπορά ισούται με

$$\tau^2 = \frac{\rho^2}{4}(\mu_{4,0} + 2\mu_{2,2} + \mu_{0,4}) - \rho(\mu_{1,3} + \mu_{3,1}) + \mu_{2,2}.$$

Μπορείτε να δείξετε ότι $\tau^2 \geq 0$;

[**Υπόδειξη/Λύση:** Όσον αφορά την μη αρνητικότητα της τ^2, αρκεί να υπολογίστε την διασπορά της τυχαίας μεταβλητής

$$T = \lambda\left\{\left(\frac{X_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{Y_1 - \mu_2}{\sigma_2}\right)^2\right\} - \left(\frac{X_1 - \mu_1}{\sigma_1}\right)\left(\frac{Y_1 - \mu_2}{\sigma_2}\right)$$

και να δώσετε κατάλληλη τιμή στην σταθερά λ. Επίσης, χρησιμοποιώντας τις (α1)-(α3), παρατηρούμε ότι ο $\widehat{\rho}_n$ γράφεται εναλλακτικά ως εξής:

$$
\begin{aligned}
\widehat{\rho}_n &= \frac{\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\left[\sum_{i=1}^n (X_i - \overline{X}_n)^2\right]^{1/2} \left[\sum_{i=1}^n (Y_i - \overline{Y}_n)^2\right]^{1/2}} \\
&= \frac{\sum_{i=1}^n X_i Y_i - n\overline{X}_n \overline{Y}_n}{\left[\sum_{i=1}^n X_i^2 - n(\overline{X}_n)^2\right]^{1/2} \left[\sum_{i=1}^n Y_i^2 - n(\overline{Y}_n)^2\right]^{1/2}} \\
&= \frac{\sum_{i=1}^n (X_i - \mu_1)(Y_i - \mu_2) - n(\overline{X}_n - \mu_1)(\overline{Y}_n - \mu_2)}{\left[\sum_{i=1}^n (X_i - \mu_1)^2 - n(\overline{X}_n - \mu_1)^2\right]^{1/2} \left[\sum_{i=1}^n (Y_i - \mu_2)^2 - n(\overline{Y}_n - \mu_2)^2\right]^{1/2}} \\
&= \frac{\sum_{i=1}^n \left(\frac{X_i - \mu_1}{\sigma_1}\right)\left(\frac{Y_i - \mu_2}{\sigma_2}\right) - n\left(\frac{\overline{X}_n - \mu_1}{\sigma_1}\right)\left(\frac{\overline{Y}_n - \mu_2}{\sigma_2}\right)}{\left[\sum_{i=1}^n \left(\frac{X_i - \mu_1}{\sigma_1}\right)^2 - n\left(\frac{\overline{X}_n - \mu_1}{\sigma_1}\right)^2\right]^{1/2} \left[\sum_{i=1}^n \left(\frac{Y_i - \mu_2}{\sigma_2}\right)^2 - n\left(\frac{\overline{Y}_n - \mu_2}{\sigma_2}\right)^2\right]^{1/2}} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n \left(\frac{X_i - \mu_1}{\sigma_1}\right)\left(\frac{Y_i - \mu_2}{\sigma_2}\right) - \left(\frac{1}{n}\sum_{i=1}^n \frac{X_i - \mu_1}{\sigma_1}\right)\left(\frac{1}{n}\sum_{i=1}^n \frac{Y_i - \mu_2}{\sigma_2}\right)}{\left[\frac{1}{n}\sum_{i=1}^n \left(\frac{X_i - \mu_1}{\sigma_1}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n \frac{X_i - \mu_1}{\sigma_1}\right)^2\right]^{1/2} \left[\frac{1}{n}\sum_{i=1}^n \left(\frac{Y_i - \mu_2}{\sigma_2}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n \frac{Y_i - \mu_2}{\sigma_2}\right)^2\right]^{1/2}}.
\end{aligned}
$$

Η τελευταία έκφραση δείχνει ότι $\widehat{\rho}_n = g(m_{1,0;n}, m_{0,1;n}, m_{2,0;n}, m_{0,2;n}, m_{1,1;n})$ όπου $g : \mathbb{R}^5 \to \mathbb{R}$ η συνάρτηση με τύπο

$$
g(x_1, x_2, x_3, x_4, x_5) = \frac{x_5 - x_1 x_2}{(x_3 - x_1^2)^{1/2} (x_4 - x_2^2)^{1/2}}.
$$

