

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
Τμήμα Επιστήμης Υπολογιστών

HY-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες-Εαρινό Εξάμηνο 2021
Διδάσκων: Π. Τσακαλίδης

Δεύτερη Σειρά Ασκήσεων

Ημερομηνία Ανάθεσης: 10/03/2021

Ημερομηνία Παράδοσης: 22/03/2021

ΣΗΜΑΝΤΙΚΕΣ ΟΔΗΓΙΕΣ

- Η παράδοση των ασκήσεών σας θα γίνεται ηλεκτρονικά, **ΑΠΟΚΛΕΙΣΤΙΚΑ** μέσω της σελίδας e-learn του μαθήματος, εδώ. Παραδόσεις με e-mail, ή οποιονδήποτε άλλον τρόπο δε θα γίνονται δεκτές και θα **μηδενίζονται**.
- Για την παράδοση, θα πρέπει να σκανάρετε ή να φωτογραφίσετε τις κόλλες των ασκήσεών σας και στη συνέχεια να φτιάξετε ένα PDF που θα περιλαμβάνει αυτές τις φωτογραφίες, δηλαδή το σύνολο της λυμένης σειράς ασκήσεων. Μπορείτε, αν προτιμάτε, να χρησιμοποιήσετε προγράμματα τύπου Latex ή Word. Για τις προγραμματιστικές ασκήσεις θα πρέπει να παραδώσετε ό,τι ζητείται στην άσκηση.
- Αριθμήστε τις σελίδες των ασκήσεών σας (από την άσκηση 1, στην άσκηση 2, στην άσκηση 3 κ.ο.κ.) και **ΣΙΓΟΥΡΕΥΤΕΙΤΕ** ότι τα σκαναρίσματα που έχετε κάνει έχουν δημιουργήσει εικόνες στις οποίες είναι **ΕΥΔΙΑΚΡΙΤΕΣ** και **ΕΥΚΟΛΟΔΙΑΒΑΣΤΕΣ** οι λύσεις σας. Θολές, και γενικότερα θορυβώδεις, εικόνες ενδέχεται να καταστήσουν αδύνατη τη διαδικασία της διόρθωσης και κατ' επέκταση να οδηγήσουν σε μηδενισμό των ασκήσεών σας.

Θέμα: Οριακά Θεωρήματα.

Άσκηση 1. Ένας στατιστικολόγος επιθυμεί να εκτιμήσει το μέσο ύψος h (σε m) ενός πληθυσμού, βασιζόμενος σε n ανεξάρτητα δείγματα $X_1, X_2, \dots, X_n$ τα οποία επιλέγονται ομοιόμορφα από ολόκληρο τον πληθυσμό. Για τον σκοπό αυτό, χρησιμοποιεί το δειγματικό μέσο $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ ως εκτιμήτρια του h , και διαλέγει τυπική απόκλιση ίση με 1 για τα δείγματα X_i .

- Να υπολογίσετε το μέγεθος του δείγματος, n , έτσι ώστε η τυπική απόκλιση του M_n να είναι το πολύ $1cm$.
- Να υπολογίσετε το μέγεθος του δείγματος, n , έτσι ώστε η **ανισότητα Chebyshev** να εγγυάται ότι η εκτίμηση είναι εντός $5cm$ από το h με πιθανότητα τουλάχιστον 0.99.

Άσκηση 2. Έστω $X_1, X_2, \dots, X_n$ ανεξάρτητες τ.μ. ομοιόμορφα κατανοημένες στο διάστημα $[-1, 1]$. Να αποδείξετε ότι η ακολουθία τ.μ. $Y_1, Y_2, \dots, Y_n$ συγκλίνει κατά πιθανότητα στο 0, για κάθε μία από τις παρακάτω περιπτώσεις:

- $Y_n = \frac{X_n}{n}$
- $Y_n = X_1 \cdot X_2 \cdot \dots \cdot X_n$ (**Υπόδειξη:** Χρησιμοποιείτε κατάλληλα την **ανισότητα Chebyshev**)

Άσκηση 3.

