

HY-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες-Εαρινό Εξάμηνο 2025
Διδάσκων: Π. Τσακαλίδης

Λύσεις Δεύτερης Σειράς Ασκήσεων

Άσκηση 1.

Ένας αστροφυσικός προσπαθεί να εκτιμήσει τη λάμψη b ενός αστεριού με ένα σφάλμα το οποίο είναι μικρότερο από 0.1 με πιθανότητα μεγαλύτερη από 95%. Το όργανο μέτρησης που διαθέτει προσθέτει θόρυβο N με μέση τιμή 0 και τυπική απόκλιση 2, ανεξάρτητα από μέτρηση σε μέτρηση. Συνεπώς, κάθε μέτρηση X_k μπορεί να μοντελοποιηθεί ως το άθροισμα της προς εκτίμηση τιμής της λάμψης b και της συνιστώσας του θορύβου N_k :

$$X_k = b + N_k$$

Ο αστρονόμος πραγματοποιεί n μετρήσεις και χρησιμοποιεί το δειγματικό μέσο όρο ως την εκτίμητρια της ποσότητας b :

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- (i) Χρησιμοποιείτε το Κεντρικό Οριακό Θεώρημα για να υπολογίσετε το ελάχιστο πλήθος των μετρήσεων n που θα επιτρέψουν στον αστροφυσικό να εκτιμήσει τη λάμψη του αστεριού με τα επίπεδα ακρίβειας και εμπιστοσύνης που αναφέρονται στην εκφώνηση. Ισχύει ότι $\Phi(1.96) = 0.975$.
- (ii) Χρησιμοποιείτε την ανισότητα Chebyshev για τον ίδιο σκοπό. Τι παρατηρείτε;

Λύση:

Έχουμε ότι η μέτρηση X_k είναι της μορφής:

$$X_k = b + N_k$$

όπου το N_k είναι θόρυβος με μέση τιμή $\mathbb{E}[N_k] = 0$ και τυπική απόκλιση $\sigma_{N_k} = 2$. Έχουμε n ανεξάρτητες μετρήσεις, οπότε η δειγματική μέση τιμή M_n εκτιμά την τιμή της λάμψης b ως εξής:

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Η τυπική απόκλιση του κάθε X_k είναι $\sigma_{X_k} = \sigma_{N_k} = 2$. Συνεπώς, η τυπική απόκλιση του μέσου όρου M_n είναι:

$$\sigma_{M_n} = \frac{\sigma_{X_k}}{\sqrt{n}} = \frac{2}{\sqrt{n}}$$

(α) Χρήση του Κεντρικού Οριακού Θεωρήματος:

Από το Κεντρικό Οριακό Θεώρημα, γνωρίζουμε ότι αν n είναι αρκετά μεγάλο, τότε ο μέσος όρος των n μετρήσεων M_n ακολουθεί μια κανονική κατανομή με μέση τιμή $\mu = b$ και τυπική απόκλιση $\sigma_{M_n} = \frac{2}{\sqrt{n}}$.

Για να εξασφαλίσουμε ότι το σφάλμα εκτίμησης της λάμψης είναι μικρότερο από 0.1 με πιθανότητα μεγαλύτερη από 95%, πρέπει να ισχύει:

$$P(|M_n - b| < 0.1) > 0.95$$

Αυτό μπορεί να αναπαρασταθεί ως:

$$P\left(\frac{|M_n - b|}{\sigma_{M_n}} < \frac{0.1}{\sigma_{M_n}}\right) > 0.95$$

Δεδομένου ότι η τυπική απόκλιση του M_n είναι $\sigma_{M_n} = \frac{2}{\sqrt{n}}$, έχουμε:

$$P\left(\left|\frac{M_n - b}{\sigma_{M_n}}\right| < \frac{0.1}{\sigma_{M_n}}\right) > 0.95$$

Αυτός ο όρος αντιστοιχεί σε μία τυπική κανονική κατανομή. Στη συνέχεια, αφού το $\frac{M_n - b}{\sigma_{M_n}}$ ακολουθεί κανονική κατανομή με μέση τιμή μηδέν και τυπική απόκλιση ίση με 1, από τη γνωστή κανονική κατανομή ξέρουμε ότι:

$$P(|Z| < z_{\alpha/2}) = 0.95$$

όπου $Z \sim N(0, 1)$ και $z_{\alpha/2} = 1.96$ είναι το κρίσιμο σημείο της κανονικής κατανομής για $P = 0.95$. Έτσι, μπορούμε να γράψουμε:

$$P\left(\left|\frac{M_n - b}{\sigma_{M_n}}\right| < 1.96\right) = 0.95$$

Με την αντίστοιχη σχέση για το σφάλμα 0.1 και τη γνωστή τυπική απόκλιση του M_n , αυτό οδηγεί στην εξίσωση:

$$\frac{0.1}{\sigma_{M_n}} = 1.96$$

και επομένως:

$$0.1 = 1.96 \cdot \frac{2}{\sqrt{n}}$$

Λύνοντας για το n :

$$\begin{aligned}\frac{0.1}{1.96} &= \frac{2}{\sqrt{n}} \\ \sqrt{n} &= \frac{2 \cdot 1.96}{0.1} \\ n &= \left(\frac{2 \cdot 1.96}{0.1}\right)^2 = \left(\frac{3.92}{0.1}\right)^2 = 1536.64\end{aligned}$$

