ΤΥΠΟΛΟΓΙΟ ΣΤΑΤΙΣΤΙΚΗΣ

1) ΜΕΤΡΑ ΘΕΣΗΣ

1) Μέση τιμή $\bar{x}$ $\quad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

ΙΔΙΟΤΗΤΕΣ

i) $x_1 = x_2 = \dots = x_N = k \Rightarrow \bar{x} = k$

ii) $y_i = x_i + \beta \Rightarrow \bar{y} = \bar{x} + \beta$

iii) $y_i = \alpha x_i \Rightarrow \bar{y} = \alpha \cdot \bar{x}$

iv) $y_i = \alpha x_i + \beta \Rightarrow \bar{y} = \alpha \bar{x} + \beta$

v) $\sum_{i=1}^{N} (x_i - \bar{x}) = 0$

vi) $\sum_{i=1}^{N} (x_i - k)^2 =$ ελάχιστο αν $k = \bar{x}$

2) Σταθμικός μέσος: $\bar{x} = \frac{\sum_{i=1}^{N} \rho_i x_i}{\sum_{i=1}^{N} \rho_i}$ $\quad \sum_{i=1}^{N} \rho_i = \rho$

$$\frac{1}{\rho}\sum_{i=1}^{N} \rho_i x_i = \sum_{i=1}^{N} \rho_i^* x_i$$

όπου $\rho_i^* = \frac{\rho_i}{\rho}$

3) Διάμεσος M $x_1, x_2, \dots x_N$

i) N = περιττός $\quad M = x_{\frac{N+1}{2}}$

ii) N = άρτιος $\quad M = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$

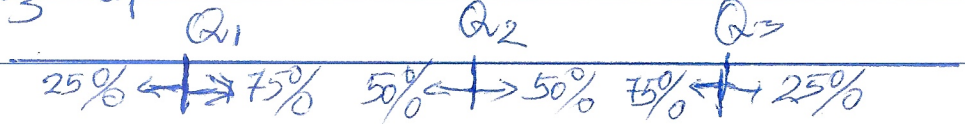
2) ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

1) **ΕΥΡΟΣ** (G) ή (R)

$$R = \max(x_i) - \min(x_i)$$

2) Τεταρτημορίων εύρος (IQR)

$$IQR = Q_3 - Q_1$$



3) Μέση απόκλιση

4) Διακύμανση (variance)

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

5) Τυπική απόκλιση (standard division)

$$s = \sqrt{s^2} \qquad s \sim \sigma$$

ΙΔΙΟΤΗΤΕΣ

i) $x_1 = x_2 = \dots = x_n$ τότε $\sigma^2 = 0$

β) $y_i = x_i + \beta$ $\Rightarrow \sigma_y^2 = \sigma_x^2$

γ) $y_i = \alpha x_i$ $\Rightarrow \sigma_y^2 = \alpha^2 \cdot \sigma_x^2$

και $\sigma_y = |\alpha| \cdot \sigma_x$

δ) $y_i = \alpha x_i + \beta$ τότε $\sigma_y = |\alpha| \cdot \sigma_x$

6) Συντελεστής μεταβλητότητας

$$CV = \frac{\sigma}{\bar{x}} = \frac{s}{\bar{x}} \cdot 100\%$$

3) Γεωμετρικός μέσος $G = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$

4) Μέσος ρυθμός μεταβολής (τ)

$$\tau = G - 1 = \sqrt[n]{\frac{x_1}{x_0}\frac{x_2}{x_1}\cdots\frac{x_n}{x_{n-1}}} - 1 = \sqrt[n]{\frac{x_n}{x_0}} - 1$$

άρα $x_n = x_0(1+\tau)^n$

5) $G \leq \bar{x}$

(γεωμ. μέσος ≤ αριθμ. μέσου)

