Θεωρία Πιθανοτήτων και Στατιστική

Διδάσκων: Επαμεινώνδας Διαμαντόπουλος
Επικοινωνία: epdiaman@ee.duth.gr

Γραφική Σύνοψη Προηγούμενου Μαθήματος

Τυχαία Μεταβλητή
X

Κατανομή Τιμών

Συναρτήσεις

Συνάρτηση Μάζας (ή Πυκνότητας) Πιθανότητας f_X
Συνάρτηση Κατανομής F_X
Ροπογεννήτρια
Πιθανογεννήτρια

Γεωμετρικά Χαρακτηριστικά Κατανομής

Αναμενόμενη Τιμή
Επικρατούσα Τιμή
Διάμεση Τιμή
Διακύμανση
Τυπική Απόκλιση
Ροπές – Κεντρικές ροπές
Συντελεστής Ασυμμετρίας
Συντελεστής Κυρτότητας

Μάθημα 4ο

Περιεχόμενα 4ου μαθήματος

- Ισόνομες τυχαίες μεταβλητές.
- Κοινή συνάρτηση πιθανότητας.
- Ανεξαρτησία τυχαίων μεταβλητών.
- Περιθώρια συνάρτηση πιθανότητας.
- Δεσμευμένη συνάρτηση πιθανότητας.
- Συνδιακύμανση – Συντελεστής συσχέτισης
- Βασικές κατανομές διακριτών τυχαίων μεταβλητών.
  - Κατανομή Bernoulli
  - Διωνυμική κατανομή
  - Υπεργεωμετρική κατανομή
  - Κατανομή Poisson
  - Γεωμετρική κατανομή (2 εκδοχές)
  - Αρνητική διωνυμική.

Γνωστικοί στόχοι 4ου μαθήματος

Στο τέλος αυτού του μαθήματος, ο φοιτητής πρέπει να είναι σε θέση :

- Για δύο τυχαίες μεταβλητές X, Y, να υπολογίζει
  - την από κοινού συνάρτηση πιθανότητας.
  - Την περιθώρια συνάρτηση πιθανότητας.
  - Τη δεσμευμένη συνάρτηση πιθανότητας.
  - Τη συνδιακύμανση και το συντελεστή συσχέτισης.
- Να αναγνωρίζει την κατανομή που ακολουθεί μία διακριτή τυχαία μεταβλητή, αξιολογώντας τη διαδικασία με την οποία αυτή λαμβάνει τιμές.
- Να μπορεί να επιλύει προβλήματα υπολογισμού πιθανοτήτων που αναφέρονται σε μία από τις βασικές διακριτές κατανομές που θα αναφερθούν.

Σχέσεις και συσχετίσεις μεταξύ δύο τυχαίων μεταβλητών

Ισόνομες Τυχαίες Μεταβλητές

Δύο τυχαίες μεταβλητές X, Y, λέγονται **ισόνομες** (identically distributed ή i.d.) όταν $F_X = F_Y$, δηλαδή

$$F_X(\kappa) = F_Y(\kappa) \text{ ή } P(X \leq \kappa) = P(Y \leq \kappa),\ \kappa \in R.$$

Σημείωση

Η αθροιστική συνάρτηση κατανομής είναι απλά μία συνάρτηση από το R στο [0, 1] και δεν συνδέεται με το πείραμα ή το δειγματοχώρο πάνω στον οποίο ορίζεται μία τυχαία μεταβλητή.

Δύο τυχαίες μεταβλητές μπορεί να είναι ισόνομες ακόμα και αν ορίζονται σε δύο τελείως διαφορετικούς δειγματοχώρους ή στον ίδιο δειγματοχώρο αλλά με διαφορετικό τρόπο.

Για παράδειγμα αν το πείραμα είναι η ρίψη 3 κερμάτων και ορίσουμε:

X = {το πλήθος των Κ} και Y = {το πλήθος των Γ}, τότε οι τ. μ. X, Y είναι ισόνομες (γιατί;) ενώ ως τυχαίες μεταβλητές αφορούν διαφορετικά αντικείμενα.

$P(X=0) = \frac{1}{8}$, $P(X=1) = \frac{3}{8}$, $P(X=2) = \frac{3}{8}$, $P(X=3) = \frac{1}{8}$.

$\Omega = \{ΚΚΚ, \ldots, ΓΓΓ\}$

Κοινή, περιθώρια και δεσμευμένη συνάρτηση πιθανότητας

$f_X(x) = P(X=x)$

Κοινή Συνάρτηση Μάζας Πιθανότητας

Ορισμός (κοινή σ.μ.π.)

Έστω δύο διακριτές τυχαίες μεταβλητές X, Y, με συναρτήσεις μάζας πιθανότητας f_X, f_Y. Η **κοινή συνάρτηση (μάζας) πιθανότητας** (joint probability mass function ή joint pmf) των X, Y είναι η συνάρτηση $f_{X,\,Y}: R \times R \to [0, 1]$ που ορίζεται ως εξής:

$$\mathbf{f_{X,\,Y}(x, y) = P(X = x, Y = y).}$$

Ιδιότητες κοινής συνάρτησης πιθανότητας

1) $0 \le f_{X,\,Y}(x, y) \le 1$
2) $\Sigma_x \Sigma_y f_{X,\,Y}(x, y) = 1.$

Κοινή Συνάρτηση Πυκνότητας Πιθανότητας

$P(X \leq a) = \int_{-\infty}^{a} f(x)\,dx$

Ορισμός (κοινή σ.π.π.)

Έστω δύο συνεχείς τυχαίες μεταβλητές X, Y. Οι X, Y ονομάζονται αμοιβαία συνεχείς, αν υπάρχει f: $R^2 \rightarrow R$, τέτοια ώστε, για κάθε $A \subseteq R^2$:

$$P\left((X, Y) \in A \right) = \iint_A f_{X,Y}(x, y)\, dxdy$$

Η συνάρτηση f_{XY} ονομάζεται **κοινή συνάρτηση (πυκνότητας) πιθανότητας** (joint probability density function ή joint pdf) των X, Y.

Ιδιότητες κοινής συνάρτησης πιθανότητας

1) $0 \leq f_{X,Y}(x, y) \leq 1$

2) $\int_R \int_R f_{X,Y}(x, y)\, dxdy = 1$

Κοινή Συνάρτηση Κατανομής

Ορισμός (κοινή σ.κ.)

Έστω δύο τυχαίες μεταβλητές X, Y, με κατανομές F_X, F_Y. Η **κοινή συνάρτηση κατανομής** (joint cumulative distribution function ή joint cdf) των X, Y είναι η συνάρτηση $F_{X, Y}$: R x R → [0, 1] που ορίζεται ως εξής:

$$F_{X, Y}(x, y) = P(X \leq x, Y \leq y).$$

Σημείωση: Ο ορισμός αυτός αφορά τόσο τις διακριτές όσο και τις συνεχείς τ.μ.

Περιθώρια Συνάρτηση Πιθανότητας

Από την κοινή συνάρτηση μάζας πιθανότητας είναι δυνατόν να ανακτηθεί η **περιθώρια (marginal) συνάρτηση πιθανότητας** των δύο επιμέρους τυχαίων μεταβλητών.

X, Y διακριτές: $f_X(x) = \Sigma_y f_{X,Y}(x, y)$, $f_Y(x) = \Sigma_x f_{X,Y}(x, y)$

$P(X=x, Y=y)$

X, Y συνεχείς: $f_X(x) = \int_R f_{X,Y}(x, y)\, dy$ $f_Y(x) = \int_R f_{X,Y}(x, y)\, dx$

Άσκηση

Δίνονται οι τ.μ. X = 0, 1, 2 και Y = 1, 2.

$P(X=x, Y=y)$

(α) Να αποδείξετε ότι η συνάρτηση $f_{X, Y}(x, y) = (x + y) / 15$, αποτελεί αποδεκτή κοινή σ.π. των X, Y.

(β) Να βρείτε την $F_{X, Y}(3/2, 1)$.

(γ) Να βρείτε τις περιθώριες σ.π. f_X, f_Y.

Λύση

(α) $0 \leq f_{X,Y}(x,y) \leq 1$ ισχύει.

$$\sum_{x=0}^{2} \sum_{y=1}^{2} f_{X,Y}(x,y) = \sum_{x=0}^{2} \sum_{y=1}^{2} \frac{x+y}{15} =$$

$$= \sum_{x=0}^{2} \left(\frac{x+1}{15} + \frac{x+2}{15} \right) = \sum_{x=0}^{2} \frac{2x+3}{15} = \frac{3}{15} + \frac{5}{15} + \frac{7}{15} = \frac{15}{15} = 1.$$

(β) $$F_{X,Y}\left(\frac{3}{2}, 1\right) = P\left(X \leq \frac{3}{2}, Y \leq 1\right) = P(X=0, Y=1) + P(X=1, Y=1)$$

$x=0$ ή 1 $\quad$ $Y=1$

$$= \frac{1}{15} + \frac{2}{15} = \frac{3}{15}.$$

άλυτη

Άσκηση

Δίνονται οι τ.μ. X = 0, 1, 2 και Y = 1, 2.

(α) Να αποδείξετε ότι η συνάρτηση $f_{X,Y}(x, y) = (x + y) / 15$, αποτελεί αποδεκτή κοινή σ.π. των X, Y.

(β) Να βρείτε την $F_{X,Y}(3/2, 1)$.

(γ) Να βρείτε τις περιθώριες σ.π. f_X, f_Y.

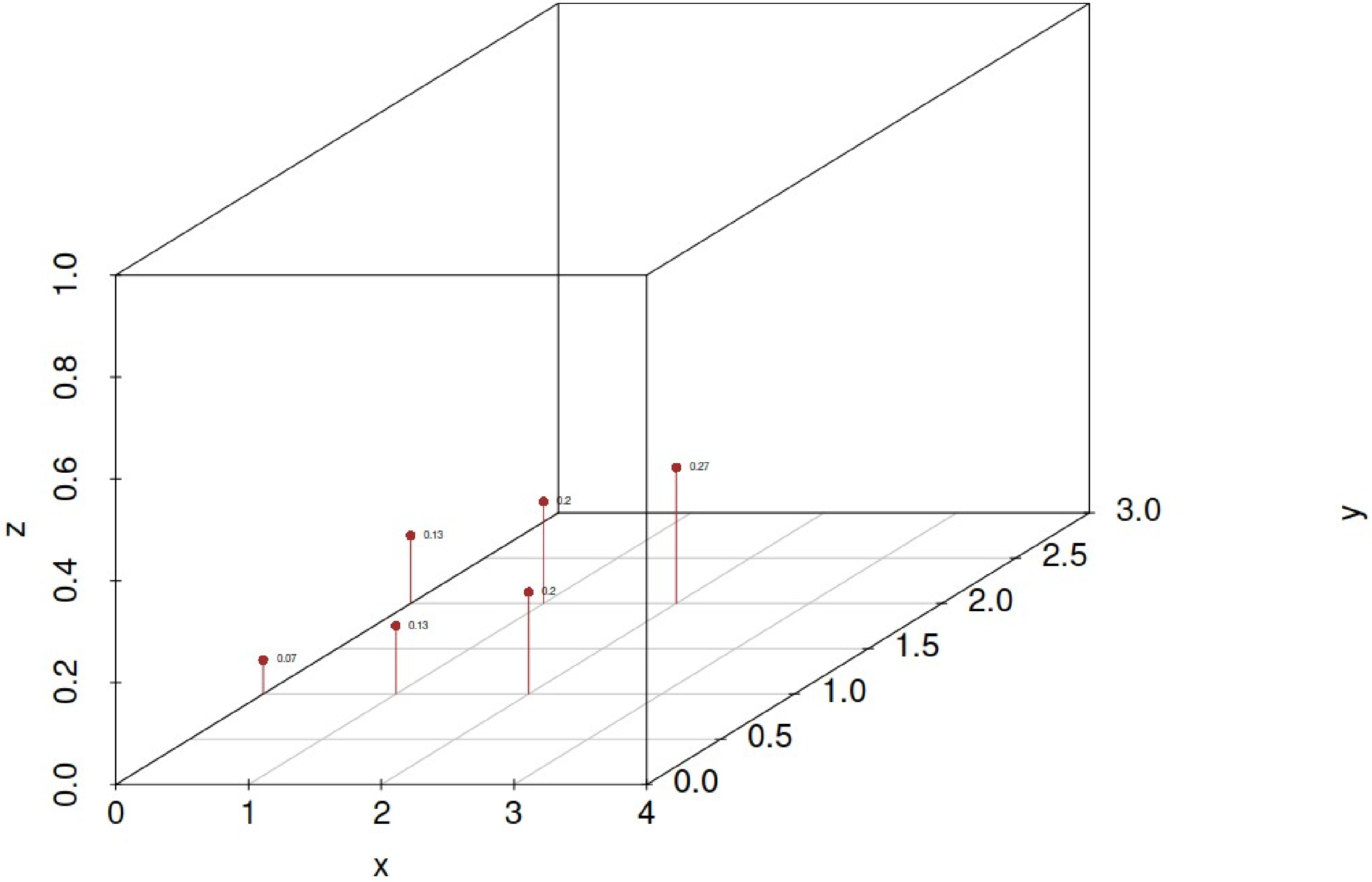
Λύση

(γ) $$f_X(x) = \sum_{y=1}^{2} f_{X,Y}(x,y) = f_{X,Y}(x,1) + f_{X,Y}(x,2) = \frac{x+1}{15} + \frac{x+2}{15} = \frac{2x+3}{15}$$

$$f_Y(y) = \sum_{x=0}^{2} f_{X,Y}(x,y) = \frac{0+y}{15} + \frac{1+y}{15} + \frac{2+y}{15} = \frac{3+3y}{15} = \frac{y+1}{5}.$$

Άσκηση

$f_{X,\,Y}(x, y) = (x + y) / 15,$



Άσκηση

Έστω δύο αμοιβαία συνεχείς τυχαίες μεταβλητές X, Y με κοινή συνάρτηση πυκνότητας πιθανότητας

$f_{X,Y}(x, y) = x + cy^2$, $0 \le x, y \le 1$ και $f_{X,Y}(x, y) = 0$, αλλού.

(α) Να δείξετε ότι c = 3/2.

(β) Να βρείτε την $P(0 \le X \le ½, 0 \le Y \le ½)$.

(γ) Να δείξετε ότι οι περιθώριες σ.π. είναι $f_X(x) = x + 1/2$, $f_Y(y) = 3/2\ y^2 + 1/2$.

Λύση

(α) $$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_0^1\int_0^1 (x+cy^2)\,dx\,dy = \int_0^1\int_0^1 x\,dx\,dy + c\int_0^1\int_0^1 y^2\,dx\,dy$$

$$= \int_0^1 \frac{1}{2}\,dy + c\int_0^1 y^2\,dy = \frac{1}{2} + c\left.\frac{y^3}{3}\right|_0^1 = \frac{1}{2} + \frac{c}{3}.$$

Πρέπει $\frac{1}{2} + \frac{c}{3} = 1 \Leftrightarrow c = \frac{3}{2}$.

άλυτη

Άσκηση

Έστω δύο αμοιβαία συνεχείς τυχαίες μεταβλητές X, Y με κοινή συνάρτηση πυκνότητας πιθανότητας

$f_{X,Y}(x, y) = x + cy^2$, $0 \leq x, y \leq 1$ και $f_{X,Y}(x, y) = 0$, αλλού.

$$f(x,y) = x + \frac{3}{2}y^2, \quad 0 \leq x, y \leq 1$$

(α) Να δείξετε ότι $c = 3/2$.

(β) Να βρείτε την $P(0 \leq X \leq ½, 0 \leq Y \leq ½)$.

(γ) Να δείξετε ότι οι περιθώριες σ.π. είναι $f_X(x) = x + 1/2$, $f_Y(y) = 3/2\, y^2 + 1/2$.

Λύση

(β)

$$P\left(0 \leq X \leq \frac{1}{2}, 0 \leq Y \leq \frac{1}{2}\right) = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} \left(x + \frac{3}{2}y^2\right) dx\,dy =$$

$$= \int_0^{\frac{1}{2}} \underbrace{\int_0^{\frac{1}{2}} x\,dx\,dy} + \frac{3}{2} \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} y^2\,dx\,dy = \frac{1}{16} + \frac{\cancel{3}}{2} \cdot \frac{1}{\cancel{48}16} = \frac{3}{2} \cdot \frac{1}{16} = \frac{3}{32}.$$

Άσκηση

Έστω δύο αμοιβαία συνεχείς τυχαίες μεταβλητές X, Y με κοινή συνάρτηση πυκνότητας πιθανότητας

$f_{X,Y}(x, y) = x + cy^2$, $0 \le x, y \le 1$ και $f_{X,Y}(x, y) = 0$, αλλού.

(α) Να δείξετε ότι $c = 3/2$.

(β) Να βρείτε την $P(0 \le X \le ½, 0 \le Y \le ½)$.

(γ) Να δείξετε ότι οι περιθώριες σ.π. είναι $f_X(x) = x + 1/2$, $f_Y(y) = 3/2\, y^2 + 1/2$.

Λύση

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy = \int_0^1 \left(x + \frac{3}{2}y^2\right)dy = \int_0^1 x\,dy + \frac{3}{2}\int_0^1 y^2\,dy$$

$$= x + \frac{3}{2}\left.\frac{y^3}{3}\right|_0^1 = x + \frac{1}{2}.$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_0^1 \left(x + \frac{3}{2}y^2\right)dx = \frac{1}{2} + \frac{3}{2}y^2.$$

άλυτη

$0 \le y \le x \le 1$ $\qquad\qquad$ $\int\int$

Άσκηση

Έστω δύο αμοιβαία συνεχείς τυχαίες μεταβλητές X, Y με κοινή συνάρτηση πυκνότητας πιθανότητας

$f_{X,Y}(x, y) = cx^2y$, $0 \le y \le x \le 1$ και $f_{X,Y}(x, y) = 0$, αλλού.

(α) Αναπαραστήστε γραφικά το σύνολο $\{(x, y) \in R^2, f_{X,Y}(x, y) \neq 0\}$.

(β) Να δείξετε ότι c = 10.

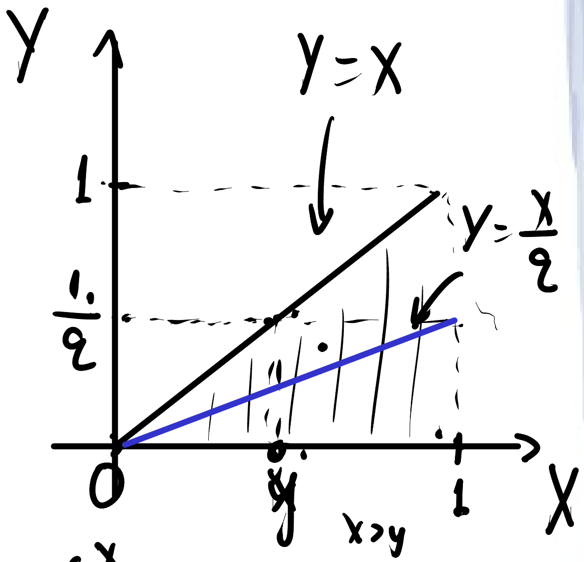
(γ) Να βρείτε τις περιθώριες συναρτήσεις πυκνότητας πιθανότητας f_X, f_Y.

(δ) Να βρείτε την P(Y ≤ X/2).

(ε) P(Y ≤ X/4 | Y ≤ X/2).

$$= \int_0^1 \int_y^1 c x^2 y \, dx \, dy$$

(α)



Λύση

(β) $$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_0^1 \int_0^x c x^2 y \, dy\, dx = c\int_0^1 x^2 \int_0^x y\, dy\, dx =$$

$$= c\int_0^1 x^2 \frac{y^2}{2}\Big|_0^x dx = c\int_0^1 \frac{x^4}{2}dx = \frac{c}{2}\,\frac{x^5}{5}\Big|_0^1 = \frac{c}{10}$$. Πρέπει $\frac{c}{10} = 1 \Leftrightarrow \boxed{c = 10}$

άλυτη

Άσκηση

Έστω δύο αμοιβαία συνεχείς τυχαίες μεταβλητές X, Y με κοινή συνάρτηση πυκνότητας πιθανότητας

$f_{X,Y}(x, y) = cx^2y$, $0 \le y \le x \le 1$ και $f_{X,Y}(x, y) = 0$, αλλού.

(α) Αναπαραστήστε γραφικά το σύνολο $\{(x, y) \in R^2, f_{X,Y}(x, y) \neq 0\}$.

(β) Να δείξετε ότι $c = 10$.

