

Ερωτήσεις κορμού θεωρητικής εξέτασης

- 1) Πως ορίζεται η δειγματική αθροιστική συνάρτηση κατανομής $\hat{F}_n(x)$; Τι μορφή έχει ως συνάρτηση στο x ; Υπολογίστε την αναμενόμενη τιμή και τη διακύμανσή της σε κάποιο σταθερό x_0 . Πως θα κατασκευάζα σημειακό (ασυμπτωτικό) διάστημα εμπιστοσύνης για το $F(x_0) = P(X \leq x_0)$ για σταθερό x_0 με τη κανονική προσέγγιση; Πως θα χρησιμοποιούσα την ανισότητα $P(|\hat{F}_n(x) - F(x)| > \varepsilon) \leq 2e^{-2n\varepsilon^2}, \forall \varepsilon, n$ για να κατασκευάσω διάστημα εμπιστοσύνης για το $F(x_0)$ για σταθερό x_0 ; Ποια ή διαφορά μεταξύ των δύο προηγούμενων ως προς τη τήρηση της ονομαστικής πιθανότητας κάλυψης? Ποιανού στατιστικού θα χρειαζόμουν τη κατανομή ώστε να μπορέσω να κατασκευάσω ζώνη εμπιστοσύνης για την F ως συνάρτηση (για όλα τα $x \in \mathbb{R}$); Γιατί; Εξηγείστε.
- 2) Δείξτε ότι: αν η $F(x)$ η αθροιστική συνάρτηση κατανομής $F(x) = P(X \leq x)$ είναι συνεχής και αντιστρέψιμη και $\hat{F}_n(x) := n^{-1} \sum_{i=1}^n I_{\{X_i \leq x\}}$ η δειγματική αθροιστική συνάρτηση κατανομής, τότε η κατανομή του στατιστικού $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)|$ δεν εξαρτάται από την $F(x)$.
- 3) Πως ορίζεται η εκτιμήτρια αντικατάστασης (plug-in) ενός συναρτησιακού $T(F)$; Πως ορίζεται η συνάρτηση επιρροής της; Πως εκτιμάται η συνάρτηση επιρροής; Τι ρόλο παίζει η συνάρτηση επιρροής στη διακύμανση της εκτιμήτριας; Πως εκτιμάται η διακύμανση της εκτιμήτριας και πως κατασκευάζουμε ασυμπτωτικά διαστήματα εμπιστοσύνης για την $T(F)$; Υπό ποιες προϋποθέσεις περιμένουμε να δουλεύουν αυτά?
- 4) Πως ορίζεται η συνάρτηση επιρροής ενός συναρτησιακού $T(F)$; Πως ερμηνεύεται στατιστικά και πως χρησιμοποιείται για το χαρακτηρισμό της ανθεκτικότητας (robustness) εκτιμητριών; Πως ορίζεται το break-down-point και πως συγκρίνεται στο ορισμό της ανθεκτικότητας εκτιμητριών με τη συνάρτηση επιρροής;

