ΠΙΘΑΝΟΤΗΤΕΣ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ

2η+3η σειρά ασκήσεων

Ονοματεπώνυμο:

Αριθμός μητρώου:

Ημερομηνία παράδοσης: Μέχρι και την Δευτέρα 20 Ιανουαρίου 2020

Σημειώστε τις ασκήσεις για τις οποίες έχετε παραδώσει λύση:

2.1	2.2	2.3	2.4	2.5
2.6	2.7	2.8	2.9	2.10
2.11	2.12	2.13	2.14	2.15
2.16	2.17	2.18	2.19	2.20

2.1) Μια συνεχής τυχαία μεταβλητή X με τιμές στο $\mathbb{R}$ έχει συνάρτηση πυκνότητας πιθανότητας (σ.π.π)

$$f(x) = \begin{cases} Cx & x \in [0,2] \\ 0, & \text{αλλιώς} \end{cases}$$

i) Να βρεθεί η σταθερά C.

ii) Να βρεθεί η συνάρτηση κατανομής πιθνανότητας $F(x)$ της X.

iii) Να σχεδιασθούν οι γραφικές παραστάσεις των $f(x)$ και $F(x)$.

iv) Να υπολογισθεί η αναμενόμενη τιμή $E(X)$ της X.

v) Να υπολογισθεί η διακύμανση $\mathrm{Var}(X)$ της X.

vi) Να υπολογισθούν οι πιθανότητες $P(X \le 1)$, $P(X > \frac{3}{2})$, $P(1 \le X < \frac{3}{2})$,

2.2) Η κατανομή της πιθανότητας ένας πελάτης που μπαίνει σε ένα κατάστημα με παγωτά να αγοράσει 1, 2, 3, 4, ή 5 μπάλες παγωτού είναι η εξής:

$X = x$	1	2	3	4	5
$P(X = x)$	0.41	0.37	0.16	0.05	0.01

i) Να υπολογισθεί η συνάρτηση κατανομής $F(x)$ της τ.μ. X: αριθμός από μπάλες παγωτού που αγοράζει ένας πελάτης.

ii) Να υπολογισθεί η αναμενόμενη τιμή και η διακύμανση της τ.μ. X.

iii) Αν κάθε μπάλα παγωτού κοστίζει 2 ευρώ, ποιό είναι το αναμενόμενο κέρδος από κάθε πελάτη;

2.3) Η τυχαία μεταβλητή X ακολουθεί την διωνυμική κατανομή $\mathrm{Binom}(n, p)$ με αναμενόμενη τιμή 20 και διακύμανση 16.

i) Να βρεθεί η παράμετρος n.

ii) Να υπολογισθεί η $E(X^2)$.

2.4) Ένας μπασκετμπολίστας όταν ρίχνει για τρίποντο πετυχαίνει σε τριπλάσιο αριθμό προσπαθειών απ' ό,τι αποτυχνάνει.

i) Να βρεθεί η πιθανότητα να ρίξει 10 φορές και να πετύχει στις 8 από αυτές.

ii) Να βρεθεί ο αναμενόμενος αριθμός πόντων που θα συγκεντρώσει αν ρίξει για τρίποντο 10 φορές.

2.5) Ο Γιώργος και ο Δημήτρης είναι συγκάτοικοι και κάθε φορά που έχουν να πλύνουν πιάτα ρίχνουν ένα νόμισμα. Όποιος φέρει πρώτος κορώνα κερδίζει και ο άλλος πλένει τα πιάτα. Ο Γιώργος παρατήρησε ότι όταν ρίχνει πρώτος κερδίζει συχνότερα, ενώ όταν ξεκινάει δεύτερος καταλήγει συχνότερα στο νεροχύτη. Υπολογίστε τις πιθανότητες και στις δυο περιπτώσεις. Είναι ο ισχυρισμός του Γιώργου σωστός;

2.6) Έστω ότι ο αριθμός των γκολ που πετυχαίνει η Εθνική ομάδα ποδοσφαίρου στα εκτός έδρας παιχνίδια ακολουθεί την κατανομή Poisson. Επίσης, είναι γνωστό ότι στα εκτός έδρας παιχνίδια η Εθνική έχει την ίδια πιθανότητα να πετύχει ένα ή δύο γκολ. Να βρεθεί η πιθανότητα στο επόμενο εκτός έδρας παιχνίδι η Εθνική να πετύχει τέσσερα γκολ.

