

ΗΥ-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες-Εαρινό Εξάμηνο 2022
Διδάσκων: Π. Τσακαλίδης

Λύσεις Δεύτερης Σειράς Ασκήσεων

Τα ερωτήματα με * είναι εκτός διδακτέας ύλης και παραθέτονται για όποιον ενδιαφέρεται.

Άσκηση 1.

- (i) Με βάση τη θεωρία, ξέρουμε ότι η τυπική απόκλιση του M_n δίνεται από τον εξής τύπο:

$$\sigma_{M_n} = \frac{1}{\sqrt{n}}$$

Με βάση την εκφώνηση της άσκησης, έχουμε:

$$\sigma_{M_n} \leq 0.1 \Rightarrow \frac{1}{\sqrt{n}} \leq 0.1 \Rightarrow \sqrt{n} \geq \frac{1}{0.1} \Rightarrow \sqrt{n} \geq 10 \Rightarrow n \geq 100$$

- (ii) Με βάση την εκφώνηση της άσκησης, θέλουμε να ισχύει:

$$P(|M_n - h| \leq 0.01) \geq 0.80$$

Κάνοντας χρήση της ανισότητας *Chebyshev*, έχουμε:

$$\begin{aligned} P(|M_n - h| \leq 0.01) &\stackrel{h=E[M_n]}{=} P(|M_n - E[M_n]| \leq 0.01) \\ &= 1 - P(|M_n - E[M_n]| \geq 0.01) \\ &\stackrel{\sigma_{M_n}^2 = \frac{1}{n}}{\geq} \stackrel{c=0.01}{1 - \frac{\frac{1}{n}}{(0.01)^2}} \end{aligned}$$

Άρα, τελικά, έχουμε:

$$\begin{aligned} P(|M_n - h| \leq 0.01) &\geq 0.80 \Rightarrow 1 - \frac{\frac{1}{n}}{(0.01)^2} \geq 0.80 \\ &\Rightarrow \frac{\frac{1}{n}}{(0.01)^2} \leq 0.20 \Rightarrow \frac{1}{n} \leq 0.00002 \\ &\Rightarrow n \geq \frac{1}{0.00002} \Rightarrow n \geq 50000 \end{aligned}$$

Άσκηση 2.

- (i) Τα σωματίδια που καταγράφονται από τον ανιχνευτή είναι τύπου A και σχηματίζουν μια σ.δ. Βερνουλι με παράμετρο p_{Aa} . Συνεπώς ο αριθμός των σωματιδίων που καταγράφονται σε ένα πλήθος 50 αφίξεων ακολουθεί Διωνυμική κατανομή με παραμέτρους $n = 50$ και $p = p_{Aa}$. Επομένως, ο μέσος αριθμός των σωματιδίων που καταγράφονται είναι $np = 50p_{Aa}$.

- (ii) Η τυχαία μεταβλητή K ισοδυναμεί με το πλήθος των προσπαθειών έως την 1η επιτυχία σε μια σ.δ. *Bernoulli* με παράμετρο p_{Aa} . Επομένως η τυχαία μεταβλητή K ακολουθεί γεωμετρική κατανομή:

$$p_K(k) = (1 - p_{Aa})^{k-1} p_{Aa}; k = 1, 2, 3, \dots$$

- (iii) Η τυχαία μεταβλητή L ισοδυναμεί με το πλήθος των δοκιμών έως την 4η επιτυχία σε μια σ.δ *Bernoulli* με παράμετρο p_{AA} . Επομένως η τ.μ. L ακολουθεί κατανομή *Pascal* τάξης 4:

$$p_L(l) = \left(\frac{l-1}{2}\right)(p_{AA})^4(1-p_{AA})^{l-4}; l = 3, 4, 5, 6, \dots$$

- (iv) * Προφανώς, η τ.μ. $N(t)$ ακολουθεί κατανομή *Poisson* με παράμετρο $2t$. Η τ.μ. M είναι το πλήθος των σωματιδίων που φτάνουν στο διάστημα $[40, 80]$ sec. Λόγω της ιδιότητας έλλειψης μνήμης της σ.δ. *Poisson*, η M ακολουθεί κατανομή *Poisson* με παράμετρο $2(80 - 40) = 80$:

$$p_M(m) = \frac{80^m}{m!} e^{-80}; m = 0, 1, 2, \dots$$

- (v) * Η σ.δ. που σχετίζεται με την $R(t)$ αφορά τις αφίξεις σωματιδίων τύπου A που καταγράφονται από τον ανιχνευτή. Αυτή είναι μια σ.δ. *Poisson* με ένταση p_{AA} σωματίδια/sec. Συνεπώς η $R(t)$ ακολουθεί κατανομή *Poisson* με παράμετρο $2p_{AA}t$: Επομένως :