Το ζητούμενο προκύπτει χρησιμοποιώντας την μέθοδο Δέλτα και το γεγονός ότι

$$
\sqrt{n}\begin{pmatrix} m_{1,0;n} & - & 0 \\ m_{0,1;n} & - & 0 \\ m_{2,0;n} & - & 1 \\ m_{0,2;n} & - & 1 \\ m_{1,1;n} & - & \rho \end{pmatrix} \xrightarrow{\mathrm{d}} N_5\left(\mathbf{0}, \begin{bmatrix} 1 & \rho & \mu_{3,0} & \mu_{1,2} & \mu_{2,1} \\ \rho & 1 & \mu_{2,1} & \mu_{0,3} & \mu_{1,2} \\ \mu_{3,0} & \mu_{2,1} & \mu_{4,0} - 1 & \mu_{2,2} - 1 & \mu_{3,1} - \rho \\ \mu_{1,2} & \mu_{0,3} & \mu_{2,2} - 1 & \mu_{0,4} - 1 & \mu_{1,3} - \rho \\ \mu_{2,1} & \mu_{1,2} & \mu_{3,1} - \rho & \mu_{1,3} - \rho & \mu_{2,2} - \rho^2 \end{bmatrix}\right).
$$

Η παραπάνω ασυμπωτική κανονικότητα είναι άμεση συνέπεια του κλασικού (πενταδιάστατου) Κεντρικού Οριακού Θεωρήματος, εφαρμοζόμενο στα ανεξάρτητα και ισόνομα τυχαία διανύσματα $\boldsymbol{W}_i := \left(\frac{X_i - \mu_1}{\sigma_1}, \frac{Y_i - \mu_2}{\sigma_2}, \left(\frac{X_i - \mu_1}{\sigma_1}\right)^2, \left(\frac{Y_i - \mu_2}{\sigma_2}\right)^2, \left(\frac{X_i - \mu_1}{\sigma_1}\right)\left(\frac{Y_i - \mu_2}{\sigma_2}\right)\right)^T$.]

2.10. Συνέχεια: z-μετασχηματισμός του Fisher.

Άν το τυχαίο δείγμα $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ προέρχεται από τη διδιάστατη κανονική κατανομή,

$$
N_2\left(\left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right], \left[\begin{array}{cc} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{array}\right]\right),
$$

να δείξετε ότι

$$
\sqrt{n}(\widehat{\rho}_n - \rho) \xrightarrow{\mathrm{d}} N\left(0, (1 - \rho^2)^2\right).
$$

Επίσης να προσδιορίσετε κατάλληλο διαπορο-σταθεροποιητικό μετασχηματισμό έτσι ώστε

$$
\sqrt{n}(g(\widehat{\rho}_n) - g(\rho)) \xrightarrow{\mathrm{d}} N(0, 1),
$$

και με βάση το μετασχηματισμό αυτό να κατασκευστεί ένα ασυμπτωτικό διάστημα εμπιστοσύνης για το ρ, με ασυμπτωτικό συντελεστή εμπιστοσύνης $1-\alpha$.

2.11. Έστω ότι $\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\theta}) \xrightarrow{\mathrm{d}} \boldsymbol{W} = (W_1, \ldots, W_k)^T$, και ας υποθέσουμε ότι οι συναρτήσεις $g_i : \mathbb{R}^k \to \mathbb{R}$ $(i = 1, 2)$ έχουν συνεχείς μερικές παραγώγους πρώτης τάξεως σε μία περιοχή του $\boldsymbol{\theta}$. Επιπροσθέτως υποθέτουμε ότι $\nabla g_1(\boldsymbol{\theta}) = \mathbf{0}$ και ότι η g_1 έχει συνεχείς παραγώγους δεύτερης τάξης σε μία περιοχή του $\boldsymbol{\theta}$. Τι μπορούμε να συμπεράνουμε για την ασυμπτωτική συμπεριφορά της

$$\sqrt{n}\begin{pmatrix} \sqrt{n}[g_1(\boldsymbol{T}_n) - g_1(\boldsymbol{\theta})] \\ g_2(\boldsymbol{T}_n) - g_2(\boldsymbol{\theta}) \end{pmatrix}?$$

2.12. Έστω $X_1, \ldots, X_n$ τυχαίο δείγμα από την Cauchy με πυκνότητα

$$f(x;\theta) = \frac{1}{\pi(1 + (x-\theta)^2)}, \quad x \in \mathbb{R}, \quad \theta \in \Theta = \mathbb{R}.$$

(α) Να δειχθεί ότι ο πληροφοριακός αριθμός του Fisher ισούται με $I(\theta) = 1/2$.