(γ) Να βρείτε τις περιθώριες συναρτήσεις πυκνότητας πιθανότητας f_X, f_Y.

(δ) Να βρείτε την $P(Y \le X/2)$.

(ε) $P(Y \le X/4 \mid Y \le X/2)$.

Λύση

(γ) $$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy = \int_0^x 10x^2y\,dy = 10x^2 \frac{y^2}{2}\Big|_0^x = 5x^4.$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx = \int_y^1 10x^2y\,dx = 10y\frac{x^3}{3}\Big|_y^1 = \frac{10y}{3}(1-y^3).$$

άλυτη

(δ) $P\left(Y \le \frac{X}{2}\right) = \int_0^1 \int_0^{\frac{x}{2}} 10\, x^2 y \, dy\, dx = \ldots$

(ε) $P\left(Y \le \frac{X}{4} \,\middle|\, Y \le \frac{X}{2}\right) = \dfrac{P\left(Y \le \frac{X}{4}, Y \le \frac{X}{2}\right)}{P\left(Y \le \frac{X}{2}\right)} =$

$$= \frac{P\left(Y \le \frac{X}{4}\right)}{P\left(Y \le \frac{X}{2}\right)} = \ldots$$

άλυτη

Ανεξαρτησία Τυχαίων Μεταβλητών

$P(AB) = P(A)P(B)$

Ανεξαρτησία Τυχαίων Μεταβλητών

Ορισμός (ανεξαρτησία τυχαίων μεταβλητών)

Δύο τ. μ. λέγονται **(στοχαστικά) ανεξάρτητες** αν τα ενδεχόμενα $\{X = x\}$ και $\{Y = y\}$ για διακριτές ΤΜ ($x, y \in R$) ή $\{X \in (\alpha, \beta)\}$ και $\{Y \in (\gamma, \delta)\}$ για συνεχείς ΤΜ ($\alpha, \beta, \gamma, \delta \in R$) είναι ανεξάρτητα μεταξύ τους.

Αν οι δύο τυχαίες μεταβλητές X, Y είναι στοχαστικά ανεξάρτητες τότε τα $\{X = x\}$ και $\{Y = y\}$ είναι ανεξάρτητα μεταξύ τους και ισχύει

$$\mathbf{f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)}$$

για κάθε $x, y \in R$.

Ανεξαρτησία Τυχαίων Μεταβλητών

Δύο τυχαίες μεταβλητές X, Y, είναι **(στοχαστικά) ανεξάρτητες** (independent) αν για κάθε x, y ϵ R, ισχύει

$$F_{X,\,Y}(x, y) = F_X(x) \cdot F_Y(y),$$

ή ισοδύναμα, όταν

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y).$$

Σημείωση

Η συνθήκη $P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$ είναι ισοδύναμη με την $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$, δηλαδή με την

$$f_{X,\,Y}(x, y) = f_X(x) \cdot f_Y(y),\ x, y \in R,$$

Η παραπάνω ισότητα ισοδυναμεί με την στοχαστική ανεξαρτησία και στην περίπτωση των συνεχών τ.μ. χωρίς ωστόσο να έχει το νόημα σημειακών πιθανοτήτων.

Ισόνομες τ.μ. $\nRightarrow$ Ανεξάρτητες τ.μ.

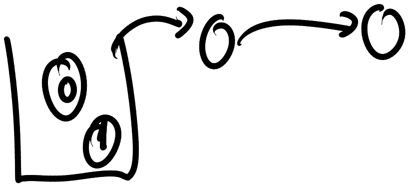
Παρατήρηση
Δύο τυχαίες μεταβλητές X, Y, μπορεί να είναι *ισόνομες* αλλά όχι *ανεξάρτητες*.

Παράδειγμα
Σε ένα σάκο υπάρχουν 2 μπάλες, μία με τον αριθμό 0 και μία με τον αριθμό 1. Εκτελούμε πείραμα με δύο επιλογές μπάλας και ορίζουμε να είναι

X: Ο αριθμός της μπάλας στην $1^{η}$ επιλογή και **Y: Ο αριθμός της μπάλας στη $2^{η}$ επιλογή.**

Αν η δειγματοληψία γίνει **με επανάθεση** τότε Ω = {00, 01, 10, 11} και υπολογίζουμε
P(X = 0) = P(X = 1) = P(Y = 0) = P(Y = 1) = 2/4 = 0,5.
Στην περίπτωση αυτή οι X, Y είναι **ισόνομες και ανεξάρτητες**.

Αν ωστόσο η δειγματοληψία γίνει **χωρίς επανάθεση**, τότε Ω = {01, 10} και υπολογίζουμε
P(X = 0) = P(X = 1) = P(Y = 0) = P(Y = 1) = 1/2 = 0,5.
Στην περίπτωση αυτή οι X, Y είναι **ισόνομες και εξαρτημένες**.



Ανεξαρτησία Τυχαίων Μεταβλητών

1,2,3,4,5,6

Παράδειγμα 1

Έστω ότι ρίχνουμε ένα ζάρι και ορίζουμε X = {1 αν η ρίψη είναι άρτιος και 0 αν είναι περιττός} και Y = {1 αν η ρίψη είναι πρώτος και 0 αν είναι σύνθετος}.

(α) Να βρεθεί η κοινή συνάρτηση μάζας πιθανότητας.

$P(X=0) = \frac{3}{6}$, $P(Y=0) = \frac{3}{6}$

(β) Είναι οι X, Y ανεξάρτητες τυχαίες μεταβλητές;

$P(X=x, Y=y)$, $x,y = 0$ ή 1.

Ρίψη	1	2	3	4	5	6
X	0	1	0	1	0	1
Y	0	1	1	0	1	0

$P(X=0, Y=0) = \frac{1}{6}$

$P(X=0, Y=1) = \frac{2}{6}$

$P(X=1, Y=0) = \frac{2}{6}$

$P(X=1, Y=1) = \frac{1}{6}$

$P(X=0, Y=0) = P(X=0)\,P(Y=0)$

$\frac{1}{6} \neq \frac{3}{6} \cdot \frac{3}{6} = \frac{1}{4} \Rightarrow X, Y$ εξαρτημένες.

Ανεξαρτησία Τυχαίων Μεταβλητών

Λύση
Καθώς, τόσο η X όσο και η Y είναι διακριτές τ.μ. μπορούμε να καταγράψουμε το σύνολο των συνδυασμών των τιμών τους

Ρίψη	1	2	3	4	5	6
X	0	1	0	1	0	1
Y	0	1	1	0	1	0

(α) Για την κοινή συνάρτηση μάζας πιθανότητας υπολογίζουμε

$P(X = 0, Y = 0) = 1/6$, $P(X = 1, Y = 0) = 2/6$,

$P(X = 0, Y = 1) = 2/6$, $P(X = 1, Y = 1) = 1/6$

(β) $P(X = 1, Y = 0) = 2/6 = 1/3$

$P(X = 1) \cdot P(Y = 0) = 3/6 \cdot 3/6 = 9/36 = 1/4$.

Οι τυχαίες μεταβλητές είναι (στοχαστικά) εξαρτημένες.

Ανεξαρτησία Τυχαίων Μεταβλητών

Παράδειγμα 2
Ρίχνουμε ένα κέρμα 2 φορές και ορίζουμε Χ = {πλήθος Κ στις δύο ρίψεις}. Ρίχνουμε το κέρμα άλλες 2 φορές και ορίζουμε Y = {πλήθος Γ στις δύο ρίψεις} (α) Είναι οι Χ, Y ανεξάρτητες τυχαίες μεταβλητές; (β) Να βρεθεί η πιθανότητα $P(X < 2, Y > 1)$

Λύση

Ανεξαρτησία Τυχαίων Μεταβλητών

Παράδειγμα 2
Ρίχνουμε ένα κέρμα 2 φορές και ορίζουμε X = {πλήθος Κ στις δύο ρίψεις}. Ρίχνουμε το κέρμα άλλες 2 φορές και ορίζουμε Y = {πλήθος Γ στις δύο ρίψεις} (α) Είναι οι X, Y ανεξάρτητες τυχαίες μεταβλητές; (β) Να βρεθεί η πιθανότητα $P(X < 2, Y > 1)$

Λύση

Είναι X = 0, 1, 2 και Y = 0, 1, 2 και

$P(X = 0) = P(Y = 0) = 1/4$, $P(X = 1) = P(Y = 1) = 1/2$, $P(X = 2) = P(Y = 2) = 1/4$.

(α) Οι X, Y λαμβάνουν τιμές από διαφορετικές ρίψεις κερμάτων άρα είναι ανεξάρτητες τυχαίες μεταβλητές.

(β) $P(X < 2, Y > 1) = P(X < 2) \cdot P(Y > 1) = 3/4 \cdot 1/4 = 3/16$.

Άσκηση

Δίνονται οι τ.μ. X = 0, 1, 2 και Y = 1, 2, με κοινή συνάρτηση πιθανότητας $f_{X,Y}(x, y) = (x + y) / 15$. Να βρείτε τις περιθώριες σ.~~π~~μ.π. f_X, f_Y. Είναι οι X, Y ανεξάρτητες;

Λύση

$$f_X(x) = \frac{2x+3}{15}, \quad f_Y(y) = \frac{1+y}{5}.$$

X, Y ανεξάρτητες αν-ν $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$

X	0	1	2	0	1	2
Y	1	1	1	2	2	2
f_{XY}	1/15	2/15	3/15	2/15	3/15	4/15
$f_X \cdot f_Y$	2/25					

$f_{X,Y}(x,y)$ || $f_X(x) f_Y(y)$

Είναι $f_{X,Y}(0,1) \neq f_X(0)\cdot f_Y(1)$

άρα X, Y εξαρτημένες.

άλυτη

Δεσμευμένη συνάρτηση κατανομής

Έστω X, Y τυχαίες μεταβλητές με κοινή συνάρτηση μάζας πιθανότητας $f_{X,Y}$. Από την κοινή συνάρτηση μάζας πιθανότητας, είναι δυνατόν να οριστεί και η **δεσμευμένη συνάρτηση μάζας πιθανότητας** της X δοθέντος του Y = y, ως εξής:

$$\mathbf{f}_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \frac{\mathbf{f}_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{\mathbf{f}_{\mathbf{Y}}(\mathbf{y})} = \frac{P(X=x, Y=y)}{P(Y=y)}$$

Επιπλέον, η **δεσμευμένη αθροιστική συνάρτηση κατανομής της X δεδομένου ότι Y = y**, ορίζεται να είναι

$$\mathbf{F_{X|Y}(x \mid y) = P(X \leq x \mid Y = y) = \Sigma_{k \leq x}\, P(X = k \mid Y = y)}$$

$$= \sum_{k \leq x} f_{X|Y}(k|Y=y)$$

Ανεξαρτησία Τυχαίων Μεταβλητών

Δύο τ. μ. είναι **(στοχαστικά) ανεξάρτητες** αν για κάθε y є R,

$$\mathbf{f_{X\,|\,Y}(x \mid y) = f_X(x)}$$

Πράγματι:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x)$$

Ασκήσεις

$X = 0$ ή 1

$Y = 0$ ή 1.

Άσκηση 1

Έστω ότι η κοινή συνάρτηση μάζας πιθανότητας των X, Y δίνεται από τις σχέσεις:

$$f_{X,Y}(0,0) = 0{,}4, \qquad f_{X,Y}(0,1) = 0{,}2, \qquad f_{X,Y}(1,0) = 0{,}1, \qquad f_{X,Y}(1,1) = 0{,}3.$$

(α) Να βρεθούν οι περιθώριες συναρτήσεις πιθανότητας των X, Y.

(β) Να βρεθεί η $f_{X|Y}(x \mid y = 1)$

Λύση

(α) $$f_X(x) = \sum_{y=0}^{1} f_{X,Y}(x,y) = f(x,0) + f(x,1)$$

$$f_X(0) = f_{X,Y}(0,0) + f_{X,Y}(0,1) = 0{,}4 + 0{,}2 = 0{,}6$$

$$f_X(1) = f_{X,Y}(1,0) + f_{X,Y}(1,1) = 0{,}1 + 0{,}3 = 0{,}4.$$

$$f_Y(0) = f(0,0) + f(1,0) = 0{,}5$$

$$f_Y(1) = f(0,1) + f(1,1) = 0{,}5.$$

(β) $$f_{X|Y}(x \mid y=1) = \frac{f_{X,Y}(x,1)}{f_Y(1)} \begin{cases} x=0 \rightarrow \dfrac{2}{5} \\ x=1 \rightarrow \dfrac{3}{5}. \end{cases}$$

Ασκήσεις

Άσκηση 2
Ο πίνακας δίνει την κοινή συνάρτηση πιθανότητας για το ζεύγος των τ.μ. (X, Y).
(α) Να βρεθούν οι περιθώριες και οι δεσμευμένες συναρτήσεις πιθανότητας.
(β) Να βρεθούν οι $P(X = Y)$, $P(X > 2Y)$, $P(X > Y)$, $P(X \geq Y \mid Y \geq 2)$.

$f_{X,Y}$	Y = 1	Y = 2	Y = 3
X = 1	1/12	1/6	0
X = 2	0	1/9	1/5
X = 3	1/18	1/4	2/15

Λύση

άλυτη

Δεσμευμένη Μέση Τιμή

Δεσμευμένη Μέση Τιμή

Ορισμός
Η **δεσμευμένη σ.π. της τ.μ. X ως προς την τιμή y_0 της τ.μ. Y** ορίζεται ως

$$f_X(x_i \mid y_0) = P(X = x_i \mid Y = y_0) = \frac{P(X = x_i, Y = y_0)}{P(Y = y_0)} = \frac{f_{X,Y}(x_1, y_0)}{f_Y(y_0)}$$

όπου **$f_{X,Y}(x, y) = P(X = x, Y = y)$** η κοινή συνάρτηση μάζας πιθανότητας των τ.μ. X, Y.

Ορισμός
Η **δεσμευμένη μέση τιμή** (conditional expected value) μιας διακριτής τυχαίας μεταβλητής X, δοθέντος της $Y = y_0$, ορίζεται ως:

$$\mathbf{E(X \mid Y = y_0) = \sum_{x_i} x_i \cdot f_X(x_i \mid y_0)}$$

$$E(X) = \sum_{x} x \cdot f_X(x) .$$

$$P(X=0 \mid Y=0) = \frac{P(X=0, Y=0)}{P(Y=0)} = \frac{1/6}{3/6} = \frac{1}{3}.$$

Δεσμευμένη Μέση Τιμή

Άσκηση 1

Έστω ότι ρίχνουμε ένα ζάρι και ορίζουμε X = {1 αν η ρίψη είναι άρτιος και 0 αν είναι περιττός} και Y = {1 αν η ρίψη είναι πρώτος και 0 αν είναι σύνθετος}.

(α) Να βρεθούν τα E(X), E(Y).

(β) Να βρεθούν τα E(X | Y = 0), E(X | Y = 1), E(Y | X = 0), E(Y | X = 1).

Λύση

(α) $$E(X) = 0 \cdot \frac{3}{6} + 1 \cdot \frac{3}{6} = \frac{1}{2}.$$

$$E(Y) = 0 \cdot \frac{3}{6} + 1 \cdot \frac{3}{6} = \frac{1}{2}$$

Ρίψη	1	2	3	4	5	6
X	0	1	0	1	0	1
Y	0	1	1	0	1	0

(β) $$E(X \mid Y=0) = 0 \cdot P(X=0 \mid Y=0) + 1 \cdot P(X=1 \mid Y=0) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

άλυτη

Δεσμευμένη Μέση Τιμή

$f_Y(y) = \frac{y+1}{5}$

Άσκηση 2
Η κοινή συνάρτηση πιθανότητας των X, Y δίνεται από τον τύπο

$$f_{X,\,Y}(x, y) = (x + y) / 15, \; x = 0, 1, 2 \text{ και } y = 1, 2.$$

Να υπολογιστούν οι δεσμευμένες μέσες τιμές E(X | Y = 1) και E(Y | X = 1).

Λύση

$$E(X|Y=1) = 0 \cdot P(X=0|Y=1) + 1 \cdot P(X=1|Y=1) + 2 \cdot P(X=2|Y=1) = \frac{1}{3} + 2 \cdot \frac{1}{2} = \frac{4}{3}$$

$$P(X=0|Y=1) = \frac{P(X=0, Y=1)}{P(Y=1)} = \frac{\frac{1}{15}}{\frac{2}{5}} = \frac{1}{6}$$

$$P(X=1|Y=1) = \frac{P(X=1, Y=1)}{P(Y=1)} = \frac{\frac{2}{15}}{\frac{2}{5}} = \frac{1}{3}, \quad P(X=2|Y=1) = \frac{\frac{3}{15}}{\frac{2}{5}} = \frac{1}{2}$$

άλυτη

Ανεξαρτησία και αναμενόμενη τιμή 1/2

Θεώρημα 1
Αν X, Y ανεξάρτητες τ.μ. τότε $E(X \mid Y = y) = E(X)$

Απόδειξη (για διακριτές τ.μ.)

Πράγματι, $E(X \mid Y = y) = \Sigma_{xi}\, x_i \cdot P(X = x_i \mid Y = y)$ (ορισμός E(X))

$= \Sigma_{xi}\, x_i \cdot P(X = x_i) \cdot P(Y = y) / P(Y = y)$ (X, Y ανεξάρτητες)

$= \Sigma_{xi}\, x_i \cdot P(X = x_i)$

$= E(X).$

$E(X+Y)=E(X)+E(Y)$.

Ανεξαρτησία και αναμενόμενη τιμή 2/2

Θεώρημα 2

Αν X, Y ανεξάρτητες τ.μ. τότε E(X · Y) = E(X) · E(Y).

Απόδειξη

Αν X, Y ανεξάρτητες, τότε $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$ και X, Y συνεχείς, τότε:

$$E(XY) = \iint_{\mathbb{R}^2} f_{XY}(x, y)\, xy\, dx\, dy = \iint_{\mathbb{R}^2} f_X(x) f_Y(y)\, xy\, dx\, dy = \int_{\mathbb{R}} f_X(x)\, x\, dx \int_{\mathbb{R}} f_Y(y)\, y\, dy = E(X)E(Y).$$

Αν X, Y διακριτές, τότε $f_{XY}(x, y) = P((X,Y) = (x,y)) = P(X = x)\ P(Y = y) = f_X(x) \cdot f_Y(y)$ και:

$$E(XY) = \sum_x \sum_y f_{XY}(x, y)\, xy = \sum_x \sum_y f_X(x) f_Y(y)\, xy = \sum_x f_X(x)\, x \sum_y f_Y(y)\, y\, dy = E(X)E(Y).$$

Σημείωση: Όπως έχουμε δει, η αντίστοιχη ιδιότητα της αναμενόμενης τιμής για το άθροισμα ισχύει χωρίς κάποια προϋπόθεση για τις τ.μ. X, Y. Δηλαδή για κάθε τ.μ. X, Y ισχύει E(X + Y) = E(X) + E(Y).

Ασκήσεις στις ανεξάρτητες τ.μ.

Άσκηση 1

Ρίχνουμε ένα κέρμα 2 φορές και ορίζουμε Χ = {πλήθος Κ στις δύο ρίψεις}. Ρίχνουμε το κέρμα άλλες 2 φορές και ορίζουμε Y = {πλήθος Γ στις δύο ρίψεις} (α) Είναι οι Χ, Y ανεξάρτητες τυχαίες μεταβλητές; (β) Να βρεθεί η πιθανότητα P(X < 2, Y > 1)

$X = 0, 1, 2$ $\quad \Omega = \{ KK, K\Gamma, \Gamma K, \Gamma\Gamma \}$, $P(X=0) = \frac{1}{4}$, $P(X=1) = \frac{2}{4}$, $P(X=2) = \frac{1}{4}$.

$Y = 0, 1, 2$ $\quad P(Y=0) = \frac{1}{4}$, $P(Y=1) = \frac{2}{4}$, $P(Y=2) = \frac{1}{4}$

(α) Προφανώς γιατί παίρνουν τιμές από διαφορετικά και ανεξάρτητα πειράματα.

(β) $P(X<2, Y>1) = P(X<2) \cdot P(Y>1) = \left(P(X=0) + P(X=1) \right) \cdot P(Y=2) =$

$= \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}$.

άλυτη

Ασκήσεις στις ανεξάρτητες τ.μ.

Άσκηση 2
Ο πίνακας δίνει την κοινή συνάρτηση πιθανότητας για το ζεύγος των τ.μ. (X, Y).
(α) Να βρεθούν οι περιθώριες συναρτήσεις πιθανότητας.
(β) Είναι οι τ.μ. X, Y ανεξάρτητες;
(γ) Να βρεθεί η δεσμευμένη κατανομή του X, δοθέντος Y = 2 και η δεσμευμένη κατανομή του Y, δοθέντος X = 3.