- 5) Τι είναι ένας M-estimator για τη θέση μιας κατανομής όταν είναι γνωστή η διακύμανση; Πως διαφοροποιείται όταν είναι άγνωστη; Πως ορίζεται α) ένας M-estimator και β) ο least trimmed squares (lts) estimator των συντελεστών μιας πολλαπλής παλινδρόμησης; Συζητείστε την ανθεκτικότητά του ανάλογα με την ψ-συνάρτηση που εμπλέκεται. Έχουν και οι δύο ψηλό breakdown-point?
- 6) Τι είναι ένας M-estimator για τη θέση μιας κατανομής όταν είναι γνωστή η διακύμανση; Πως ο M-estimator για τη θέση μιας κατανομής όταν είναι άγνωστη η διακύμανση; Πως ορίζεται ο M-estimator για τη κλίμακα σ της κατανομής ανεξάρτητων $Y_i \sim N(0, \sigma)$;
- 7) Πως εκτιμάμε τη μεροληψία μιας εκτιμήτριας με Jackknife; Πως κάνουμε διόρθωση της μεροληψίας της εκτιμήτριας και με τι αποτέλεσμα (καταφέρνουμε να μηδενίσουμε τη μεροληψία της;)
- 8) Πως εκτιμάμε τη διακύμανση μιας εκτιμήτριας με Jackknife; Είναι πάντα συνεπής αυτή η εκτιμήτρια;
- 9) Ποιός είναι ο σκοπός του bootstrap; Τι είναι το ιδανικό bootstrap; Γιατί το ιδανικό bootstrap προσεγγίζει το στόχο του μόνο όταν ο αριθμός των αρχικών μας δεδομένων μας τείνει στο άπειρο? Εξηγείστε πως μπορούμε να προσεγγίσουμε το ιδανικό bootstrap με resampling των αρχικών δεδομένων. Πόσο καλή μπορεί να είναι αυτή η προσέγγιση;
- 10) Εξηγείστε τη κατασκευή κανονικών (normal) - bootstrap διαστημάτων εμπιστοσύνης. Αιτιολογείστε γιατί να έχουν πιθανότητα κάλυψης που πλησιάζει την ονομαστική;
- 11) Εξηγείστε τη κατασκευή basic (pivot) - bootstrap διαστημάτων εμπιστοσύνης. Αιτιολογείστε γιατί να έχουν πιθανότητα κάλυψης που πλησιάζει την ονομαστική; Τι πλεονέκτημα έχουν έναντι των κανονικών (normal) - bootstrap διαστημάτων εμπιστοσύνης;
- 12) Εξηγείστε τη κατασκευή Student - bootstrap διαστημάτων εμπιστοσύνης. Αιτιολογείστε γιατί να έχουν πιθανότητα κάλυψης που πλησιάζει την ονομαστική; Τι πλεονέκτημα έχουν έναντι των basic - bootstrap διαστημάτων εμπιστοσύνης; Γιατί δεν χρησιμοποιούμε πάντα τα student αντί για τα basic;
- 13) Εξηγείστε τη κατασκευή percentile - bootstrap διαστημάτων εμπιστοσύνης. Αιτιολογείστε γιατί να έχουν πιθανότητα κάλυψης που πλησιάζει την ονομαστική; Υπό ποια βασική προϋπόθεση θα

δώσουν ασυμπτωτικά σωστά αποτελέσματα; Ποιό είναι το βασικό τους πλεονέκτημα έναντι π.χ. των Student- bootstrap διαστημάτων;

- 14) Πως ορίζεται το ιστόγραμμα $\hat{f}_{n,h}(x)$ με σταθερό εύρος διαστήματος h ; Αν η μεροληψία του δίνεται προσεγγιστικά από $E[\hat{f}_{n,h}(x)] - f(x) \approx c_1 f'(x)h$, για c_1 κατάλληλο, και η διακύμανσή του από $\text{var}[\hat{f}_{n,h}(x)] \approx f(x)/(nh)$, συζητείστε το ρόλο του h στη συμπεριφορά αυτών των δύο ποσοτήτων. Τι συμβαίνει για πολύ μικρά και τι για πολύ μεγάλα h ; Πως ορίζεται και πως υπολογίζεται ένα θεωρητικά "βέλτιστο" εύρος διαστήματος h_n^* . Πώς θα εξαρτάται αυτό από το μέγεθος δείγματος n ;
- 15) Αιτιολογήστε την επιλογή εύρους διαστήματος h για ένα ιστόγραμμα με cross-validation: Τι στοχεύουμε να εκτιμήσουμε, προκειμένου να το ελαχιστοποιήσουμε; Πως χρησιμοποιούμε εκτιμήτριες που εξαιρούν την i - παρατήρηση και γιατί;
- 16) Εξηγήστε τη κατασκευή σημειακών (για σταθερό x) κανονικών ασυμπτωτικών διαστημάτων εμπιστοσύνης για τη πυκνότητα $f(x)$ με βάση το ιστόγραμμα. Δίνεται ότι η εκτιμήτρια είναι ασυμπτωτικά κανονική με $E[\hat{f}_{n,h}(x)] - f(x) \approx c_1 f'(x)h$ και $\text{var}[\hat{f}_{n,h}(x)] \approx f(x)/(nh)$. Υπό ποιές προϋποθέσεις πρόκειται για διαστήματα για την $E[\hat{f}_{n,h}(x)]$ και υπό ποιές προϋποθέσεις πρόκειται για διαστήματα για την ίδια την $f(x)$.
- 17) Εξηγήστε τη κατασκευή ομοιόμορφων (για όλα τα x ταυτόχρονα) διαστημάτων εμπιστοσύνης για την για την αναμενόμενη τιμή του ιστογράμματος με τη χρήση Bootstrap.
- 18) Πως ορίζεται η "εκτιμήτρια της πυκνότητας βασισμένη σε πυρήνα" $\hat{f}_{n,h}(x)$ για κάποιο πυρήνα $K(x)$, που είναι διάφορος του 0 στο $[-1,1]$; Σε τι διαφοροποιείται από το ιστόγραμμα, όταν ο πυρήνας είναι απλά η χαρακτηριστική συνάρτηση στο $[-1,1]$: $I_{[-1,1]}(x)$; Γιατί προτιμώ πυρήνες που είναι συνεχείς συναρτήσεις, με κορυφή στο $x=0$, που φθίνουν όταν το x απομακρύνεται από το 0; Πως εξαρτάται η εικόνα της εκτιμήτριας από το εύρος του πυρήνα h , και πως σχετίζεται αυτό από την εξάρτηση της διακύμανσης και της μεροληψίας της εκτιμήτριας από το h ;