2.7) Μια από τις πιο διάσημες εμφανίσεις της κατανομής Poisson δόθηκε από τον Ρώσο μαθηματικό Ladislaus Bortkiewicz (1868 – 1931) ως μοντέλο του αριθμού των θανάτων από κλωτσιά αλόγου των αξιωματικών του ιππικού. Ο Bortkiewicz συνέλεξε στατιστικά στοιχεία για τους θανάτους αυτούς από 10 μονάδες ιππικού του πρωσικού στρατού σε μια περίοδο 20 ετών (από το 1875 μέχρι το 1894).

Στον επόμενο πίνακα δίδονται τα στατιστικά που συνέλεξε ο Bortkiewicz για τις 10 μονάδες του ιππικού (με λατινικούς αριθμούς) αναλυτικά ανά έτος (εμφανίζονται τα δύο τελευταία ψηφία του).

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
II	-	-	-	2	-	2	-	-	1	1	-	-	2	1	1	-	-	2	-	-
III	-	-	-	1	1	1	2	-	2	-	-	-	1	-	1	2	1	-	-	-
IV	-	1	-	1	1	1	1	-	-	-	-	1	-	-	-	-	1	1	-	-
V	-	-	-	-	2	1	-	-	1	-	-	1	-	1	1	1	1	1	1	-
VII	1	-	1	-	-	-	1	-	1	1	-	-	2	-	-	2	1	-	2	-
VIII	1	-	-	-	1	-	-	1	-	-	-	-	1	-	-	-	1	1	-	1
IX	-	-	-	-	-	2	1	1	1	-	2	1	1	-	1	2	-	1	-	-
X	-	-	1	1	-	1	-	2	-	2	-	-	-	-	2	1	3	-	1	1
XIV	1	1	2	1	1	3	-	4	-	1	-	3	2	1	-	2	1	1	-	-
XV	-	1	-	-	-	-	-	1	-	1	1	-	-	-	2	2	-	-	-	-

Τα παραπάνω δεδομένα, που αφορούν $20 \cdot 10 = 200$ υπηρεσιακά χρόνια, συνοψίζονται στον επόμενο πίνακα.

αριθμός θανάτων ανά μονάδα ανά υπηρεσιακό έτος	0	1	2	3	4	5 και πάνω
αριθμός υπηρεσιακών ετών	109	65	22	3	1	0

Ο Bortkiewicz έδειξε ότι ο αριθμός των θανάτων ανά μονάδα ανά υπηρεσιακό έτος ακολουθεί την κατανομή Poisson. Μπορείτε να επαληθεύσετε τον ισχυρισμό του Bortkiewicz;

(Υπόδειξη: Να υπολογισθεί ο μέσος αριθμός θανάτων ανά μονάδα ανά υπηρεσιακό έτος με βάση τα στοιχεία του συνοπτικού πίνακα.)

2.8) Ο χρόνος εξυπηρέτησης ενός πελάτη στα ταμεία μιας τράπεζας ακολουθεί την εκθετική κατανομή με μέση τιμή 3.5 λεπτά. Να βρεθεί η πιθανότητα ο χρόνος που θα περιμένετε για να εξυπηρετηθεί ο προηγούμενος από εσάς πελάτης να είναι:

i) μεγαλύτερος από πέντε λεπτά.

ii) ανάμεσα σε δύο και τέσσερα λεπτά.