(β) Να κατασκευαστεί ασυμπτωτικά κανονική αποδοτική εκτιμήτρια για το θ.

(γ) Να βρεθεί η σχετική (απόλυτη) ασυμπτωτική αποδοτικότητα της εκτιμήτριας $X_{[\frac{n+1}{2}]:n}$ (δειγματική διάμεσος) σε σχέση με την εκτιμήτρια που βρήκατε στο ερώτημα (β).

2.13. Έστω $X_1, \ldots, X_n$ τυχαίο δείγμα από την κατανομή Logistic με πυκνότητα

$$f(x;\theta) = \frac{e^{-(x-\theta)}}{(1 + e^{-(x-\theta)})^2}, \quad x \in \mathbb{R}, \quad \theta \in \Theta = \mathbb{R}.$$

(α) Να δειχθεί ότι ο πληροφοριακός αριθμός του Fisher ισούται με $I(\theta) = 1/3$.

(β) Να κατασκευαστεί ασυμπτωτικά κανονική αποδοτική εκτιμήτρια για το θ.

(γ) Να βρεθεί η σχετική (απόλυτη) ασυμπτωτική αποδοτικότητα της εκτιμήτριας $X_{[\frac{n+1}{2}]:n}$ (δειγματική διάμεσος) σε σχέση με την εκτιμήτρια που βρήκατε στο ερώτημα (β).

(δ) Να βρεθεί η σχετική (απόλυτη) ασυμπτωτική αποδοτικότητα του δειγματικού μέσου σε σχέση με την εκτιμήτρια που βρήκατε στο ερώτημα (β).

(ε) Να βρεθεί η σχετική ασυμπτωτική αποδοτικότητα της δειγματικής διαμέσου σε σχέση με τον δειγματικό μέσο.

2.14*. Έστω $X_1, \ldots, X_n$ τυχαίο δείγμα από την κατανομή Γάμμα, $G(a, 1)$, με πυκνότητα

$$f(x;a) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x}, \quad x \in \mathbb{R}, \quad a \in \Theta = (0, \infty).$$

Θεωρούμε την συνάρτηση $\psi(a) := \frac{\Gamma'(a)}{\Gamma(a)} = \frac{\partial}{\partial a} \log \Gamma(a)$, $a > 0$, όπου $\Gamma(a) := \int_0^\infty x^{a-1} e^{-x} dx$. Αποδείξτε τα εξής:

(α) $\psi(a) = \mathbb{E}[\log X_1]$ και $\psi(1) = -\gamma$ (γ η σταθερά Euler).

(β) Η συνάρτηση $\Gamma \in C^\infty(0, \infty)$ (δηλ. είναι απείρως διαφορίσιμη).

(γ) Η συνάρτηση $\psi \in C^\infty(0, \infty)$.

(δ) $\frac{\Gamma^{(n)}(a)}{\Gamma(a)} = \mathbb{E}[(\log X_1)^n]$.

(ε) $\psi'(a) = \mathrm{Var}[\log(X_1)]$ και, συνεπώς, $\psi'(a) > 0$ (και η ψ είναι γνήσια αύξουσα).

(στ) $\psi(a+1) - \psi(a) = \frac{1}{a}$ και, συνεπώς, $\psi(n) = 1 + \frac{1}{2} + \ldots + \frac{1}{n-1} + \psi(1) \to \infty$.

(ζ) $\psi(a) < \log(a)$, $a > 0$.

(η) $\lim_{a\to 0_+} \psi(a) = -\infty$, $\lim_{a\to+\infty} \psi(a) = +\infty$ και $a\psi'(a) > 1$ για κάθε $a > 0$.

(θ) Να δειχθεί ότι ο πληροφοριακός αριθμός του Fisher ισούται με $I(a) = \psi'(a)$.

(ι) Να δειχθεί ότι ο MLE του a είναι μοναδικός και δίνεται από τον τύπο

$$\widehat{a}_n = \psi^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \log(X_i)\right).$$

(ια) Να δειχθεί ότι η σχετική (απόλυτη) ασυμπτωτική αποδοτικότητα του δειγματικού μέσου σε σχέση με τον MLE ισούται με $\frac{1}{a\psi'(a)} < 1$.