$f_{X,\,Y}$		X				
		1	2	3	4	
	1	0,06	0,12	0,24	0,18	
Y	2	0,02	0	0,08	0,06	
	3	0	0,08	0,16	0	

άλυτη

– Συνδιακύμανση

– Συντελεστής συσχέτισης

Συνδιακύμανση Τυχαίων Μεταβλητών

Η συνδιακύμανση (covariance) είναι μία στατιστική ποσότητα με την οποία ποσοτικοποιείται το είδος της συμμεταβολής των δύο μεταβλητών, δηλαδή το είδος της μεταβολής στις τιμές μίας συνεχούς τυχαίας μεταβλητής καθώς μία άλλη μεταβάλλεται.

Ο υπολογισμός της έχει νόημα για ζεύγη μεταβλητών που ορίζονται στο ίδιο πείραμα, άρα οι τιμές τους προσδιορίζονται μαζί με την εξέλιξη του πειράματος.

Ορισμός

Αν X και Y είναι δύο τυχαίες μεταβλητές τότε η **συνδιακύμανση Cov(X, Y)** ορίζεται ως

$$\mathbf{Cov}(\mathbf{X},\mathbf{Y}) = \mathbf{E}[(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})\cdot(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})],\ \boldsymbol{\mu}_{\mathbf{X}} = \mathbf{E}(\mathbf{X}),\ \boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{E}(\mathbf{Y}).$$

$$Cov(X,Y) = \sum_{x_i}\sum_{y_j}(x_i-\mu_x)\cdot(y_i-\mu_y)\cdot P(X=x_i, Y=y_j).$$

Συνδιακύμανση Τυχαίων Μεταβλητών

Ιδιότητες

1. $Cov(X, Y) = Cov(Y, X)$.

2. $Cov(X, Y) = E(XY) – \mu_X \cdot \mu_Y$

Απόδειξη: $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY - \mu_X \cdot Y - \mu_Y \cdot X + \mu_X \cdot \mu_Y)$

$= E(XY) – \mu_X \cdot E(Y) – \mu_Y \cdot E(X) + \mu_X \mu_Y = E(XY) – \mu_X \cdot \mu_Y$

3. Αν X, Y ανεξάρτητες τυχαίες μεταβλητές τότε $Cov(X, Y) = 0$.

Απόδειξη: $Cov(X, Y) = E(XY) – \mu_X \cdot \mu_Y = E(X)E(Y) – \mu_X \cdot \mu_Y = \mu_X \cdot \mu_Y – \mu_X \cdot \mu_Y = 0$.

4. $Cov(X, X) = Var(X)$

Απόδειξη: $Cov(X, X) = E(X^2) – \mu_X^2 = Var(X)$.

5. $Cov(\alpha X + \beta Y, Z) = \alpha \cdot Cov(X, Z) + \beta \cdot Cov(Y, Z)$

6. $Var(\alpha X + \beta Y) = \alpha^2 Var(X) + \beta^2 Var(Y) + 2\alpha\beta Cov(X, Y)$.

7. $Var(\alpha X – \beta Y) = \alpha^2 Var(X) + \beta^2 Var(Y) – 2\alpha\beta Cov(X, Y)$.

8. Αν X, Y ανεξάρτητες τ.μ. τότε $Var(X + Y) = Var(X) + Var(Y)$.

Συνδιακύμανση Τυχαίων Μεταβλητών

Παράδειγμα
Δύο τυχαίες μεταβλητές, X, Y παίρνουν τιμές X = 1, 2, 3 και Y = 1, 2. Η κοινή συνάρτηση μάζας πιθανότητας παρουσιάζεται στον πίνακα. Να βρεθεί η συνδιακύμανση των X, Y.

Λύση

$f_{X,Y}(x, y)$		$f_X(x)$			Σύνολο
		X = 1	X = 2	X = 3	
$f_X(y)$	Y = 1	0,2	0,2	0	0,4
	Y = 2	0,1	0,2	0,3	0,6
Σύνολο		0,3	0,4	0,3	1

$E(Y) = \mu_y$

$\mu_x \cdot E(Y) = \mu_x \mu_y$

$$Cov(X, Y) = E\left[(X-\mu_x)(Y-\mu_y)\right] = E\left(XY - \mu_x \cdot Y - \mu_y \cdot X + \mu_x \mu_y\right) =$$

$$= E(XY) - \cancel{\mu_x E(Y)} - \mu_y \cdot E(X) + \cancel{\mu_x \mu_y} = E(XY) - \mu_x \mu_y .$$

Συνδιακύμανση Τυχαίων Μεταβλητών

Παράδειγμα

Δύο τυχαίες μεταβλητές, X, Y παίρνουν τιμές X = 1, 2, 3 και Y = 1, 2. Η κοινή συνάρτηση μάζας πιθανότητας παρουσιάζεται στον πίνακα. Να βρεθεί η συνδιακύμανση των X, Y.

Λύση

$f_{X,\,Y}(x, y)$		$f_X(x)$			Σύνολο
		X = 1	X = 2	X = 3	
$f_X(y)$	Y = 1	0,2	0,2	0	0,4
	Y = 2	0,1	0,2	0,3	0,6
Σύνολο		0,3	0,4	0,3	1

$$E(XY) = 1\cdot 1\cdot 0{,}2 + 1\cdot 2\cdot 0{,}1 + 2\cdot 1\cdot 0{,}2 + 2\cdot 2\cdot 0{,}2 + 3\cdot 1\cdot 0 + 3\cdot 2\cdot 0{,}3 = 3{,}4.$$

$$E(X) = 1\cdot 0{,}3 + 2\cdot 0{,}4 + 3\,0{,}3 = 2$$

$$E(Y) = 1\cdot 0{,}4 + 2\cdot 0{,}6 = 1{,}6$$

$$Cov(X,Y) = 3{,}4 - 2\cdot 1{,}6 = 0{,}2.$$

Συνδιακύμανση Τυχαίων Μεταβλητών

Παράδειγμα
Δύο τυχαίες μεταβλητές, X, Y παίρνουν τιμές X = 1, 2, 3 και Y = 1, 2. Η κοινή συνάρτηση μάζας πιθανότητας παρουσιάζεται στον πίνακα. Να βρεθεί η συνδιακύμανση των X, Y.

Λύση

$f_{X,\,Y}(x, y)$	X = 1	X = 2	X = 3	Σύνολο	
Y = 1	0,25	0,2	0	0,4	$f_Y(y)$
Y = 2	0,05	0,2	0,3	0,6	
Σύνολο	0,3	0,4	0,3	1	
	$f_X(x)$				

Υπολογίζουμε,

$E(X) = \Sigma_x x \cdot f_X(x) = 1 \cdot 0,3 + 2 \cdot 0,4 + 3 \cdot 0.3 = 2$,

$E(Y) = \Sigma_y y \cdot f_Y(y) = 1 \cdot 0,4 + 2 \cdot 0,6 = 1,6$,

$E(XY) = \Sigma_x \Sigma_y x \cdot y \cdot f_{X,\,Y}(x, y) = 1 \cdot 1 \cdot 0,25 + 1 \cdot 2 \cdot 0,05 + 2 \cdot 1 \cdot 0,2 + 2 \cdot 2 \cdot 0,2 + 3 \cdot 1 \cdot 0 + 3 \cdot 2 \cdot 0,3 =$

$= 0,25 + 0,1 + 0,4 + 0,8 + 0 + 1,8 = 3,35$

Υπολογίζουμε $Cov(X, Y) = E(XY) – E(X)E(Y) = 3,35 – 3,2 = 0,15$.

Συντελεστής Συσχέτισης Pearson

Ο συντελεστής συσχέτισης του Pearson είναι το κατάλληλο στατιστικό για την ανίχνευση της γραμμικής σχέσης δύο ποσοτικών μεταβλητών, για συνεχείς ή αριθμητικές διακριτές μεταβλητές

$$\rho_{X,Y} = \mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

Ισοδύναμα: $$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2)-\mu_X^2}\cdot\sqrt{E(Y^2)-\mu_Y^2}}$$

$-1 \le \rho_{X,Y} \le 1.$

Συντελεστής Συσχέτισης Pearson

X, Y ανεξάρτητες $\quad Cov(X,Y) = E(X\cdot Y) - \mu_X \mu_Y = E(X)\cdot E(Y) - \mu_X \cdot \mu_Y = 0.$

Αν δύο μεταβλητές είναι ανεξάρτητες, τότε cov(X, Y) = 0 και $\rho_{X,\,Y} = 0$ επίσης. Το αντίθετο δεν ισχύει, δηλαδή μπορεί δύο μεταβλητές να είναι γραμμικά ασυσχέτιστες ($\rho_{X,\,Y} = 0$) και να είναι εξαρτημένες.

Άσκηση

Έστω, X = -1, 0, 1 με $f_X(-1) = f_X(0) = f_X(1) = 1/3$ και Y = 1, αν X = 0 και Y = 0, αν X = -1, 1. Προφανώς οι X, Y είναι εξαρτημένες. Να δείξετε ότι $\rho_{X,\,Y} = 0$.

Υπόδειξη
Μελετήστε το γινόμενο XY.

$$\mu_X = E(X) = -1\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3} = 0, \quad \mu_Y = 0\cdot\frac{2}{3} + 1\cdot\frac{1}{3} = \frac{1}{3}$$

$$Cov(X,Y) = E(XY) - \mu_X\cdot\mu_Y = 0 - 0\cdot\frac{1}{3} = 0.$$

$$E(XY) = (-1)\cdot 0\cdot f_{XY}(-1,0) + 0\cdot 1\cdot f_{X,Y}(0,1) + 1\cdot 0\cdot f_{X,Y}(1,0) = 0.$$

άλυτη

Συντελεστής Συσχέτισης Pearson

Άσκηση

Έστω ότι οι τ.μ. X_1, X_2 είναι ανεξάρτητες με αναμενόμενη τιμή 0 και διακύμανση σ^2. Να βρεθεί ο συντελεστής συσχέτισης Pearson των τ.μ. $X_1 + 2X_2$ and $4X_1 - 3X_2$.

Λύση

άλυτη

Άσκηση εμπέδωσης / επανάληψης

Στον πίνακα παρουσιάζεται η κοινή συνάρτηση μάζας πιθανότητας $f_{X,\,Y}$.

(α) Να αποδείξετε ότι $f_X(1) = 0,3$, $f_X(2) = 0,5$, $f_X(3) = 0,2$.

(β) Να σχεδιάσετε το διάγραμμα της αθροιστικής συνάρτησης κατανομής F_X.

(γ) Να αποδείξετε ότι $f_Y(0) = 0,6$, $f_Y(1) = 0,4$.

(δ) Δείξτε ότι $M_Y(t) = 0,4e^t + 0,6$, $G_Y(t) = 0,4z + 0,6$.

(ε) Δείξτε ότι $E(X) = 1,9$, $E(X^2) = 4,1$, $\mathrm{Var}\, X = 0,49$, $\sigma_X = 0,7$.

(ζ) Δείξτε ότι $E(Y) = 0,4$, $E(Y^2) = 0,4$, $\mathrm{Var}\, Y = 0,24$, $\sigma_Y = 0,49$.

(η) Δείξτε ότι $E(XY) = 0,8$ και $\mathrm{Cov}(X, Y) = 0,04$.

(θ) Δείξτε ότι $\mathrm{Corr}(X, Y) = 0,117$.

(ι) Δείξτε ότι οι τ.μ. X, Y είναι εξαρτημένες.

$f_{X,\,Y}$	X=1	X=2	X=3
Y=0	0,2	0,3	0,1
Y=1	0,1	0,2	0,1

Άσκηση εμπέδωσης / επανάληψης

$f_{X,\,Y}$	X=1	X=2	X=3	f_Y
Y=0	0,2	0,3	0,1	0,6
Y=1	0,1	0,2	0,1	0,4
f_X	0,3	0,5	0,2	1

(α) – (γ): Από τον πίνακα.

(δ) $M_Y(t) = E(e^{Yt}) = 0{,}4e^t + 0{,}6$, $G_Y(t) = E(z^Y) = 0{,}4z + 0{,}6$.

(ε)
$E(X) = 1\cdot f_X(1) + 2\cdot f_X(2) + 3\cdot f_X(3) = 1\cdot 0{,}3 + 2\cdot 0{,}5 + 3\cdot 0{,}2 = 1{,}9$

$E(X^2) = 1^2\cdot f_X(1) + 2^2\cdot f_X(2) + 3^2\cdot f_X(3) = 1\cdot 0{,}3 + 4\cdot 0{,}5 + 9\cdot 0{,}2 = 4{,}1$,

$\mathrm{Var}\, X = E(X^2) – E(X)^2 = 4{,}1 – 1{,}9^2 = 0{,}49$, $\sigma_X = 0{,}49^{½} = 0{,}7$.

(ζ) ανάλογα με το (ε) $E(Y) = 0{,}4$, $E(Y^2) = 0{,}4$, $\mathrm{Var}\, Y = 0{,}24$, $\sigma_Y = 0{,}49$.

(η) $E(XY) = 1\cdot 0\cdot f_{XY}(1, 0) + 2\cdot 0\cdot f_{XY}(2, 0) + 3\cdot 0\cdot f_{XY}(3, 0) + 1\cdot 1\cdot f_{XY}(1, 1) + 2\cdot 1\cdot f_{XY}(2, 1) + 3\cdot 1\cdot f_{XY}(3, 1) =$
$1\cdot 0\cdot 0{,}2 + 2\cdot 0\cdot 0{,}3 + 3\cdot 0\cdot 0{,}1 + 1\cdot 1\cdot 0{,}1 + 2\cdot 1\cdot 0{,}2 + 3\cdot 1\cdot 0{,}1 = 0{,}8$.
$\mathrm{Cov}(X, Y) = E(XY) – E(X) \cdot E(Y) = 0{,}04$.
(θ) $\mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y) / (\sigma_X \cdot \sigma_Y) = 0{,}04 / (0{,}7\cdot 0{,}49) = 0{,}117$.

(ι) $f_{XY}(1, 0) = 0{,}2 \neq 0{,}18 = f_X(1)\cdot f_Y(0)$ άρα οι X, Y είναι εξαρτημένες.

Γραφική Σύνοψη

Ζεύγος Τυχαίων Μεταβλητών
X, Y

Συναρτήσεις

Κοινή Συνάρτηση Μάζας Πιθανότητας $f_{X, Y}$
Κοινή Συνάρτηση Κατανομής $F_{X, Y}$
Περιθώριες Συναρτήσεις Πιθανότητας f_X, f_Y.
Δεσμευμένη Συνάρτηση Πιθανότητας $f_{X|Y}$.

Περιγραφή Σχέσης

Ισόνομες ($F_X = F_Y$)
Ανεξάρτητες ($F_{X, Y} = F_X F_Y$)
Συνδιακύμανση Cov(X, Y)
Συσχέτιση Corr(X, Y)

Δεσμευμένη Μέση Τιμή
$E(X \mid Y = y)$

Συνήθεις κατανομές διακριτών μεταβλητών

Συνήθεις κατανομές διακριτών μεταβλητών

κατανομή	συνάρτηση πιθανότητας	παράμετροι	μέση τιμή	διακύμανση
Bernoulli	$p(1-p)$	$p \in (0,1)$	p	$p(1-p)$
Διωνυμική	$\binom{n}{k} p^k (1-p)^{n-k}$	$p \in (0,1), n \in \mathbb{N}$	np	$np(1-p)$
Αρνητική διωνυμική	$\binom{r+k-1}{r-1} p^r (1-p)^k$	$p \in (0,1), r \in \mathbb{N}$	$r\frac{p}{1-p}$	$r\frac{1-p}{p^2}$
Γεωμετρική	$p(1-p)^{n-1}$	$p \in (0,1)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Υπεργεωμετρική	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$n, N, K \in \mathbb{N},\ n, K \leq N$	$\frac{nK}{N}$	$\frac{nK(N-K)(N-n)}{N^2(N-1)}$
Poisson	$\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}$	$\lambda \in \mathbb{R}_+$	λ	λ

Κατανομή Bernoulli

Κατανομή Bernoulli

Η **κατανομή Μπερνούλλι** (Bernoulli) είναι μια συνάρτηση κατανομής διακριτής τυχαίας μεταβλητής. Με αυτήν περιγράφεται ένα τυχαίο πείραμα με δύο πιθανά αποτελέσματα (επιτυχία - αποτυχία) και πιθανότητα επιτυχίας p.

Πιο συγκεκριμένα, αν X είναι η τ.μ. που παίρνει τιμές 0 (αποτυχία) ή 1 (επιτυχία) και $X \sim \text{Bernoulli}(p)$ (γράφουμε και $X \sim B(1, p)$) τότε:

$$P(X = 1) = p \text{ και } P(X = 0) = q = 1 - p.$$

Η κατανομή Bernoulli αξιοποιείται για την μοντελοποίηση πειραμάτων με δυο πιθανά αποτελέσματα.

Κατανομή Bernoulli

Άσκηση
Να βρεθεί η αναμενόμενη τιμή, η διακύμανση και η τυπική απόκλιση της X ~ Bernoulli(p).

$X \sim Bernoulli(p)$. $\quad X = 0, 1 \quad P(X=0) = 1-p, \quad P(X=1) = p$.

$E(X) = 0 \cdot (1-p) + 1 \cdot p = p$. $\quad E(X^2) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$.

$Var(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$.

$\sigma = \sqrt{Var(X)} = \sqrt{p(1-p)}$.

Bernoulli: Γεωμετρικά χαρακτηριστικά

X ~ Bernoulli(p): $P(X = 1) = p$ και $P(X = 0) = q = 1 - p$.

Αναμενόμενη τιμή: $EX = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = p$.

Διακύμανση: $\mathrm{Var}X = EX^2 - (EX)^2 = p - p^2 = p \cdot (1 - p) = p \cdot q$.

Τυπική απόκλιση: $\sigma = \sqrt{p \cdot (1 - p)}$

Ασυμμετρία: $\operatorname{skew}(X) = \gamma = \dfrac{\mu_3}{\sigma^3} = \dfrac{1-2p}{\sqrt{p(1-p)}}$

Κυρτότητα: $\operatorname{kurt}(X) = \alpha = \dfrac{\mu_4}{\sigma^4} = \dfrac{1 - 3p(1-p)}{p(1-p)}$

Διωνυμική κατανομή

$C(n,k) = \binom{n}{k}$

Διωνυμική κατανομή

$8A: (1-p)^8 \quad 7A, 1E \quad (1-p)^7 \cdot p$

Άσκηση

Έστω ένα πείραμα που εκτελείται 8 φορές, ανεξάρτητα η μία από την άλλη και ότι σε κάθε μία επανάληψη υπάρχει πιθανότητα επιτυχίας p και αποτυχίας 1 – p. Αν X είναι το πλήθος των επιτυχιών, να βρεθεί η πιθανότητα P(X = 3).

Λύση

$X = 0, 1, 2, \ldots, 8, \quad P(X=3) = \binom{8}{3} \cdot \frac{1}{2^8} = \frac{\frac{8!}{3! \cdot 5!}}{2^8} = \frac{\frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1}}{2^8} = \frac{\cancel{56}}{\cancel{256}}$

$\underline{A}\ \underline{A}\ \underline{A}\ \underline{A}\ \underline{A}\ \underline{A}\ \underline{A}\ \underline{A} \rightarrow X=0$

$\underline{A}\ \underline{A}\ \underline{A}\ \underline{A}\ \underline{A}\ \underline{A}\ \underline{A}\ \underline{E} \rightarrow X=1$

$\underline{E}\ \underline{E}\ \underline{E}\ \underline{E}\ \underline{E}\ \underline{E}\ \underline{E}\ \underline{E} \rightarrow X=8$

Λάθος, γιατί οι διαφορετικές διατάξεις δεν είναι απλά ενδεχόμενα. Έχουν διαφορετικές πιθανότητες εμφάνισης.

Διωνυμική κατανομή

Άσκηση

Έστω ένα πείραμα που εκτελείται 8 φορές, ανεξάρτητα η μία από την άλλη και ότι σε κάθε μία επανάληψη υπάρχει πιθανότητα επιτυχίας p και αποτυχίας 1 – p. Αν X είναι το πλήθος των επιτυχιών, να βρεθεί η πιθανότητα P(X = 3).

