Εργαστήριο Μαθηματικών & Στατιστικής

$1^{η}$ Πρόοδος στο Μάθημα Στατιστική

4/12/09

Α΄ ΣΕΙΡΑ ΘΕΜΑΤΩΝ

$1^{ο}$ Θέμα (20 Μονάδες)

Από τα ασθενή ζώα μιας κτηνοτροφικής μονάδας, ποσοστό 10% έχει προσβληθεί από την ασθένεια Α, 30% από την ασθένεια Β και 60% από την ασθένεια Γ. Από τα ζώα που έχουν προσβληθεί από την ασθένεια Α, θεραπεύεται το 40%, από την ασθένεια Β το 80% και από την ασθένεια Γ το 90%. **α)** Επιλέγουμε τυχαία ένα ζώο (από τα ασθενή). Ποια είναι η πιθανότητα, το ζώο αυτό να θεραπευθεί. **β)** Επιλέγουμε τυχαία ένα ζώο και διαπιστώνουμε ότι θεραπεύθηκε. Ποια είναι η πιθανότητα, το ζώο αυτό να είχε προσβληθεί από την ασθένεια Α. **γ)** Τα ενδεχόμενα *«το ζώο έχει προσβληθεί από την ασθένεια Α»* και *«το ζώο θεραπεύθηκε»* είναι ανεξάρτητα ή εξαρτημένα; (εξηγείστε γιατί).

$2^{ο}$ Θέμα (45 Μονάδες)

Ι. 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:

$$f(x)=\begin{cases} c\cdot x^{3}(1-x), & 0<x<1 \\ 0, & \text{αλλού.} \end{cases}$$

α) Να βρεθεί η τιμή της σταθεράς c. **β)** Να βρεθούν η μέση τιμή $\mu = E(X)$ και η διακύμανση $\sigma^2 = V(X)$ της X. **γ)** Επιλέγουμε τυχαία ένα φιαλίδιο. Ποια είναι η πιθανότητα η περιεκτικότητα του φιαλιδίου σε ουσία Α να είναι το πολύ 0.7. **δ)** Παίρνουμε ένα τυχαίο δείγμα 60 φιαλιδίων. Ποια είναι η πιθανότητα η μέση περιεκτικότητα σε ουσία Α των 60 φιαλιδίων να είναι το πολύ 0.7. *(Θεωρείστε ότι οι περιεκτικότητες των φιαλιδίων σε ουσία Α είναι μεταξύ τους ανεξάρτητες).*

ΙΙ. Να αναφέρετε τρεις πηγές δειγματοληπτικών σφαλμάτων. Είναι δυνατόν τα δειγματοληπτικά σφάλματα να αποφευχθούν; Εξηγείστε γιατί.

$3^{ο}$ Θέμα (35 Μονάδες)

Σε μια πολύ μεγάλη μονάδα θερμοκηπίων έχει εγκατασταθεί σύστημα αυτόματου ποτίσματος. Έχει παρατηρηθεί ότι ο αριθμός X_t των ελαττωματικών μπεκ σε μήκος σωλήνα t, περιγράφεται ικανοποιητικά από μια στοχαστική διαδικασία *Poisson* με μέσο αριθμό ελαττωματικών μπεκ, 3 μπεκ ανά κομμάτι σωλήνα (το κάθε κομμάτι σωλήνα είναι μεγάλου μήκους). **α)** Να βρεθούν οι πιθανότητες: **i.** σε ένα κομμάτι σωλήνα να υπάρχουν τουλάχιστον δύο ελαττωματικά μπεκ **ii.** σε μισό κομμάτι σωλήνα να υπάρχει το πολύ ένα ελαττωματικό μπεκ **iii**. σε μήκος σωλήνα όσο δύο κομμάτια να υπάρχουν τουλάχιστον τρία ελαττωματικά μπεκ **iv.** από τρία τυχαία επιλεγμένα κομμάτια σωλήνα, να υπάρχει το πολύ ένα κομμάτι με τουλάχιστον δύο ελαττωματικά μπεκ. **β)** Σε μήκος σωλήνα όσο τρία κομμάτια, ποιος είναι ο πιθανότερος αριθμός ελαττωματικών μπεκ.

Διάρκεια εξέτασης 2 ώρες και 30 λεπτά.
Ευχόμαστε επιτυχία!