19) Δείξτε ότι η αναμενόμενη τιμή της εκτιμήτριας της πυκνότητας βασισμένης σε πυρήνα

$$\hat{f}_{n,h}(x) := \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right), \text{ δίνεται από } E[\hat{f}_{n,h}(x)] = \int K(u)f(x+hu)du \text{ και ότι αυτή}$$

ακολουθώς προσεγγίζεται με $f(x) + \frac{1}{2}f''(x)h^2 \int u^2 K(u)du$, αν ο πυρήνας είναι συμμετρικός,

έχει $\int K(u)du = 1$ και αν επιπλέον $h \rightarrow 0$.

20) Εξηγήστε τη κατασκευή σημειακών (για σταθερό x) κανονικών ασυμπτωτικών διαστημάτων εμπιστοσύνης για τη πυκνότητα $f(x)$ με βάση την εκτιμήτρια της πυκνότητας βασισμένης σε πυρήνα. Δίνεται ότι η εκτιμήτρια είναι ασυμπτωτικά κανονική με

$$E[\hat{f}_{n,h}(x)] \approx f(x) + \frac{1}{2}f''(x)h^2 \int u^2 K(u)du \text{ και } \text{var}[\hat{f}_{n,h}(x)] \approx f(x) \int K^2(u)du / (nh).$$

Υπό ποιές προϋποθέσεις πρόκειται για διαστήματα για την $E[\hat{f}_{n,h}(x)]$ και υπό ποιές προϋποθέσεις πρόκειται για διαστήματα για την ίδια την $f(x)$;

21) Πως ορίζεται το θεωρητικά βέλτιστο h^* μιας εκτιμήτριας της πυκνότητας βασισμένης σε πυρήνα; Τι στοχεύουμε να εκτιμήσουμε, προκειμένου να το ελαχιστοποιήσουμε; Δίνεται ότι η εκτιμήτρια έχει

$$E[\hat{f}_{n,h}(x)] \approx f(x) + \frac{1}{2}f''(x)h^2 \int u^2 K(u)du \text{ και } \text{var}[\hat{f}_{n,h}(x)] \approx f(x) \int K^2(u)du / (nh).$$

Πως υπολογίζεται από αυτά το h^* ; Τι τάξη έχει το ελάχιστο μέσο τετραγωνικό σφάλμα συναρτήσει του μεγέθους του δείγματος;

22) Αιτιολογήστε την επιλογή εύρους διαστήματος h για μια εκτιμήτριας της πυκνότητας βασισμένης σε πυρήνα με cross-validation: Τι στοχεύουμε να εκτιμήσουμε, προκειμένου να το ελαχιστοποιήσουμε; Πως χρησιμοποιούμε εκτιμήτριες που εξαιρούν την i - παρατήρηση και γιατί;

23) Εξηγήστε τη κατασκευή ομοιόμορφων (για όλα τα x ταυτόχρονα) διαστημάτων εμπιστοσύνης για την αναμενόμενη τιμή της εκτιμήτριας της πυκνότητας βασισμένης σε πυρήνα με τη χρήση bootstrap.

24) Πως ορίζεται η εκτιμήτρια Nadaraya-Watson της συνάρτησης παλινδρόμησης $r(x_0) := E(Y|X = x_0)$; Ποια η διαφορά της από το regressogram για σταθερό πυρήνα; Πως ορίζεται η εκτιμήτρια τοπικής γραμμικής παλινδρόμησης (loess) της $r(x_0)$; Εξηγήστε με ένα

σχήμα γιατί η δεύτερη συμπεριφέρεται καλύτερα από τη πρώτη στα άκρα (όρια, boundaries) του design, δηλαδή κοντά στο μέγιστο και το ελάχιστο των x_i . Πως γράφεται η κάθε μια ως γραμμικός εξομαλυντής ("linear smoother"). Εξηγήστε τον υπολογισμό των σταθμών.