2.9) Μια εταιρεία τηλεφώνων εκτιμά ότι η μέση διάρκεια ζωής της μπαταρίας τους κάτω από κανονικές συνθήκες ακολουθεί την εκθετική κατανομή με μέση τιμή 4 χρόνια. Αν η εταιρεία έχει την δυνατότητα να αντικαταστήσει δωρεάν (μέσα στην εγγύηση) την μπαταρία μέχρι το 40% των συσκευών που θα πωληθούν στην αγορά, ποιος πρέπει να είναι ο μέγιστος χρόνος εγγύησης που θα δοθεί για την μπαταρία;

2.10) Το 45% των καπνιστών μιας πόλης προτιμούν την μάρκα τσιγάρων A. Αν πάρουμε ένα τυχαίο δείγμα 100 καπνιστών, ποια είναι η πιθανότητα να πλειοψηφούν στο δείγμα αυτοί που προτιμούν την A;

2.11) Μια αλυσίδα αποτελέται από 4 κρίκους. Το όριο αντοχής κάθε κρίκου ακολουθεί την κανονική κατανομή με μέσο $\mu = 60$ κιλά και διακύμανση $\sigma^2 = 25$. Αν η αλυσίδα χρησιμοποιηθεί για να σηκώσει ένα φορτίο βάρους 62 κιλών, ποια είναι η πιθανότητα να σπάσει;

2.12) Για να αποφοιτήσει ένας φοιτητής πρέπει να περάσει 48 μαθήματα. Έστω ότι η βαθμολογία του σε κάθε μάθημα που περνά είναι τυχαία μεταβλητή X που λαμβάνει τις τιμές 5, 6, 7, 8, 9 και 10, κάθε μια με πιθανότητα 1/6.

i) Να υπολογισθούν τα $E(X)$ και $\mathrm{Var}(X)$.

ii) Να βρεθεί προσεγγιστικά με χρήση του κεντρικού οριακού θεωρήματος η πιθανότητα ο βαθμός πτυχίου να είναι μεγαλύτερος από 8.

2.13) Ένας blogger θέλει να εμφανίζει ότι το site του έχει μεγάλη επισκεψιμότητα και χρησιμοποιεί ένα μετρητή επισκεψιμότητας, ο οποίος αυξάνει από τις πραγματικές επισκέψεις στο site αλλά και αυτόματα χρησιμοποιώντας μια γεννήτρια ψευδοτυχαίων αριθμών η οποία παράγει κάθε λεπτό, με ίση πιθανότητα, ένα φυσικό αριθμό από το 1 εως το 10, ο οποίος προστίθεται στον μετρητή επισκεψιμότητας.

i) Να βρεθεί μέσος αριθμός των "επισκεψεων" στο site του blogger μέσα σε 1 μέρα.

ii) Αν σε 1 μέρα το site είχε 8000 επισκέψεις (πραγματικές και εικονικές). Να βρεθεί ο ελάχιστος αριθμός των πραγματικών επισκέψεων με πιθανότητα τουλάχιστον 99%.

2.14) Έστω ότι οι ανεξάρτητες τ.μ. $X_1, X_2, \ldots, X_n$ έχουν μέση τιμή $E(X_k) = k$ και διακύμανση $\mathrm{Var}(X_k) = \dfrac{k}{4}$ για κάθε $k \in [n]$. Θεωρούμε την τ.μ. $S_n = X_1 + X_2 + \cdots + X_n$.