Δίνονται οι παρακάτω τιμές της συνάρτησης κατανομής της τυποποιημένης κανονικής κατανομής:
Φ(0)=0.5, Φ(0.5)=0.6915, Φ(0.75)=0.7734, Φ(0.8)=0.7881, Φ(0.83)=0.7967, Φ(1)=0.8413, Φ(1.24)=0.8925, Φ(1.25)=0.8944, Φ(1.5)=0.9332, Φ(1.96)=0.9750, Φ(2.24)=0.9875, Φ(2.33)=0.9900.
Επίσης, δίνονται οι τιμές:
$e^{-7}\approx 0.00091$, $e^{-9}\approx 0.00012$, $e^{-6}\approx 0.0025$, $e^{-3}\approx 0.0498$, $e^{-2.5}\approx 0.0821$, $e^{-1.5}\approx 0.2231$

<u>Ενδεικτικές απαντήσεις (Α΄ Σειρά)</u>

1ο Θέμα

Έστω Α το ενδεχόμενο: *το ζώο έχει προσβληθεί από την ασθένεια Α*, Β το ενδεχόμενο: *το ζώο έχει προσβληθεί από την ασθένεια Β*, Γ το ενδεχόμενο: *το ζώο έχει προσβληθεί από την ασθένεια Γ* και Θ το ενδεχόμενο: *το ζώο θεραπεύθηκε*.

Δίνονται οι πιθανότητες: $P(A) = 0.1$, $P(B) = 0.3$, $P(\Gamma) = 0.6$, $P(\Theta / A) = 0.4$, $P(\Theta / B) = 0.8$ και $P(\Theta / \Gamma) = 0.9$.

α) Ζητείται η πιθανότητα: $P(\Theta)$. Από το θεώρημα της ολικής πιθανότητας έχουμε: $P(\Theta) = P(\Theta / A)P(A) + P(\Theta / B)P(B) + P(\Theta / \Gamma)P(\Gamma) = 0.82$.

β) Ζητείται η δεσμευμένη πιθανότητα $P(A / \Theta)$. Από τον τύπο του Bayes έχουμε:

$$P(A / \Theta) = \frac{P(\Theta / A)P(A)}{P(\Theta)} = 0.0488.$$

γ) $P(A / \Theta) = 0.0488 \neq P(A) = 0.1$ άρα τα ενδεχόμενα Α και Θ είναι εξαρτημένα.

2ο Θέμα

Ι. α) Πρέπει

$$\int_0^1 c \cdot x^3 (1 - x)\, dx = 1 \Rightarrow c \int_0^1 (x^3 - x^4)\, dx = 1 \Rightarrow c \left[\frac{x^4}{4} - \frac{x^5}{5} \right]_0^1 = 1 \Rightarrow$$

$$c \left[\frac{1}{4} - \frac{1}{5} \right] = 1 \Rightarrow c = 20$$

β) $E(X) = \int_0^1 x \cdot 20 \cdot x^3 (1 - x)\, dx = 20 \int_0^1 (x^4 - x^5)\, dx = 20 \left[\frac{x^5}{5} - \frac{x^6}{6} \right]_0^1 = \frac{2}{3} = 0.67$

$$\sigma^2 = E(X^2) - [E(X)]^2 = \int_0^1 x^2 \cdot 20 \cdot x^3 (1 - x) dx - \left(\frac{2}{3} \right)^2 = 20 \left[\frac{x^6}{6} - \frac{x^7}{7} \right]_0^1 - \left(\frac{2}{3} \right)^2 = \frac{2}{63} = 0.032$$

γ) Ζητάμε την πιθανότητα $P(X \le 0.7) = 20 \int_0^{0.7} x^3 (1 - x)\, dx = \ldots = 0.5282$.

δ) Έστω X_i η ποσοστιαία περιεκτικότητα σε ουσία Α του φιαλιδίου *i* ($i = 1, 2, \ldots, 60$). Προφανώς ζητάμε την πιθανότητα $P(\overline{X} \le 0.70)$ όπου $\overline{X} = \frac{X_1 + X_2 + \ldots + X_{60}}{60}$. Οι τυχαίες μεταβλητές X_i είναι ανεξάρτητες και ισόνομες με μέση τιμή $\mu_i = E(X_i) = 0.67$ και διασπορά $\sigma_i^2 = 0.032$, και επειδή το *n* είναι αρκετά μεγάλο, από το Κ.Ο.Θ. προκύπτει ότι η $\overline{X}$ προσεγγίζεται ικανοποιητικά από την κανονική κατανομή $N(0.67,\ \frac{0.032}{60})$ ή $N(0.67,\ 0.02^2)$. Επομένως η ζητούμενη πιθανότητα είναι: $P(\overline{X} \le 0.70) = P(Z \le \frac{0.70 - 0.67}{0.02}) = P(Z \le 1.5) = \Phi(1.5) = 0.9332$.