6) Διάμεσος σε ομαδοποιημένες παρατηρήσεις ($M, Q_1, Q_2, P_{k\%}$)

$$M = x_{i1} + \delta\frac{\frac{N}{2} - F_{i-1}}{f_i} \quad (1)$$

$$Q_1 = x_{i1} + \delta\frac{\frac{N}{4} - F_{i-1}}{f_i} \quad (2)$$

$$Q_2 = x_{i1} + \delta\frac{\frac{2N}{4} - F_{i-1}}{f_i} \quad (3)$$

$$P_{k\%} = x_{i1} + \delta\frac{\frac{kN}{100} - F_{i-1}}{f_i} \quad (4)$$

x_{i1} = είναι η αρχή του i υποδιαστήματος που περιέχει την διάμεσο

δ = το εύρος των ίσων διαστημάτων

f_i = η συχνότητα του i διαστήματος

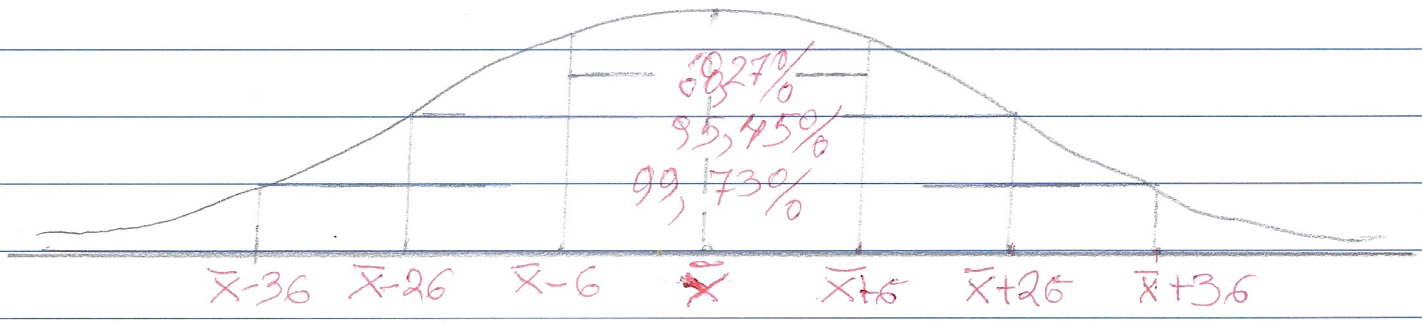
F_{i-1} = η αθρ. συχν. του προηγούμενου υποδιαστήματος

7) Τυποποιημένη τιμή της τιμής x_i

$$z_i = \frac{x_i - \bar{x}}{\sigma_x}$$

8) $z = \frac{x - \bar{x}}{\sigma_x} \Leftrightarrow \bar{z} = 0 \quad \sigma_z^2 = 1$

9) Κανονική κατανομή



10) Σε εμφρά ασύμμετρες κατανομές

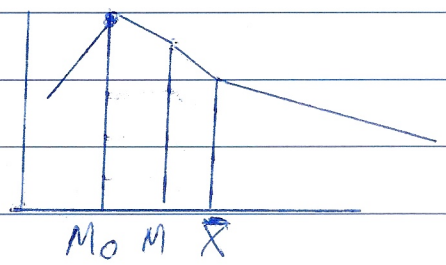
$$\bar{x} - M_o \approx 3(\bar{x} - M)$$

όπου $\bar{x}$: μέση τιμή M_o: επικρατούσα τιμή

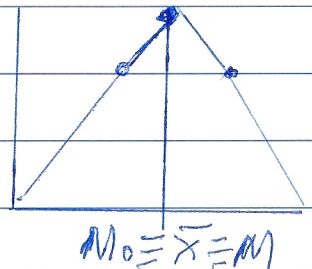
M: διάμεσος

11) Κριτήριο κατά Pearson

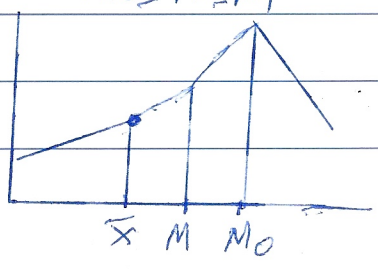
- $\bar{x} - M_o > 0$ θετική ασυμμετρία



- $\bar{x} - M_o = 0$ έχουμε συμμετρία



- $\bar{x} - M_o < 0$ αρνητική ασυμμετρία



12) Κριτήριο κατά Pearson

$$S_p(M_o) = \frac{\bar{X} - M_o}{\sigma_x} = \frac{3(\bar{X} - M)}{\sigma_x} \quad S_p(M) \text{ Συντελεστής ασυμμετρίας κατά Pearson}$$