(ιβ) Με χρήση του Κεντρικού Οριακού Θεωρήματος και της μεθόδου Δέλτα να δειχθεί απεθείας ότι $\sqrt{n}(\widehat{a}_n - a) \xrightarrow{\mathrm{d}} N(0, 1/\psi'(a))$.

2.15*. Έστω $X_1, \ldots, X_n$ τυχαίο δείγμα από την κατανομή Γάμμα, $G(a,\lambda)$, με πυκνότητα

$$f(x;(a,\lambda)) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \quad x \in \mathbb{R}, \quad (a,\lambda) \in \Theta = (0,\infty)^2.$$

Θεωρούμε την συνάρτηση $h(a) := \log(a) - \psi(a)$, $a > 0$. Αποδείξτε τα εξής:

(α) Η h είναι γνήσια φθίνουσα και διαφορίσιμη με όρια ∞ και 0 καθώς το a τείνει στα $0+$ και $+\infty$, αντίστοιχα.

(β) Ο πίνακας πληροφορίας του Fisher ισούται με

$$I(a,\lambda) = \frac{1}{\lambda^2}\begin{pmatrix} \lambda^2\psi'(a) & -\lambda \\ -\lambda & a \end{pmatrix},$$

και είναι θετικά ορισμένος.

(γ) Η MLE των a, λ είναι μοναδική και δίνεται από τους τύπους

$$\widehat{a}_n = h^{-1}\left(\log(\overline{X}_n) - \frac{1}{n}\sum_{i=1}^{n}\log(X_i)\right), \quad \widehat{\lambda}_n = \frac{\widehat{a}_n}{\overline{X}_n}.$$

(δ) Να βρεθεί η ασυμπτωτική κατανομή της $\sqrt{n}(\widehat{a}_n - a, \widehat{\lambda}_n - \lambda)^T$ με απευθείας χρήση του Κεντρικού Οριακού Θεωρήματος και της μεθόδου Δέλτα.

(ε) Να συγκριθεί η MLE των a, λ με την αντίστοιχη ροποεκτιμήτρια των παραμέτρων, $\widetilde{a}_n = \frac{\overline{X}_n^2}{M_{2,n}}$, $\widetilde{\lambda}_n = \frac{\overline{X}_n}{M_{2,n}}$.

(στ) Ας υποθέσουμε ότι η πραγματική τιμή της παραμέτρου λ είναι $\lambda = 1$. Να δειχθεί ότι η σχετική ασυμπτωτική αποδοτικότητα της MLE του a που βρήκαμε στο (γ) σε σχέση με την MLE της άσκησης 2.14 (τότε ήταν γνωστό ότι $\lambda = 1$) ισούται με $1 - \frac{1}{a\psi'(a)} < 1$.

2.16. Συντελεστής συσχέτισης από διδιάστατα κανονικά δεδομένα με γνωστούς μέσους και διασπορές.
Το τυχαίο δείγμα $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ προέρχεται από την τυποποιημένη διδιάστατη κανονική κατανομή,

$$N_2\left(\left[\begin{array}{c}0\\0\end{array}\right],\left[\begin{array}{cc}1&\rho\\\rho&1\end{array}\right]\right),$$

με πυκνότητα

$$f(x,y)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right),\quad (x,y)\in\mathbb{R}^2,\quad \rho\in\Theta=(-1,1).$$

(α) Να δείξετε ότι $(X_1,Y_1)\stackrel{\mathrm{d}}{=}(Z_1,\rho Z_1+\sqrt{1-\rho^2}Z_2)$, όπου οι Z_1,Z_2 είναι ανεξάρτητες κανονικές $N(0,1)$, ώστε να υπολογίσετε (εύκολα) τις ροπές

$$\mu_{i,j}=\mathbb{E}(X_1^iY_1^j),\quad i,j\geq 0,\quad 1\leq i+j\leq 4.$$

(β) Να δείξετε ότι $I(\rho)=\frac{1+\rho^2}{(1-\rho^2)^2}$.

(γ) Να δείξετε ότι η εξίσωση πιθανοφάνειας γράφεται ως

$$\rho(1-\rho^2)+(1+\rho^2)\frac{1}{n}\sum_{i=1}^n X_iY_i-\rho\left(\frac{1}{n}\sum_{i=1}^n X_i^2+\frac{1}{n}\sum_{i=1}^n Y_i^2\right)=0,$$

και να αποδείξετε ότι με πιθανότητα 1 έχει τουλάχιστον μία λύση στο διάστημα $(-1,1)$.