ΙΙ. Πηγές δειγματοληπτικών σφαλμάτων είναι η μεταβλητότητα του δείγματος, το μέγεθος του δείγματος και η επιλογή σχεδίου δειγματοληψίας. Τα δειγματοληπτικά σφάλματα είναι αναπόφευκτα γιατί είναι αναπόφευκτη η μεταβλητότητα στον πληθυσμό.

3^{o} Θέμα

α) Ισχύει $P(X_t = x) = e^{-3\cdot t}\dfrac{(3\cdot t)^x}{x!},\quad x = 0,1,2,...$

i. $P(X_1 \geq 2) = 1 - P(X_1 < 2) = 1 - P(X_1 = 0) - P(X_1 = 1) = 1 - e^{-3}\dfrac{3^0}{0!} - e^{-3}\dfrac{3^1}{1!} = 1 - 4e^{-3}$

ii. $P(X_{1/2} \leq 1) = P(X_{1/2} = 0) + P(X_{1/2} = 1) = e^{-3/2}\dfrac{(3/2)^0}{0!} + e^{-3/2}\dfrac{(3/2)^1}{1!} = \dfrac{5}{2}e^{-3/2}$

iii. $P(X_2 \geq 3) = 1 - P(X_2 < 3) = 1 - P(X_2 = 0) - P(X_2 = 1) - P(X_2 = 2) =$

$= 1 - e^{-6}\dfrac{6^0}{0!} - e^{-6}\dfrac{6^1}{1!} - e^{-6}\dfrac{6^2}{2!} = 1 - 25e^{-6}$

iv. Έστω Υ ο αριθμός κομματιών σωλήνα από τα τρία που επελέγησαν τυχαία που στο καθένα υπάρχουν τουλάχιστον δύο ελαττωματικά μπεκ. Προφανώς $Y \sim B(3,\, 1 - 4e^{-3})$ και επομένως η ζητούμενη πιθανότητα είναι,

$P(Y \leq 1) = \binom{3}{0}\cdot(1 - 4e^{-3})^0 \cdot \left(4e^{-3}\right)^3 + \binom{3}{1}\cdot(1 - 4e^{-3})^1 \cdot \left(4e^{-3}\right)^2 = ...$

β) Προφανώς $X_3 \sim P(9)$ άρα οι πιθανότερες τιμές είναι οι $x_0 = 9$ και $x_0' = 8$.

Εργαστήριο Μαθηματικών & Στατιστικής

$1^{η}$ Πρόοδος στο Μάθημα Στατιστική

4/12/09

Β΄ ΣΕΙΡΑ ΘΕΜΑΤΩΝ

$1^{ο}$ Θέμα (20 Μονάδες)

Το 5% των ζώων μιας κτηνοτροφικής μονάδας έχει προσβληθεί από μια ασθένεια. Από τα ζώα που έχουν προσβληθεί από την ασθένεια, ποσοστό 80% παρουσιάζει ένα συγκεκριμένο σύμπτωμα ενώ από τα ζώα που δεν έχουν προσβληθεί, ένα ποσοστό 10% παρουσιάζει επίσης το σύμπτωμα αυτό. **α)** Επιλέγουμε τυχαία ένα ζώο από την κτηνοτροφική μονάδα. Ποια είναι η πιθανότητα το ζώο να παρουσιάζει το συγκεκριμένο σύμπτωμα. **β)** Επιλέγουμε τυχαία ένα ζώο από την κτηνοτροφική μονάδα και διαπιστώνουμε ότι παρουσιάζει το συγκεκριμένο σύμπτωμα. Ποια είναι η πιθανότητα το ζώο αυτό να έχει προσβληθεί από την ασθένεια. **γ)** Τα ενδεχόμενα *«το ζώο έχει προσβληθεί από την ασθένεια»* και *«το ζώο παρουσιάζει το συγκεκριμένο σύμπτωμα»* είναι ανεξάρτητα ή εξαρτημένα; (εξηγείστε γιατί).