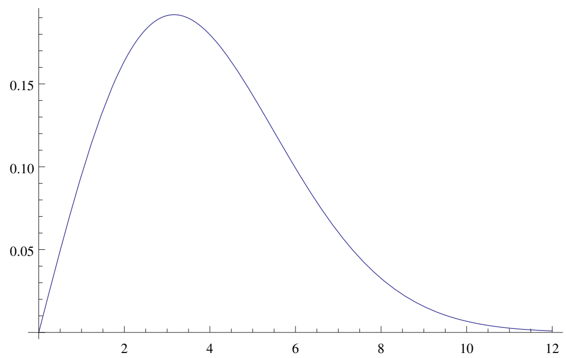
i) Να βρεθεί η μέση τιμή μ και η διακύμανση σ^2 της S_n.

ii) Με τη βοήθεια της ανισότητας του Chebyshev να βρεθεί ένα διάστημα της μορφής $[\mu - \theta, \mu + \theta]$ στο οποίο η S_n λαμβάνει τιμές με πιθανότητα τουλάχιστον 60%.

iii) Να εξηγηθεί γιατί δεν μπορούμε να θεωρήσουμε ότι η S_n προσεγγίζεται από την κανονική κατανομή.

2.15) Έστω X μια συνεχής τυχαία μεταβλητή που ακολουθεί την κατανομή Rayleigh με παράμετρο $k > 0$ έχει συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} kxe^{-\frac{kx^2}{2}}, & x \geq 0 \\ 0, & \text{αλλιώς} \end{cases}$$



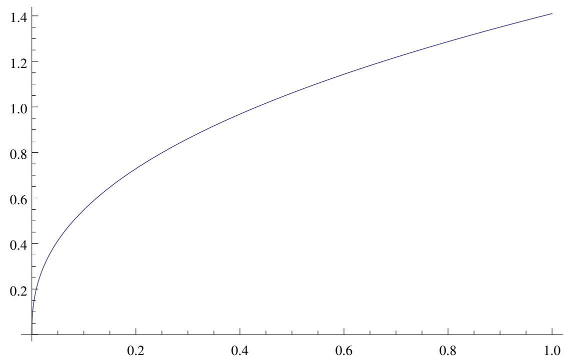
i) Να δειχθεί ότι η συνάρτηση κατανομής της X ισούται με $F(x) = 1 - e^{-\frac{kx^2}{2}}$, $x \geq 0$.

ii) Να βρεθεί με την μέθοδο του αντίστροφου μετασχηματισμού πως μπορούμε να παράγουμε τυχαία δείγματα της μεταβλητής X αν μπορούμε να παράγουμε τυχαία δείγματα της μεταβλητής U που ακολουθεί την ομοιόμορφη κατανομή στο $[0, 1]$.

iii) Να παραχθούν 30 τυχαία δείγματα της τ.μ. X όταν $k = 1/10$.

2.16) Έστω $x_1, x_2, \ldots, x_n$ ένα τυχαίο δείγμα που προέρχεται από πληθυσμό με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} (k+1)x^k, & \text{αν } 0 \leq x \leq 1 \\ 0, & \text{αλλιώς} \end{cases}, \text{ όπου } k > -1.$$



i) Να βρεθεί με τη μέθοδο της μέγιστης πιθανοφάνειας μια εκτιμήτρια για την παράμετρο k.

ii) Με βάση την παραπάνω εκτιμήτρια, ποιά είναι η εκτίμηση του k αν το δείγμα αποτελούνταν (σε αύξουσα σειρά) από τις επόμενες 10 τιμές 0.108, 0.206, 0.401, 0.417, 0.603, 0.657, 0.675, 0.726, 0.778, 0.895; (Μπορείτε να δείτε την πραγματική τιμή του k αν παρατηρήσετε την γραφική παράσταση της σ.π.π που συνοδεύει την άσκηση.)

2.17) Σε μια επιδημία γρίπης ερωτήθηκαν τυχαία 380 άτομα εκ των οποίων 140 απάντησαν ότι αρρώστησαν από την γρίπη.

i) Να βρεθεί ένα 95% διάστημα εμπιστοσύνης για το ποσοστό των ατόμων που έχει προσβάλλει η γρίπη.

ii) Να βρεθεί ο ελάχιστος αριθμός ατόμων n που πρέπει να ερωτηθεί ώστε να βρεθεί ένα 95% διάστημα εμπιστοσύνης για το ποσοστό των ατόμων που έχει προσβάλλει η γρίπη με σφάλμα το πολύ ±3%.