(δ) Δείξτε ότι η $\widetilde{\rho}_n:=\frac{1}{n}\sum_{i=1}^n X_iY_i$ είναι ασυμπτωτικά κανονική εκτιμήτρια του ρ.

(ε) Βρείτε ασυμπτωτικά αποδοτική εκτιμήτρια, $\widehat{\rho}_n$, του ρ, και συγκρίνετε την $\widetilde{\rho}_n$ με την $\widehat{\rho}_n$.

(στ) Να κατασκευστεί ένα ασυμπτωτικό διάστημα εμπιστοσύνης για το ρ, με ασυμπτωτικό συντελεστή εμπιστοσύνης $1-\alpha$.

2.17.* Συντελεστής συσχέτισης από διδιάστατα κανονικά δεδομένα με άγνωστους μέσους και διασπορές.
Το τυχαίο δείγμα $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ $(n\geq 3)$ προέρχεται από την διδιάστατη κανονική κατανομή,

$$N_2\left(\left[\begin{array}{c}\mu_1\\\mu_2\end{array}\right],\left[\begin{array}{cc}\sigma_1^2&\rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2&\sigma_2^2\end{array}\right]\right),$$

με όλες τις παραμέτρους άγνωστες δηλαδή

$$\boldsymbol{\theta}:=(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)\in\Theta=\mathbb{R}\times\mathbb{R}\times(0,\infty)\times(0,\infty)\times(-1,1).$$

(α) Να δείξετε ότι με πιθ. 1 οι εξισώσεις πιθανοφάνειας έχουν μοναδική λύση την

$$\widehat{\boldsymbol{\theta}}_n=\left(\begin{array}{c}\widehat{\mu}_{1n}\\\widehat{\mu}_{2n}\\\widehat{\sigma}_{1n}^2\\\widehat{\sigma}_{2n}^2\\\widehat{\rho}_n\end{array}\right)=\left(\begin{array}{c}\overline{X}_n\\\overline{Y}_n\\\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X}_n)^2\\\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y}_n)^2\\\frac{\sum_{i=1}^n(X_i-\overline{X}_n)(Y_i-\overline{Y}_n)}{\left[\sum_{i=1}^n(X_i-\overline{X}_n)^2\right]^{1/2}\left[\sum_{i=1}^n(Y_i-\overline{Y}_n)^2\right]^{1/2}}\end{array}\right).$$

[Σχόλιο για την μοναδικότητα της λύσεως: Αν θέσουμε $S_{xx}=\sum_{i=1}^n (X_i-\overline{X}_n)^2$, $S_{xy}=\sum_{i=1}^n (X_i-\overline{X}_n)(Y_i-\overline{Y}_n)$, $S_{yy}=\sum_{i=1}^n (Y_i-\overline{Y}_n)^2$, τότε, επειδή $n\geq 3$, η τυχαία μεταβλητή $T=S_{xx}S_{yy}-S_{xy}^2$ είναι μεγαλύτερη του μηδενός με πιθ. 1. Αυτό έπεται από το γεγονός ότι $T\geq 0$ (από την ανισότητα Cauchy-Schwarz) και $T=0$ αν και μόνο αν τα n σημεία του επιπέδου $\boldsymbol{\xi_i}:=(X_i-\overline{X}_n, Y_i-\overline{Y}_n)^T$ είναι συνευθειακά. Προφανώς, το ενδεχόμενο {τα $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n$ είναι συνευθειακά} περιέχεται στο ενδεχόμενο $\{X_1Y_2+X_2Y_3+X_3Y_1=X_1Y_2+X_2Y_1+X_3Y_2\}$, το οποίο έχει πιθανότητα 0. Τώρα μπορούμε να δούμε ότι οι εξισώσεις έχουν μοναδική λύση όταν $S_{xx}S_{yy}>S_{xy}^2$.]

