Εισαγωγή στην Οικονομετρία : 4η Διάλεξη, 19/3/2024

Επανάληψη στις Πιθανότητες:

Παράδειγμα:

Έστω ενδεχόμενα A και B ενός δειγματικού χώρου S. Να γραφούν οι τύποι για τις παρακάτω περιπτώσεις χρησιμοποιώντας τις πιθανότητες $P(A)$, $P(B)$ και $P(A \cap B)$.

i) Να συμβεί είτε το A είτε το B είτε κ' τα δύο
ii) Να — " — " — " — αλλά όχι κ' τα δύο.
iii) Να συμβεί τουλάχιστον ένα από τα A κ' B.
iv) Να συμβεί το πολύ ένα από τα A κ' B.

Λύση

i) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

ii) $P\big((A \cap B^c) \cup (A^c \cap B)\big) \overset{\text{ξένα}}{\underset{\text{ενδεχόμενα}}{=}} P(A \cap B^c) + P(A^c \cap B) =$

$= P(A) + P(B) - 2P(A \cap B) \overset{\text{εάν}}{\underset{A,B \text{ ανεξάρτητα}}{=}} P(A)\big(1 - P(B)\big) + P(B)\big(1 - P(A)\big)$

$\quad \big(\underbrace{P(A \cup B)}_{\text{τουλάχιστον ένα}} - \underbrace{P(A \cap B)}_{\text{κ' τα δύο}}\big)$ αλλά όχι κ' τα δύο

iii) $P(\text{τουλάχιστον ένα}) = P(A \cup B)$

iv) $P(\text{το πολύ ένα}) = P(A^c \cap B) + P(A \cap B^c) + P(A^c \cap B^c)$

* Boole's inequality (ανισότητα): $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$

Παράδειγμα: Έστω μια βάση δεδομένων Ελλήνων φορολογούμενων και A_1, A_2, A_3 τα ενδεχόμενα μια τυχαία εγγραφή στην βάση (φορολογούμενος) να επιλέγεται σε φορολογική απάτη, τελωνειακή απάτη και τα 2 αντίστοιχα. Έστω $P(A_1)=0.0002$ $P(A_2)=0.0008$, $P(A_3) = 0.0005$. Ποια η πιθανότητα ένας τυχαία επιλεγμένος φορολογούμενος να ανήκει σε μια τουλάχιστον από τις 3 κατηγορίες.

Λύση: $P\left(\bigcup_{i=1}^{3} A_i\right) \leq P(A_1)+P(A_2)+P(A_3) = 0.0015$

* Bonferroni's Inequality: $P(A \cap B) \geq P(A) + P(B) - 1$

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum P(A_i) - (n-1)$$

Παράδειγμα: Μια εταιρεία παροχής υπηρεσιών διατηρεί μια βάση δεδομένων πελατών με καθυστερημένη πληρωμή. Έστω $P(A_1) = 0.95$ και $P(A_2)=0.97$ η πιθανότητα ένας τυχαία επιλ. πελάτης να καθυστερεί σε πληρωμές για υπηρεσίες κατηγορίας A_1 και A_2 αντίστοιχα. Ποια η πιθ. ένας τυχαία επιλ. να κάνει καθυστέρηση πληρωμών και για υπηρεσίες A_1 κ' A_2.

Λύση $P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1 = 0.92$

Τυχαίες Μεταβλητές:

↳ Τυχαία μεταβλητή (τ.μ.) είναι μια συνάρτηση από τον δειγματικό χώρο S στους πραγματικούς αριθμούς.

$$X: S \longrightarrow \mathbb{R}$$

π.χ. $X(\{K\}) = 0$
$X(\{\Gamma\}) = 1$.

↳ Διακριτή τυχαία μεταβλητή: Παίρνει τιμές σε αριθμήσιμο σύνολο $\{x_1, x_2, \ldots\}$

↳ Κατανομή διακριτής (τ.μ.): Οι πιθανότητες

$$f(x_i) = P(X = x_i) = p_i \quad \text{τ.ω.} \quad p_i \geq 0 \quad \text{και} \quad \sum_{i=1}^{\infty} p_i = 1$$

★ Για ένα ενδεχόμενο A: $P(X \in A) = \sum_{x_i \in A} P(X = x_i)$

Παράδειγμα: Ένα δίκαιο νόμισμα ρίχνεται 3 φορές (ανεξάρτητες ρίψεις).

α) Να οριστεί ~~...~~ η τ.μ. X ως τον αριθμό K στις 3 ρίψεις και να βρεθεί η κατανομή της
β) $P(X \geq 1) = ?$
γ) $P(X > 1) = ?$

Λύση

α) $S = \{KKK, K\Gamma\Gamma, KK\Gamma, \Gamma KK, K\Gamma K, \Gamma\Gamma K, \Gamma K\Gamma, \Gamma\Gamma\Gamma\}$

τότε $X(\{\Gamma\Gamma\Gamma\}) = 0$, $X(\{\Gamma K \Gamma\}) = 1$,
$X(\{\Gamma\Gamma K\}) = 1$, $X(\{K\Gamma K\}) = 2$ ...