13) Κριτήριο κατά Bowley

$$S_b = \frac{(Q_3 - M) - (M - Q_1)}{Q_3 - Q_1} \quad S_b = \text{Συντελεστής ασυμμετρίας κατά Bowley}$$

14) $S_b < S_p$ (Το S_b μικρότερο έναντι S_p)

ΚΥΡΤΩΣΗ

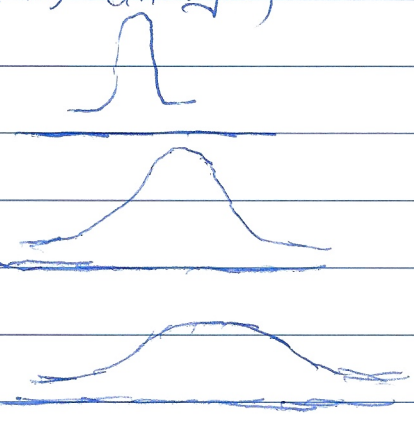
15) Συντελεστής κατά Pearson

$$b_2 = \frac{\frac{\sum f_i (x_i^k - \bar{x})^4}{\sum f_i}}{\sigma^4}$$

f_i = συχνότητα $\bar{X}$ = μέση τιμή

x_i^k = το κέντρο του i διαστήματος σ = τυπική απόκλιση

- $b_2 > 3$ η κατανομή είναι λεπτόκυρτη
- $b_2 = 3$ » » » μεσόκυρτη
- $b_2 < 3$ » » » πλατύκυρτη



16) Αν $-0,3 < S_p(M) < 0,3$ η κατανομή θεωρείται συμμετρική (κανονική) και ισχύουν

$S_p(M) = S_p(M_o) = S_b = 0$ (16i) $\bar{X} = M = M_o$ (16ii)

ΤΥΠΟΛΟΓΙΟ ΠΙΘΑΝΟΤΗΤΩΝ

1) $B=(A\cap B)\cup(A'\cap B)$

2) $P(B)=P(A\cap B)+P(A'\cap B)$

3) $P(A')=1-P(A)$

4) $P(\emptyset)=0$

5) Αν $A\subseteq B \Rightarrow P(B)=P(A)+P(A'\cap B)$

6) Αν $A\subseteq B \Rightarrow P(A)\leq P(B)$

7) $\left(\bigcup_{i=1}^{n}A_i\right)'=\bigcap_{i=1}^{n}A_i'$

8) $\left(\bigcap_{i=1}^{n}A_i\right)'=\bigcup_{i=1}^{n}A_i'$

} ΤΥΠΟΙ DE MORGAN

9) $P\left(\bigcup_{i=1}^{n}A_i\right)=1-P\left(\bigcap_{i=1}^{n}A_i'\right)$

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

11) $P(A\cup B)\leq P(A)+P(B)$

12) $P(A\cup B)=P(A\cap B')+P(A\cap B)+P(A'\cap B)$

13) $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - P\left(\bigcap_{i=1}^{n} A_i\right)$

$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$

14) $\boxed{P(B/A) = \frac{P(A \cap B)}{P(A)}}$ Δεσμευμένη πιθανότητα.

15) $\boxed{P(A_i/B) = \frac{P(B/A_i) \cdot P(A_i)}{\sum_{k=1}^{n} \left[P(A_k) \cdot P(B/A_k)\right]}}$

ΤΥΠΟΣ BAYES

16) A, B ανεξάρτητα αν

i) $P(B/A) = P(B)$ ή

ii) $P(A \cap B) = P(A) \cdot P(B)$

17) A, B, C ανεξάρτητα αν

$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

ΣΥΝΔΥΑΣΤΙΚΗ ΑΝΑΛΥΣΗ

1) Αν $|S|=m$ και $|T|=n$ τότε $|S\times T|=m\times n$

2) $|S^n|=|S|^n$

3) $P(S,n)=(S)_n=S\cdot(S-1)\cdot(S-2)\dots(S-n+1)$ διατάξεις
(επιλογή τυχαίου δείγματος η αντιπροσώπων από S χωρίς επανάθεση)