- 25) Έστω μια εκτιμήτρια της συνάρτησης παλινδρόμησης $r(x_0) := E(Y|X = x_0)$ που γράφεται ως γραμμικός εξομαλυντής ("linear smoother"). Δείξτε ότι η μεροληψία της υπολογίζεται ως:

$$\text{bias}[\hat{r}_{n,h}(x_0)] = r'(x_0) \sum_{i=1}^n l_i(x_0)(x_i - x_0) + \frac{r''(x_0)}{2} \sum_{i=1}^n l_i(x_0)(x_i - x_0)^2 + \dots$$

Ποιός όρος του

αναπτύγματος της μεροληψίας μηδενίζεται για την Loess, αλλά όχι για την Nadaraya-Watson, και τι συνέπεια έχει αυτό;

- 26) Έστω ότι για κάποια εκτιμήτρια $\hat{r}_{n,h}(x_0)$ της συνάρτησης παλινδρόμησης $r(x_0) := E(Y|X = x_0)$ έχουμε $E[\hat{r}_{n,h}(x)] - f(x) \approx c_1(x)h^2$ και $\text{var}[\hat{r}_{n,h}(x)] \approx c_2(x)/(nh)$ για κατάλληλα $c_1(x)$ και $c_2(x)$ και εύρος πυρήνα h . Υπολογίστε το θεωρητικά βέλτιστο h_n^* που ελαχιστοποιεί το αναμενόμενο μέσο τετραγωνικό σφάλμα $AEL(h) := E\left[n^{-1} \sum_{i=1}^n (\hat{r}_{n,h}(x_i) - r(x_i))^2\right]$. Ποιάς τάξης είναι το $AEL(h_n^*)$;

- 27) Αιτιολογήστε την επιλογή με cross-validation του εύρους του πυρήνα h για μια εκτιμήτρια $\hat{r}_{n,h}(x_0)$ της συνάρτησης παλινδρόμησης $r(x_0) := E(Y|X = x_0)$: ποια ποσότητα στοχεύουμε να εκτιμήσουμε, προκειμένου να την ελαχιστοποιήσουμε; Πως χρησιμοποιούμε εκτιμήτριες που εξαιρούν την i - παρατήρηση και γιατί; Ποια είναι η γενικευμένη cross-validation (GCV) και πως αιτιολογείται; Πως φαίνεται ότι είναι ισοδύναμη με ελαχιστοποίηση του AIC?

- 28) Με ποιούς δύο τρόπους εκτιμάμε τη (συνάρτηση) διακύμανσης $\sigma(x_i)$ σε ένα μοντέλο $Y_i = r(x_i) + \sigma(x_i) \cdot \varepsilon_i$ με ετεροσκεδαστικότητα;

- 29) Έστω μια εκτιμήτρια της συνάρτησης παλινδρόμησης $r(x_0) := E(Y|X = x_0)$ που γράφεται ως γραμμικός εξομαλυντής ("linear smoother"). Πως εκτιμάμε τη διακύμανσή της και πως κατασκευάζουμε σημειακά διαστήματα εμπιστοσύνης για την $\bar{r}(x_0) = E[\hat{r}(x_0)]$; Πότε είναι αυτά ταυτόχρονα και διαστήματα εμπιστοσύνης για την $r(x_0)$;