Άρα, το ελάχιστο πλήθος μετρήσεων είναι περίπου:

$$n \approx 1537$$

(β) **Χρήση της Ανισότητας Chebyshev:**

Η ανισότητα Chebyshev μας λέει ότι για οποιαδήποτε τυχαία μεταβλητή X με μέση τιμή μ και τυπική απόκλιση σ , ισχύει:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Αντιστρόφως, για την πιθανότητα να έχουμε σφάλμα μικρότερο από 0.1, έχουμε:

$$P(|M_n - b| < 0.1) \geq 1 - \frac{1}{k^2}$$

όπου k είναι το k που αντιστοιχεί στο σφάλμα 0.1. Λύνοντας για το k :

$$P(|M_n - b| < 0.1) \geq 0.95$$

$$1 - \frac{1}{k^2} = 0.95$$

$$\frac{1}{k^2} = 0.05$$

$$k = \frac{1}{\sqrt{0.05}} \approx 4.472$$

Από την ανισότητα *Chebyshev*, γνωρίζουμε ότι:

$$0.1 = k \cdot \frac{2}{\sqrt{n}}$$

Άρα:

$$0.1 = 4.472 \cdot \frac{2}{\sqrt{n}}$$

Λύνοντας ως προς το n :

$$\sqrt{n} = \frac{4.472 \cdot 2}{0.1} = \frac{8.944}{0.1} = 89.44$$

$$n = (89.44)^2 \approx 8015$$

Άρα, το πλήθος των μετρήσεων που απαιτείται για την ανισότητα *Chebyshev* είναι περίπου:

$$n \approx 8015$$

Άσκηση 2.

Έστω $X_1, X_2, \dots, X_n$ ανεξάρτητες τ.μ. ομοιόμορφα κατανομημένες 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:

(i) $Y_n = \frac{X_n}{n}$

(ii) $Y_n = X_1 \cdot X_2 \cdot \dots \cdot X_n$ (**Υπόδειξη:** Χρησιμοποιείτε κατάλληλα την *ανισότητα Chebyshev*)

Λύση:

Εφόσον ζητείται να αποδειχθεί ότι η ακολουθία τ.μ. $Y_1, Y_2, \dots, Y_n$ συγκλίνει κατά πιθανότητα στο 0, θα πρέπει να αποδείξουμε ότι ισχύει η σχέση:

$$\lim_{n \rightarrow +\infty} P(|Y_n - 0| \geq \epsilon) = \lim_{n \rightarrow +\infty} P(|Y_n| \geq \epsilon) = 0 \Rightarrow P(|Y_n| \geq \epsilon) \rightarrow 0$$

(i) Για κάθε $\epsilon > 0$ έχουμε:

$$P(|Y_n - 0| \geq \epsilon) = P(|Y_n| \geq \epsilon) = 0$$

για όλα τα n με $\frac{1}{n} < \epsilon$. Άρα, τελικά είναι: $P(|Y_n| \geq \epsilon) \rightarrow 0$.

(ii) Εφόσον οι $X_1, X_2, \dots, X_n$ είναι ανεξάρτητες τ.μ., έχουμε:

- $E[Y_n] = E[X_1] \cdot E[X_2] \cdot \dots \cdot E[X_n] = 0 \cdot 0 \cdot \dots \cdot 0 = 0$
- $VAR[Y_n] = E[Y_n^2] - (E[Y_n])^2 = E[Y_n^2] - 0^2 = E[Y_n^2] = E[X_1^2] \cdot E[X_2^2] \cdot \dots \cdot E[X_n^2] = \frac{(1-(-1))^2}{12} \cdot \frac{(1-(-1))^2}{12} \cdot \dots \cdot \frac{(1-(-1))^2}{12} = \frac{4}{12} \cdot \frac{4}{12} \cdot \dots \cdot \frac{4}{12} = \left(\frac{4}{12}\right)^n = \left(\frac{1}{3}\right)^n$

Από τις παραπάνω σχέσεις, παρατηρούμε ότι ισχύει: $VAR[Y_n] \rightarrow 0$.