4) $P(S,S)=S!$ (Μεταθέσεις S αντικειμένων)

5) $P(S,n)=\frac{S!}{(S-n)!}$

6) Αν διαλέξουμε τ αντικείμενα από τα S χωρίς επανάθεση και χωρίς να μας ενδιαφέρει η διάταξή τους τότε έχουμε συνδυασμούς και θέλω C_τ^S ή $\binom{S}{\tau}$

$$C_\tau^S=\binom{S}{\tau}=\frac{S!}{\tau!\,(S-\tau)!}$$

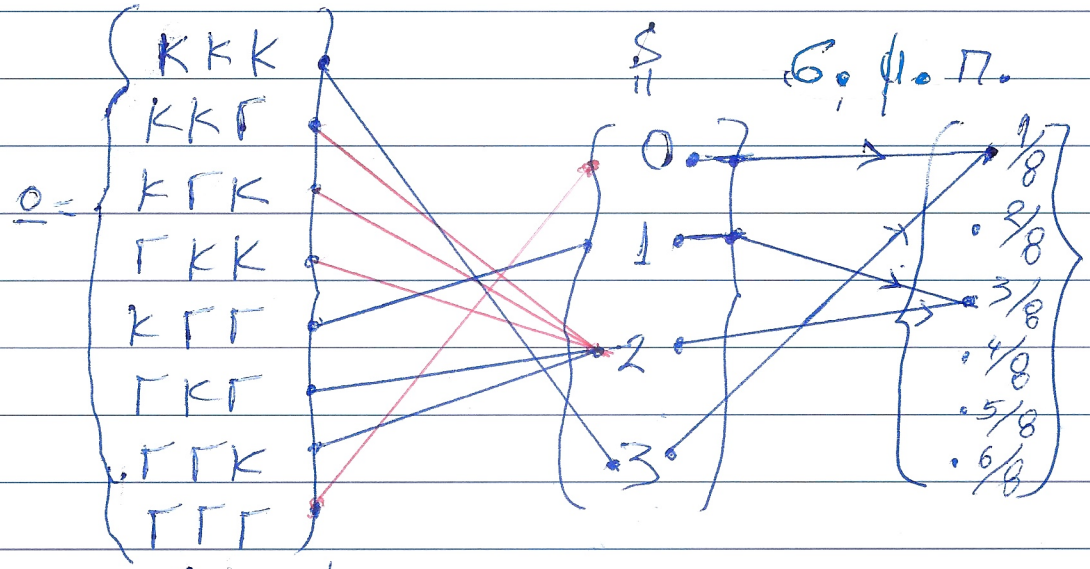
7) Αν $\epsilon_1,\epsilon_2,\dots,\epsilon_n$ γεγονότα τότε:

$$P(\epsilon_1\cap\epsilon_2\cap\dots\cap\epsilon_n)=P(\epsilon_1)\cdot P\left(\frac{\epsilon_2}{\epsilon_1}\right)\cdot P\left(\frac{\epsilon_3}{\epsilon_1\cap\epsilon_2}\right)\cdots P\left(\frac{\epsilon_n}{\epsilon_1\cap\epsilon_2\cap\dots\cap\epsilon_{n-1}}\right)$$

8) Αν $\epsilon_1,\epsilon_2,\dots,\epsilon_n$ ανεξάρτητα τότε

$$P(\epsilon_1\cap\epsilon_2\cap\dots\cap\epsilon_n)=P(\epsilon_1)\cdot P(\epsilon_2)\cdots P(\epsilon_n)$$

9) Επεξήγηση συνάρτησης μάζας πιθανότητας (δ.μ.π.)