Λύση

A: 1-p
E: p

A: 1-p
E: p

$$P(X=3) = \binom{8}{3} \cdot (1-p)^5 \cdot p^3 .$$

$$\binom{8}{3} \cdot \begin{cases} A\ E\ A\ E\ E\ A\ A\ A \\ \vdots \\ E\ E\ E\ A\ A\ A\ A\ A \end{cases}$$

A E A E E A A A :

$$(1-p) \cdot p \cdot (1-p) \cdot p \cdot p \cdot (1-p) \cdot (1-p) \cdot (1-p)$$

$$= (1-p)^5 \cdot p^3 .$$

Διωνυμική κατανομή

Γέννεση της Διωνυμικής Κατανομής

Έστω ένα πείραμα που εκτελείται n φορές, ανεξάρτητα η μία από την άλλη και ότι σε κάθε μία επανάληψη υπάρχει πιθανότητα επιτυχίας p και αποτυχίας 1 – p. Αν X είναι το πλήθος των επιτυχιών, να βρεθεί η πιθανότητα P(X = κ), κ = 0, 1, 2, ..., n.

Λύση

Διωνυμική κατανομή

Έστω ότι ένα πείραμα επαναλαμβάνεται n φορές και κάθε φορά υπάρχει σταθερή πιθανότητα επιτυχίας p. Τότε, αν η τ.μ. X μετράει το πλήθος επιτυχιών στα n ανεξάρτητα πειράματα, είναι

$$P(X=\kappa) = \binom{n}{\kappa} p^{\kappa}(1-p)^{n-\kappa},\ \kappa = 0, 1, 2, ..., n.$$

Λέμε ότι η διακριτή τ.μ. X ακολουθεί τη **διωνυμική κατανομή** (binomial distribution) και γράφουμε **X ~ B(n, p)**.

Σημείωση
Έστω ότι σε ένα σάκο υπάρχουν N αντικείμενα από τα οποία τα K έχουν μία ιδιότητα και προσδιορίζουν την “επιτυχία” στην επιλογή. Αν επιλέξουμε n αντικείμενα *με επανατοποθέτηση* και X = {πλήθος επιτυχιών στις n επιλογές} τότε X ~ B(n, p), όπου p = K/N.

Διωνυμική: Γεωμετρικά χαρακτηριστικά

X ~ B(n, p): $P(X = \kappa) = \binom{n}{\kappa} p^{\kappa} (1 - p)^{n-\kappa}, \kappa = 0, 1, 2, ..., n.$

Αναμενόμενη τιμή: $EX = n \cdot p.$

Διακύμανση: $VarX = n \cdot p \cdot q = n \cdot p \cdot (1 - p).$

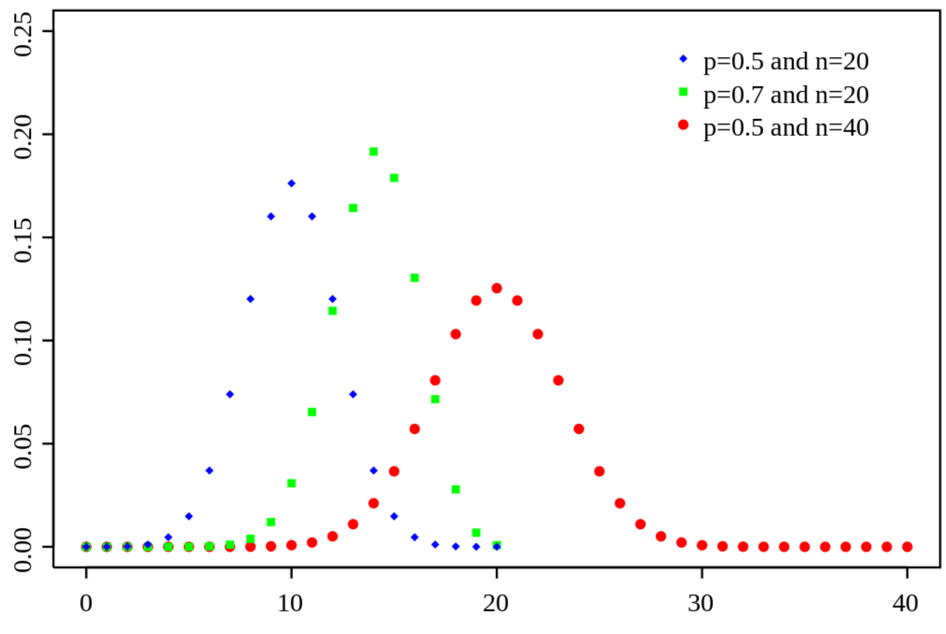
Τυπική απόκλιση: $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$

Ασυμμετρία: $skew(X) = \gamma = \frac{\mu_3}{\sigma^3} = \frac{1 - 2p}{\sqrt{np(1 - p)}}$

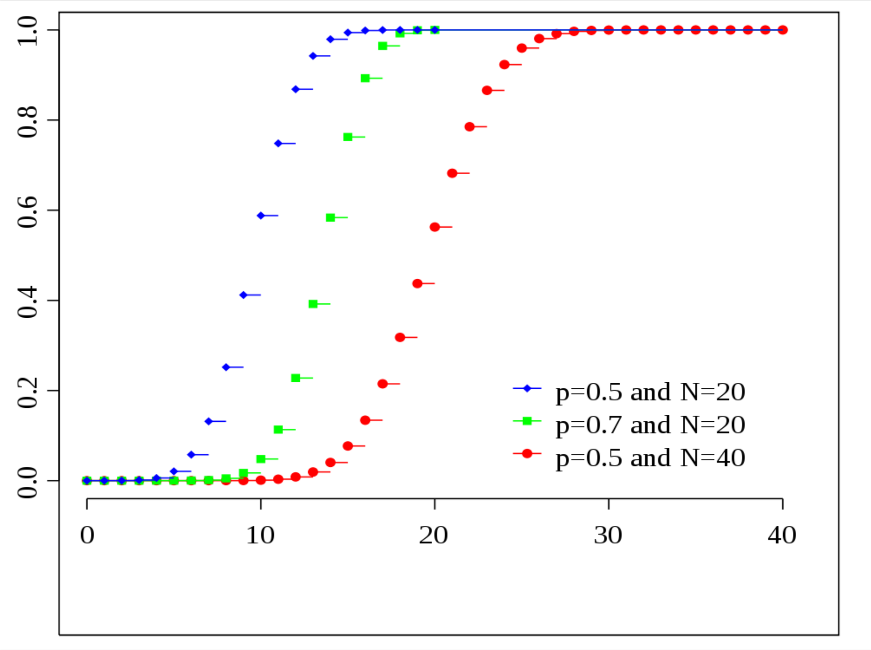
Κυρτότητα: $kurt(X) = \alpha = \frac{\mu_4}{\sigma^4} = \frac{1 - 6p(1 - p)}{np(1 - p)}$

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

Διωνυμική κατανομή



Συνάρτηση πιθανότητας P(X = i), i = 0, …, n



Συνάρτηση κατανομής F(x) = P(X ≤ x), x є R

Πιθανογεννήτρια συνάρτηση διωνυμικής

Παράδειγμα

Αν $X \sim B(n, p)$ τότε $P(X = \kappa) = (n \text{ ανά } \kappa)\, p^{\kappa} (1 - p)^{n-\kappa}$ και

$$G_X(z) = E(z^X) = \Sigma_{\kappa = 0, 1, \ldots} P(X = \kappa) \cdot z^{\kappa}$$

$$= \Sigma_{\kappa = 0, 1, \ldots}\ C(n, \kappa)\, p^{\kappa} (1 - p)^{n-\kappa} \cdot z^{\kappa}$$

$$= \Sigma_{\kappa = 0, 1, \ldots}\ C(n, \kappa)\, (zp)^{\kappa} (1 - p)^{n-\kappa}$$

$$= (zp + 1 - p)^n$$

$$= (1 + (z - 1)\, p)^n.$$

Άρα, $G_X(z) = (1 + (z - 1)\, p)^n$ για **κάθε $z \in C$**, ειδικότερα,

$G_X(x) = (1 + (x - 1)\, p)^n$, για **κάθε $x \in R$**.

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \cdot b^{n-k}$$

Ροπογεννήτρια συνάρτηση διωνυμικής

Άσκηση

Αν X ~ B(n, p) να δείξετε ότι $M_X(t) = E(e^{tx}) = (q + pe^t)^n$.

Απόδειξη

$$X \sim B(n,p) \Rightarrow P(X=k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}, \quad k=0,1,2,\dots,n.$$

$$M_X(t) = E(e^{tx}) = \sum_{k=0}^{n} e^{t\cdot k} \cdot P(X=k) = \sum_{k=0}^{n} e^{t\cdot k} \cdot \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k \cdot (1-p)^{n-k} = (1-p+pe^t)^n = (q+pe^t)^n.$$

Αναγνώριση διωνυμικής κατανομής

Ένα τυχαίο πείραμα είναι διωνυμικό πείραμα και υποστηρίζει τον ορισμό μίας διωνυμικής τυχαίας μεταβλητής όταν:

- έχει σταθερό αριθμό (n) δοκιμών.
- κάθε δοκιμή είναι ανεξάρτητη από τις άλλες.
- κάθε δοκιμή έχει μόνο δύο πιθανά αποτελέσματα, «επιτυχία» και «αποτυχία».
- η πιθανότητα επιτυχίας p για κάθε δοκιμή, είναι σταθερή και το συμπλήρωμα της είναι η πιθανότητα αποτυχίας $q = 1 – p$.

$X \sim B(\nu, p)$

Ασκήσεις στη διωνυμική κατανομή

1. Αν X ~ B(3, 0,5) να βρεθεί η συνάρτηση μάζας πιθανότητας της X.

Λύση

$$X \sim B(\underline{\underline{3}}, 0.5) \Rightarrow X = 0, 1, 2, 3 \quad \text{με} \quad P(X=\kappa) = \binom{3}{\kappa} 0.5^{\kappa} \cdot (1-0.5)^{3-\kappa}$$

$$P(X=\kappa) = \binom{3}{\kappa} 0.5^{3}.$$

$$P(X=0) = \binom{3}{0} 0.5^3 = 0{,}5^3 = 0{,}125$$

$$P(X=1) = \binom{3}{1} \cdot 0.5^3 = 3 \cdot 0.5^3 = 0{,}375$$

$$P(X=2) = \binom{3}{2} \cdot 0{,}5^3 = 3 \cdot 0.5^3 = 0{,}375$$

$$P(X=3) = \binom{3}{3} \cdot 0.5^3 = 0.5^3 = 0{,}125.$$

$X \sim B(5, 0.6)$

Ασκήσεις στη διωνυμική κατανομή

2. Γνωρίζουμε ότι το 60% των φοιτητών της σχολής ζει μόνιμα στην πόλη της Ξάνθης. Επιλέγουμε τυχαία 5 φοιτητές από τον κατάλογο όλων των φοιτητών που σπουδάζουν στη σχολή. Να βρεθεί η πιθανότητα:

(α) Οι 3 να ζουν μόνιμα στην πόλη.

(β) Τουλάχιστον 4 να ζουν μόνιμα στην πόλη.

(γ) Το πολύ 2 να ζουν μόνιμα στην πόλη.

Λύση

$p = 0,6$

$\underline{0,6}\ \underline{0,6}\ \underline{0,6}\ \underline{0,4}\ \underline{0,4} : 0,6^3 \cdot 0,4^2$

X = πλήθος φοιτητών στους 5 που ζουνε μόνιμα στην Ξάνθη. = 0, 1, 2, 3, 4, 5.

(α) $P(X=3) = \binom{5}{3} \cdot 0,6^3 \cdot 0,4^2$

(β) $P(X \geq 4) = P(X=4) + P(X=5) = \binom{5}{4} \cdot 0,6^4 \cdot 0,4 + \binom{5}{5} \cdot 0,6^5 = \ldots$

(γ) $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = \binom{5}{0} \cdot 0,6^0 \cdot 0,4^5 + \binom{5}{1} \cdot 0,6 \cdot 0,4^4 + \binom{5}{2} 0,6^2 \cdot 0,4^3 = \ldots$

10 ανεξάρτητες επαναλ. $p=\frac{18}{37}$: επιτυχία, X: πλήθος επιτ. στις 10 $\sim B(10, \frac{18}{37})$.

Ασκήσεις στη διωνυμική κατανομή

3. Ένας παίκτης παίζει ρουλέτα και ποντάρει συνεχώς στο μαύρο χρώμα. Η πιθανότητα να έρθει μαύρο σε ένα οποιοδήποτε γύρισμα της ρουλέτας είναι ίση με 18/37. Να βρεθεί η πιθανότητα να κερδίσει 4 φορές σε 10 γυρίσματα της ρουλέτας.

Λύση

$$\overset{18}{\underline{M}}\cdot\overset{18}{\underline{M}}\cdot\overset{18}{\underline{M}}\cdot\overset{18}{\underline{M}}\cdot\overset{19}{\underline{K}}\cdot\overset{19}{\underline{K}}\cdot\overset{19}{\underline{K}}\cdot\overset{19}{\underline{K}}\;\overset{19}{\underline{K}}\;\overset{19}{\underline{K}} : \left(\frac{18}{37}\right)^4\cdot\left(1-\frac{18}{37}\right)^6$$

37^{10}

$$\underline{K}\;\underline{K}\;\underline{K}\;\underline{K}\;\underline{K}\;\underline{K}\;\underline{M}\;\underline{M}\;\underline{M}\;\underline{M}$$

X = πλήθος M στις 10 ρίψεις

$$P(X=4)=210\cdot\left(\frac{18}{37}\right)^4\left(1-\frac{18}{37}\right)^6.$$

$$\binom{10}{4} \text{ τρόποι} = \frac{10!}{4!\cdot 6!} = \frac{10\cdot \cancel{9}^{3}\cdot\cancel{8}\cdot 7}{\cancel{4}\cdot\cancel{3}\cdot\cancel{2}} = 210$$

Ασκήσεις στη διωνυμική κατανομή

4. Μία εταιρεία λιπασμάτων ισχυρίζεται ότι το λίπασμα της αυξάνει την παραγωγή του 80% των χωραφιών που εφαρμόζεται. Ένας συνεταιρισμός, χρησιμοποίησε το λίπασμα της εταιρείας σε 7 χωράφια και παρατήρησαν αύξηση στην παραγωγή στα 3 από αυτά. Πόσο πιθανό ήταν να συμβεί αυτό, αν ο ισχυρισμός της εταιρείας είναι σωστός;

Λύση

Ισχυρισμός H_0: 80% των χωραφιών με αύξηση, $X \sim B(7, 0.8)$

X = πλήθος χωραφιών με αύξηση μεταξύ των 7: $0, 1, 2, .., 7$

$$P(X=3) = \binom{7}{3} \cdot 0.8^3 \cdot 0.2^4 = \frac{7!}{3! \cdot 4!} 0.8^3 0{,}2^4 = 0{,}029$$

Ασκήσεις στη διωνυμική κατανομή

5. Πόσα παιδιά πρέπει να αποκτήσει μια οικογένεια ώστε να έχει ένα τουλάχιστον αγόρι και ένα τουλάχιστον κορίτσι με πιθανότητα μεγαλύτερη από α) 95%, β) 99%

Λύση

X = κορίτσια στα n παιδιά, $n \geq 2$.

$X = 0, 1, 2, \ldots, n$.

$X \sim B(n, 0.5)$

$$P(1 \text{ τουλ. } A + 1 \text{ τουλ. } K) = P(1 \leq X < n) = 1 - P(X=n) - P(X=0) =$$

$$= 1 - \binom{n}{n} \cdot 0.5^n - \binom{n}{0} 0.5^n = 1 - 2 \cdot 0.5^n.$$

(α) $1 - 2 \cdot 0.5^n > 0{,}95 \Leftrightarrow 2 \cdot 0.5^n < 0{,}05 \Leftrightarrow 0{,}5^n < 0{,}025 \Leftrightarrow n \cdot \ln 0{,}5 < \ln 0{,}025$

$\Leftrightarrow n > \ln 0{,}025 / \ln 0{,}5 = 5{,}32.$

$X \sim Bernoulli(p), X = 0, 1. \quad E(X) = p.$

Ασκήσεις στη διωνυμική κατανομή

6. Ένα test πολλαπλής επιλογής αποτελείται από 20 ερωτήσεις. Για κάθε ερώτηση, υπάρχουν 4 πιθανές απαντήσεις και μια μόνο από τις οποίες είναι σωστή. Κάθε σωστή απάντηση παίρνει 1 μονάδα. Να βρεθεί ο μέσος αναμενόμενος βαθμός των φοιτητών που απαντούν στην τύχη.

Λύση

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

$X = \{$πλήθος σωστών αποκρίσεων στις $20\} : 0, 1, .., 20$

20 ανεξάρτητες επιλογές, $p = 0.25$ πιθανότητα "επιτυχίας"

$X = \{$πλήθος επιτυχιών στις $n = 20\} \sim B(20, 0.25)$

$$E(X) = n \cdot p = 20 \cdot 0.25 = 5.$$

Ασκήσεις στη διωνυμική κατανομή

κάποια   Παίρνουμε 4.

7. Έχουμε 4 tablet, τα οποία έχουν όλα ίδια πιθανότητα p να είναι ελαττωματικά. Γνωρίζουμε ότι η πιθανότητα να μην υπάρχει ελαττωματικό tablet είναι ίση με την πιθανότητα να είναι όλα ελαττωματικά. Αν X = {πλήθος ελαττωματικών tablet}, τότε:

(α) Να βρεθεί η συνάρτηση πιθανότητας της τ.μ. X
(β) Να βρεθεί η μέση τιμή και η τυπική απόκλιση της κατανομής.

Λύση

$X = 0, 1, 2, 3, 4 \sim B(4, p) = B(4, 0.5)$

$P(X=0) = P(X=4) \Leftrightarrow \binom{4}{0} \cdot p^0 \cdot (1-p)^4 = \binom{4}{4} \cdot p^4 \cdot (1-p)^0 \Leftrightarrow$

$\Leftrightarrow (1-p)^4 = p^4 \Leftrightarrow 1-p = \pm p \begin{cases} 1-p = p \Leftrightarrow p = \frac{1}{2} \\ 1-p = -p : \text{Αδύν.} \end{cases}$

$E(X) = np = 2$

$Var(X) = npq = 4 \cdot 0.5 \cdot 0.5 = 1, \ \sigma = \sqrt{Var X} = 1.$

Υπεργεωμετρική κατανομή

Υπεργεωμετρική κατανομή

Παράδειγμα (γέννεση της υπεργεωμετρικής)

Στο Λόττο επιλέγουμε 6 αριθμούς από 49. Να βρεθεί η πιθανότητα να προβλέψουμε τους 4 αριθμούς που θα κληρωθούν, επιλέγοντας 6 αριθμούς.

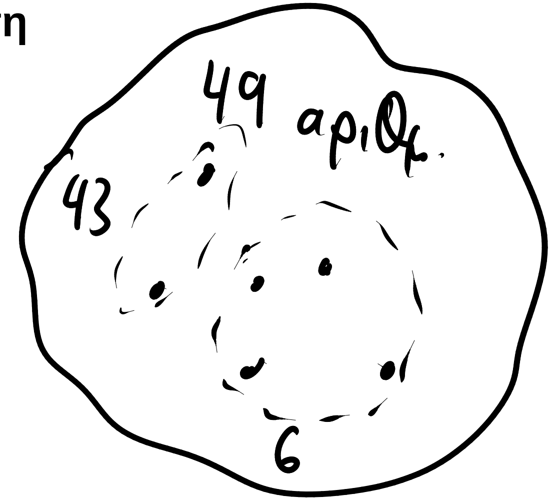
Υπόδειξη: Η σωστή πρόβλεψη των 4 αριθμών από τους 6 που κληρώνονται μπορεί να θεωρηθεί μία διαδικασία που υλοποιείται σε δύο διαδοχικά βήματα.

Στο $1^{ο}$ βήμα επιλέγονται οι 4 νικηφόροι αριθμοί από τους 6 που κληρώθηκαν.

Στο $2^{ο}$ βήμα επιλέγονται οι 2 μη νικηφόροι αριθμοί από τους 43.