Εφόσον όλες οι Y_n έχουν κοινή μέση τιμή (ίση με 0), κάνοντας χρήση της *ανισότητας Chebyshev*, έχουμε:

$$P(|Y_n - \mu_{Y_n}| \geq c) \leq \frac{\sigma_{Y_n}^2}{c^2} \xrightarrow[\sigma_{Y_n}^2 = (\frac{1}{3})^n]{\mu_{Y_n} = 0} P(|Y_n - 0| \geq c) \leq \frac{(\frac{1}{3})^n}{c^2} \xrightarrow{c=\epsilon} P(|Y_n| \geq \epsilon) \leq \frac{(\frac{1}{3})^n}{\epsilon^2}$$

Άρα, τελικά είναι: $P(|Y_n| \geq \epsilon) \rightarrow 0$.

Άσκηση 3.

- (i) Στην εκπομπή 'Με Αρετή και Τόλμη' της ΕΡΤ, ένας έμπειρος πιλότος ανέφερε ότι η διάρκεια την οποία μπορεί να έχει 'κλειδωμένο' στο στόχαστρο του ένα έχθρικό αεροσκάφος είναι κατά μέσο όρο 17.5 δευτερόλεπτα, με τυπική απόκλιση ± 4 δευτερόλεπτα. Να υπολογίσετε την πιθανότητα σε 400 ανεξάρτητες καταδιώξεις, ο συνολικός χρόνος των 'κλειδωμένων αεροσκαφών' να μην ξεπερνά τις 2 ώρες.
- (ii) Παρακολουθείτε το τηλεπαιχνίδι 'Μια στις δέκα!' το οποίο αποτελείται από 50 ερωτήσεις multiple-choice, η κάθε μία με 10 δυνατές απαντήσεις. Για κάθε σωστή απάντηση, ο παίκτης παίρνει 3 βαθμούς, και για κάθε λάθος χάνει 1 βαθμό, ενώ για να κερδίσει κάποιο χρηματικό έπαθλο χρειάζεται να συγκεντρώσει μέσο όρο βαθμολογίας (> 0.5). Γνωρίζετε ότι ο φίλος σας ο Κώστας που συμμετέχει στο επεισόδιο απαντάει όλες τις ερωτήσεις στην τύχη, ανεξάρτητα τη μία από την άλλη. Να υπολογίσετε την πιθανότητα ο Κώστας να φύγει με χρηματικό έπαθλο.

Λύση

- (i) Έστω X_i ο χρόνος που έχουμε ένα κλειδωμένο αεροσκάφος την φορά i , $i = 1, 2, \dots, 400$. Με βάση τα δεδομένα της άσκησης, έχουμε ότι:

- Οι X_i είναι ανεξάρτητες τ.μ.
- $E[X_i] = 17.5$ και $\sigma_{X_i} = 4$

Άρα, για τη ζητούμενη πιθανότητα-με χρήση του *Κεντρικού Οριακού Θεωρήματος*-έχουμε:

$$\begin{aligned} P(S_n < 2 \cdot 60 \cdot 60) &= P(S_n < 7200) \stackrel{n=400}{=} P\left(\frac{S_{400} - 400 \cdot 17.5}{4\sqrt{400}} \leq \frac{7200 - 400 \cdot 17.5}{4\sqrt{400}}\right) \\ &= P\left(\frac{S_{400} - 7000}{80} \leq \frac{7200 - 7000}{80}\right) = P\left(\frac{S_{400} - 7000}{80} \leq \frac{200}{80}\right) \\ &= \Phi(2.5) = 0.9938 \end{aligned}$$

- (ii) Δεδομένου ότι ο Κώστας απαντά κάθε ερώτηση στην τύχη, η πιθανότητα να επιλέξει την σωστή απάντηση είναι $\frac{1}{10}$ (αφού κάθε multiple-choice ερώτηση έχει 10 δυνατές απαντήσεις).

Ορίζουμε τις τ.μ. X_i ($i = 1, 2, \dots, 50$) ως εξής:

$$X_i = \begin{cases} 3 & \text{με πιθανότητα } p = \frac{1}{10} \\ -1 & \text{με πιθανότητα } 1 - p = \frac{9}{10} \end{cases}$$

Για τις τ.μ. X_i , έχουμε:

- $E[X_i] = 3 \cdot \frac{1}{10} - 1 \cdot \frac{9}{10} = -\frac{6}{10} = -\frac{3}{5}$
- $VAR[X_i] = E[X_i^2] - (E[X_i])^2 = 3^2 \cdot \frac{1}{10} + (-1)^2 \cdot \frac{9}{10} - \frac{6}{10} = \frac{12}{10} = \frac{6}{5}$