10) Αθροιστική συνάρτηση κατανομής τυχαίας μεταβλητής X

$$F_X(\alpha) = P(X \le \alpha) = \sum_{\{x_i : x_i \le \alpha\}} p(x = x_i)$$

Ιδιότητες

i) $P(\alpha < X \le b) = F_X(b) - F_X(a)$

ii) $F(-\infty) = 0$

iii) $F(+\infty) = 1$

iv) $F_X(x)$ συνάρτηση συνεχής από δεξιά και ... μη φθίνουσα

$\lim_{x \to \alpha^+} F(x) = F(\alpha)$

11) Μέση τιμή ή μαθηματική ελπίδα σε διακριτή τυχαία μεταβλητή με σ.μ.π $p(x)$

$$E(X) = \frac{\sum_{\{x:\, p(x)>0\}} x \cdot p(x)}{\underbrace{\sum_{\{x:\, p(x)>0\}} p(x)}_{1}} = \sum_{\{x:\, p(x)>0\}} x \cdot p(x)$$

12) Νόμος αφηρημένου στατιστικού

$$E(g(x)) = \sum_{i=1}^{\infty} g(x_i) \cdot \underbrace{p_x(x_i)}_{P(x=x_i)}$$

13) Ιδιότητες της μέσης τιμής $E(x)$

i) $E(b) = b$ $\quad b =$ σταθερό

ii) $E(ax) = a \cdot E(x)$ $\quad a =$ σταθερό

iii) $E(ax+b) = a \cdot E(x) + b$

14) $\mathrm{Var}(x) = V(x) = E(x-\mu)^2 = E(x - E(x))^2$

με $E(x) = \mu$.

15) $\boxed{\mathrm{Var}(x) = E(x^2) - [E(x)]^2 \quad !!!}$

16) Ιδιότητες της διακύμανσης

i) $Var(X) > 0$

ii) $Var(\theta) = 0$ $\quad \theta =$ σταθερό

iii) $Var(X+\theta) = Var(X)$

iv) $Var(aX) = a^2 Var(X)$

v) $Var(aX+\theta) = a^2 \cdot Var(X)$!!!

ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ

$$SD = \sqrt{Var(X)}$$

$$SD = \sqrt{Var(aX+\theta)} = \sqrt{a^2 Var(X)} = |a| \sqrt{Var(X)}$$

ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ

I) Κατανομή Διωνυμική ($X \sim Bin(n, p)$)

$X_i = 1$ επιτυχία $P(X_i = 1) = p$

$X_i = 0$ αποτυχία $P(X_i) = 1-p$

Έστω το π.τ. n φορές έχω k επιτυχίες $n-k$ αποτυχίες

πιθανότητα $\underbrace{p \cdot p \cdot p \cdots p}_{k} \underbrace{(1-p)(1-p)\cdots(1-p)}_{n-k} = p^k \cdot (1-p)^{n-k}$

και αφού έχουμε $\binom{n}{k}$ τέτοιες ακολουθίες (η σειρά δεν μας ποιάζει)

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Άσκηση Αθλητής που υποχρεώνει συμβόλαιο με πιθανότητα $p=0,6$ θα επισκεφθεί σήμερο 5 υποψήφιους παίχτες

Ποιά η πιθανότητα να υπογράψει το πολύ 4 συμβόλαια

Λύση $b(x, n, p) = b(5, 5, 0,6)$

$$P_{ΖΗΤ} = 1 - P(X=5) = 1 - b(5, 5, 0,6) =$$

$$= 1 - \binom{5}{5} \cdot 0,6^5 \cdot 0,4^0 = 1 - 0,078 = 0,922.$$

II) Κατανομή Bernoulli $X \sim Bernoulli(p)$

Είναι ειδική περίπτωση της διωνυμικής για $n=1$

δηλ $Bin(1, p)$.

Σε διωνυμική έχω:

1) $E(X) = np$

2) $Var(X) = npq$

III) Ομοιόμορφη διακριτή κατανομή.

Η τυχαία μεταβλητή (τμ) παίρνει κ ισοπίθανες τιμές για $x=1,2,\ldots,\kappa$ λέμε ότι ακολουθεί ομοιόμορφη κατανομή με παράμετρο κ $X \sim U(\kappa)$

$$E(X)=\frac{\kappa+1}{2} \qquad V(X)=\frac{\kappa^2-1}{12}$$

Άσκηση Από 6 αριθμημένους μήλους 1, 2, ..., 6 κόβεται ένας. Αν η τ.μ. X δίνει τον αριθμό που θα μήλου θα βρεθεί να βρεθεί $E(X)=;$ $V(X)=;$