Λύση

Πιθανές τυχαίες 6άδες: $\binom{49}{6}$



$$\frac{\binom{6}{4}\cdot\binom{43}{2}}{\binom{49}{6}}$$

49

13 5

5 13

$\underline{49}\ \underline{48}\ \underline{\cdot}\ \underline{\cdot}\ \underline{\cdot}\ \underline{\cdot}$

$\underline{7}\ \underline{13}\ \underline{48}\ \underline{5}\ \underline{10}\ \underline{22}$

$$49 \cdot 48 \cdots 44 = \frac{49!}{43!}$$

$$\frac{49!}{43!\,6!} = \binom{49}{6}$$

Υπεργεωμετρική κατανομή

Παράδειγμα (γέννεση της υπεργεωμετρικής)
Στο Λότto επιλέγουμε 6 αριθμούς από 49. Να βρεθεί η πιθανότητα να προβλέψουμε τους 4 αριθμούς που θα κληρωθούν, επιλέγοντας 6 αριθμούς.
Υπόδειξη: Η σωστή πρόβλεψη των 4 αριθμών από τους 6 που κληρώνονται μπορεί να θεωρηθεί μία διαδικασία που υλοποιείται σε δύο διαδοχικά βήματα.
Στο 1ο βήμα επιλέγονται οι 4 νικηφόροι αριθμοί από τους 6 που κληρώθηκαν.
Στο 2ο βήμα επιλέγονται οι 2 μη νικηφόροι αριθμοί από τους 43.
Λύση
X = {πλήθος σωστών επιλογών από τους 6 που κληρώθηκαν}

$$P(X=4)=\frac{\text{ευνοϊκές}}{\text{συνολικές}}=\frac{\text{ευνοϊκές } 1^{\text{ο}} \text{ βήμα}\cdot\text{ευνοϊκές } 2^{\text{ο}} \text{ βήμα}}{\text{συνολικές}}=\frac{\binom{6}{4}\binom{49-6}{6-4}}{\binom{49}{6}}=0{,}000968=0{,}1\%$$

Υπεργεωμετρική κατανομή

49 6

Από έναν πληθυσμό με N στοιχεία, τα K έχουν μία ιδιότητα (επιτυχία). Εμείς, παίρνουμε ένα δείγμα n στοιχείων *χωρίς επανάθεση* και ορίζουμε

6

X = {πλήθος επιτυχιών στις n επιλογές}. Η τ.μ. X έχει συνάρτηση πιθανότητας

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Λέμε ότι η X ακολουθεί την υπεργεωμετρική κατανομή Hypergeometric(N, K, n) (X ~ HG(N, k, n)).

Σημείωση
Η διωνυμική και η υπεργεωμετρική μοιάζουν στο γεγονός πως προκύπτουν σε πείραμα δειγματοληψίας n αντικειμένων από ένα σωρό. Η διαφορά τους είναι ότι στην υπεργεωμετρική τα αντικείμενα είναι πεπερασμένα (N) σε πλήθος και η πιθανότητα επιτυχίας αλλάζει κάθε φορά, ενώ στη διωνυμική είναι απεριόριστα και η πιθανότητα επιτυχίας είναι σταθερή (p) σε όλες τις επιλογές.

Υπεργεωμετρική: Γεωμετρικά χαρακτηριστικά

Y ~ HG(N, K, n): $$P\left(X = k\right) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Αναμενόμενη τιμή: $$EX = \frac{n \cdot K}{N}.$$

Διακύμανση: $$VarX = \frac{n \cdot K \cdot \left(N - K\right) \cdot \left(N - n\right)}{N^2 \cdot \left(N - 1\right)}.$$

Υπεργεωμετρική Κατανομή

Παράδειγμα 1
Σε ένα σάκο είναι 45 κόκκινοι και 5 πράσινοι βόλοι. Βγάζουμε χωρίς επανάθεση 10 βόλους. Ποια είναι η πιθανότητα οι 4 από αυτούς να είναι πράσινοι;

Λύση

45 Κ
5 Π
$\rightarrow$ 10 βόλους

$X = \{$πράσινοι βόλοι στους 10$\}$

$$P(X=4) = \frac{\binom{5}{4}\cdot\binom{45}{6}}{\binom{50}{10}}$$

$$X \sim HG(50, 5, 10)$$

Υπεργεωμετρική Κατανομή

Παράδειγμα 1
Σε ένα σάκο είναι 45 κόκκινοι και 5 πράσινοι βόλοι. Βγάζουμε χωρίς επανάθεση 10 βόλους. Ποια είναι η πιθανότητα οι 4 από αυτούς να είναι πράσινοι;

Λύση

Αν X = {πλήθος πράσινων βόλων στους 10} τότε X ~ HG(50, 5, 10):

$$P(X = 4) = \frac{\binom{5}{4}\binom{50-5}{10-4}}{\binom{50}{10}} = 0{,}0039 = 0{,}39\%.$$

Υπεργεωμετρική Κατανομή

Παράδειγμα 2 (Λόττο updated)
Στο Λόττο κληρώνονται 6 αριθμοί από 49. Παίζουμε μία στήλη 6 αριθμών. Αν X = {το πλήθος των αριθμών που προβλέπουμε σωστά}, τότε να βρεθούν οι πιθανότητες (α) $P(X = 6)$ (β) $P(X = 5)$, (γ) $P(X = 4)$.
Λύση

Υπεργεωμετρική Κατανομή

Παράδειγμα 2 (Λόττο updated)
Στο Λόττο κληρώνονται 6 αριθμοί από 49. Παίζουμε μία στήλη 6 αριθμών. Αν X = {το πλήθος των αριθμών που προβλέπουμε σωστά}, τότε να βρεθούν οι πιθανότητες (α) P(X = 6) (β) P(X = 5), (γ) P(X = 4).
Λύση
Τους 49 αριθμούς μπορούμε να τους αντιληφθούμε ως ένα σάκο με βόλους από τους οποίους οι 43 είναι πράσινοι και οι 6 είναι κόκκινοι. Αν X = {πλήθος κόκκινων βόλων} τότε X ~ HG(49, 6, 6) και

$$P(X = k) = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}, \quad k = 0, 1, 2, 3, 4, 5, 6.$$

Η διωνυμική ως οριακή περίπτωση της υπεργεωμετρικής

Η μοναδική διαφορά στο πείραμα μεταξύ της διωνυμικής και της υπεργεωμετρικής είναι το γεγονός πως στην πρώτη περίπτωση υπάρχει επανάθεση ενώ στη δεύτερη όχι.

Είναι αντιληπτό πως όσο μεγαλύτερο είναι το πλήθος των διαθέσιμων επιλογών τόσο μικρότερη είναι η επιρροή της επανάθεσης στην πιθανότητα επιτυχίας μίας επιλογής.

Πράγματι, αν $p = K / N$ παραμένει σταθερό καθώς το $N \to \infty$, τότε $K = Np$, $N – K = N(1 – p)$ και αποδεικνύεται ότι

$$\lim_{N \to \infty} \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}.$$

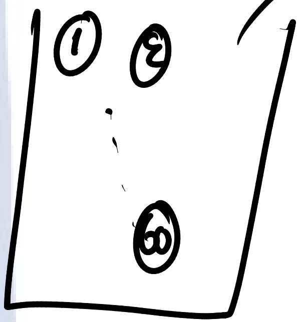
Σημείωση: Η απόδειξη μπορεί να βρεθεί εδώ:
https://math.stackexchange.com/questions/330553/proof-that-the-hypergeometric-distribution-with-large-n-approaches-the-binomia

$\frac{5!}{2!\cdot 3!}$

Ασκήσεις στην υπεργεωμετρική κατανομή

1. Σε μία κάλπη υπάρχουν 60 αριθμημένοι βόλοι, από 1 έως 60. Παίρνουμε 5 βόλους και έστω X = {το πλήθος των βόλων με αριθμό που διαιρείται με το 3}. Να βρεθεί η πιθανότητα P(X = 2), αν (α) η επιλογή γίνει με επανατοποθέτηση, (β) η επιλογή γίνει χωρίς επανατοποθέτηση.

Λύση



→ 5 βόλους $X = \{$βόλοι που διαιρούνται με το 3$\}$

(α) $P(X=2) = \binom{5}{2}\cdot\left(\frac{1}{3}\right)^2\cdot\left(1-\frac{1}{3}\right)^3 = 10\cdot\frac{1}{9}\cdot\frac{8}{27} = \frac{80}{243} = 0{,}329$

→ 5

Με επανατοποθέτηση: Σταθερό $p = \frac{20}{60} = \frac{1}{3}$.

$X \sim B(5, \frac{1}{3})$

Ασκήσεις στην υπεργεωμετρική κατανομή

1. Σε μία κάλπη υπάρχουν 60 αριθμημένοι βόλοι, από 1 έως 60. Παίρνουμε 5 βόλους και έστω X = {το πλήθος των βόλων με αριθμό που διαιρείται με το 3}. Να βρεθεί η πιθανότητα P(X = 2), αν (α) η επιλογή γίνει με επανατοποθέτηση, (β) η επιλογή γίνει χωρίς επανατοποθέτηση.

Λύση

(β) $X \sim HG(60, 20, 5)$

$$\boxed{\begin{array}{l} 40\ A \\ 20\ E \end{array}} \to O$$

$$P(X=2) = \frac{\binom{20}{2}\binom{40}{3}}{\binom{60}{5}} = \frac{\frac{20!}{2!\cdot 18!}\cdot\frac{40!}{3!\cdot 37!}}{\frac{60!}{5!\cdot 55!}} =$$

$$= \frac{190\cdot\frac{\overset{20}{\cancel{40}}\cdot\overset{13}{\cancel{39}}\cdot 38}{\cancel{3}\cdot\cancel{2}}}{\frac{60\cdot 59\cdot 58\cdot 57\cdot 56}{5\cdot 4\cdot 3\cdot 2}} = 0,344$$

Ασκήσεις στην υπεργεωμετρική κατανομή

2. Σε ένα δοχείο υπάρχουν 100 λαμπτήρες από τους οποίους οι 20 είναι ελαττωματικοί. Επιλέγουμε 5 λάμπες χωρίς επανατοποθέτηση και έστω X = {το πλήθος των ελαττωματικών λαμπτήρων}. Να βρεθεί η πιθανότητα $P(X = 2)$, (α) με ακρίβεια από την υπεργεωμετρική κατανομή,
(β) προσεγγιστικά από τη διωνυμική κατανομή.

Λύση

Ασκήσεις στην υπεργεωμετρική κατανομή

3. Στην αίθουσα υπάρχουν 100 φοιτητές από τους οποίους οι 70 έχουν αναπτύξει ανοσία στον Κορονοϊό είτε λόγω νόσησης είτε λόγω εμβολιασμού. Επιλέγονται 20 φοιτητές για να συμμετέχουν ως θεατές σε ένα θεατρικό έργο. Ποια είναι η πιθανότητα στους 20 φοιτητές οι 17 να έχουν ανοσία;

Λύση

A: Ανοσία, B = όχι ανοσία.

$$\boxed{\begin{array}{l} 70\,A \\ 30\,B \end{array}}$$

X = {πλήθος φοιτητών στους 20 με ανοσία}

$X \sim HG(100, 70, 20)$

$$P(X=17) = \frac{\binom{70}{17}\cdot\binom{30}{3}}{\binom{100}{20}}.$$

Κατανομή Poisson

Κατανομή Poisson

Έστω ότι κάποιες αφίξεις συμβαίνουν **εντελώς τυχαία** στο χρόνο και **ανεξάρτητα μεταξύ τους** με ρυθμό **λ αφίξεις / μονάδα χρόνου**. Έστω, επίσης

X = {πλήθος αφίξεων στο διάστημα (0, 1] } = 0, 1, 2, ….

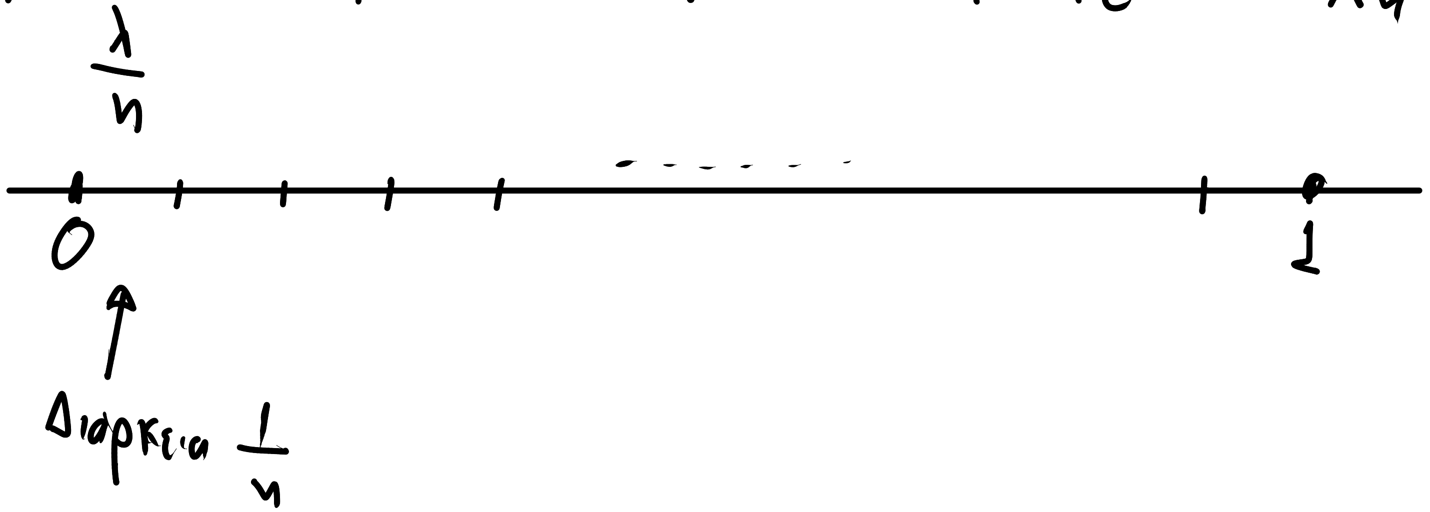
Αποδεικνύεται ότι:

$$\mathbf{P\left(X = k\right) = e^{-\lambda}\frac{\lambda^{k}}{k!},\ \ k = 0, 1, 2, ...}$$

Την κατανομή αυτή την ονομάζουμε **κατανομή Poisson** με παράμετρο λ και γράφουμε **X ~ Poisson(λ)** ή απλά **X ~ P(λ)**.

λ: μέσο πλήθος συμβάντων στη μονάδα του χρόνου.

Y = Συμβάντα στη μονάδα του χρόνου $= X_1 + X_2 + \dots + X_n \sim B\left(n, \frac{\lambda}{n}\right)$



$X_i = 1$ αν υπάρχει συμβάν στο i μέρος, $i = 1, 2, \dots n$. και $X_i = 0$ αν δεν συμβεί.

$X_i \sim \text{Bernoulli}\left(\frac{\lambda}{n}\right)$

$X \sim \text{Bernoulli}(p)$, $EX = p$.

$$Y \sim B\left(n, \frac{\lambda}{n}\right) \Rightarrow P(Y=k) = \binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1-\frac{\lambda}{n}\right)^{n-k}, \quad k=0,1,..,n$$

$$\lim_{n\to+\infty} \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \cdot \left(1-\frac{\lambda}{k}\right)^{n-k} = \lim_{n\to+\infty} \frac{n!}{k!\cdot(n-k)!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lim_{n\to+\infty} \frac{\overbrace{\cancel{n\cdot(n-1)\cdots(n-k+1)}}^{n^k+O(n^{k-1})}}{k!} \cdot \frac{\lambda^k}{\cancel{n^k}} \left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} \lim_{n\to+\infty} \left(1-\frac{\lambda}{n}\right)^{n-k}$$

Είναι $\lim_{n\to+\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1$ και $\lim_{n\to+\infty}\left(1+\frac{-\lambda}{n}\right)^n = e^{-\lambda}$, άρα $P(Y=k) = e^{-\lambda}\frac{\lambda^k}{k!}$

$X \sim Poisson(\lambda)$, $P(X=k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$

$k = 0, 1, 2, \ldots$

π.χ. $\lambda = 10$ πελάτες/ώρα $X \sim Poisson(10)$

$P(X=5) = e^{-10} \cdot \frac{10^5}{5!} = 0,038$. $P(X=10) = e^{-10} \frac{10^{10}}{10!} = 0,125$

$P(X=12) = e^{-10} \frac{10^{12}}{12!} = 0,095$.

Η γέννεση της κατανομής Poisson ως όριο δοκιμασιών Bernoulli

Απόδειξη (1/2)

Επιλέγουμε n ϵ N και διαμερίζουμε το διάστημα (0, 1] σε n υποδιαστήματα μήκους **$\delta = 1/n$** > 0:

$$(0, \delta], (\delta, 2\delta], (2\delta, 3\delta], \ldots (1 - \delta, 1],$$

Σε κάθε ένα διάστημα αναμένονται **$\lambda \cdot \delta = \lambda/n$** αφίξεις. Αν $X_i = 1$ όταν υπάρχει άφιξη στην i διαμέριση του (0, 1] και $X_i = 0$ όταν δεν υπάρχει, τότε $EX_i = \lambda/n$, και

$$\mathbf{X_i \sim Bernoulli(p_\delta)},\ i = 1, 2, \ldots, n,\ \text{με}\ \mathbf{p_\delta = \lambda/n}.$$

Για τις μεταβλητές X_i, $i = 1, 2, \ldots, n$, σημειώνουμε ότι:
(α) Είναι στοχαστικά ανεξάρτητες (τυχαίες αφίξεις)
(β) $X_1 + X_2 + \ldots + X_n$ = {σύνολο αφίξεων στο $(0, \delta] \cup (\delta, 2\delta] \cup (2\delta, 3\delta] \cup \ldots \cup (1 - \delta, 1]$ }
(γ) $E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n) = n \cdot \lambda \cdot \delta = \lambda$

Αν $Y_n = X_1 + X_2 + \ldots + X_n$ τότε $Y_n \sim B(n, p_\delta)$.

Σημείωση
Ο λόγος λ/n, για n αρκούντως μεγάλο είναι μικρότερο της μονάδας και μπορεί να θεωρηθεί πιθανότητα.

Η γέννεση της κατανομής Poisson ως όριο δοκιμασιών Bernoulli

Απόδειξη (2/2)

Καθώς $\delta = 1/n$, είναι $\lim_{n\to\infty}\delta = \lim_{n\to\infty}1/n = 0$ και η διαμέριση $(0, \delta], (\delta, 2\delta], (2\delta, 3\delta], \ldots (1 - \delta, 1]$ προσεγγίζει την σημειακή ανάλυση του διαστήματος (0, 1], επιτρέποντας με τον τρόπο αυτό την καταγραφή των αφίξεων σε οποιαδήποτε χρονική στιγμή του διαστήματος. Δηλαδή,

$$\mathbf{\lim_{n\to\infty} Y_n = \lim_{n\to\infty}(X_1 + X_2 + \ldots + X_n) = \{\text{σύνολο αφίξεων στο } (0, 1]\} = X}$$

Είναι

$$P(X = k) = \lim_{n\to\infty} P_{Y_n}(Y_n = k) = \lim_{n\to\infty} C(n, k)\ p^k_\delta\,(1 - p_\delta)^{n-k} = e^{-\lambda}\lambda^k / k!,\ k = 0, 1, 2, \ldots, (*)$$

δηλαδή, αποδείχθηκε ότι:

$$\mathbf{P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!},\ \ k = 0, 1, 2, \ldots}$$

(*) Η απόδειξη της τελευταίας ισότητας γίνεται στις επόμενες 4 διαφάνειες.