Άρα, η πιθανότητα να συγκεντρώσει ο Κώστας μέσο όρο βαθμολογίας (> 0.5):

$$\begin{aligned} P\left(\frac{S_n}{n} > 0.5\right) &\stackrel{n=50}{=} P(S_{50} > 50 \cdot 0.5) = P(S_{50} > 25) \\ &= P\left(\frac{S_{50} - 50 \cdot (-\frac{3}{5})}{50\sqrt{\frac{6}{5}}} > \frac{25 - 50 \cdot (-\frac{3}{5})}{50\sqrt{\frac{6}{5}}}\right) \\ &= P\left(\frac{S_{50} + 30}{50\sqrt{\frac{6}{5}}} > \frac{25 + 30}{50\sqrt{\frac{6}{5}}}\right) = P\left(\frac{S_{50} + 30}{50\sqrt{\frac{6}{5}}} > \frac{55}{50\sqrt{\frac{6}{5}}}\right) \\ &= P\left(\frac{S_{50} + 30}{50\sqrt{\frac{6}{5}}} > \frac{11}{10\sqrt{\frac{6}{5}}}\right) = 1 - P\left(\frac{S_{50}}{50\sqrt{\frac{6}{5}}} \leq \frac{11}{10\sqrt{\frac{6}{5}}}\right) \\ &= 1 - \Phi\left(\frac{11}{10\sqrt{\frac{6}{5}}}\right) = 1 - \Phi(1.0042) = 1 - 0.8423 = 0.1577 \end{aligned}$$

Άσκηση 4.

Ένα εργοστάσιο παράγει X_n αντικείμενα την ημέρα n , όπου X_n είναι ανεξάρτητες και όμοια κατα-
νεμημένες 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 $N \geq 220$.

Λύση:

- (i) Έστω $S_n = X_1 + X_2 + \dots + X_n$ ο συνολικός αριθμός αντικειμένων που παράγονται σε n μέρες. Με βάση τα δεδομένα της άσκησης (X_n είναι ανεξάρτητες και όμοια κατανεμημένες τ.μ. με μέση τιμή ίση με 5 και διασπορά ίση με 9), έχουμε:

•

$$\begin{aligned} E[S_n] &= E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n] \\ &= 5 + 5 + \dots + 5 = 5n \end{aligned}$$

•

$$\begin{aligned} VAR[S_n] &= VAR[X_1 + X_2 + \dots + X_n] = VAR[X_1] + VAR[X_2] + \dots + VAR[X_n] \\ &= 9 + 9 + \dots + 9 = 9n \end{aligned}$$

•

$$\sigma_{S_n} = \sqrt{VAR[S_n]} = \sqrt{9n} = 3\sqrt{n}$$

Άρα, για τη ζητούμενη προσέγγιση-με χρήση του *Κεντρικού Οριακού Θεωρήματος*-έχουμε:

$$\begin{aligned} P(S_{100} < 440) &= P(S_{100} \leq 439.5) \stackrel{n=100}{=} P\left(\frac{S_{100} - 500}{30} \leq \frac{439.5 - 500}{30}\right) \\ &= P\left(\frac{S_{100} - 500}{30} \leq -2.0167\right) = \Phi(-2.0167) \\ &= 1 - \Phi(2.0167) = 1 - 0.9781 = 0.0219 \end{aligned}$$

- (ii) Εφαρμόζοντας το *Κεντρικό Οριακό Θεώρημα*, έχουμε:

$$\begin{aligned} P(X_1 + X_2 + \dots + X_n \geq 200 + 5n) &\leq 0.05 \Rightarrow P(S_n \geq 200 + 5n) \leq 0.05 \\ \Rightarrow P\left(\frac{S_n - 5n}{3\sqrt{n}} \geq \frac{200 + 5n - 5n}{3\sqrt{n}}\right) &\leq 0.05 \Rightarrow P\left(\frac{S_n - 5n}{3\sqrt{n}} \geq \frac{200}{3\sqrt{n}}\right) \leq 0.05 \\ \Rightarrow 1 - P\left(\frac{S_n - 5n}{3\sqrt{n}} \leq \frac{200}{3\sqrt{n}}\right) &\leq 0.05 \Rightarrow 1 - \Phi\left(\frac{200}{3\sqrt{n}}\right) \leq 0.05 \\ \Rightarrow \Phi\left(\frac{200}{3\sqrt{n}}\right) &\geq 0.95 \Rightarrow \Phi\left(\frac{200}{3\sqrt{n}}\right) \stackrel{\text{Normal Tables}}{\geq} \Phi(1.65) \stackrel{\Phi(1.65) \approx 0.95}{\geq} \Phi(1.65) \\ \Rightarrow \frac{200}{3\sqrt{n}} &\geq 1.65 \Rightarrow \sqrt{n} \leq 40.4040 \\ \Rightarrow n &\leq 1632.483216 \Rightarrow n \leq 1632 \end{aligned}$$

- (iii) Το γεγονός $N \geq 220$ (δηλαδή, ότι χρειάζονται τουλάχιστον 220 ημέρες ώστε ο συνολικός αριθμός παραγόμενων αντικειμένων να υπερβαίνει τα 1000) είναι το ίδιο με το γεγονός $S_{219} \leq 1000$ (δηλαδή, ότι δεν παρήχθησαν περισσότερα από 1000 αντικείμενα τις πρώτες 219 ημέρες).