$$E[R(iT)] = 2p_{AA}iT = \text{var}(R(iT))$$

Επίσης :

$$E[R((k+1)T) - R(kT)] = 2p_{AA}(k+1)T - 2p_{AA}kT = 2p_{AA}T$$

Και

$$\text{var}(R((k+1)T) - R(kT)) = 2p_{AA}T$$

- (vi) * Η τ.μ. $R(iT)$ μπορεί να εκφραστεί ως το άθροισμα i το πλήθος ανεξάρτητων και όμοια κατανομημένων τ.μ. *Poisson* με παράμετρο $2p_{AA}T$:

$$R(iT) = \sum_{k=1}^i [R(kT) - R((k-1)T)]$$

Με άμεση εφαρμογή του ΚΟΘ έχουμε ότι :

$$F_{R(iT)}(r) \approx \Phi\left(\frac{r - i2p_{AA}T}{\sqrt{2p_{AA}Ti}}\right).$$

Άσκηση 3.

- (i) Ας θεωρήσουμε ότι S είναι οι μέρες που δεν έχει χιονίσει καθόλου. Η S είναι μια διωνιμική τ.μ. με παραμέτρους $n = 45$ και $p = 0.95$ έτσι ώστε $E[X] = 50 * 0.95$ και $\sigma_S = \sqrt{50 * 0.95 * 0.05} = 1.54$. Χρησιμοποιώντας κανονική προσέγγιση στην διωνιμική κατανομή, βρίσκουμε

$$P(S \geq 45) = P\left(\frac{S - 47.5}{1.54} \geq \frac{45 - 47.5}{1.54}\right) \approx 1 - \Phi(-1.62) = \Phi(1.62) = 0.9474.$$

Χρησιμοποιώντας την *deMoivre - Laplace* προσέγγιση

$$P(S \geq 45) = P(S \geq 44.5) = P\left(\frac{S - 47.5}{1.54} \geq \frac{44.5 - 47.5}{1.54}\right) \approx 1 - \Phi(-1.95) = \Phi(1.95) = 0.9744.$$

- (ii) Η τ.μ. S είναι διωνιμική με παράμετρο $p = 0.95$. Όμως, η τ.μ. $50 - S$ είναι επίσης διωνιμική με παράμετρο $p = 0.05$. Δεδομένου ότι η προσέγγιση *Poisson* είναι ακριβώς στο όριο μικρών p και μεγάλων n , θα δώσει πιό ακριβή αποτελέσματα εάν εφαρμοστεί στην $50 - S$ με παράμετρο $\lambda = 50 * 0.05 = 2.5$

$$\begin{aligned} P(S \geq 45) &= P(50 - S \leq 5) \\ &= \sum_{k=0}^5 P(n - S = k) \\ &= \sum_{k=0}^5 e^{-\frac{k}{2.5}} \frac{k!}{k!} \\ &= 0.958 \end{aligned}$$

Η ακριβής πιθανότητα είναι

$$\sum_{k=0}^5 \binom{50}{k} 0.05^k 0.95^{50-k} = 0.962$$

Επομένως, η προσέγγιση *Poisson* είναι πιο κοντά. Αυτό συμφωνεί με την διαίσθηση ότι η κανονική προσέγγιση στην διωνιμική κατανομή δουλεύει καλά όταν το p είναι κοντά το 0.5 ή το n είναι πολύ μεγάλο.

Άσκηση 4.*

Δεδομένου ότι οι μεταβλητές X_n είναι ανεξάρτητες και πανομοιότυπα κατανεμημένες με συνάρτηση πυκνότητας πιθανότητας f η πιθανότητα δυο να είναι ίδιες είναι μηδενική. Για $n \geq 2$, κάθε μια από τις μεταβλητές X_{n-1}, X_n, X_{n+1} μπορεί να είναι η μέγιστη με την ίδια πιθανότητα. Άρα η πιθανότητα το μέγιστο να εμφανιστεί στην μεταβλητή με δείκτη n είναι ίση με $1/3$. Όμως οι μεταβλητές X_n και X_{n+1} είναι εξαρτημένες. Εάν το μέγιστο εμφανιστεί στην μεταβλητή με δείκτη n , η μεταβλητή με δείκτη $n+1$ δεν μπορεί να είναι μέγιστη. Ορίζοντας μια νέα τ.μ Y_n , όπου $Y_n = 1$ όταν η τ.μ X_n εμφανίζει μέγιστο στον δείκτη n και $Y_n = 0$ σε κάθε άλλη περίπτωση. Γνωρίζουμε ότι $EY_n = P(Y_n = 1) = 1/3$ για όλα τα $n \geq 2$, αλλά οι διαδοχικές μεταβλητές Y_n είναι εξαρτημένες. Όμως $\{Y_2, Y_5, Y_8, \dots\}$, $\{Y_3, Y_6, Y_9, \dots\}$ και $\{Y_4, Y_7, Y_{10}, \dots\}$ είναι τρεις όμοια κατανεμημένες σειρές τ.μ. Για κάθε μια, μπορούμε να εφαρμόσουμε το ισχυρό θεώρημα των μεγάλων αριθμών