Απόδειξη του $\lim_{n\to\infty} C(n, k)\ p^k_\delta\ (1 – p_\delta)^{n – k} = e^{-\lambda}\ \lambda^k / k!$ (1/4)

Από τον ορισμό του ο αριθμός e είναι ίσος με

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$$

Ένα χρήσιμο όριο που προκύπτει από τον ορισμό είναι το

$$e^\lambda = \lim_{n \to \infty} \left(1 + \frac{\lambda}{n}\right)^n, \lambda \in \mathbb{R}.$$

Απόδειξη του $\lim_{n\to\infty} C(n, k)\ p^k_\delta\ (1 - p_\delta)^{n-k} = e^{-\lambda}\ \lambda^k / k!$ (2/4)

$$e^{\lambda} = \lim_{n\to\infty}\left(1 + \frac{\lambda}{n}\right)^n,\ \ \lambda \in \mathbb{R}.$$

Απόδειξη

Θέτουμε $y_n = \left(1 + \frac{\lambda}{n}\right)^n$, άρα $\ln y_n = n\ln\left(1+\frac{\lambda}{n}\right) = \dfrac{\ln\left(1+\frac{\lambda}{n}\right)}{\frac{1}{n}}$

Είναι $\lim\limits_{n\to\infty} \ln y_n = \lim\limits_{n\to\infty} \dfrac{\ln\left(1+\frac{\lambda}{n}\right)}{\frac{1}{n}} = \lim\limits_{n\to\infty} \dfrac{\frac{1}{1+\frac{\lambda}{n}}\left(-\frac{\lambda}{n^2}\right)}{-\frac{1}{n^2}} = \lambda.$

Άρα $\lim\limits_{n\to\infty}\left(1 + \frac{\lambda}{n}\right)^n = \lim\limits_{n\to\infty} y_n = \lim\limits_{n\to\infty} e^{\ln y_n} = e^{\lambda}.$

Απόδειξη του $\lim_{n\to\infty} C(n, k)\ p^k_\delta (1 - p_\delta)^{n-k} = e^{-\lambda} \lambda^k / k!$ (3/4)

Θεώρημα

Έστω $Y_n \sim B(n, p_n)$, $p_n = p(n)$ και έστω ότι υπάρχει $\lambda > 0$ τέτοιο ώστε $\lim_{n\to\infty} n \cdot p_n = \lambda$. Τότε

$$\lim_{n\to\infty} P_{Y_n}(Y_n = \kappa) = e^{-\lambda} \lambda^\kappa / \kappa!,\ \kappa = 0, 1, 2, \ldots$$

Απόδειξη

Αν $Y_n \sim B(n, p_n)$, τότε $P_{Y_n}(X = \kappa) = (n \text{ ανά } \kappa)\ p_n^\kappa (1 - p_n)^{n-\kappa}$. Καθώς, $\lim_{n\to\infty} n \cdot p_n = \lambda$, είναι $|n \cdot p_n - \lambda| < \varepsilon$ ή $|p_n - \lambda / n| < \varepsilon/n$, δηλαδή $\lim_{n\to\infty}(p_n - \lambda / n) = 0$. Είναι:

$$\begin{aligned}
\lim_{n\to\infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} &\simeq \lim_{n\to\infty} \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{n^k + O\left(n^{k-1}\right)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}
\end{aligned}$$

Απόδειξη του $\lim_{n\to\infty} C(n, k)\ p^k_\delta (1 - p_\delta)^{n-k} = e^{-\lambda} \lambda^k / k!$ (4/4)

Θεώρημα

Έστω $Y_n \sim B(n, p_n)$, $p_n = p(n)$ και έστω ότι υπάρχει $\lambda > 0$ τέτοιο ώστε $\lim_{n\to\infty} n \cdot p_n = \lambda$. Τότε

$$\lim_{n\to\infty} P_{Y_n}(Y_n = \kappa) = e^{-\lambda} \lambda^\kappa / \kappa!,\ \kappa = 0, 1, 2, \dots$$

Απόδειξη

Καθώς, $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$ και $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1$ (γιατί;), βρίσκουμε

$$\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{\lambda^k e^{-\lambda}}{k!}.$$

Κατανομή Poisson

Συνοψίζοντας:

X ~ Poisson(λ), αν για κάποιο $\lambda > 0$, $P(X = \kappa) = e^{-\lambda} \lambda^{\kappa} / \kappa!$, $\kappa = 0, 1, 2, \ldots$

Η κατανομή Poisson εφαρμόζεται σε φαινόμενα με διακριτή συχνότητα και σταθερή πιθανότητα να συμβούν στη μονάδα μέτρησης χρόνου ή χώρου.

Η παράμετρος λ συμβολίζει την αναμενόμενη τιμή της X στην μονάδα μέτρησης.

Κατανομή Poisson

Παραδείγματα τ.μ. που ακολουθούν την Poisson κατανομή:

- οι τηλεφωνικές κλήσεις που φτάνουν σε ένα σύστημα (X: πλήθος κλήσεων, λ: μέσο πλήθος κλήσεων).
- τα φωτόνια που φτάνουν σε ένα τηλεσκόπιο (X: πλήθος φωτονίων, λ: μέσο πλήθος φωτονίων που φτάνουν).
- ο αριθμός των μεταλλάξεων σε ένα κλώνο του DNA ανά μονάδα μήκους (X: πλήθος μεταλλάξεων, λ: μέσο πλήθος μεταλλάξεων που παρατηρούνται).
- Ο αριθμός των γκολ σε αγώνες ποδοσφαίρου (X: πλήθος γκολ, λ: μέσο πλήθος γκολ σε ανάλογους αγώνες).

Κατανομή Poisson

Άσκηση

Να αποδείξετε ότι αν X ~ Poisson(λ) τότε E(X) = λ.

Απόδειξη

$$X \sim Poisson(\lambda) \Longrightarrow P(X=k) = e^{-\lambda} \cdot \frac{\lambda^{k}}{k!}, \quad k=0,1,\ldots$$

$$E(X) = \sum_{k=0}^{+\infty} k \cdot P(X=k) = \sum_{k=0}^{+\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^{k}}{k!} = e^{-\lambda} \cdot \sum_{k=1}^{+\infty} \frac{\lambda^{k}}{(k-1)!} =$$

$$k-1=0, \quad k-1=t, \quad k=t+1$$

$$= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^{k+1}}{k!} = e^{-\lambda} \cdot \lambda \cdot \sum_{k=0}^{+\infty} \frac{\lambda^{k}}{k!} = e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \underline{\underline{\lambda}}.$$

Poisson: Γεωμετρικά χαρακτηριστικά

Y ~ Poisson(λ): $P(X = \kappa) = e^{-\lambda} \lambda^{\kappa} / \kappa!$, κ = 0, 1, 2,…

Αναμενόμενη τιμή: $EX = \lambda$.

Διακύμανση: $\mathrm{Var}X = \lambda$.

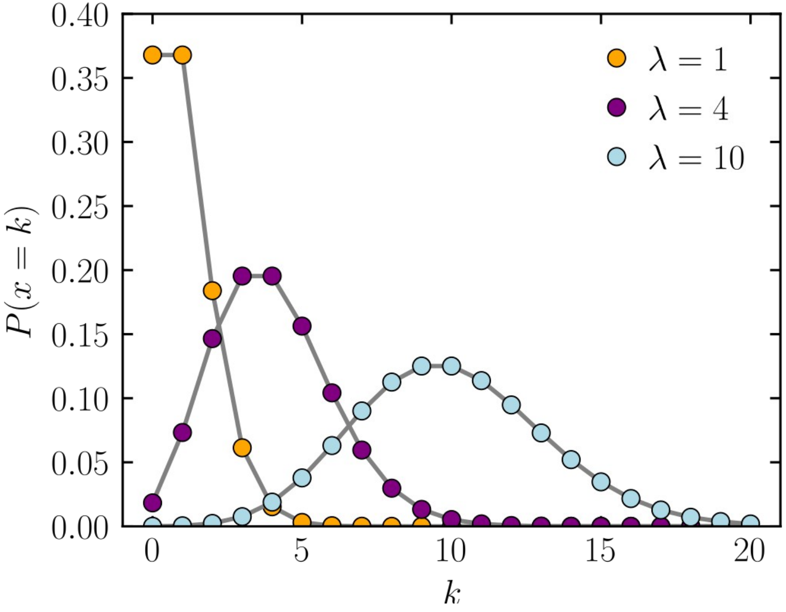
Τυπική απόκλιση: $\sigma = \sqrt{\lambda}$

Ασυμμετρία: $\mathrm{skew}(X) = \gamma = \dfrac{\mu_3}{\sigma^3} = \dfrac{1}{\sqrt{\lambda}}$

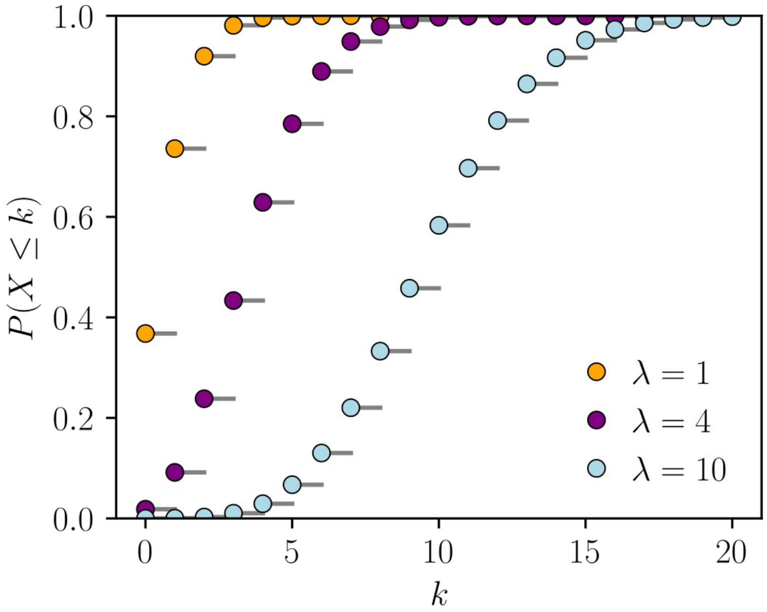
Κυρτότητα: $\mathrm{kurt}(X) = \alpha = \dfrac{\mu_4}{\sigma^4} = \dfrac{1}{\lambda}$

Σημείωση: Από τον τρόπο ορισμού της η Poisson έχει πάντα θετική ασυμμετρία (ουρά προς τα δεξιά)

Κατανομή Poisson

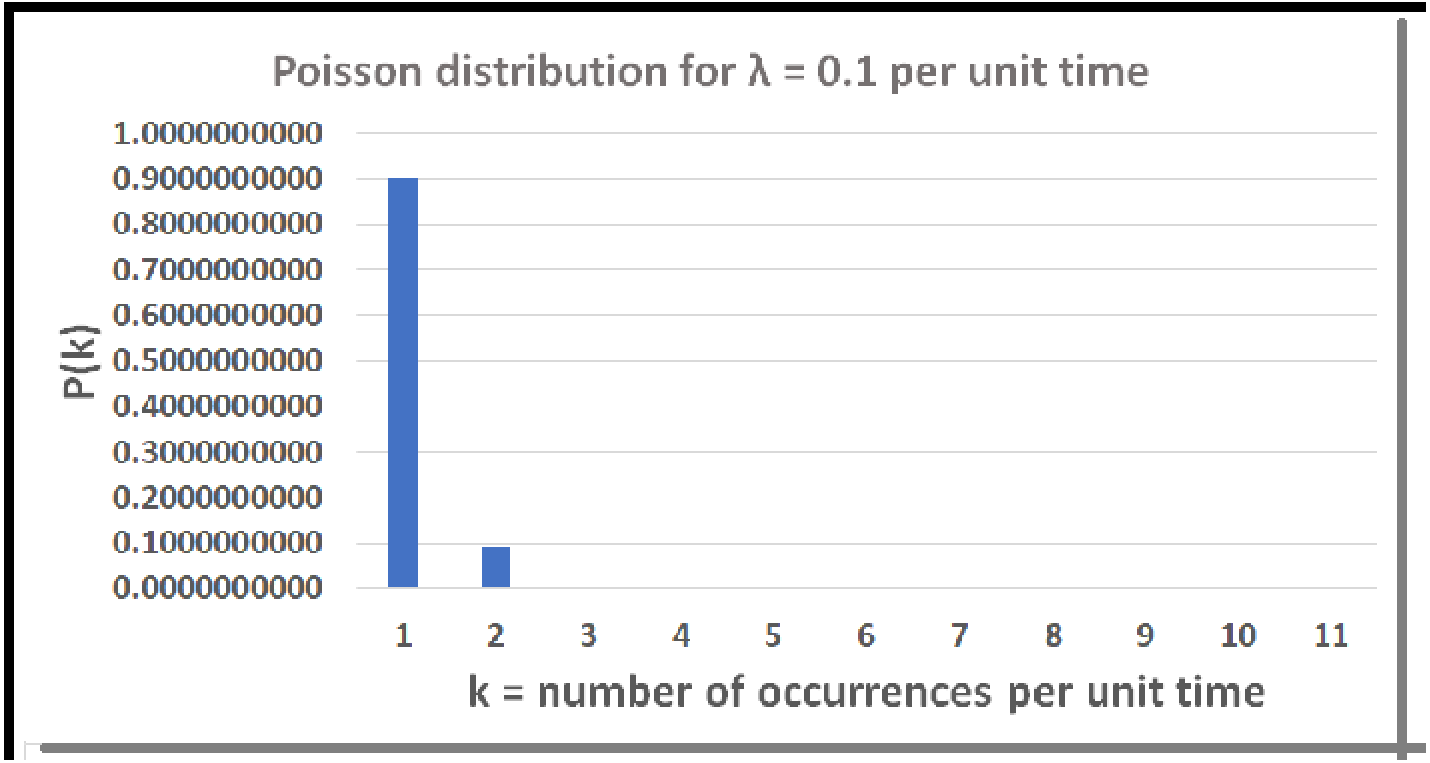


Συνάρτηση πιθανότητας $P(X = \kappa) = e^{-\lambda} \lambda^{\kappa} / \kappa!$, $\kappa = 0, 1, 2, \ldots$



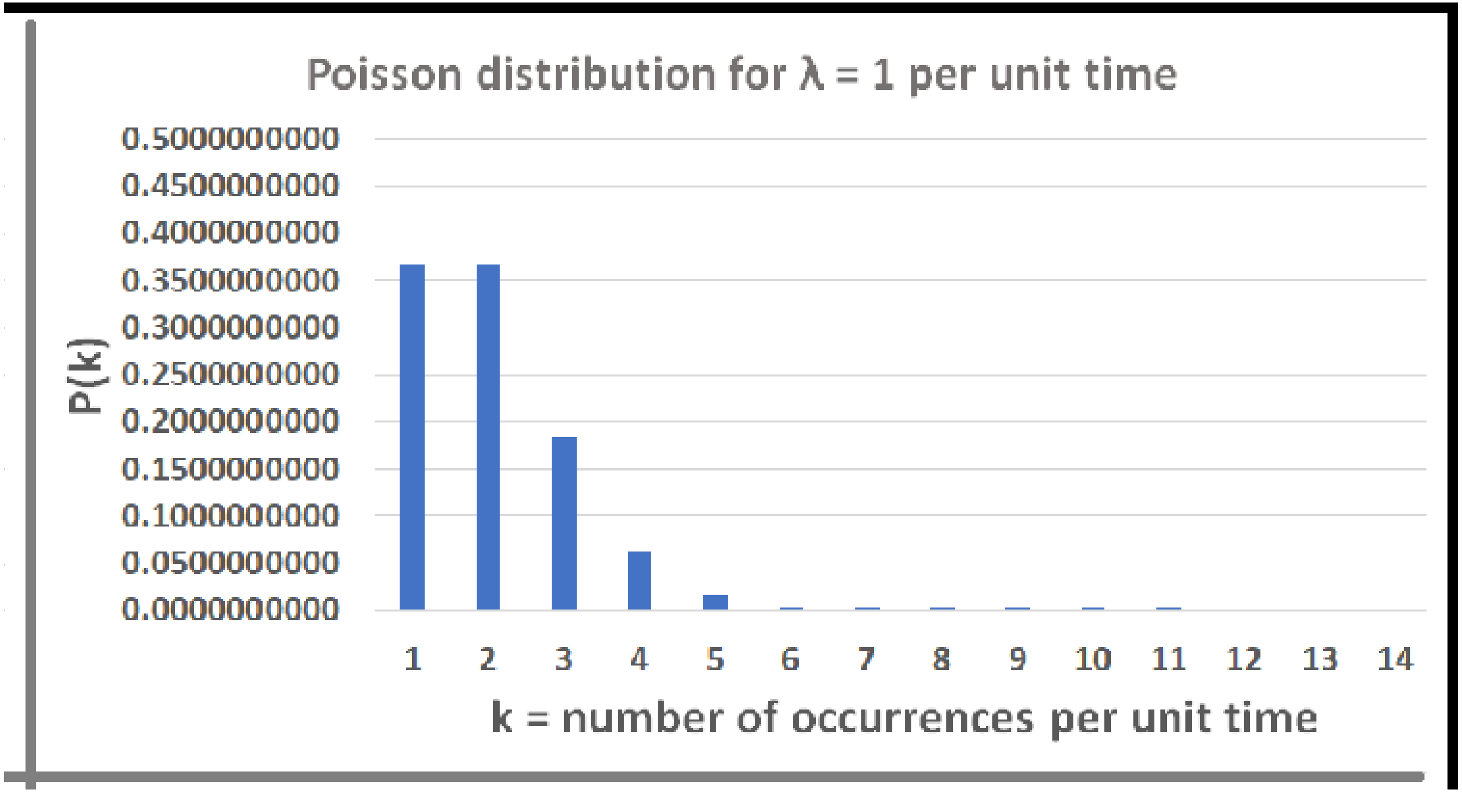
Συνάρτηση κατανομής $F(x) = P(X \leq x)$, $x \in R$

$\lambda = 0{,}1$: $P(X = \kappa) = e^{-0{,}1} \cdot 0{,}1^{\kappa} / \kappa!$, $\kappa = 0, 1, \ldots$



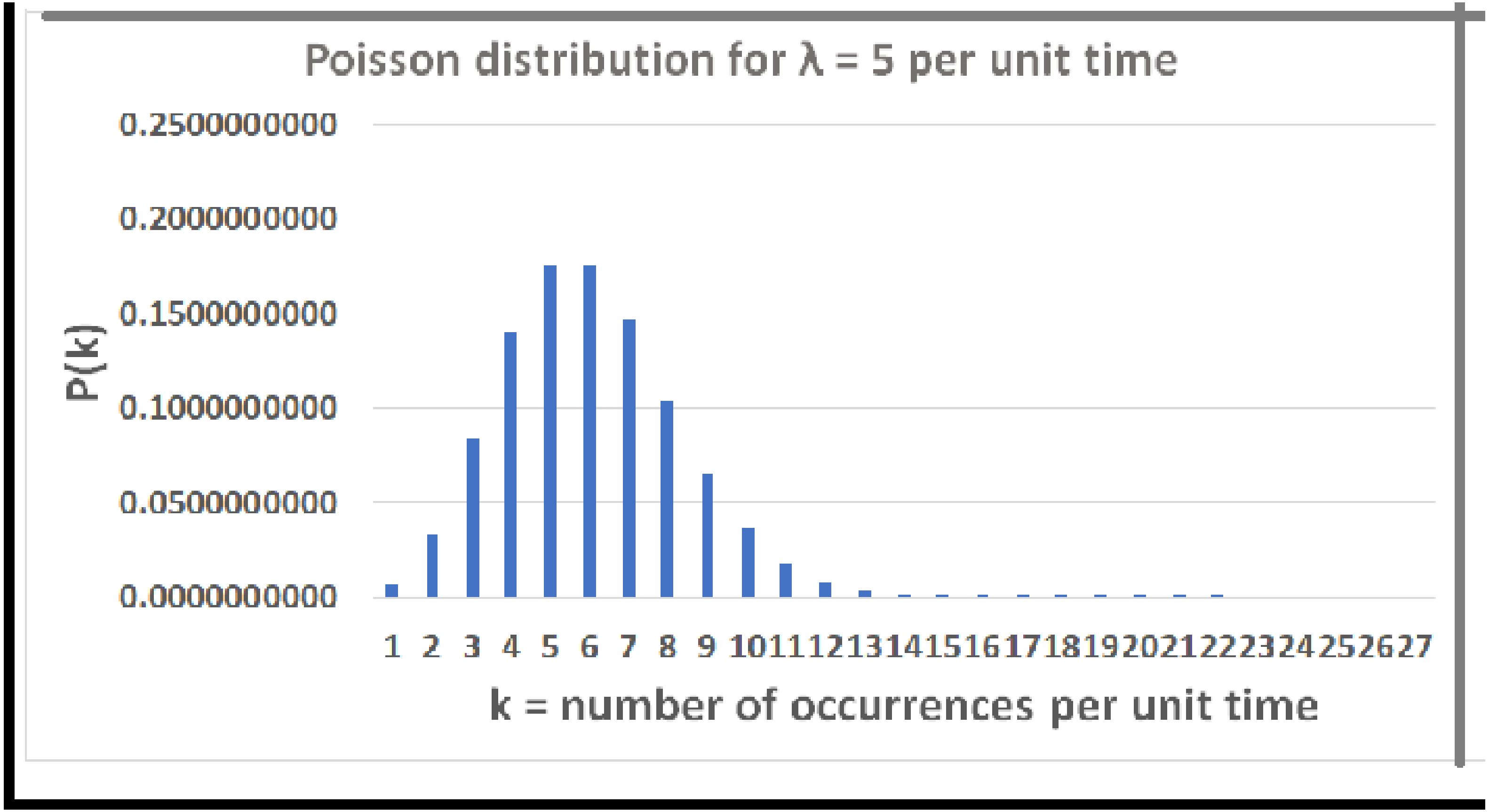
Πηγή: https://towardsdatascience.com/the-poisson-process-everything-you-need-to-know-322aa0ab9e9a