Έτσι, εφαρμόζοντας το Κεντρικό Οριακό Θεώρημα, έχουμε:

$$\begin{aligned} P(N \geq 220) &= P(S_{219} \leq 1000) = P\left(\frac{S_{219} - 5 \cdot 219}{3\sqrt{219}} \leq \frac{1000 - 5 \cdot 219}{3\sqrt{219}}\right) \\ &= P\left(\frac{S_{219} - 1095}{3\sqrt{219}} \leq \frac{1000 - 1095}{3\sqrt{219}}\right) = P\left(\frac{S_{219} - 1095}{3\sqrt{219}} \leq \frac{-95}{3\sqrt{219}}\right) \\ &= P\left(\frac{S_{219} - 1095}{3\sqrt{219}} \leq -2.1398\right) = \Phi(-2.1398) \\ &= 1 - \Phi(2.1398) = 1 - 0.9838 = 0.0162 \end{aligned}$$

Άσκηση 5.

Η μεταβολή της στάθμης ενός ταμιευτήρα σε ένα υδρολογικό έτος ακολουθεί ομοιόμορφη κατανομή στο $[0.5, 0.8]$. Οι τιμές της μεταβολής αναφέρονται σε μέτρα.

1. Να υπολογίσετε προσεγγιστικά την πιθανότητα μετά από 25 υδρολογικά έτη, ο ταμιευτήρας να έχει επιφάνεια η οποία ανέβηκε τουλάχιστον κατά 16μ, θεωρώντας ότι αρχίσαμε να μετράμε από το φετινό υδρολογικό έτος.
2. Να βρείτε τον ελάχιστο αριθμό υδρολογικών ετών που πρέπει να περάσουν, ώστε με πιθανότητα τουλάχιστον 80% η στάθμη του ταμιευτήρα να ανέβηκε τουλάχιστον κατά 16μ (σε σχέση με φέτος).

Λύση

Μέσος όρος και τυπική απόκλιση της ετήσιας μεταβολής

Έστω $X \sim U(0.5, 0.8)$, όπου X η ετήσια μεταβολή της στάθμης. Ο μέσος όρος της κατανομής είναι:

$$E[X] = \frac{a+b}{2} = \frac{0.5+0.8}{2} = 0.65$$

και η διακύμανση είναι:

$$\text{Var}(X) = \frac{(b-a)^2}{12} = \frac{(0.8-0.5)^2}{12} = \frac{0.09}{12} = 0.0075$$

Άρα, η τυπική απόκλιση είναι:

$$\sigma_X = \sqrt{0.0075} \approx 0.0866$$

(α) Πιθανότητα να ανέβει η στάθμη τουλάχιστον κατά 16μ σε 25 έτη

Η συνολική μεταβολή μετά από $n = 25$ έτη είναι το άθροισμα των ανεξάρτητων τυχαίων μεταβλητών:

$$S_n = X_1 + X_2 + \dots + X_n$$

Το άθροισμα ακολουθεί προσεγγιστικά κανονική κατανομή λόγω του κεντρικού οριακού θεωρήματος:

$$S_n \sim \mathcal{N}(nE[X], n\sigma_X^2)$$

Δηλαδή:

$$S_{25} \sim \mathcal{N}(25 \times 0.65, 25 \times 0.0075) = \mathcal{N}(16.25, 0.1875)$$

Η πιθανότητα που ζητείται είναι:

$$P(S_{25} \geq 16) = P\left(\frac{S_{25} - 16.25}{\sqrt{0.1875}} \geq \frac{16 - 16.25}{\sqrt{0.1875}}\right)$$

Υπολογίζουμε την κανονικοποιημένη μεταβλητή:

$$Z = \frac{16 - 16.25}{\sqrt{0.1875}} = \frac{-0.25}{0.433} \approx -0.577$$

Από τον πίνακα της κανονικής κατανομής:

$$P(Z \geq -0.577) = 1 - P(Z \leq -0.577) \approx 1 - \Phi(0.577) = 1 - 0.282 = 0.718$$