$$n^{-1} \sum_{k=1}^n Y_{2+3k} \rightarrow 1/3,$$

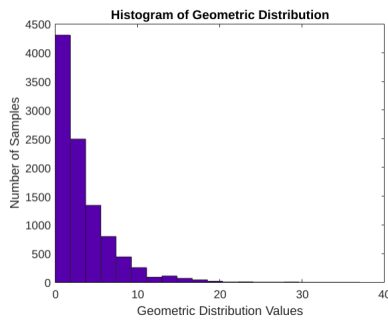
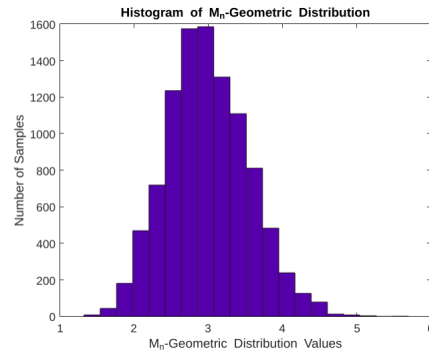
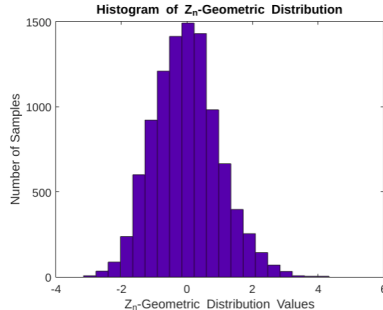
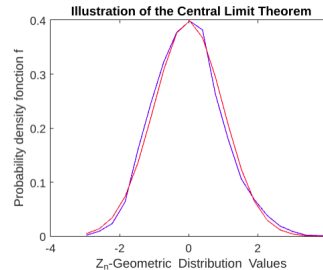
όσο το n τείνει στο άπειρο με πιθανότητα 1.

$$n^{-1} \sum_{k=1}^n Y_k \rightarrow 1/3,$$

όσο το n τείνει στο άπειρο με πιθανότητα 1.

Άσκηση 5.

Στο Σχήμα 1 (α) απεικονίζεται το ιστόγραμμα της ομοιόμορφης τ.μ. X , για $N = 10000$ δείγματα και 20 bins. Στο Σχήμα 1 (β) φαίνεται το ιστόγραμμα του δειγματικού μέσου όρου M_n , για $n = 20$ ανεξάρτητες ομοιόμορφες τ.μ. $X_1, X_2, \dots, X_n$ που μοιάζει πλέον, όπως περιμέναμε από τη θεωρία, να ακολουθεί κανονική κατανομή γύρω από τη μέση τιμή των X_i . Το Σχήμα 1 (γ) αναπαριστά το ιστόγραμμα του μετασχηματισμού Z_n : Πρώτα έχουμε αφαιρέσει από την S_n , ώστε να προκύψει η τ.μ. $S_n - n\mu$ με μηδενική μέση τιμή και κατόπιν διαιρέσαμε με $\sigma\sqrt{n}$ ώστε η διασπορά να γίνει ίση με τη μονάδα. Τώρα, στο παραγόμενο ιστόγραμμα βλέπουμε στην πράξη το Κεντρικό Οριακό Θεώρημα, που μας λέει ότι η ασυμπτωτική μορφή της Z_n τείνει οριακά στην Κανονική κατανομή, $Z_n \sim N(0, 1)$. Τέλος, στο Σχήμα 1 (δ) βλέπουμε τις εκτιμήσεις της pdf του δειγματικού μέσου όρου M_n με χρήση ιστογράμματος και κανονικής κατανομής, όπου γίνεται αντιληπτή η καλή προσέγγιση μεταξύ των δύο κατανομών.