Λύση

$X \sim U(6)$ άρα $E(X)=\frac{6+1}{2}=3,5$ $V(X)=\frac{6^2-1}{12}=\frac{35}{12}$

IV) Γεωμετρική κατανομή (Geometric)

$P_X(x) = P(X=x)$ = αριθμός αποτυχιών $(1-p)$ πριν την πρώτη επιτυχία (p)

π.χ $P(X=3)$ = πιθανότητα πριν την 1η επιτυχία να έχουμε 3 αποτυχίες

άρα $P_X(x) = P(X=x) = \begin{cases} p\cdot(1-p)^x & \text{αν } x=0,1,2,\dots \\ 0 & \text{αν όχι.} \end{cases}$

$X \sim Geometric \quad E(x) = \frac{1}{p} \quad Var(x) = \frac{1-p}{p^2}$

Άσκηση Μπάλες μαύρες M και άσπρες N επιλέγονται τυχαία μέχρι να εμφανιστεί μία μαύρη δειγματοληψία με επανάθεση.

Ποια η πιθανότητα ότι χρειάζονται η δοκιμές

$p = \frac{M}{M+N}$ X = αριθμός των δοκιμών μέχρι να εμφανιστεί μαύρη μπάλα $X \sim G(p)$

$S_X = \{1,2,3,\dots\}$

$$P(X=n) = (1-p)^{n-1}\cdot p = \frac{M}{M+N}\cdot\left(1-\frac{M}{M+N}\right)^{n-1} = \frac{M}{M+N}\cdot\frac{N^{n-1}}{(M+N)^{n-1}}$$

$$= \frac{M\cdot N^{n-1}}{(M+N)^n}$$

V) Υπεργεωμετρική κατανομή
Επιλογή χωρίς επανάθεση → η πιθανότητα αλλάζει
γιατί αλλάζει ο δ.χ.

n = επιλογή φυτών
N = μπάλες
m = γενικές
$M-m$ = μαύρες
Ζητάω πιθανότητα
να έχω i γενικές μπάλες

$$P(X=i) = \frac{\binom{m}{i}\cdot\binom{N-M}{n-i}}{\binom{N}{n}}$$

$$X \sim Hypergeom(N, m, n)$$

Άσκηση Μοιράζουν 5 φύλλα από τα 52 της τράπουλας
Ποιά η πιθανότητα να έχω 2 Κλαγγάδες

$m = 4$ (αγγάδες)
$N = 52$ (όλα τα φύλλα)
$n = 5$ (επιλέγω 5)
$i = 2$ (οι αγγάδες που θέλω να έχω)

$$P(X=2) = \frac{\binom{4}{2}\cdot\binom{48}{3}}{\binom{52}{5}}$$

VI) Κατανομή Poisson

Μετράει γεγονότα ανά μονάδα χρόνου, εμβαδού κ.ο.κ

$$P_X(x) = P(X=x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} \quad x = 0, 1, 2, \ldots$$

λ = μέσο όρο των συμβάντων ανά μονάδα μέτρησης

π.χ. αξίζεις σε αεροδρόμιο $E(X) = V(X) = \lambda$.

α) Μ.Ο. 1 ξέβγ το δεντό σούζε $\lambda = 1$

ανα ώρα $\lambda' = 60$

β) Μ.Ο 1 αξίζη το 5ζεπτο $\Rightarrow \lambda_1 = 1$

αν α ώρα $\Rightarrow \lambda_1' = 12$

ανά 2,5 ζεπτα $\Rightarrow \lambda_1'' = \frac{1}{2}$

ανά 1 ζεπτό $\Rightarrow \lambda_1''' = \frac{1}{5}$

γ) Μ.Ο 3 ατυχήματα την ημέρα $\Rightarrow \lambda = 3$

Την εβδομάδα $\Rightarrow \lambda = 21$.

