Μάθημα: Στατιστική (Κωδ. 105)

Διδάσκων: Γιώργος Κ. Παπαδόπουλος

3. Τυχαίες μεταβλητές-Κατανομές

Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Τυχαία μεταβλητή X, στο δειγματικό χώρο Ω	Μια πραγματική συνάρτηση που αντιστοιχίζει τα στοιχεία του δειγματικού χώρου Ω στο σύνολο των πραγματικών αριθμών τέτοια ώστε για κάθε διάστημα πραγματικών αριθμών I, το σύνολο $\{\omega \in \Omega : X(\omega) \in I\}$ να είναι ενδεχόμενο του Ω. Το σύνολο τιμών μιας τυχαίας μεταβλητής X συμβολίζεται με R_X.
Συνάρτηση κατανομής F, τυχαίας μεταβλητής X	Η πραγματική συνάρτηση με τύπο $F(x) = P(X \le x) = P(\{\omega \in \Omega : X(\omega) \le x\})$, $-\infty < x < +\infty$
Ιδιότητες της συνάρτησης κατανομής F	α) $0 \le F(x) \le 1$, για κάθε $x \in R$. β) Είναι αύξουσα συνάρτηση στο *R*. γ) $\lim_{x \to +\infty} F(x) = 1$ και $\lim_{x \to -\infty} F(x) = 0$· δ) Είναι δεξιά συνεχής
Διακριτή τυχαία μεταβλητή	Το σύνολο τιμών της είναι πεπερασμένο ή αριθμησίμως άπειρο.
Συνάρτηση πιθανότητας f διακριτής τ.μ. X με σύνολο τιμών R_X	$f : R \to [0,1]$ *με* $f(x) = \begin{cases} 0, & \text{για κάθε } x \notin R_X \\ P(X = x), & \text{για κάθε } x \in R_X \end{cases}$
Ιδιότητες της συνάρτησης πιθανότητας f μιας διακριτής τ.μ. X με $R_X = \{x_1, x_2, \ldots, x_\nu, \ldots\}$	*α)* $f(x) = 0$, *για κάθε* $x \notin R_X$ *β)* $f(x_i) \ge 0$, *για κάθε* $i = 1, 2, \ldots$ *γ)* $f(x_1) + f(x_2) + \ldots + f(x_\nu) + \ldots = 1$
Πιθανότητα υποσυνόλου A του $R_X = \{x_1, x_2, \ldots, x_\nu, \ldots\}$	$P(X \in A) = \sum_{i : x_i \in A} f(x_i)$
Μέση τιμή μ διακριτής τ.μ. X με $R_X = \{x_1, x_2, \ldots, x_\nu, \ldots\}$	$\mu = E(X) = \sum_{x \in R_x} x f(x)$
Μέση τιμή συνάρτησης διακριτής τ.μ.	$E[g(X)] = \sum_{x \in R_X} g(x) f(x)$
Συνεχής τ.μ. και συνάρτηση πυκνότητας f συνεχούς τ.μ.	Μια τ.μ. *X* λέγεται συνεχής αν υπάρχει μια μη αρνητική πραγματική συνάρτηση *f* τέτοια ώστε $P(X \in A) = \int_A f(x)dx$ για κάθε (μετρήσιμο) υποσύνολο *A* των πραγματικών. Η συνάρτηση *f* ονομάζεται συνάρτηση πυκνότητας της *X*.
Ιδιότητες της συνάρτησης πυκνότητας f συνεχούς τ.μ.	α) $f(x) \ge 0$, για κάθε $x \in (-\infty, +\infty)$. β) $\int_{-\infty}^{+\infty} f(x)dx = 1$
Υπολογισμός πιθανοτήτων για συνεχείς τ.μ.	$P(\alpha < X < \beta) = P(\alpha \le X < \beta) = P(\alpha < X \le \beta) = P(\alpha \le X \le \beta) =$ $= \int_\alpha^\beta f(x)dx = F(\beta) - F(\alpha)$ $P(X = x) = 0$ για κάθε $x \in (-\infty, +\infty)$.
Μέση τιμή μ συνεχούς τ.μ., X	$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x)dx$
Μέση τιμή συνάρτησης συνεχούς τ.μ.	$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x)dx$