Οπότε η τ.μ. παίρνει τις τιμές $X = \overset{x_1}{0}, \overset{x_2}{1}, \overset{x_3}{2}, \overset{x_4}{3}$

Για την κατανομή της X πρέπει να βρούμε

τις $P(X=0)$, $P(X=1)$, $P(X=2)$, $P(X=3)$

$$\frac{1}{8} \quad \frac{3}{8} \quad \frac{3}{8} \quad \frac{1}{8}$$

* $\sum_{i=1}^{4} P(X=x_i) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$

β) $P(X \geq 1) = \sum_{x_i \in A} P(X=x_i) = \sum_{i=1}^{3} P(X=i) =$

* $(A = \{1, 2, 3\})$ $= P(X=1) + P(X=2) + P(X=3) = 7/8$

γ) $P(X > 1) = P(X=2) + P(X=3) = 4/8$

* Παρατήρηση : $P(X>1) \neq P(X \geq 1)$

↳ Συνεχείς τ.μ. : Παίρνουν τιμές σε διάστημα πραγμ. αριθμών

$$P(X \in A) = \int_A f(x)\,dx$$

↳ συνάρτηση πυκνότητας πιθανότητας (probability density function) p.d.f.

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Παράδειγμα: Έστω $f(x) = \begin{cases} 1, & \text{εάν } x \in [0,1] \\ 0, & \text{αλλιώς} \end{cases}$

α) Να βρεθεί $P(X \geq 0,5)$

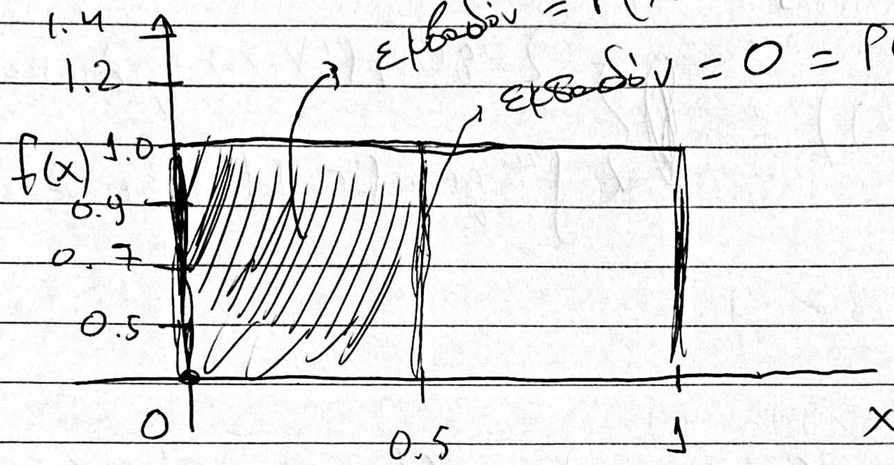
$\hookrightarrow \int_{0.5}^{1} 1\,dx = x\Big|_{0.5}^{1} = 1 - 0.5 = 0.5$

β) $P(X = 0,5)$

$\hookrightarrow = 0$

εμβαδόν $= P(X \in [0, 0,5])$

εμβαδόν $= 0 = P(X = 0.5)$



→ Αθροιστική Συνάρτηση Κατανομής (cumulative distribution function (cdf)).

$$F(x) = P(X \leq x)$$

Ιδιότητες: 1) $\lim_{x \to -\infty} F(x) = 0$ και $\lim_{x \to \infty} F(x) = 1$

2) $F(x)$ είναι μη-φθίνουσα (αύξουσα) συνάρτηση

3) $F(x)$ είναι δεξιά συνεχής

Παρατήρηση: Η pdf ορίζει μοναδικά την cdf και αντίστροφα. ⑥

↳ Μέση τιμή τ.μ. X (expected value)

$$\mu = E(X) = \begin{cases} \sum_{x_i} x_i P(X=x_i), & \text{εάν } X \text{ διακριτή} \\ \int_{-\infty}^{\infty} x f(x), & \text{εάν } X \text{ συνεχής} \end{cases}$$

Επίσης, $E(g(X)) = \begin{cases} \sum_{x_i} g(x_i) P(X=x_i), & \text{διακριτή} \\ \int_{-\infty}^{\infty} g(x) f(x) dx, & \text{συνεχής} \end{cases}$

Ιδιότητες:

1) $E(a g_1(x) + b g_2(x) + c) = aE(g_1(x)) + bE g_2(x) + c$

2) Εάν $g(x) \geq 0 \ \forall x \Rightarrow E(g(x)) \geq 0$

3) $g_1(x) \geq g_2(x) \ \forall x \Rightarrow E g_1(x) \geq E g_2(x)$

- Ροπή κατανομής: $\mu^n = E X^n$ είναι η n-οστή ροπή
- Κεντρική ροπή: $E(X-\mu)^n$
- Διασπορά = (n=2) κεντρική ροπή =

$= \sigma^2 = Var X = E(X - E(X))^2 = EX^2 - (EX)^2$

- Τυπική Απόκλιση: $\sigma = \sqrt{Var X}$.

Παρατήρηση: Η τυπική απόκλιση "μετράει" πόσο "απλωμένες" είναι οι τιμές της τ.μ. X από την μέση τιμή της

Παρατήρηση: Εάν μια τ.μ. X έχει διασπορά 0 τότε η τ.μ. παίρνει μόνο μια τιμή με πιθανότητα 1.

→ Κορυφή (mode) κατανομής: Η τιμή x για την οποία η $f(x)$ παίρνει την μέγιστη τιμή της.

Παρατηρήσεις:

i) Η κορυφή της κατανομής μπορεί να είναι παραπάνω από μια → πολυκόρυφη κατανομή (multimodal distribution).

ii) Εάν $x_1, x_2, \ldots, x_n$ ένα δείγμα με διαφορετικές τιμές η κορυφή είναι η τιμή που εμφανίζεται περισσότερο.

↳ Skewness με κατανομή μέσω της τρίτης κεντρικής ροπής

$$a_3 = E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right) = \frac{E(X-\mu)^3}{\sigma^3}$$