Άθλημα. Αριθμός δειγμάτων ακολουθεί Poisson με παράμετρο $\lambda = 2$ ανα εβδομάδα (2 δείγματα την εβδομάδα)

Να βρεθεί η πιθανότητα να γίνουν τουλάχιστον 3 δείγματα στις επόμενες 2 εβδομάδες

Λύση $X \sim$ Poisson $(\lambda = 2)$ $\quad y$ = αριθμ. δειγμάτων σε 2 εβδ. $(\lambda' = 2\lambda = 4)$

$$P(y \geq 3) = 1 - P(y < 3) = 1 - P(y=0) - P(y=1) - P(y=2) =$$

$$= 1 - \frac{4^0 \cdot e^{-4}}{0!} - \frac{4^1 e^{-4}}{1!} - \frac{4^2 \cdot e^{-4}}{2!} = 1 - e^{-4} - 4e^{-4} - \frac{16 \cdot e^{-4}}{2}$$

$$= 1 - e^{-4}(1 + 4 + 8) = 1 - 13e^{-4} = 0,7619$$

Από κοινού κατανομή

ΠΙΝΑΚΑΣ

x \ y	0	1	$P_X(x)$
0	1/8	0	1/8
1	2/8	1/8	3/8
2	1/8	2/8	3/8
3	0	1/8	1/8
$P_Y(y)$	4/8	4/8	1

→ περιθωριακή συνάρτηση $P_X(x)$

↓ περιθωριακή συνάρτηση $P_Y(y)$

$P(3,1) = P[X=3, Y=1] = \frac{1}{8}$

$P(x,y)$ ⇝ Περιθωριακή συνάρτηση μάζας πιθανότητα

$P(x,y) \geq 0 \quad \sum_{\{(x,y)\in S_{XY}} P(x,y) = 1$

α) Δεσμευμένη (Π) πιθανότητα δοθέντος $Y=y_0$ $\forall x \in S_X$ δίνεται

$P(X=x / Y=y_0) = \frac{P(x,y_0)}{P_Y(y_0)}$ π.χ $P(X=0/Y=0) = \frac{P(0,0)}{P_Y(0)} = \frac{\frac{1}{8}}{\frac{4}{8}} =$

$= \frac{1}{4}$

β) Η δεσμευμένη δ.μ.π δοθέντος $X=x_0$ $\forall y \in S_Y$ δίδεται ως

$P(Y=y/X=x_0) = \frac{P(x_0,y)}{P_X(x_0)}$ π.χ $P(Y=0/X=0) = \frac{P(0,0)}{P_X(0)} = \frac{\frac{1}{8}}{\frac{4}{8}} =$

$= \frac{1}{4}$

ΙΔΙΟΤΗΤΕΣ ΤΗΣ ΣΥΝΔΙΑΚΥΜΑΝΣΗΣ

i) $\sigma_{xy} = E[(x-\mu_x)\cdot(y-\mu_y)] = \sum_{\{(x,y)\in S_{xy}\}} (x-\mu_x)(y-\mu_y)\cdot p(x,y)$

$\sigma_{xy} = cov(x,y)$

i) $cov(x,y) = cov(y,x)$

ii) $cov(ax,by) = ab\cdot cov(x,y)$ $\quad a, b$ σταθερές

iii) $cov(x,x) = E(x-\mu_x)(x-\mu_x) = E(x-\mu_x)^2 = V(x)$ διακύμανση

iv) x, y ανεξάρτητες $\Rightarrow cov(x,y)=0$
αλλά $cov(x,y)=0 \Rightarrow x, y$ ανεξάρτητες

v) $cov(x,y) = E(xy) - E(x)\cdot E(y)$ $\quad (\le 0,5)$
ΣΥΧΝΕΣ ΤΙΣΗ

$\rho_{xy} = corr(x,y) = \dfrac{cov(x,y)}{\sqrt{var(x)}\cdot\sqrt{var(y)}} = \dfrac{\sigma_{xy}}{\sigma_x\,\sigma_y}$

Ιδιότητες συσχέτισης

i) $-1 \le \rho_{xy} \le 1$

ii) $\rho_{xy} = 1$ αν $y = ax+b$ $\quad a>0$
$\rho_{xy} = -1$ αν $y = ax+b$ $\quad a<0$

iii) $\rho_{xy} = 0 \Leftrightarrow \sigma_{xy} = 0$

iv) $\rho_{xy} = 0 \not\Rightarrow x, y$ ανεξάρτητα γενικά.