- 30) Έστω μια εκτιμήτρια της συνάρτησης παλινδρόμησης $r(x_0) := E(Y|X = x_0)$ που γράφεται ως γραμμικός εξομαλυντής ("linear smoother"). Πως κατασκευάζουμε σημειακά διαστήματα εμπιστοσύνης για την $\bar{r}(x_0) = E[\hat{r}(x_0)]$ με τη χρήση Bootstrap;
- 31) Έστω κόμβοι ξ_1, \dots, ξ_K . Ποιές "συναρτήσεις βάσης" $h_1(x), \dots, h_j(x)$ πρέπει να χρησιμοποιήσω ώστε να πάρω **τετραγωνικά** "regression splines"? Πως υπολογίζεται η $\hat{\mathbf{r}}_n := [\hat{r}_n(x_1), \dots, \hat{r}_n(x_n)]^T$, πως η αναμενόμενη τιμή $E[\hat{r}_n(x_i)]$ και η διακύμανση $\text{var}[\hat{r}_n(x_i)]$;
- 32) Συζητείστε τη penalized παλινδρόμηση που ελαχιστοποιεί την $\sum_{i=1}^n (Y_i - r(x_i))^2 + \lambda \int [r''(x)]^2 dx$. Γιατί χρειάζεται ο τελευταίος όρος; Ποιος είναι ο ρόλος του λ ; Τι σχέση έχουν με αυτήν τα smoothing splines και τι διαφορά έχουν τα τελευταία από τα "regression splines"; Έστω $B_1(x), \dots, B_n(x)$ μία βάση για smoothing splines. Δίνεται ότι αν $\hat{r}_n(x) = \sum_{j=1}^n \beta_j B_j(x)$ τότε $\int [r''(x)]^2 = \boldsymbol{\beta}^T \boldsymbol{\Omega} \boldsymbol{\beta}$ για $\boldsymbol{\beta} := (\beta_1, \dots, \beta_n)^T$ και γνωστό πίνακα $\boldsymbol{\Omega}$. Πως γράφεται η $\hat{r}_n(x)$ ως γραμμικός εξομαλυντής; Πως διαλέγουμε την παράμετρο λ στη πράξη?
- 33) Ποια συνάρτηση μεγιστοποιώ για να ορίσω την εκτιμήτρια τοπικής πιθανοφάνειας (Local likelihood); Πως συγκρίνεται με την συνάρτηση που μεγιστοποιώ για να ορίσω την κλασσική εκτιμήτρια των Γενικευμένων γραμμικών μοντέλων (όπου x συνεχής); Σε τι στατιστικό μπορώ να στηρίξω τον έλεγχο μεταξύ των δύο μοντέλων (δε ζητούνται η κατανομή και τύποι για τους βαθμούς ελευθερίας)?
- 34) Πως θα κάναμε τοπική παλινδρόμηση με πολλές ανεξάρτητες μεταβλητές; Πως ορίζονται τα Γενικευμένα Αθροιστικά Μοντέλα και πως συγκρίνονται με τη τοπική παλινδρόμηση σε πολλές μεταβλητές (πλεονεκτήματα-μειονεκτήματα); Πως εκτιμώνται με το "backfitting" αλγόριθμο;
- 35) Πως ορίζονται τα δένδρα παλινδρόμησης και πως διαλέγουμε το μέγεθός τους στη πράξη?

Ερωτήσεις θεωρητικής εξέτασης εργαστηρίου (rank tests)

- 36) Έστω ένα δείγμα $\{X_1, \dots, X_n\}$ από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, και $R_i := \#\{k \mid |X_k| \leq |X_i|\}$. Ορίζουμε $T_n := \sum_{i=1}^n R_i \mathbf{I}_{\{X_i > 0\}}$. Ποια μηδενική ελέγχουμε με αυτό το στατιστικό, έναντι ποιών εναλλακτικών; Διαλέξτε μια μονόπλευρη εναλλακτική: θα την απορρίπτω για μικρές ή για μεγάλες τιμές της T_n και γιατί; Υπό ποιες προϋποθέσεις μπορώ να κάνω το παραπάνω έλεγχο; Ποιος είναι ο αντίστοιχος παραμετρικός έλεγχος και ποιος από τους δυο είναι καλύτερος;
- 37) Έστω ένα δείγμα $\{X_1, \dots, X_n\}$ από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που προέρχονται από συμμετρική κατανομή. Σε ποια δύο στατιστικά βασίζονται οι μη-παραμετρικοί έλεγχοι που προτείνονται για την μηδενική ότι η διάμεσος είναι το 0 (δε ζητούνται η αναμενόμενη τιμή και η διακύμανση του στατιστικού)? Την εναλλακτική ότι η διάμεσος είναι μικρότερη του μηδενός την απορρίπτουμε για μεγάλες ή για μικρές τιμές του στατιστικού και γιατί; Υπό ποιες προϋποθέσεις μπορούμε να κάνουμε αυτούς τους ελέγχους? Πως συγκρίνονται οι έλεγχοι αυτοί με το κλασσικό t-test (πλεονεκτήματα & μειονεκτήματα);
- 38) Έστω X_1, \dots, X_m i.i.d από F και ανεξάρτητες Y_1, \dots, Y_n i.i.d από G , όπου F, G συνεχείς. Υποθέτουμε επιπλέον ότι η G προέρχεται από μετατόπιση (θέσης) της F : $G(t) = F(t - \Delta)$ (location model). Θέλουμε να ελέγξουμε τη μηδενική $H_0: \Delta = 0$ έναντι της εναλλακτικής $H_1: \Delta > 0$. Σε ποιο στατιστικό βασίζεται ο μη-παραμετρικός έλεγχος που προτείνεται (δε ζητούνται η αναμενόμενη τιμή και η διακύμανση); Θα απορρίπτουμε τη μηδενική για μικρές ή για μεγάλες τιμές του στατιστικού και γιατί; Πως συγκρίνεται ο έλεγχος με το κλασσικό t-test (πλεονεκτήματα & μειονεκτήματα);
- 39) Έστω X_1, \dots, X_m i.i.d από F και ανεξάρτητες Y_1, \dots, Y_n i.i.d από G , όπου F, G συνεχείς και συμμετρικές. Θέλουμε να ελέγξουμε τη μηδενική της ισότητας των διαμέσων των F, G έναντι της εναλλακτικής η διάμεσος της F να είναι μεγαλύτερη από τη διάμεσο της G , χωρίς να υποθέσουμε ισότητα των διακυμάνσεων. Σε ποιο στατιστικό βασίζεται ο μη-παραμετρικός έλεγχος που προτείνεται (δε ζητούνται η αναμενόμενη τιμή και η διακύμανση); Θα απορρίπτω τη μηδενική για μικρές ή για μεγάλες τιμές του στατιστικού και γιατί; Τι αλλάζει στο εκάστοτε στατιστικό