Ιδιότητες της μέσης τιμής τ.μ. *(διακριτής ή συνεχούς)*	*α)* $E[\alpha_1 g_1(X) + \alpha_2 g_2(X) + \ldots + \alpha_k g_k(X)] =$ $= \alpha_1 E[g_1(X)] + \alpha_2 E[g_2(X)] + \ldots + \alpha_k E[g_k(X)]$ *β)* $E(\alpha X + \beta) = \alpha E(X) + \beta$ *γ)* $E(\alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k) =$ $= \alpha_1 E(X_1) + \alpha_2 E(X_2) + \ldots + \alpha_k E(X_k)$
Παράμετροι διασποράς τ.μ. *(διακριτής ή συνεχούς)*	*α) Διακύμανση,* σ^2, $Var(X)$ $\sigma^2 = Var(X) = E[(X-\mu)^2]$ *β) Τυπική απόκλιση,* σ, $\sqrt{Var(X)}$ $\sigma = \sqrt{Var(X)} = \sqrt{E[(X-\mu)^2]}$
Εναλλακτικός τύπος υπολογισμού της διακύμανσης τ.μ. *(διακριτής ή συνεχούς)*	$\sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$
Ιδιότητες της διακύμανσης τ.μ. *(διακριτής ή συνεχούς)*	α) $\sigma^2_{\alpha X+\beta} = Var(\alpha X + \beta) = \alpha^2 Var(X)$ και $\sigma_{\alpha X+\beta} = \lvert\alpha\rvert \sigma_X$ β) Αν $X_1,\ X_2, \ldots, X_k$ **ανεξάρτητες** τ.μ. $Var(\alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k) =$ $= \alpha_1^2 Var(X_1) + \alpha_2^2 Var(X_2) + \ldots + \alpha_k^2 Var(X_k)$
Τυχαίο δείγμα μεγέθους ν	ν ανεξάρτητες τυχαίες μεταβλητές $X_1, X_2, \ldots, X_\nu$ οι οποίες ακολουθούν την ίδια κατανομή
Δειγματικός μέσος	$\overline{X} = \frac{1}{\nu}\sum_{i=1}^{\nu} X_i$
Μέση τιμή και Διακύμανση δειγματικού μέσου	$E(\overline{X}) = \mu_{\overline{X}} = \mu$, $Var(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2 / \nu$
Στατιστικός πληθυσμός	Η κατανομή των (δυνατών) τιμών μιας τυχαίας μεταβλητής.

Προβλήματα και Ασκήσεις

1. Έστω *X* τυχαία μεταβλητή που εκφράζει τις ημερήσιες πωλήσεις συγκεκριμένου τύπου αρτοσκευασμάτων γνωστής αλυσίδας αρτοποιείων. Στον πίνακα που ακολουθεί φαίνεται ο ημερήσιος αριθμός πωλήσεων τις τελευταίες 1500 ημέρες (τα πέντε περίπου τελευταία χρόνια). Πιο συγκεκριμένα, στην πρώτη γραμμή φαίνεται ο ημερήσιος αριθμός των πωλήσεων (στρογγυλοποιημένος στην πλησιέστερη χιλιάδα) και στη δεύτερη ο αριθμός των ημερών που η εταιρεία πέτυχε τις αντίστοιχες πωλήσεις.

Αριθμός πωλήσεων x	5	6	7	8	9	
Αριθμός ημερών που επετεύχθησαν x πωλήσεις	300	450	300	300	150	1500

Αφού βρείτε (εκτιμήσετε) τη συνάρτηση πιθανότητας και τη συνάρτηση κατανομής της *X*, να υπολογίσετε την πιθανότητα η αλυσίδα αρτοποιείων να πουλήσει σήμερα α) λιγότερα από 5 χιλιάδες αρτοσκευάσματα β) περισσότερα από 9 χιλιάδες αρτοσκευάσματα γ) το πολύ 9 χιλιάδες αρτοσκευάσματα δ) τουλάχιστον 8 χιλιάδες αρτοσκευάσματα δεδομένου ότι ήδη έχει πουλήσει περισσότερα από 6 χιλιάδες. Επίσης να υπολογίσετε και να ερμηνεύσετε τη μέση τιμή και τη διακύμανση της *X*.

2. Μια διακριτή τυχαία μεταβλητή *X* έχει συνάρτηση πιθανότητας *f* που δίνεται από τον πίνακα αντιστοιχίας

x	0	1	2	3	4
$f(x)$	1/16	4/16	6/16	c	1/16

Να βρεθούν α) η τιμή της σταθεράς *c* β) η πιθανότητα $P(0 \le X < 4)$ γ) η δεσμευμένη πιθανότητα $P(X \ge 3 / 1 < X < 4)$ δ) η *E(X)* και ε) η *Var(X)*.

3. Ο αριθμός των επισκέψεων των γεωπόνων-ελεγκτών σε μια γεωργική μονάδα ανά έτος είναι διακριτή τυχαία μεταβλητή *X* με συνάρτηση πιθανότητας

x	0	1	2	3
$f(x)$	0.1	0.6	0.2	?

α) Βρείτε την πιθανότητα οι ελεγκτές σε ένα έτος να επισκεφθούν τη μονάδα 3 φορές ακριβώς. β) Βρείτε τον αναμενόμενο αριθμό επισκέψεων των ελεγκτών στην μονάδα σε ένα έτος. γ) Υπολογίστε το διάστημα $\mu \pm 2\sigma$ και βρείτε την πιθανότητα η τυχαία μεταβλητή *X* να πάρει τιμή στο διάστημα αυτό.