$2^{ο}$ Θέμα (45 Μονάδες)

Ο χρόνος ζωής ενός ευαίσθητου προϊόντος εκτός ψυγείου (δηλαδή ο χρόνος μη αλλοίωσής του) είναι συνεχής τυχαία μεταβλητή X (σε ώρες) με συνάρτηση πυκνότητας:

$$f(x)=\begin{cases} c\cdot x, & 0<x\le 2 \\ c\cdot(4-x), & 2<x\le 4. \end{cases}$$

α) Να βρεθεί η τιμή της σταθεράς c. **β)** Να βρεθούν η μέση τιμή $\mu=E(X)$ και η διακύμανση $\sigma^2=V(X)$ της X. **γ)** Ποια είναι η πιθανότητα ο χρόνος ζωής του προϊόντος, να είναι μεταξύ μίας και δύο ωρών. **δ)** Παίρνουμε ένα τυχαίο δείγμα 60 προϊόντων. Ποια είναι η πιθανότητα ο μέσος χρόνος ζωής των 60 προϊόντων να είναι μεταξύ μίας και δύο ωρών. *(Θεωρείστε ότι οι χρόνοι ζωής των προϊόντων είναι μεταξύ τους ανεξάρτητοι).*

$3^{ο}$ Θέμα (35 Μονάδες)

(Ι) Έχει παρατηρηθεί ότι το επίπεδο γλυκόζης στα άτομα ενός συγκεκριμένου πληθυσμού είναι τυχαία μεταβλητή η οποία ακολουθεί κανονική κατανομή με μέση τιμή $\mu=90\,mg/lt$ και τυπική απόκλιση $\sigma=10mg/lt$. **α)** Τι ποσοστό του πληθυσμού έχει επίπεδο γλυκόζης μεταξύ 85 και 100 *mg/lt*. **β)** Ποιο είναι εκείνο το επίπεδο γλυκόζης ξ, για το οποίο ισχύει ότι: μεγαλύτερο επίπεδο από ξ έχει μόνο το 1% του πληθυσμού. **γ)** Το επίπεδο γλυκόζης ενός ατόμου του πληθυσμού κρίνεται ως φυσιολογικό αν βρίσκεται σε εκείνο το συμμετρικό διάστημα γύρω από τη μέση τιμή που περιέχει το 95% των επιπέδων γλυκόζης του συγκεκριμένου πληθυσμού. Προσδιορίστε τα φυσιολογικά όρια του επιπέδου γλυκόζης των ατόμων του συγκεκριμένου πληθυσμού. **δ)** Επιλέγουμε τυχαία 5 άτομα από τον συγκεκριμένο πληθυσμό. Ποια είναι η πιθανότητα το πολύ 2 από αυτά να έχουν επίπεδο γλυκόζης εντός των φυσιολογικών ορίων (όπως αυτά ορίσθηκαν στο ερώτημα γ). **(ΙΙ)** Είναι λογικό η μέση τιμή της κατανομής *Poisson* να είναι ίση με τη διασπορά της; Εξηγείστε.

Διάρκεια εξέτασης 2 ώρες και 30 λεπτά.
Ευχόμαστε επιτυχία!

Δίνονται οι παρακάτω τιμές της συνάρτησης κατανομής της τυποποιημένης κανονικής κατανομής:
Φ(0)=0.5, Φ(0.5)=0.6915, Φ(0.75)=0.7734, Φ(0.8)=0.7881, Φ(0.83)=0.7967, Φ(1)=0.8413, Φ(1.24)=0.8925, Φ(1.25)=0.8944, Φ(1.5)=0.9332, Φ(1.96)=0.9750, Φ(2.24)=0.9875, Φ(2.33)=0.9900.
Επίσης, δίνονται οι τιμές:
$e^{-7}\approx 0.00091$, $e^{-9}\approx 0.00012$, $e^{-6}\approx 0.0025$, $e^{-3}\approx 0.0498$, $e^{-2.5}\approx 0.0821$, $e^{-1.5}\approx 0.2231$

<u>Ενδεικτικές απαντήσεις (Β΄ Σειρά)</u>

1ο Θέμα

Έστω Α το ενδεχόμενο: *το ζώο έχει προσβληθεί από την ασθένεια*, και Π το ενδεχόμενο: *το ζώο παρουσιάζει το συγκεκριμένο σύμπτωμα.*

Δίνονται οι πιθανότητες: $P(A)=0.05$, $P(\Pi/A)=0.8$ και $P(\Pi/A')=0.1$.

α) Ζητείται η πιθανότητα: $P(\Pi)$. Από το θεώρημα της ολικής πιθανότητας έχουμε: $P(\Pi)=P(\Pi/A)P(A)+P(\Pi/A')P(A')=0.135$. **β)** Ζητείται η δεσμευμένη πιθανότητα $P(A/\Pi)$. Από τον τύπο του Bayes έχουμε:

$$P(A/\Pi)=\frac{P(\Pi/A)P(A)}{P(\Pi)}=0.2963.$$

β) $P(A/\Pi)=0.2963\neq P(A)=0.05$ άρα τα ενδεχόμενα Α και Π είναι εξαρτημένα.