(β) Υπολογίστε την πίνακα πληροφορίας $I(\boldsymbol{\theta})$. Συγκεκριμένα, δείξτε ότι

$$I(\boldsymbol{\theta})=\frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} & 0 & 0 & 0\\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} & 0 & 0 & 0\\ 0 & 0 & \frac{2-\rho^2}{4\sigma_1^4} & \frac{-\rho^2}{4\sigma_1^2\sigma_2^2} & \frac{-\rho}{2\sigma_1^2}\\ 0 & 0 & \frac{-\rho^2}{4\sigma_1^2\sigma_2^2} & \frac{2-\rho^2}{4\sigma_2^4} & \frac{-\rho}{2\sigma_2^2}\\ 0 & 0 & \frac{-\rho}{2\sigma_1^2} & \frac{-\rho}{2\sigma_2^2} & \frac{1+\rho^2}{1-\rho^2} \end{pmatrix}.$$

(γ) Υπολογίζοντας και τον αντίστροφό του συμπεράνατε ότι

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})\xrightarrow{\mathrm{d}} N_5\left(\mathbf{0},\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & 0 & 0 & 0\\ \rho\sigma_1\sigma_2 & \sigma_2^2 & 0 & 0 & 0\\ 0 & 0 & 2\sigma_1^4 & 2\rho^2\sigma_1^2\sigma_2^2 & \rho(1-\rho^2)\sigma_1^2\\ 0 & 0 & 2\rho^2\sigma_1^2\sigma_2^2 & 2\sigma_2^4 & \rho(1-\rho^2)\sigma_2^2\\ 0 & 0 & \rho(1-\rho^2)\sigma_1^2 & \rho(1-\rho^2)\sigma_2^2 & (1-\rho^2)^2 \end{bmatrix}\right).$$

[Παρατηρήστε ότι η εκτιμήτρια μέγιστης πιθανοφάνειας του ρ ταυτίζεται με τον δειγματικό συντελεστή συσχέτισης της άσκησης 2.9, η δε οριακή διασπορά της δίδεται στην άσκηση 2.10 (όταν τα δεδομένα είναι κανονικά). Συμπεράνατε ότι η σχετική αποδοτικότητα της εκτιμήτριας του ρ σε σχέση με την εκτιμήτρια της προηγούμενης άσκησης (τότε ήταν γνωστές οι μέσες τιμές και οι διασπορές) ισούται με $\frac{1}{1+\rho^2}\leq 1$.]

(δ) Μπορείτε να αποδείξετε το (γ) απευθείας με χρήση του Κεντρικού Οριακού Θεωρήματος και της μεθόδου Δέλτα;

2.18. Κανονικά δεδομένα με γνωστό συντελεστή μεταβλητότητας.
Το τυχαίο δείγμα $X_1,X_2,\ldots,X_n$ προέρχεται από την κανονική κατανομή $N(\lambda_0\theta,\theta^2)$, $\theta\in\Theta=(0,\infty)$, $\lambda_0>0$ γνωστό.

(α) Να δείξετε ότι η εξίσωση πιθανοφάνειας έχει μοναδική λύση την

$$\widehat{\theta}_n=\frac{1}{2}\left[\left((\lambda_0\overline{X}_n)^2+\frac{4}{n}\sum_{i=1}^n X_i^2\right)^{1/2}-\lambda_0\overline{X}_n\right].$$

(β) Να βρείτε την οριακή κατανομή της $\sqrt{n}(\widehat{\theta}_n-\theta)$.

(γ) Χρησιμοποιώντας κατάλληλο κυρτό συνδυασμό των $\overline{X}_n$ και $\sqrt{\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X}_n)^2}$ να βρείτε μία δεύτερη ασυμτωτικά αποδοτική ακολουθία εκτιμητριών.

2.19. Το τυχαίο δείγμα $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ προέρχεται από την k-διάστατη κατανομή Bernoulli

$$\mathbb{P}(\boldsymbol{X}_1=\boldsymbol{x})=p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}(1-p_1-\ldots-p_k)^{1-x_1-\cdots-x_k},\quad \boldsymbol{x}\in\{0,1\}^k,\quad \sum_{i=1}^n x_i\leq 1.$$

Εδώ ο παραμετρικός χώρος είναι το ανοικτό σύνολο

$$\Theta = \left\{ (p_1, \ldots, p_k) : p_i > 0 \ (i = 1, 2, \ldots, k), \quad \sum_{i=1}^{k} p_i < 1 \right\},$$

και για απλότητα θέτουμε $p_{k+1} = 1 - p_1 - \ldots - p_k > 0$. Επίσης γράφουμε $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ik})^T$ για τις συντεταγμένες του διανύσματος $\boldsymbol{X}_i$ (ή όλες είναι 0 ή μία ακριβώς είναι 1 και οι άλλες 0).

(α) Υπολογίστε την πίνακα πληροφορίας $I(\boldsymbol{\theta})$ και τον αντίστροφό του.