$\lambda = 1$: $P(X = \kappa) = e^{-1} / \kappa!$, $\kappa = 0, 1, \ldots$



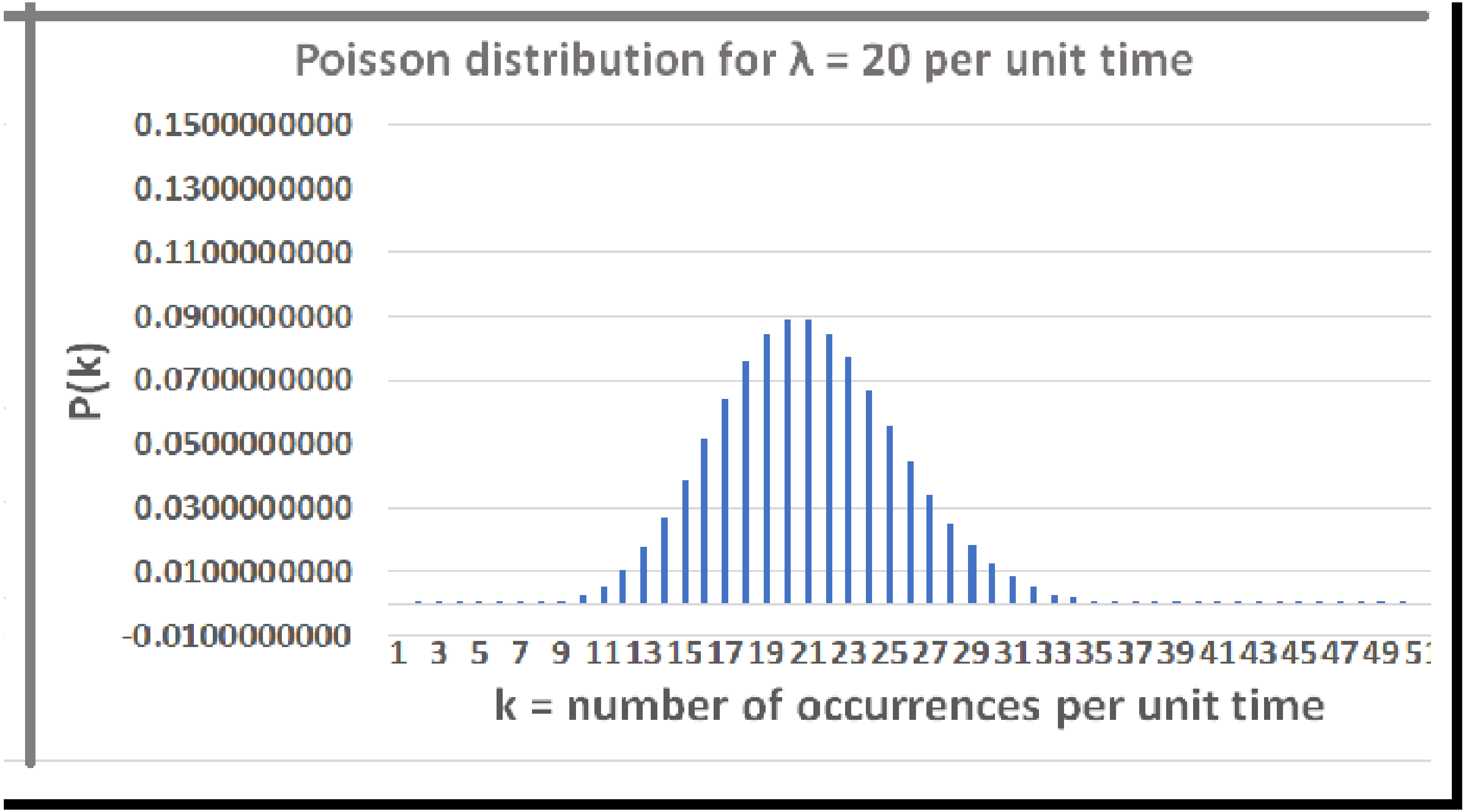
Πηγή: https://towardsdatascience.com/the-poisson-process-everything-you-need-to-know-322aa0ab9e9a

$\lambda = 5$: $P(X = \kappa) = e^{-5} \cdot 5^{\kappa} / \kappa!$, $\kappa = 0, 1, \ldots$



Πηγή: https://towardsdatascience.com/the-poisson-process-everything-you-need-to-know-322aa0ab9e9a

$\lambda = 20$: $P(X = \kappa) = e^{-20} \cdot 20^{\kappa} / \kappa!$, $\kappa = 0, 1, \ldots$



Πηγή: https://towardsdatascience.com/the-poisson-process-everything-you-need-to-know-322aa0ab9e9a

Παράδειγμα 1

Ο αριθμός των σωματιδίων που εκπέμπει μια ραδιενεργός πηγή ανά δευτερόλεπτο ακολουθεί την κατανομή Poisson με μέσο λ = 5. Να υπολογιστούν οι πιθανότητες: P(X = 0), P(X = 1), P(X = 2) και P(X ≥ 3).

Λύση

$X \sim Poisson(5)$.

$P(X=0) = e^{-5} \cdot \frac{5^0}{0!} = e^{-5} = 0{,}007$.

$P(X=1) = e^{-5} \frac{5^1}{1!} = 0{,}034$

$P(X=2) = e^{-5} \frac{5^2}{2!} = 0{,}084$

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

$= 0{,}875$.

Παράδειγμα 1

Ο αριθμός των σωματιδίων που εκπέμπει μια ραδιενεργός πηγή ανά δευτερόλεπτο ακολουθεί την κατανομή Poisson με μέσο $\lambda = 5$. Να υπολογιστούν οι πιθανότητες: $P(X = 0)$, $P(X = 1)$, $P(X = 2)$ και $P(X \geq 3)$.

Λύση

Είναι $X \sim \text{Poisson}(5)$ άρα $P(X = \kappa) = e^{-5} \cdot 5^{\kappa} / \kappa!$. Υπολογίζουμε:

- $P(X = 0) = e^{-5} \cdot 5^0 / 0! = e^{-5} = 0{,}0067 = 0{,}67\%$.
- $P(X = 1) = e^{-5} \cdot 5^1 / 1! = 5 \cdot e^{-5} = 0{,}0337 = 3{,}37\%$.
- $P(X = 2) = e^{-5} \cdot 5^2 / 2! = 25 \cdot e^{-5} / 2 = 0{,}0842 = 8{,}42\%$.
- $P(X \geq 3) = 1 – P(X < 3) = 1 – [P(X = 0) + P(X = 1) + P(X = 2)] = 0{,}8753 = 87{,}5\%$.

Παράδειγμα 2

Έστω ότι η τ.μ. X ακολουθεί την κατανομή Poisson με παράμετρο λ. Αν ισχύει $P(X \le 1) = 4 \cdot P(X = 2)$, να προσδιοριστεί η παράμετρος λ και να υπολογιστούν οι πιθανότητες $P(X < 3)$, $P(X \ge 1 \mid X < 3)$.

Λύση

$$P(X\le 1) = 4\cdot P(X=2) \Leftrightarrow P(X=0)+P(X=1)=4P(X=2) \Leftrightarrow$$

$$\Leftrightarrow \cancel{e^{-\lambda}}\cdot\frac{\lambda^0}{0!} + \cancel{e^{-\lambda}}\cdot\frac{\lambda^1}{1!} = 4\cdot\cancel{e^{-\lambda}}\cdot\frac{\lambda^2}{2!} \Leftrightarrow 1+\lambda = 2\lambda^2$$

$$\Leftrightarrow 2\lambda^2 - \lambda - 1 = 0, \quad \lambda_{1,2} = \frac{1\pm 3}{4} \begin{cases} 1 \Rightarrow \boxed{\lambda = 1} \\ -\frac{1}{2} \text{ απορρ.} \end{cases}$$

$$\Delta = (-1)^2 - 4\cdot 2(-1) = 9$$

Παράδειγμα 2

Έστω ότι η τ.μ. X ακολουθεί την κατανομή Poisson με παράμετρο λ. Αν ισχύει $P(X \leq 1) = 4 \cdot P(X = 2)$, να προσδιοριστεί η παράμετρος λ και να υπολογιστούν οι πιθανότητες $P(X < 3)$, $P(X \geq 1 \mid X < 3)$.

Λύση

Είναι $X \sim Poisson(\lambda)$ άρα $P(X = \kappa) = e^{-\lambda} \cdot \lambda^{\kappa} / \kappa!$. Τώρα, $P(X \leq 1) = 4 \cdot P(X = 2) \leftrightarrow$

$\leftrightarrow P(X = 0) + P(X = 1) = 4 \cdot P(X = 2) \leftrightarrow e^{-\lambda} + e^{-\lambda} \cdot \lambda = 4\, e^{-\lambda} \cdot \lambda^2 / 2 \leftrightarrow 2\lambda^2 - \lambda - 1 = 0 \leftrightarrow \lambda = -1/2$ ή $\lambda = 1$.

Καθώς, το λ αποτελεί ρυθμό εμφάνισης γεγονότων, πρέπει να είναι θετικός αριθμός. Άρα $\lambda = 1$.

$P(X < 3) = P(X=0) + P(X=1) + P(X=2) = e^{-1} \cdot 1^0 / 0! + e^{-1} \cdot 1^1 / 1! + e^{-1} \cdot 1^2 / 2! = e^{-1} \cdot 5 / 2 = 0{,}9196 \approx 92\%$

$P(X \geq 1 \mid X < 3) = P(X \geq 1, X < 3) / P(X < 3) = P(X = 1$ ή $X = 2) / P(X < 3)$

$= [P(X = 1) + P(X = 2)] / P(X < 3) = 3/5 = 0{,}6 = 60\%$.

Πιθανογεννήτρια συνάρτηση Poisson

Παράδειγμα

Αν X ~ Poisson(λ) τότε $P(X = \kappa) = \lambda^{\kappa} \cdot e^{-\lambda}/ \kappa!$, $\kappa = 0, 1, 2, \ldots$, και

$$G_X(z) = E(z^X) = \Sigma_{\kappa = 0, 1, \ldots} P(X = \kappa) \cdot z^{\kappa}$$

$$= \Sigma_{\kappa = 0, 1, \ldots} \lambda^{\kappa} \cdot e^{-\lambda}/ \kappa! \cdot z^{\kappa}$$

$$= e^{-\lambda} \cdot \Sigma_{\kappa = 0, 1, \ldots} (z\lambda)^{\kappa}/ \kappa!$$

$$= e^{-\lambda} \cdot e^{z\lambda}$$

$$= e^{(z - 1) \lambda}.$$

Άρα, $G_X(z) = e^{(z - 1) \lambda}$ για **κάθε z є C**, ειδικότερα, $G_X(x) = e^{(x - 1) \lambda}$, για **κάθε x є R**.

Ροπογεννήτρια συνάρτηση Poisson

Άσκηση

Να δείξετε ότι αν X ~ Poisson(λ) τότε $M_X(t) = E(e^{tX}) = e^{\lambda(e^t - 1)}$

Απόδειξη

Άθροισμα ανεξάρτητων Poisson μεταβλητών

Πρόταση

Το άθροισμα δύο **ανεξάρτητων** τυχαίων μεταβλητών $X_1 \sim \text{Poisson}(\lambda_1)$ και $X_2 \sim \text{Poisson}(\lambda_2)$ είναι ομοίως κατανομή Poisson με παράμετρο $\lambda_1 + \lambda_2$, δηλαδή, αν $X \sim X_1 + X_2$, και X_1, X_2 ανεξάρτητες τυχαίες μεταβλητές τότε $X \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Απόδειξη

Αρκεί να δείξουμε ότι $P(X = z) = e^{-\lambda}\cdot\lambda^z/z!$, για κάθε $z \in N$ και $\lambda = \lambda_1 + \lambda_2$. Πράγματι, είναι

$$
\begin{aligned}
P(X = z) &= \Sigma_{j = 0, 1, \ldots, z} P(X_1 = j, X_2 = z - j) \\
&= \Sigma_{j = 0, 1, \ldots, z} P(X_1 = j) \cdot P(X_2 = z - j) \qquad \text{(οι } X_1, X_2\text{, είναι ανεξάρτητες τ.μ.)} \\
&= \Sigma_{j = 0, 1, \ldots, z} e^{-\lambda 1}\cdot\lambda_1^{\,j}/j! \cdot e^{-\lambda 2}\cdot\lambda_2^{\,z - j}/(z - j)! \\
&= \Sigma_{j = 0, 1, \ldots, z} e^{-(\lambda 1+\lambda 2)}\cdot\lambda_1^{\,j} \cdot \lambda_2^{\,z - j}/[j!\cdot(z - j)!] \\
&= e^{-(\lambda 1+\lambda 2)}\cdot\Sigma_{j = 0, 1, \ldots, z}(z \text{ ανά } j) \cdot \lambda_1^{\,j} \cdot \lambda_2^{\,z - j}/z! \\
&= e^{-(\lambda 1+\lambda 2)}\cdot(\lambda_1 + \lambda_2)^z \\
&= e^{-\lambda}\cdot\lambda^z/z!
\end{aligned}
$$

Άθροισμα ανεξάρτητων Poisson μεταβλητών

Πρόταση
Το άθροισμα δύο **ανεξάρτητων** τυχαίων μεταβλητών $X_1 \sim \text{Poisson}(\lambda_1)$ και $X_2 \sim \text{Poisson}(\lambda_2)$ είναι ομοίως κατανομή Poisson με παράμετρο $\lambda_1 + \lambda_2$, δηλαδή, αν $X \sim X_1 + X_2$, και X_1, X_2 ανεξάρτητες τυχαίες μεταβλητές τότε $X \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Παρατηρήσεις
1. Η υπόθεση της (στοχαστικής) ανεξαρτησίας των X_1, X_2 είναι ουσιαστική. Αν οι X_1, X_2 δεν είναι ανεξάρτητες τότε $\text{Cov}(X_1, X_2) > 0$ και κατά συνέπεια θα είναι

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + 2\text{Cov}(X_1, X_2) + \text{Var}(X_2) > \text{Var}(X_1) + \text{Var}(X_2) = E(X_1 + X_2),$$

δηλαδή, η τ.μ. $X_1 + X_2$ δεν μπορεί να ακολουθεί την κατανομή Poisson στην οποία $\mu = \sigma^2$.

2. Η παραπάνω πρόταση δεν αποκλείει το ενδεχόμενο το άθροισμα δύο **εξαρτημένων** Poisson μεταβλητών να είναι Poisson μεταβλητή. Ένα παράδειγμα αναφέρεται στη δημοσίευση
[Jacod, J. (1975). Two Dependent Poisson Processes Whose Sum Is Still a Poisson Process. Journal of Applied Probability, 12(1), 170–172. https://doi.org/10.2307/3212423]

Ασκήσεις στην κατανομή Poisson

1. (α) Αν $X \sim \text{Poisson}(2)$ να βρεθεί η $P(X = 2)$. (β) Αν $X \sim \text{Poisson}(0,5)$ να βρεθεί η $P(X = 1)$

Λύση

Ασκήσεις στην κατανομή Poisson

2. Οι αφίξεις οχημάτων σε ένα πρατήριο καυσίμων αποτελούν διαδικασία Poisson με αναμενόμενο πλήθος 2 οχήματα κάθε 1 λεπτό. Το 80% των οχημάτων είναι Ι.Χ. αυτοκίνητα και το 20% φορτηγά.
α) Ποια η πιθανότητα να μην εμφανιστεί κανένα όχημα στη διάρκεια ενός πεντάλεπτου;
β) Έστω ότι έρχονται 10 οχήματα στο πρατήριο. Ποια η πιθανότητα από τα 10 οχήματα τα 4 να είναι φορτηγά;
γ) Ποια η πιθανότητα σε ένα πεντάλεπτο να έχουμε ακριβώς 4 Ι.Χ. και 2 φορτηγά;
δ) Δεδομένου ότι δεν έχει εμφανιστεί φορτηγό τα τελευταία 5 λεπτά, υπολογίστε την πιθανότητα ότι δεν θα εμφανιστεί κανένα φορτηγό και στα επόμενα 5 λεπτά.

Λύση

$\lambda = 2$ οχημ./min

(α) $X = \{$πλήθος οχημάτων στα 5min$\}$.

$X \sim Poisson(10)$.

$$P(X=0) = e^{-10} \cdot \frac{10^0}{0!} = 0{,}00004$$

(β) $Y = \{$πλήθος φορτηγών στα 10 οχήματα$\}$; $P(Y=4)$

$Y \sim B(10, 0.2) \Rightarrow P(Y=4) = \binom{10}{4} \cdot 0.2^4 \cdot 0.8^6 = 0{,}088$

$\binom{10}{4} = \frac{10!}{4! \cdot 6!} = \frac{10 \cdot \cancel{9}^{3} \cdot 8 \cdot 7}{4 \cdot \cancel{3} \cdot \cancel{2}} = 210$

(γ) $P(X=6) = e^{-10} \cdot \frac{10^6}{6!} = 0{,}063.$

$Z = \{$φορτηγά στα 6 οχήματα$\} \sim B(6, 0.2)$, $P(Z=2) = \binom{6}{2} \cdot 0.2^2 \cdot 0.8^4 = 0{,}246.$

$\binom{6}{2} = \frac{6!}{2! \cdot 4!} = 15.$

$P(X=6, Z=2) = P(Z=2 \mid X=6) \cdot P(X=6) =$

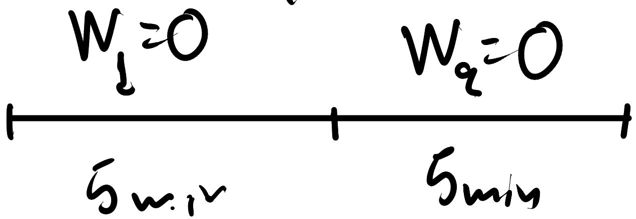
$= 0{,}246 \cdot 0{,}063 = 0{,}015.$

$P(AB) = P(B \mid A) \cdot P(A).$

(δ) $W = \{\text{φορτηγά ανά } 5min\} \sim Poisson(2)$.

10 οχήματα / 5min
Το 20% είναι φορτηγ
άρα $\lambda = 0{,}2 \cdot 10 = 2$ φ/5m

$$P(W=0) = e^{-2} \cdot \frac{2^0}{0!} = 0{,}135$$



$$P(W_2=0 \mid W_1=0) = \frac{P(W_2=0, W_1=0)}{P(W_1=0)} =$$

$$= \frac{P(W_2=0)\cancel{P(W_1=0)}}{\cancel{P(W_1=0)}} = P(W_2=0)$$

Ασκήσεις στην κατανομή Poisson

3. Το πλήθος σωματιδίων που εκπέμπονται από μία ραδιενεργή πηγή είναι τυχαία μεταβλητή που ακολουθεί την κατανομή Poisson. Αν η πιθανότητα εκπομπής κανενός σωματιδίου στη μονάδα του χρόνου είναι 1/3 τότε να βρεθεί η πιθανότητα εκπομπής 2 ή περισσοτέρων σωματιδίων.

Λύση

$X \sim Poisson(\lambda)$.

$$P(X=0)=\frac{1}{3} \Leftrightarrow e^{-\lambda}\cdot\frac{\lambda^0}{0!}=\frac{1}{3} \Leftrightarrow e^{-\lambda}=\frac{1}{3} \Leftrightarrow -\lambda=\ln\frac{1}{3}$$

$$\Leftrightarrow \lambda=\ln 3$$

$$P(X\geq 2)=1-P(X=0)-P(X=1)=1-e^{-\lambda}\cdot\frac{\lambda^0}{0!}-e^{-\lambda}\cdot\frac{\lambda^1}{1!}=1-\frac{1}{3}-\frac{1}{3}\ln 3$$

$$=\frac{2}{3}-\frac{\ln 3}{3}.$$

Ασκήσεις στην κατανομή Poisson

4. Η πιθανότητα ένα άτομο να παρουσιάσει σοβαρές παρενέργειες σε ένα εμβόλιο είναι 0,002. Εάν κάνουν το εμβόλιο 2000 άτομα, ποια είναι η πιθανότητα να έχουν σοβαρές παρενέργειες
(α) ακριβώς 2 άτομα (β) τουλάχιστον 3 άτομα.

Λύση

Ασκήσεις στην κατανομή Poisson

5. Γνωρίζουμε ότι σε μία περιοχή υπάρχουν 4 φωλιές πουλιών ανά στρέμμα. Να βρεθεί η πιθανότητα σε ένα χωράφι 5 στρεμμάτων να υπάρχουν 8 φωλιές πουλιών.