Άρα, η πιθανότητα είναι:

$$P(S_{25} \geq 16) \approx 0.718$$

(β) Ελάχιστος αριθμός ετών ώστε η πιθανότητα να είναι τουλάχιστον 80%

Ζητούμε το ελάχιστο n ώστε:

$$P(S_n \geq 16) \geq 0.8$$

Με τον ίδιο συλλογισμό:

$$S_n \sim \mathcal{N}(nE[X], n\sigma_X^2)$$

Αντικαθιστούμε:

$$S_n \sim \mathcal{N}(0.65n, 0.0075n)$$

Η αντίστοιχη κανονικοποιημένη μεταβλητή είναι:

$$Z = \frac{16 - 0.65n}{\sqrt{0.0075n}}$$

Από τον πίνακα της κανονικής κατανομής, θέλουμε:

$$P(Z \geq z) = 0.8 \Rightarrow P(Z \leq z) = 0.2$$

Από τους πίνακες, το z που ικανοποιεί αυτή τη συνθήκη είναι περίπου -0.842 , δηλαδή:

$$\frac{16 - 0.65n}{\sqrt{0.0075n}} = -0.842$$

Λύνουμε ως προς n :

$$16 - 0.65n = -0.842\sqrt{0.0075n}$$

Υψώνουμε στο τετράγωνο:

$$(16 - 0.65n)^2 = 0.708 \cdot n$$

Επιλύοντας αριθμητικά, βρίσκουμε ότι το ελάχιστο n που ικανοποιεί την ανίσωση είναι:

$$n = 29$$

Άσκηση 6. Σε αυτή την άσκηση σας ζητείται να προσομοιώσετε μια εφαρμογή του Κεντρικού Οριακού Θεωρήματος, χρησιμοποιώντας την *Python*. Ακολουθήστε τα παρακάτω βήματα:

1. Δημιουργήστε $N = 1200$ ψευδοτυχαία δείγματα μιας ομοιόμορφης τυχαίας μεταβλητής X στο διάστημα $[0, 1]$ και μιας εκθετικής τυχαίας μεταβλητής Y με παράμετρο $\lambda = 1$. (Υπόδειξη: Χρησιμοποιείτε τη συνάρτηση `numpy.random.uniform()` και `numpy.random.exponential()`).
2. Απεικονίστε τα ιστογράμματα των τυχαίων μεταβλητών X και Y , με $n_{bins} = 25$. (Υπόδειξη: Χρησιμοποιείτε τη συνάρτηση `matplotlib.pyplot.hist()`).

- Δημιουργήστε $N = 1200$ ψευδοτυχαία δείγματα από $n = 25$ ομοιόμορφες τυχαίες μεταβλητές $X_1, X_2, \dots, X_n$ στο διάστημα $[0, 1]$ και από $n = 25$ εκθετικές τυχαίες μεταβλητές $Y_1, Y_2, \dots, Y_n$ με παράμετρο $\lambda = 1$.
- Υπολογίστε 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.
- Ορίστε τις μεταβλητές $Z_{X_n} = \frac{M_{X_n}-\mu_X}{\sigma_X/\sqrt{n}}$ και $Z_{Y_n} = \frac{M_{Y_n}-\mu_Y}{\sigma_Y/\sqrt{n}}$, όπου μ_X και σ_X είναι η μέση τιμή και η τυπική απόκλιση της ομοιόμορφης κατανομής, ενώ μ_Y και σ_Y της εκθετικής. Απεικονίστε τα ιστογράμμά τους.
- Εκτιμήστε τις συναρτήσεις πυκνότητας πιθανότητας (*PDFs*) των M_{X_n} και M_{Y_n} μέσω ιστογραμμάτων. Υπολογίστε την πυκνότητα κάθε *bin* με τη σχέση:

$$p = \frac{\text{bin_counts}}{\text{bin_length} \cdot N}$$

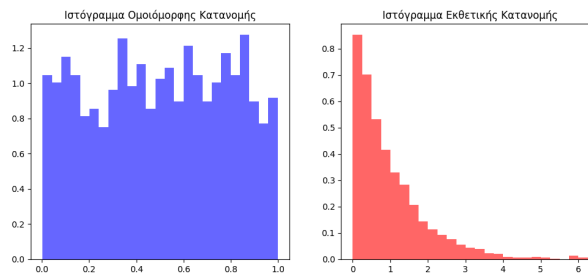
- Εκτιμήστε τις συναρτήσεις πυκνότητας πιθανότητας των M_{X_n} και M_{Y_n} μέσω κανονικών κατανομών (*Gaussian*). Χρησιμοποιήστε τη συνάρτηση `scipy.stats.norm.pdf()` για την εκτίμηση της κανονικής πυκνότητας.
- Απεικονίστε σε κοινά διαγράμματα τις εκτιμήτριες των συναρτήσεων πυκνότητας πιθανότητας των M_{X_n} και M_{Y_n} και συγκρίνετέ τις με τις αντίστοιχες θεωρητικές *Gaussian* κατανομές.
- Πειραματιστείτε με διαφορετικές τιμές των N , n_{bins} και n (π.χ., $N = 10000$, $n_{bins} = 100$, $n = 100$) και καταγράψτε τις παρατηρήσεις σας.