(β) Δείξτε ότι οι εξισώσεις πιθανοφάνειας έχουν μοναδική λύση την

$$\widehat{p}_i = \overline{X}_i, \quad i = 1, 2, \ldots, k,$$

όπου $\overline{X}_i = \frac{Y_1}{n}$, $Y_i = Y_{ni} = X_{1i} + X_{2i} + \ldots + X_{ni}$ ο συνολικός αριθμός επιτυχιών κατηγορίας i.

(γ) Συμπεράνατε ότι οι έλεγχοι Wald και Rao-scores συμπίπτουν. Συγκεκριμένα δείξτε ότι υπό την $H_0 : p_1 = p_1^0, \ldots, p_k = p_k^0$, οι δύο έλεγχοι οδηγούν στην στατιστική συνάρτηση χ^2 του Pearson:

$$W_n = R_n = \sum_{i=1}^{k+1} \frac{(Y_i - np_i^0)^2}{np_i^0},$$

όπου $p_{k+1}^0 = 1 - p_1^0 - \cdots - p_k^0$, $Y_{k+1} = n - Y_1 - \cdots - Y_k$.

(δ) Δείξτε ότι το κριτήριο λόγου πιθανοφανειών γράφεται ως

$$2\Delta_n = 2 \sum_{i=1}^{k+1} Y_i \log \frac{Y_i}{np_i^0},$$

όπου $p_{k+1}^0 = 1 - p_1^0 - \cdots - p_k^0$, $Y_{k+1} = n - Y_1 - \cdots - Y_k$. Επομένως, οι τρεις κρίσιμες περιοχές του ελέγχου $H_0 : \boldsymbol{p} = \boldsymbol{p}^0$, ασυμπτωτικού επιπέδου εμπιστοσύνης α, είναι οι

$$K : \sum_{i=1}^{k+1} \frac{(Y_i - np_i^0)^2}{np_i^0} \geq \chi_k^2(\alpha) \ (\text{για Wald και Rao}) \quad K : 2 \sum_{i=1}^{k+1} Y_i \log \frac{Y_i}{np_i^0} \geq \chi_k^2(\alpha) \ (\text{για LR}).$$

2.20. (Συνέχεια της άσκησης 2.18.) Το τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ προέρχεται από την κανονική κατανομή $N(\lambda_0\theta, \theta^2)$, $\theta \in \Theta = (0, \infty)$, $\lambda_0 > 0$ γνωστό. Να γίνει ο έλεγχος της $H_0 : \theta = \theta_0$ ($\theta_0 > 0$) χρησιμοποιώντας τα κριτήρια Wald, Rao και LR.

2.21. (Συνέχεια της άσκησης 2.16.) Το τυχαίο δείγμα $(X_1, Y_1), \ldots, (X_n, Y_n)$ προέρχεται από την διδιάστατη τυποποιημένη κανονική με διάνυσμα μέσων τιμών $\mathbf{0} = (0, 0)^T$ και πίνακα διασποράς

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad -1 < \rho < 1.$$

Να γίνει ο έλεγχος της $H_0 : \rho = 0$ χρησιμοποιώντας τα κριτήρια Wald, Rao και LR.

2.22*. (Συνέχεια της άσκησης 2.17*.) Το τυχαίο δείγμα $(X_1, Y_1), \ldots, (X_n, Y_n)$ προέρχεται από την διδιάστατη κανονική με διάνυσμα μέσων τιμών $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ και πίνακα διασποράς

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \sigma_1 > 0, \ \sigma_2 > 0, \ -1 < \rho < 1.$$

Να γίνει ο έλεγχος της

$$H_0: \ \mu_1 = \mu_1^0, \ \mu_2 = \mu_2^0, \ \sigma_1 = \sigma_1^0, \ \sigma_2 = \sigma_2^0, \ \rho = \rho^0$$

χρησιμοποιώντας τα κριτήρια Wald, Rao και LR.

2.23. (Οικογένεια θέσης-κλίμακος) Υποθέτουμε ότι η $f_0(x)$ είναι μία μονοδιάστατη γνήσια θετική και παραγωγίσιμη πυκνότητα πιθανότητας. Θεωρούμε το τυχαίο δείγμα $X_1, X_2, \ldots X_n$, προερχόμενο από την οικογένεια

$$\left\{ f(x;\boldsymbol{\theta}) = \frac{1}{\theta_2} f_0\left(\frac{x-\theta_1}{\theta_2}\right) \right\}$$

όπου $\boldsymbol{\theta} = (\theta_1, \theta_2)^T \in \Theta = \mathbb{R} \times (0, \infty)$.