Λύση

4 φωλιές/στρέμμα. $X = \{$φωλιές στα 5 στρέμματα$\} \sim \text{Poisson}(20)$

$$P(X=8) = e^{-20} \cdot \frac{20^8}{8!} = 0{,}001.$$

Ασκήσεις στην κατανομή Poisson

6. Σε ένα βιβλίο 400 σελίδων υπάρχουν 200 τυπογραφικά λάθη.
(α) Ποιο είναι το μέσο πλήθος λαθών ανά σελίδα;
(β) Ποια είναι η πιθανότητα σε μία σελίδα να μην υπάρχουν λάθη;
(γ) Ποιο είναι το πλήθος σελίδων του βιβλίου που περιμένουμε να μην έχει τυπογραφικά λάθη;

Λύση

(α) $\frac{200}{400} = 0,5$ λάθη / σελίδα

(β) $X = \{$πλήθος λαθών σε μία σελίδα$\} \sim Poisson(0.5)$

$$P(X=0) = e^{-0,5} \cdot \frac{0,5^0}{0!} = 0,607$$

(γ) $Y = \{$σελίδες χωρίς λάθη$\} \sim B(400, 0.607) \Rightarrow E(Y) = 400 \cdot 0,607 = 242,8$

$X \sim Poisson(\lambda)$, $Y \sim Poisson(\mu)$, X, Y ανεξ. $\Rightarrow X+Y \sim Poisson(\lambda+\mu)$.

Ασκήσεις στην κατανομή Poisson

7. Μία φαρμακευτική αποθήκη εξυπηρετεί 10 φαρμακεία στο νομό της Ξάνθης. Το μέσο πλήθος των συσκευασιών ενός φαρμάκου που πωλείται από κάθε ένα από τα φαρμακεία δίνεται στον παρακάτω πίνακα:

Φαρμακείο	1	2	3	4	5	6	7	8	9	10
Πλήθος	2,1	3,5	7	1	0,4	5	2,6	1	3	4,4

Οι πωλήσεις κάθε ενός φαρμακείου είναι ανεξάρτητες. Ποια είναι η πιθανότητα, μία ημέρα να ζητηθούν από την αποθήκη 40 συσκευασίες από το φάρμακο;

Λύση

$$Y = X_1 + X_2 + \dots + X_{10}, \quad P(Y=40) = e^{-30}\frac{30^{40}}{40!} = 0{,}014.$$

$$X_i \sim Poisson(\lambda_i) \Rightarrow Y \sim Poisson(\lambda_1+\lambda_2+\dots+\lambda_{10}) = Poisson(30)$$

Γεωμετρική κατανομή

Γεωμετρική κατανομή (εισαγωγή)

Άσκηση 1: Καταμέτρηση <u>δοκιμασιών</u> μέχρι την πρώτη επιτυχία
Ένα πείραμα έχει δύο πιθανές εκδοχές, επιτυχία με πιθανότητα p και αποτυχία με πιθανότητα 1 – p.
Ορίζουμε X = {το πλήθος των **δοκιμασιών** μέχρι την πρώτη επιτυχία}.
(α) Να βρείτε τις πιθανές τιμές της X.
(β) Να βρείτε τις πιθανότητες P(X = 1), P(X = 2), P(X = 3).
(γ) Ποιος είναι ο γενικός τύπος για την P(X = n);

Λύση

(α) $X = 1, 2, \ldots$ (β) $P(X=1) = p$, $P(X=2) = (1-p)\cdot p$

$$P(X=3) = (1-p)^2 \cdot p$$

(γ) $P(X=n) = (1-p)^{n-1} \cdot p$

Γεωμετρική κατανομή (εισαγωγή)

Άσκηση 2: Καταμέτρηση <u>αποτυχιών</u> μέχρι την πρώτη επιτυχία
Ένα πείραμα έχει δύο πιθανές εκδοχές, επιτυχία με πιθανότητα p και αποτυχία με πιθανότητα 1 – p.
Ορίζουμε Y = {το πλήθος των **αποτυχιών** μέχρι την πρώτη επιτυχία}.
(α) Να βρείτε τις πιθανές τιμές της Y.
(β) Να βρείτε τις πιθανότητες P(Y = 0), P(Y = 1), P(Y = 2).
(γ) Ποιος είναι ο γενικός τύπος για την P(Y = n);

Λύση

(α) $Y = 0, 1, 2, \ldots$

$P(Y=0) = p$

$P(Y=1) = (1-p)\cdot p$

$P(Y=2) = (1-p)^2\cdot p$

(γ) $P(Y=n) = (1-p)^n\cdot p$

Γεωμετρική κατανομή (2 εκδοχές)

Με τον όρο “**γεωμετρική κατανομή**” περιγράφονται οι εξής δύο διακριτές κατανομές πιθανοτήτων:

- Η κατανομή πιθανότητας του πλήθους X є N* των δοκιμών Bernoulli(p) που απαιτούνται για να επιτευχθεί η πρώτη επιτυχία. Εάν η πιθανότητα επιτυχίας σε κάθε δοκιμή είναι p, τότε η πιθανότητα ότι η n – οστή δοκιμή είναι η πρώτη επιτυχία είναι

$$\mathbf{P(X = n) = (1 - p)^{n-1} \cdot p}, \; n = 1, 2, \ldots \; (E(X) = 1/p, \; Var(X) = (1 - p)/p^2) \qquad X \sim G_T(p)$$

- Η κατανομή πιθανότητας του πλήθους Y = X − 1 (є N) αποτυχιών πριν από την πρώτη επιτυχία. Στην περίπτωση αυτή η συνάρτηση πιθανότητας της Y είναι

$$\mathbf{P(Y = n) = P(X = n + 1) = (1 - p)^{n} \cdot p}, \; n = 0, 1, \ldots \; (E(Y) = (1 - p)/p, \; Var(Y) = (1 - p)/p^2) \qquad Y \sim G_F(p)$$

Και στις δύο περιπτώσεις γράφουμε **X ~ Geometric(p)** ή **Y ~ Geometric(p)** αντίστοιχα και θα διευκρινίζεται από το πλαίσιο αναφοράς το ακριβές είδος της κατανομής.

Σημείωση: Η προέλευση της ονομασίας είναι φανερή, καθώς και στις δύο περιπτώσεις η ακολουθία των πιθανοτήτων αποτελεί γεωμετρική πρόοδο.

$$X \sim G_T(p) \Rightarrow P(X=n) = (1-p)^{n-1} \cdot p$$

$X = 1, 2, \ldots$

$$E(X) = \sum_{n=1}^{+\infty} n \cdot P(X=n) = \sum_{n=1}^{+\infty} n \cdot (1-p)^{n-1} \cdot p = -p \sum_{n=1}^{+\infty} \left[(1-p)^n\right]'$$

$$= -p \left(\sum_{n=0}^{+\infty} (1-p)^n\right)' = -p \left(\frac{1}{1-(1-p)}\right)' = -p \left(\frac{1}{p}\right)' =$$

$$= -p \left(-\frac{1}{p^2}\right) = \frac{1}{p}.$$

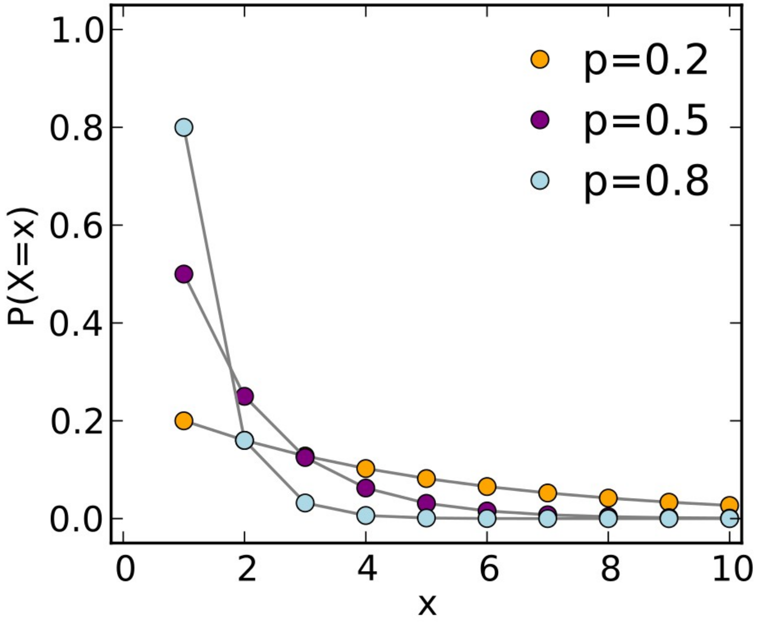
$$\sum_{n=0}^{+\infty} a_n x^n = f(x)$$

$$\left(a_0 + a_1 x + a_2 x^2 + \cdots\right)' = f'(x)$$

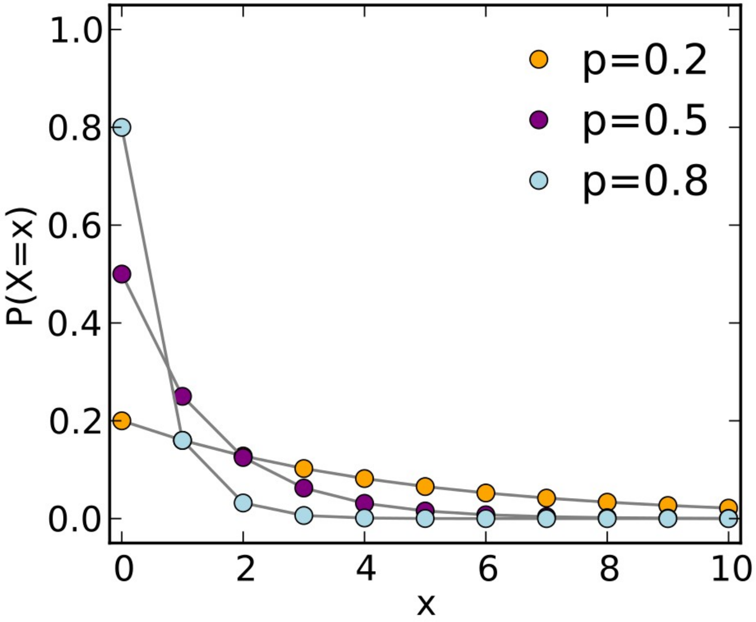
$$1 \cdot a_1 + 2 a_2 x + \cdots = f'(x)$$

$$\sum_{n=1}^{+\infty} n \cdot a_n \cdot x^{n-1}$$

Γεωμετρική κατανομή (2 εκδοχές)



Συνάρτηση πιθανότητας $P(X = n) = (1 - p)^{n-1} \cdot p$, $n = 0, 1, \dots$



Συνάρτηση πιθανότητας $P(X = n) = (1 - p)^{n} \cdot p$, $n = 0, 1, \dots$

Γεωμετρική: Γεωμετρικά χαρακτηριστικά

$Y \sim G(p)$: **$P(Y = n) = P(X = n + 1) = (1 - p)^n \cdot p,\ n = 0, 1, \ldots$**

Αναμενόμενη τιμή: $E\cancel{X}Y = (1 - p) / p$.

Διακύμανση: $Var\cancel{X}Y = (1 - p) / p^2$.

Τυπική απόκλιση: $\sigma = \sqrt{1 - p} / p$

Ασυμμετρία: $skew(\cancel{X}Y) = \gamma = \dfrac{\mu_3}{\sigma^3} = \dfrac{2 - p}{\sqrt{1 - p}}$

Κυρτότητα: $kurt(\cancel{X}Y) = \alpha = \dfrac{\mu_4}{\sigma^4} = \dfrac{p^2 - 9p + 9}{1 - p}$

Πιθανογεννήτρια συνάρτηση γεωμετρικής

Παράδειγμα

Αν X ~ Geometric(p) τότε $P(X = \kappa) = (1 - p)^\kappa \cdot p$, $\kappa = 0, 1, 2, \ldots$, και

$G_X(z) = E(z^X) = \Sigma_{\kappa = 0, 1, \ldots} (1 - p)^\kappa \cdot p \cdot z^\kappa$

$\qquad = p \cdot \Sigma_{\kappa = 0, 1, \ldots} [z(1 - p)]^\kappa$

$\qquad = p / [1 - z(1 - p)]$

Άρα, $G_X(z) = p / [1 - z(1 - p)]$ **για κάθε $|z(1 - p)| < 1$**, δηλαδή **$|z| < 1 / (1 - p)$**.

Σημείωση

Άθροισμα γεωμετρικής προόδου $\Sigma_{\kappa = 0, 1, \ldots} \alpha^\kappa = 1 + \alpha + \alpha^2 + \ldots = 1/(1 - \alpha)$ για $|\alpha| < 1$.

Ασκήσεις στη γεωμετρική κατανομή

1. Σε ένα τυχερό παιχνίδι η πιθανότητα επιτυχίας είναι 40%. Ποια είναι η πιθανότητα να κερδίζεις συνεχώς μέχρι το πέμπτο παιχνίδι; Ποιο είναι το αναμενόμενο πλήθος παιχνιδιών που θα κερδίζει ο παίκτης μέχρι να αποτύχει;

Λύση

(α) $P(\text{νίκη μέχρι το } 5^{o} \text{ παιχνίδι}) = 0{,}4 \cdot 0{,}4 \cdot 0{,}4 \cdot 0{,}4 \cdot 0{,}4 = 0{,}4^5 =$

$= 0{,}01$

(β) $Y = \{\text{επιτυχίες μέχρι την } 1^{η} \text{ αποτυχία}\} \sim G_F(0{,}6)$

$$E(Y) = \frac{1-p}{p} = \frac{0{,}4}{0{,}6} = 0{,}67.$$

Ασκήσεις στη γεωμετρική κατανομή

2. Ένας παίκτης ρίχνει βελάκια στο στόχο. Γνωρίζουμε ότι η πιθανότητα να πετύχει το κέντρο είναι 0,17. Ποια είναι η πιθανότητα να χρειαστεί 8 βολές μέχρι να χτυπήσει κέντρο; Ποιο είναι το αναμενόμενο πλήθος βολών που θα κάνει ο παίκτης μέχρι να πετύχει το κέντρο;

Λύση

(α) $P(\{8 \text{ βολές μέχρι επιτυχία}\}) = (1-0,17)^7 \cdot 0,17 = 0,046$

(β) $X = \{\text{πλήθος βολών μέχρι την επιτυχία}\} \sim G_T(0,17)$.

$E(X) = \frac{1}{p} = \frac{1}{0,17} = 5,88$ βολές. $\quad\Big|\quad E(Y) = \frac{1-p}{p} = \frac{1-0,17}{0,17}$.

$Y = \{\text{πλήθος βολών εκτός κέντρου μέχρι την επιτυχία}\} \sim G_F(0,17)$

Αρνητική διωνυμική κατανομή

Αρνητική διωνυμική κατανομή

Άσκηση

Αν X = {πλήθος ρίψεων ζαριού μέχρι να έρθουν 3 εξάρια} να βρεθούν:

(α) Οι πιθανές τιμές της X

(β) Η πιθανότητα P(X = 5)

Λύση

(α) $X = 3, 4, \ldots$

$\binom{4}{2}$

$\underline{6}\ \underline{6}\ \underline{\text{όχι } 6}\ \underline{\text{όχι } 6}\ \underline{6}\ : \frac{1}{6}\cdot\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}$

$P(X=5) = \binom{4}{2}\left(\frac{1}{6}\right)^3\cdot\left(\frac{5}{6}\right)^2$

Αρνητική διωνυμική κατανομή

Έστω ότι ένα πείραμα επαναλαμβάνεται, έχει δύο πιθανά αποτελέσματα “Επιτυχία” και “Αποτυχία” και η πιθανότητα επιτυχίας είναι σταθερή και ίση με p.

Αν η τ.μ. X εκφράζει το πλήθος επαναλήψεων που θα απαιτηθούν μέχρι να συμβούν r επιτυχίες, τότε X = r, r+1, r+2, … και η τ.μ. X έχει συνάρτηση πιθανότητας

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

και λέμε ότι η X ακολουθεί την αρνητική διωνυμική κατανομή NB(r, p) (X ~ NB(r, p)).

$n-1$

p

Σημείωση: Αν X ~ NB(1, p) τότε η X εκφράζει το πλήθος προσπαθειών έως την 1η επιτυχία που είναι η γεωμετρική κατανομή (1η εκδοχή: $P(X = n) = p(1 - p)^n$)

Αρνητική διωνυμική: Γεωμετρικά χαρακτηριστικά

Y ~ NB(r, p): $P\left(X = n\right) = \binom{n-1}{r-1} p^{r}\left(1 - p\right)^{n-r}$

Αναμενόμενη τιμή: EX = rp / (1 – p). $\frac{r\cdot(1-p)}{p}$

Διακύμανση: VarX = rp / (1 – p)^2.

Τυπική απόκλιση: $\sigma = \frac{\sqrt{rp}}{1-p}$

Ασυμμετρία: $\text{skew}\left(X\right) = \gamma = \frac{\mu_3}{\sigma^3} = \frac{1+p}{\sqrt{pr}}$

Κυρτότητα: $\text{kurt}\left(X\right) = \alpha = \frac{\mu_4}{\sigma^4} = \frac{6}{r} + \frac{1-p^2}{pr} + 3$

Ασκήσεις στην αρνητική διωνυμική

1. Μια εταιρεία πετρελαίου γνωρίζει ότι κάθε γεώτρηση που κάνει σε μία περιοχή, έχει 20% πιθανότητα να βρει πετρέλαιο.

(α) Ποια είναι η πιθανότητα να βρει πετρέλαιο στην τρίτη γεώτρηση που θα ανοίξει;

(β) Ποια είναι η πιθανότητα να έχει τρεις επιτυχείς ανιχνεύσεις πετρελαίου στις 7 πρώτες γεωτρήσεις που θα κάνει;

(γ) Ποιο είναι το μέσο πλήθος γεωτρήσεων που πρέπει να κάνει η εταιρεία αν θέλει να έχει τρεις επιτυχείς ανιχνεύσεις πετρελαίου;

(α) $(1-0,2)^2 \cdot 0,2$

(β) $X \sim B(7, 0.2)$, $P(X=3) = \ldots$

(γ) $Y \sim NB(3, 0.2)$. $E(Y) = \dfrac{3 \cdot (1-0,2)}{0,2} = 12$.

Ασκήσεις στην αρνητική διωνυμική

2. Εκτελούμε το εξής πείραμα:

Πείραμα

Ανακατεύουμε μια τράπουλα και αναποδογυρίζουμε το επάνω φύλλο. Βάζουμε ξανά την κάρτα στην τράπουλα και ανακατεύουμε. Επαναλαμβάνουμε αυτή τη διαδικασία μέχρι να πάρουμε 2 άσσους.

Ερώτηση

Αν με X συμβολίζεται το πλήθος των δοκιμασιών μέχρι να πετύχουμε το στόχο μας, τότε ακολουθεί η X την αρνητική διωνυμική κατανομή;

Ασκήσεις στην αρνητική διωνυμική

3. Ένας επιθετικός σκοράρει με πιθανότητα 68% όταν εκτελεί σουτ προς το αντίπαλο τέρμα. Γνωρίζουμε ότι σε έναν αγώνα εκτέλεσε 10 σουτ.

(α) Ποια είναι η πιθανότητα να έβαλε 9 γκολ;

(β) Αν σε κάθε αγώνα σουτάρει 10 φορές, τότε ποιο είναι το αναμενόμενο πλήθος γκολ που περιμένουμε αυτός να βάζει σε κάθε αγώνα;

Λύση

Ασκήσεις στην αρνητική διωνυμική

4. Για μία τυχαία μεταβλητή X, γνωρίζουμε ότι ακολουθεί την αρνητική διωνυμική κατανομή και πως $P(X = 0) = 0{,}2397410$, $P(X = 1) = 0{,}1038878$, $P(X = 2) = 0{,}0398236$

Να βρεθεί η μέση τιμή και η διακύμανση της X.

Λύση

Ασκήσεις στην αρνητική διωνυμική

5. Γνωρίζουμε ότι η τ.μ. X ακολουθεί την αρνητική διωνυμική κατανομή με αναμενόμενη τιμή 0,36 και διακύμανση 1,44. Να βρεθεί η P(X = 3).

Λύση

Ασκήσεις στην αρνητική διωνυμική

6. Ποια είναι η πιθανότητα, ρίχνοντας ένα νόμισμα να πάρουμε 4 Κορώνες πριν από 3 Γράμματα;
Λύση