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

Λύση:

Το Κεντρικό Οριακό Θεώρημα (ΚΟΘ) αναφέρει ότι, για μεγάλα μεγέθη δείγματος, η κατανομή των δειγματικών μέσων προσεγγίζει την κανονική κατανομή ανεξάρτητα από την αρχική κατανομή των δεδομένων. Σε αυτήν την εργασία, εξετάζουμε την εφαρμογή του ΚΟΘ στις ομοιόμορφες και εκθετικές κατανομές.

- Ιστόγραμμα των Αρχικών Κατανομών Παρακάτω φαίνεται η κατανομή των αρχικών δεδομένων, τόσο για την ομοιόμορφη όσο και για την εκθετική κατανομή.

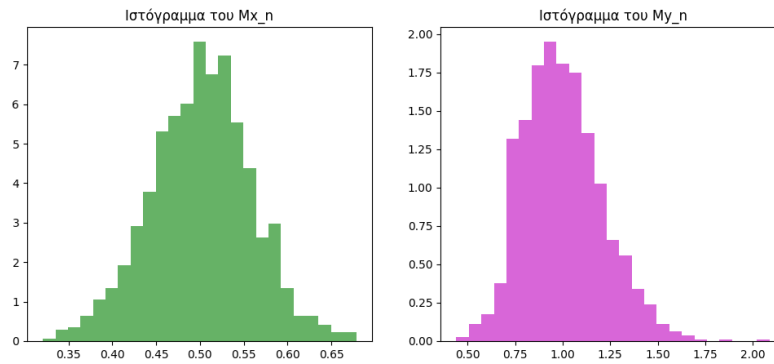
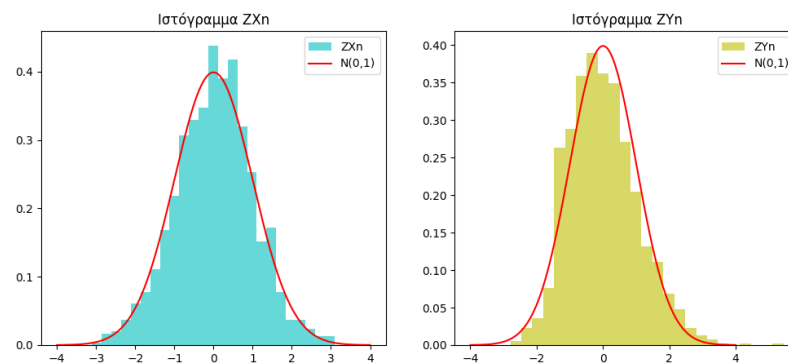


Σχήμα 1: Ιστόγραμμα Ομοιόμορφης και Εκθετικής Κατανομής

- Ιστόγραμμα των Δειγματικών Μέσων
Με τον υπολογισμό του μέσου όρου δείγματος από τις δύο κατανομές, παίρνουμε τα ακόλουθα ιστογράμματα.
- Κανονικοποιημένες Μεταβλητές
Οι κανονικοποιημένες μεταβλητές Z_{X_n} και Z_{Y_n} ορίζονται ως:

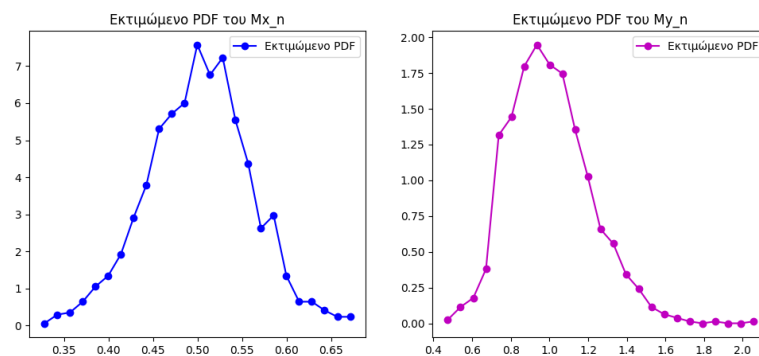
$$Z_{X_n} = \frac{M_{X_n} - \mu_X}{\sigma_X/\sqrt{n}}, \quad Z_{Y_n} = \frac{M_{Y_n} - \mu_Y}{\sigma_Y/\sqrt{n}} \quad (1)$$

Τα ιστογράμματα των μεταβλητών αυτών συγκρίνονται με την κανονική κατανομή $N(0, 1)$.