συγκριτικά με τη περίπτωση που υποθέτω επιπλέον πως $G(t) = F(t - \Delta)$; Πως συγκρίνεται ο έλεγχος με το κλασσικό t-test (πλεονεκτήματα & μειονεκτήματα);

40) Έστω X_1, \dots, X_m i.i.d από F και ανεξάρτητες Y_1, \dots, Y_n i.i.d από G . Υποθέτουμε επιπλέον ότι

$\frac{X_i - \theta_1}{\eta_1}$ έχουν την ίδια κατανομή με τα $\frac{Y_i - \theta_2}{\eta_2}$ (location-scale model). Θέλουμε να ελέγξουμε τη

μηδενική $H_0 : \eta_1 = \eta_2$ έναντι της εναλλακτικής $H_1 : \eta_1 > \eta_2$. Αν $\theta_1 = \theta_2$, σε ποιο στατιστικό βασίζεται ο μη-παραμετρικός έλεγχος που προτείνεται (δε ζητούνται η αναμενόμενη τιμή και η διακύμανση); Τι κάνουμε αν θέλουμε να ελέγξουμε τη μηδενική $H_0 : \eta_1 = \eta_2$ χωρίς να υποθέσουμε πως $\theta_1 = \theta_2$; Πως συγκρίνεται ο έλεγχος κλασσικό F-test (πλεονεκτήματα & μειονεκτήματα);

41) Έστω X_1, \dots, X_m i.i.d από F και ανεξάρτητες Y_1, \dots, Y_n i.i.d από G , όπου F, G συνεχείς. Θέλουμε να ελέγξουμε τη μηδενική ότι $F = G$ έναντι της εναλλακτικής οι F, G να διαφέρουν μεταξύ τους, χωρίς λοιπόν να περιορίζομαι σε εναλλακτικές που αφορούν μετατόπιση θέσης ή και κλίμακας. Σε ποιο μη-παραμετρικό στατιστικό μπορούμε να βασίσουμε ένα τέτοιο έλεγχο; Θα απορρίπτω τη μηδενική για μικρές ή για μεγάλες τιμές του στατιστικού και γιατί;

42) Έστω έχουμε δείγμα $(X_1, Y_1), \dots, (X_n, Y_n)$ από διμεταβλητή συνεχή κατανομή με περιθώριες F_x, F_y . Θέλουμε να ελέγξουμε τη μηδενική ότι η X είναι ανεξάρτητη από τη Y . Σε ποια δύο στατιστικά βασίζονται οι μη-παραμετρικοί έλεγχοι που προτείνονται (δε ζητούνται η αναμενόμενη τιμή και η διακύμανση); Πότε προκρίνονται αυτοί οι έλεγχοι του ελέγχου που βασίζεται στο κλασσικό μη παραμετρικό συντελεστή συσχέτισης;