- (i) Από εμπειρικές μετρήσεις παρατηρούμε ότι η διάρκεια εκτέλεσης ενός αλγορίθμου είναι κατά μέσο όρο 17.5 δευτερόλεπτα, με τυπική απόκλιση ± 4 δευτερόλεπτα. Να υπολογίσετε την πιθανότητα σε 400 ανεξάρτητες χρήσεις του αλγορίθμου, ο συνολικός χρόνος εκτέλεσης να μην ξεπερνά τις 2 ώρες.
- (ii) Ο Γιάννης είναι φοιτητής του Τμήματος Επιστήμης Υπολογιστών, και στο πρόσφατο χειμερινό εξάμηνο εξετάστηκε επιτυχώς στο μάθημα ΗΥ-217 (Πιθανότητες). Στο εαρινό εξάμηνο, αποφασίζει να παρακολουθήσει το μάθημα ΗΥ-317 (Εφαρμοσμένες Στοχαστικές Διαδικασίες), χωρίς όμως να συμμετέχει στις διαλέξεις ούτε και να διαβάζει, βασιζόμενος στις γνώσεις που απέκτησε από το ΗΥ-217. Ως εκ τούτου, σχεδιάζει να δει την επίδοσή του στην πρόοδο του μαθήματος, ώστε να καθορίσει την περαιτέρω στάση του.

Το διαγώνισμα της προόδου του μαθήματος θα αποτελείται από 50 ερωτήσεις multiple-choice, η κάθε μία με 10 δυνατές απαντήσεις. Για κάθε σωστή απάντηση, ο εξεταζόμενος παίρνει 3 βαθμούς, και για κάθε λάθος χάνει 1 βαθμό. Ο Γιάννης απαντάει όλες τις ερωτήσεις στην τύχη, ανεξάρτητα τη μία από την άλλη. Να υπολογίσετε την πιθανότητα ο Γιάννης να πάρει προβιβασμό μέσο όρο βαθμολογίας (> 0.5) στην πρόοδο του ΗΥ-317.

Άσκηση 4. Ένα εργοστάσιο παράγει X_n αντικείμενα την ημέρα n , όπου X_n είναι ανεξάρτητες και όμοια κατανομημένες τ.μ. με μέση τιμή ίση με 5 και διασπορά ίση με 9.

- (i) Να υπολογίσετε μία προσέγγιση του ότι ο συνολικός αριθμός αντικειμένων που παράγονται σε 100 ημέρες είναι λιγότερος από 440.
- (ii) Να υπολογίσετε το μέγεθος του δείγματος, n , έτσι ώστε να ισχύει η σχέση:

$$P(X_1 + X_2 + \dots + X_n \geq 200 + 5n) \leq 0.05$$

- (iii) Έστω N η πρώτη ημέρα κατά την οποία ο συνολικός αριθμός παραγόμενων αντικειμένων υπερβαίνει τα 1000. Να υπολογίσετε μία προσέγγιση της πιθανότητας ότι $N \geq 220$.

Άσκηση 5. Σε αυτή την άσκηση σας ζητείται να προσομοιώσετε μια εφαρμογή του *Κεντρικού Οριακού Θεωρήματος*, χρησιμοποιώντας το Matlab. Ακολουθήστε τα παρακάτω βήματα:

- (i) Δημιουργίστε $N = 1000$ ψευδοτυχαία δείγματα μιας ομοιόμορφης τ.μ. X στο διάστημα $[0, 1]$ και μίας εκθετικής τ.μ. Y με παράμετρο $\lambda = 1$. (**Υπόδειξη:** Χρησιμοποιείστε την συνάρτηση `random()` με κατάλληλα ορίσματα για κάθε κατανομή.)
- (ii) Απεικονίστε τα ιστογράμματα των τ.μ. X και Y , για $n.bins = 20$. (**Υπόδειξη:** Χρησιμοποιείστε την συνάρτηση `histogram()` με κατάλληλα ορίσματα.)
- (iii) Δημιουργίστε $N = 1000$ ψευδοτυχαία δείγματα από $n = 20$ ομοιόμορφες τ.μ. $X_1, X_2, \dots, X_n$ στο διάστημα $[0, 1]$ και από $n = 20$ εκθετικές τ.μ. $Y_1, Y_2, \dots, Y_n$ με παράμετρο $\lambda = 1$. (**Υπόδειξη:** Χρησιμοποιείστε την συνάρτηση `random()` με κατάλληλα ορίσματα για κάθε κατανομή.)
- (iv) Θεωρείστε τους δειγματικούς μέσους όρους $M_{X_n} = \frac{X_1 + X_2 + \dots + X_n}{n}$ και $M_{Y_n} = \frac{Y_1 + Y_2 + \dots + Y_n}{n}$ και απεικονίστε τα ιστογράμμά τους για $n.bins = 20$. Τι παρατηρείτε; (**Υπόδειξη:** Χρησιμοποιείστε τις συναρτήσεις `mean()` και `histogram()` με κατάλληλα ορίσματα.)