Σχήμα 2: Ιστόγραμμα των Δειγματικών Μέσων (M_{X_n} και M_{Y_n})Σχήμα 3: Ιστόγραμμα των ZX_n και ZY_n

4. Εκτίμηση Πυκνότητας Πιθανότητας

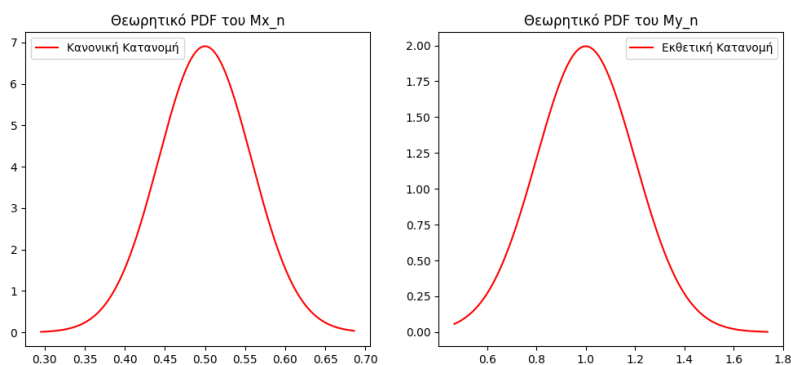
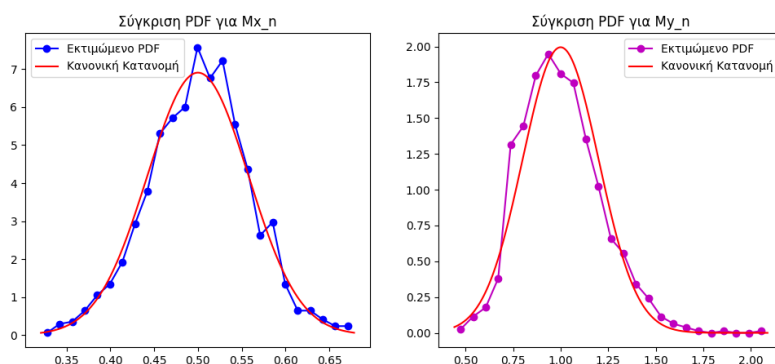
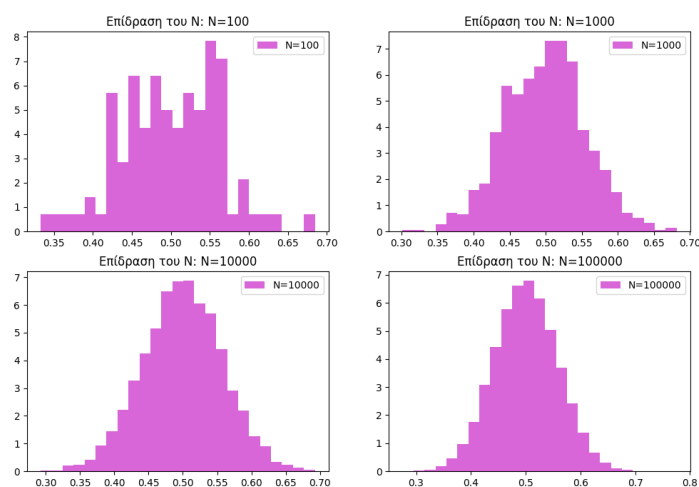
Εκτιμούμε την πυκνότητα πιθανότητας (PDF) από το δείγμα και συγκρίνουμε με την θεωρητική κατανομή.

Σχήμα 4: Εκτιμώμενο PDF των Δειγματικών Μέσων

5. Σύγκριση Εκτιμώμενου και Θεωρητικού PDF Συγκρίνουμε το εκτιμώμενο PDF με το θεωρητικό PDF της κανονικής κατανομής.

6. Πειραματισμός με Μεγαλύτερο Δείγμα

Αυξάνοντας το πλήθος των δειγμάτων και το μέγεθος του δείγματος, βλέπουμε ακόμα πιο ομαλή προσέγγιση της κανονικής κατανομής.

Σχήμα 5: Θεωρητικό PDF των Δειγματικών ΜέσωνΣχήμα 6: Σύγκριση Εκτιμώμενου και Θεωρητικού PDF Σχήμα 7: Πειραματισμός με $N = [100, 1000, 10000, 100000]$

Τα αποτελέσματα επιβεβαιώνουν το Κεντρικό Οριακό Θεώρημα. Καθώς το μέγεθος δείγματος αυξάνεται, η κατανομή των δειγματικών μέσων προσεγγίζει την κανονική κατανομή, ανεξάρτητα από την αρχική κατανομή των δεδομένων.