Να δειχθεί ότι ο πίνακας πληροφορίας είναι

$$I(\theta_1, \theta_2) = \frac{1}{\theta_2^2} \left(\begin{array}{cc} \int_{-\infty}^{\infty} \frac{(f_0'(x))^2}{f_0(x)} dx & \int_{-\infty}^{\infty} f_0'(x) \left(1 + \frac{x f_0'(x)}{f_0(x)}\right) dx \\ \int_{-\infty}^{\infty} f_0'(x) \left(1 + \frac{x f_0'(x)}{f_0(x)}\right) dx & \int_{-\infty}^{\infty} f_0(x) \left(1 + \frac{x f_0'(x)}{f_0(x)}\right)^2 dx \end{array} \right),$$

με την προϋπόθεση φυσικά ότι τα ολοκληρώματα είναι πεπερασμένα. Να συμπεράνετε ότι αν η f_0 είναι συμμετρική ($f_0(x) = f_0(-x)$) τότε ο πίνακας πληροφορίας είναι διαγώνιος. Παρατηρήστε ότι αν $\lim_{x\to\pm\infty} f_0(x) = 0$ τότε τα μη διαγώνια στοιχεία του πίνακα πληροφορίας γράφονται ως $I_{12} = \frac{1}{\theta_2^2} \int_{-\infty}^{\infty} x \frac{(f_0'(x))^2}{f_0(x)} dx$.

2.24. Θεωρούμε τυχαίο δείγμα $X_1, X_2, \ldots X_n$ με πυκνότητα $f(x;\boldsymbol{\theta}) = \frac{\exp\left(-\frac{x-\theta_1}{\theta_2}\right)}{\theta_2\left(1+\exp\left(-\frac{x-\theta_1}{\theta_2}\right)^2\right)}$ όπου $\boldsymbol{\theta} = (\theta_1, \theta_2)^T \in \Theta = \mathbb{R} \times (0, \infty)$ (οικογένεια θέσης-κλίμακος της logistic).

(α) Να δειχθεί ότι ο πίνακας πληροφορίας είναι

$$I(\theta_1, \theta_2) = \frac{1}{\theta_2^2} \left(\begin{array}{cc} \frac{1}{3} & 0 \\ 0 & \frac{3+\pi^2}{9} \end{array} \right).$$

(β) Να γίνει ο έλεγχος της $H_0: \theta_1 = \theta_1^0, \theta_2 = \theta_2^0$ χρησιμοποιώντας τα κριτήρια Wald, Rao και LR.

2.25. Θεωρούμε τυχαίο δείγμα $X_1, X_2, \ldots X_n$ με πυκνότητα $f(x;\boldsymbol{\theta}) = \frac{\theta_2/\pi}{(x-\theta_1)^2+\theta_2^2}$ όπου $\boldsymbol{\theta} = (\theta_1, \theta_2)^T \in \Theta = \mathbb{R} \times (0, \infty)$ (οικογένεια θέσης-κλίμακος της Cauchy). Να δειχθεί ότι $I(\theta_1, \theta_2) = \frac{1}{2\theta_2^2} I_2$ (I_2 ο μοναδιαίος 2×2) και να γίνει ο έλεγχος της $H_0: \theta_1 = \theta_1^0, \theta_2 = \theta_2^0$ χρησιμοποιώντας τα κριτήρια Wald, Rao και LR.

2.26. Θεωρούμε τυχαίο δείγμα $X_1, X_2, \ldots X_n$ από την $N(\theta_1, \theta_2^2)$ όπου $\boldsymbol{\theta} = (\theta_1, \theta_2)^T \in \Theta = \mathbb{R} \times (0, \infty)$ (οικογένεια θέσης-κλίμακος της $N(0, 1)$). Να δειχθεί ότι $I(\theta_1, \theta_2) = \frac{1}{\theta_2^2} \left(\begin{array}{cc} 1 & 0 \\ 0 & 2 \end{array} \right)$ και να γίνει ο έλεγχος της $H_0: \theta_1 = \theta_1^0, \theta_2 = \theta_2^0$ χρησιμοποιώντας τα κριτήρια Wald, Rao και LR. [Σημειώστε ότι ο πίνακας πληροφορίας δεν ισούται με τον πίνακα πληροφορίας $I(\mu, \sigma^2)$ της άσκησης 1.14. Γιατί συμβαίνει αυτό;]