4. Ο αριθμός των μηχανημάτων που πουλάει μια έκθεση γεωργικών μηχανημάτων σε μια εβδομάδα είναι διακριτή τυχαία μεταβλητή *X* με συνάρτηση πιθανότητας που δίνεται από τον τύπο

$$f(x) = \begin{cases} cx, & x = 1,2,3,4,5 \\ c(10-x), & x = 6,7,8,9. \end{cases}$$

α) Να βρεθεί η τιμή της σταθεράς *c*. β) Ποια είναι η πιθανότητα να πουληθούν σε μια εβδομάδα i) ακριβώς 7 μηχανήματα ii) λιγότερα από 4 μηχανήματα iii) περισσότερα από 4 μηχανήματα iv) άρτιος αριθμός μηχανημάτων v) περισσότερα από 5 μηχανήματα γνωρίζοντας ότι έχουν πουληθεί τουλάχιστον 3. γ) Πόσα μηχανήματα αναμένεται να πουλάει η έκθεση ανά εβδομάδα.

5. Αν η τυπική απόκλιση μιας τυχαίας μεταβλητής X είναι 15.7, να βρεθεί η τυπική απόκλιση της τυχαίας μεταβλητής $Y = 5/9(X - 32)$. (Η X εκφράζει τιμές θερμοκρασίας σε βαθμούς Fahrenheit και η Y τις αντίστοιχες τιμές σε βαθμούς Celsius).

6. Ο χρόνος ζωής (σε εκατοντάδες ώρες) ενός εξαρτήματος εργαστηριακού οργάνου περιγράφεται από μια συνεχή τυχαία μεταβλητή X με συνάρτηση πυκνότητας
$$f(x) = \begin{cases} \dfrac{c}{x^3} & 1 \le x \le 3 \\ 0 & \text{αλλού.} \end{cases}$$
Να βρεθεί α) η μέση τιμή $E(X)$ και η διακύμανση $Var(X)$ της X β) η πιθανότητα να λειτουργήσει το εξάρτημα περισσότερο από 200 ώρες γ) η πιθανότητα να λειτουργήσει το εξάρτημα από 200 έως 250 ώρες και δ) η πιθανότητα να σταματήσει το εξάρτημα να λειτουργεί εντός των επόμενων 50 ωρών αν είναι γνωστό ότι έχει λειτουργήσει 200 ώρες.

7. Η μηνιαία κατανάλωση πετρελαίου (σε kgal) μιας μονάδας θερμοκηπίων κατά τους χειμερινούς μήνες είναι τυχαία μεταβλητή X με συνάρτηση πυκνότητας
$$f(x) = \begin{cases} c(1-x)^4 & 0 \le x \le 1 \\ 0 & \text{αλλού.} \end{cases}$$
Να βρεθεί α) η μέση τιμή $E(X)$ και η διακύμανση $Var(X)$ της X β) η πιθανότητα να εξαντληθεί το πετρέλαιο κατά την διάρκεια κάποιου μήνα όταν για τον μήνα αυτό έγινε προμήθεια 0.9kgal και γ) η ελάχιστη ποσότητα πετρελαίου που πρέπει να έχει στη διάθεση της η μονάδα στην αρχή κάθε μήνα ώστε η πιθανότητα να εξαντληθεί το πετρέλαιο μέσα στο μήνα να είναι 0.05.

8. Ένα φυτοφάρμακο διατίθεται σε φιαλίδια και περιέχει, μεταξύ άλλων, μια συγκεκριμένη χημική ουσία Α. Η περιεκτικότητα (ποσοστιαία) κάθε φιαλιδίου σε ουσία Α είναι συνεχής τυχαία μεταβλητή X με συνάρτηση πυκνότητας
$$f(x) = \begin{cases} cx^3(1-x) & 0 < x < 1 \\ 0 & \text{αλλού.} \end{cases}$$
α) Να βρεθεί η τιμή της σταθεράς c. β) Να βρεθούν η μέση τιμή και η διακύμανση της X. γ) Επιλέγουμε τυχαία ένα φιαλίδιο. Ποια είναι η πιθανότητα η περιεκτικότητα του φιαλιδίου να είναι το πολύ 0.7.

9. Ο χρόνος ζωής ενός ευαίσθητου προϊόντος εκτός ψυγείου είναι συνεχής τυχαία μεταβλητή X (σε ώρες) με συνάρτηση πυκνότητας
$$f(x) = \begin{cases} cx & 0 < x \le 2 \\ c(4-x) & 2 < x \le 4. \end{cases}$$
α) Να βρεθεί η τιμή της σταθεράς c. β) Να βρεθούν η μέση τιμή και η διακύμανση της X. γ) Ποια είναι η πιθανότητα ο χρόνος ζωής του προϊόντος, να είναι μεταξύ μίας και δύο ωρών.