- (v) Ορίστε τις τ.μ. $Z_{Xn} = \frac{X_1+X_2+\dots+X_n-n\mu_X}{\sigma_X\sqrt{n}} = \frac{S_{Xn}-n\mu_X}{\sigma_X\sqrt{n}}$ και $Z_{Yn} = \frac{Y_1+Y_2+\dots+Y_n-n\mu_Y}{\sigma_Y\sqrt{n}} = \frac{S_{Yn}-n\mu_Y}{\sigma_Y\sqrt{n}}$ και δώστε τα ιστογράμμά τους για $n_bins = 20$. Άλλαξε κάτι στις κατανομές; (**Υπόδειξη**: Για μία ομοιόμορφη τ.μ. έχουμε $\mu = \frac{a+b}{2}$ και $\sigma^2 = \frac{1}{12}(a-b)^2$, όπου a, b τα όρια του δοθέντος διαστήματος. Για μία εκθετική τ.μ. έχουμε $\mu = \frac{1}{\lambda}$ και $\sigma^2 = \frac{1}{\lambda^2}$.)
- (vi) Εκτιμήστε τις συναρτήσεις πυκνότητας πιθανότητας (σππ) των δειγματικών μέσων όρων M_{Xn} και M_{Yn} , μέσω ιστογραμμάτων. (**Υπόδειξη**: Χρησιμοποιείτε την συνάρτηση `hist()` με κατάλληλα ορίσματα για κάθε κατανομή, ώστε να λάβετε τόσο τον αριθμό των στοιχείων σε κάθε bin (counts) όσο και τη θέση του κάθε bin στον άξονα x'x (centers). Ακολουθώντας, εκτιμήστε την "πυκνότητα-εμβαδόν" κάθε bin μέσω της σχέσης: $p = \frac{bin_counts}{bin_length \cdot N}$.)
- (vii) Εκτιμήστε τις συναρτήσεις πυκνότητας πιθανότητας (σππ) των δειγματικών μέσων όρων M_{Xn} και M_{Yn} , μέσω κανονικών (gaussian) κατανομών. (**Υπόδειξη**: Χρησιμοποιείτε την συνάρτηση `pdf()` με κατάλληλα ορίσματα για τις παραμέτρους των δειγματικών μέσων (M_{Xn} και M_{Yn}) κάθε κατανομής.)
- (viii) Απεικονίστε σε κοινά διαγράμματα της εκτιμήτριες των συναρτήσεων πυκνότητας πιθανότητας (σππ) των δειγματικών μέσων (M_{Xn} και M_{Yn}) που βρήκατε στα παραπάνω ερωτήματα. Τι παρατηρείτε; (**Υπόδειξη**: Χρησιμοποιείτε την συνάρτηση `plot()` με κατάλληλα ορίσματα.)
- (ix) "Παίξτε" με τις μεταβλητές N , n_bins και n και καταγράψτε τις παρατηρήσεις σας σε κάθε περίπτωση. Τι αλλάζει στις κατανομές; Υπάρχουν αλλαγές στις διασπορές; Δώστε τα αντίστοιχα ιστογράμματα. (Π.χ. $N = 10000, n_bins = 100, n = 100$.)

Παραθέστε τον κώδικά σας, τα ζητούμενα γραφήματα και μία μικρή αναφορά με τις παρατηρήσεις σας για τα αποτελέσματα.