2.18) Να βρεθούν 95% και 99% διαστήματα εμπιστοσύνης για την μέση τιμή μ ενός πληθυσμού στις παρακάτω περιπτώσεις:

i) Ο πληθυσμός ακολουθεί την κανονική κατανομή με διακύμανση $\sigma^2 = 10$ και ένα τυχαίο δείγμα μεγέθους $n = 100$ έδωσε δειγματική μέση τιμή $\overline{\mu} = 187$.

ii) Ο πληθυσμός ακολουθεί την κανονική κατανομή και ένα τυχαίο δείγμα μεγέθους $n = 28$ έδωσε δειγματική μέση τιμή $\overline{\mu} = 187$ και δειγματική διακύμανση $s^2 = 9.3$.

iii) Ο πληθυσμός ακολουθεί την κανονική κατανομή και ένα τυχαίο δείγμα μεγέθους $n = 121$ έδωσε δειγματική μέση τιμή $\overline{\mu} = 187$ και δειγματική διακύμανση $s^2 = 9.6$.

iv) Ένα τυχαίο δείγμα μεγέθους $n = 121$ έδωσε δειγματική μέση τιμή $\overline{\mu} = 187$ και δειγματική διακύμανση $s^2 = 9.6$.

2.19) Προκειμένου να μελετηθεί η επίδραση της κληρονομικότητας στην εμφάνιση υπέρτασης μετρήθηκε η αρτηριακή πίεση σε δύο ανεξάρτητες ομάδες ατόμων. Η πρώτη ομάδα που αποτελείται από 8 άτομα των οποίων οι γονείς ήταν υπερτασικοί εμφάνισε μέση τιμή υπέρτασης $\overline{x_1} = 104.5$. Η δεύτερη ομάδα που αποτελείται από 10 άτομα των οποίων οι γονείς δεν ήταν υπερτασικοί εμφάνισε μέση τιμή υπέρτασης $\overline{x_2} = 97.3$. Να βρεθεί ένα 95% διάστημα εμπιστοσύνης για την διαφορά $\mu_1 - \mu_2$. Να εξετασθεί η υπόθεση $H_0 : \mu_1 = \mu_2$ έναντι της υπόθεσης $H_1 : \mu_1 \neq \mu_2$ σε επίπεδο σημαντικότητας $a = 0.05$ όταν υποθέτουμε ότι η υπέρταση και στις δύο ομάδες ατόμων ακολουθεί την κανονική κατανομή με διακυμάνσεις $\sigma_1^2 = 70$ και $\sigma_2^2 = 22$ αντίστοιχα.

2.20) Προκειμένου να εξετασθεί η αποτελεσματικότητα ενός νέου εμβολίου επιλέχθηκαν τυχαία 150 ασθενείς και στους 80 από αυτούς δόθηκε το νέο εμβόλιο, ενώ στους υπόλοιπους 70 δόθηκε ένα εικονικό εμβόλιο (placebo). Μετά από μια εβδομάδα βρέθηκε ότι βελτιώθηκε η υγεία 48 ασθενών που πήραν το εικονικό εμβόλιο και 56 ασθενών που πήραν το νέο εμβόλιο.

i) Μπορούμε σε επίπεδο σημαντικότητας $a = 5\%$ να αποδείξουμε ότι η νέα θεραπεία με το νέο εμβόλιο είναι προτιμότερη;

ii) Να κατασκευασθεί ένα 95% διάστημα εμπιστοσύνης για την διαφορά των ποσοστών p_1, p_2 όπου p_1 είναι η πιθανότητα να θεραπευτεί ένας ασθενής χρησιμοποιώντας εικονικό εμβόλιο και p_2 η αντίστοιχη πιθανότητα για το νέο εμβόλιο.