2ο Θέμα

α) Πρέπει $\int_0^2 c\cdot x\,dx+\int_2^4 c\cdot(4-x)\,dx=1\Rightarrow c\cdot\left[\frac{x^2}{2}\right]_0^2+4\cdot c[x]_2^4-c\cdot\left[\frac{x^2}{2}\right]_2^4=1\Rightarrow c=\frac{1}{4}$

β) $E(X)=\int_0^2 x\cdot\frac{1}{4}\cdot x\,dx+\int_2^4 x\cdot\frac{1}{4}\cdot(4-x)\,dx=...=2$

$$\sigma^2=E(X^2)-[E(X)]^2=\int_0^2 x^2\cdot\frac{1}{4}\cdot x\,dx+\int_2^4 x^2\,\frac{1}{4}\cdot(4-x)\,dx-2^2=...=0.67$$

γ) Ζητάμε την πιθανότητα $P(1<X<2)=\frac{1}{4}\int_1^2 x\,dx=...=\frac{3}{8}$.

δ) Έστω X_i ο χρόνος ζωής του i προϊόντος ($i=1,2,...,60$). Προφανώς ζητάμε την πιθανότητα $P(1<\overline{X}<2)$ όπου $\overline{X}=\frac{X_1+X_2+...+X_{60}}{60}$. Οι τυχαίες μεταβλητές X_i είναι ανεξάρτητες και ισόνομες με μέση τιμή $\mu_i=2$ και διασπορά $\sigma_i^2=0.67$, και επειδή το *n* είναι αρκετά μεγάλο, από το Κ.Ο.Θ. προκύπτει ότι η $\overline{X}$ προσεγγίζεται ικανοποιητικά από την κανονική κατανομή $N(2,\ \frac{0.67}{60})$ ή $N(2,\ 0.11^2)$. Επομένως η ζητούμενη πιθανότητα είναι:

$$P(1<\overline{X}<2)=P(\frac{1-2}{0.11}<Z<\frac{2-2}{0.11})=P(-9.1<Z<0)=\Phi(0)-\Phi(-9.1)=\Phi(0)-1+\Phi(9.1)=0.5$$

3ο Θέμα

Έστω X το επίπεδο γλυκόζης στα άτομα του συγκεκριμένου πληθυσμού. Δίνεται ότι $X\sim N(90,\ 10^2)$.

α) Ζητείται η πιθανότητα

$$P(85<X<100)=P(\frac{85-90}{10}<Z<\frac{100-90}{10})=P(-0.5<Z<1)=\Phi(1)-\Phi(-0.5)=$$
$$=\Phi(1)-[1-\Phi(0.5)]=\Phi(1)-1+\Phi(0.5)=0.5328$$

β) Πρέπει: $P(X>\xi)=0.01$. Άρα

$$P(\frac{X-90}{10}>\frac{\xi-90}{10})=0.01\Leftrightarrow P(Z>\frac{\xi-90}{10})=0.01\Leftrightarrow\Phi(\frac{\xi-90}{10})=0.99\Leftrightarrow\frac{\xi-90}{10}=2.33$$

άρα $\xi=113.3$

γ) Έστω c τέτοιο ώστε:

$$P(90-c<X<90+c)=0.95 \Leftrightarrow P(\frac{90-c-90}{10}<Z<\frac{90+c-90}{10})=0.95$$

$$\Leftrightarrow P(\frac{-c}{10}<z<\frac{c}{10})=0.95 \Leftrightarrow 2\Phi(\frac{c}{10})-1=0.95 \Leftrightarrow \Phi(\frac{c}{10})=0.975 \Leftrightarrow \frac{c}{10}=1.96$$

άρα $c=19.6$ και επομένως τα φυσιολογικά όρια είναι 70.4 *mg/lt* το κατώτερο και 109.6 *mg/lt* το ανώτερο.

δ) Έστω Y ο αριθμός των ατόμων, από τα πέντε που επελέγησαν τυχαία που έχουν επίπεδο γλυκόζης εντός των φυσιολογικών ορίων. Προφανώς $Y \sim B(5,\ 0.95)$ και επομένως η ζητούμενη πιθανότητα είναι,

$$P(Y\le 2)=\binom{5}{0}\cdot(0.95)^0\cdot(0.05)^5+\binom{5}{1}\cdot(0.95)^1\cdot(0.05)^4+\binom{5}{2}\cdot(0.95)^2\cdot(0.05)^3=\ldots$$

II. Είναι λογικό αφού η Poisson, ως οριακή κατανομή της διωνυμικής με p πολύ μικρό (άρα $1-p\approx 1$), αναμένεται να έχει μέση τιμή np και διασπορά $np(1-p)\approx np$.