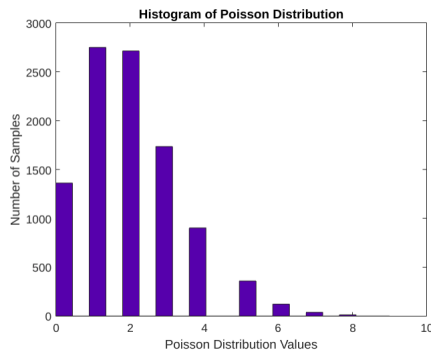
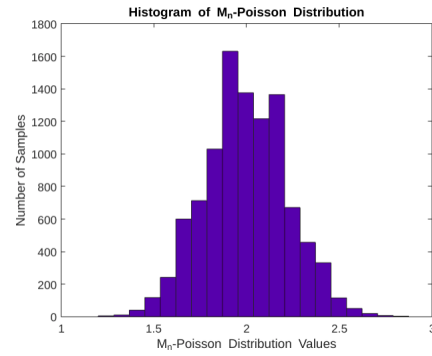
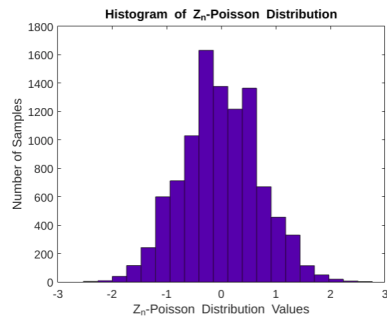
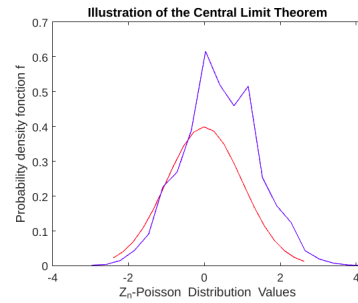
(α) Η Γεωμετρική τ.μ. X .(β) Ιστόγραμμα του δειγματικού μέσου όρου M_{Xn} .(γ) Ιστόγραμμα της τ.μ. Z_{Xn} .(δ) Εκτίμηση της pdf του δειγματικού μέσου όρου Z_{Xn} .

Σχήμα 1: Τα ζητούμενα γραφήματα για την ομοιόμορφη Τ.Μ. X για $N = 1000$, $n = 20$, $n_{bins} = 20$.

Στο Σχήμα 2 αναπαριστούνται τα ανάλογα συμπεράσματα εξάγονται και για τις τ.μ. $Y_1, Y_2, \dots, Y_n$.

Έχει ενδιαφέρον να παρατηρήσουμε αλλαγές στα παραπάνω διαγράμματα για διαφορετικές τιμές των παραμέτρων. Καθώς αυξάνουμε τον αριθμό των μεταβλητών αφήνοντας τις υπόλοιπες παραμέτρους ως έχουν, παρατηρούμε ότι η διασπορά της Z_n μειώνεται και έχει επανέλθει στην τιμή 1 προσεγγίζοντας την κανονική κατανομή. Επίσης παρατηρούμε ότι η διασπορά των κατανομών των εκτιμήσεων της pdf του δειγματικού μέσου όρου Z_n με χρήση ιστογράμματος και κανονικής κατανομής, μειώνεται όσο ο αριθμός των τ.μ. αυξάνει, διατηρώντας μια καλή προσέγγιση μεταξύ τους.

Όσο αυξάνουμε τον αριθμό των δειγμάτων αφήνοντας τις υπόλοιπες παραμέτρους ως έχουν, η μόνη διαφορά που παρατηρούμε στα παραγόμενα ιστογράμματα είναι προφανώς μόνο η αύξηση στη συχνότητα εμφάνισης στην εκάστοτε περιοχή τιμών (άξοντας y). Δεν παρατηρείται καμία αλλαγή ως προς τη μέση τιμή ή τη διασπορά. Επίσης η αντίστοιχη εκτίμηση της pdf συγκλίνει περισσότερο στην κανονική.

(α) Η εκθετική τ.μ. Y .(β) Ιστόγραμμα του δειγματικού μέσου όρου M_{Yn} .(γ) Ιστόγραμμα της τ.μ. Z_{Yn} .(δ) Εκτίμηση της pdf του δειγματικού μέσου όρου Z_{Yn} .Σχήμα 2: Τα ζητούμενα γραφήματα για την εκθετική Τ.Μ. Y για $N = 1000$, $n = 20$, $n_{bins} = 20$.