

ΦΡΟΝΤΙΣΤΗΡΙΟ #4

ΥΠΕΝΟΤΗΜΙΣΗ ΘΕΩΡΙΑΣ

Ορισμός Έστω X, Y δύο τ.μ. με ρέσες τιμές μ_X, μ_Y αντίστοιχα.

Ορίζουμε ως συνδιακύμανση ή συνδιασπορά (covariance) των X και Y και συμβολίζουμε με $\text{Cov}(X, Y)$ την ποσότητα:

$$\text{Cov}(X, Y) := E[(X - \mu_X)(Y - \mu_Y)]$$

Ιδιότητες της συνδιακύμανσης

$$\text{Ισχύει ότι: } \text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Η παραπάνω ιδιότητα χρησιμοποιείται συχνά αντί του ορισμού για τον υπολογισμό της συνδιακύμανσης. Αν X, Y είναι ανεξάρτητες τ.μ. τότε $\text{Cov}(X, Y) = 0$. Το αντίστροφο δεν ισχύει.

Δηλαδή, αν $\text{Cov}(X, Y) = 0$, δεν έπεται ότι οι X, Y είναι ανεξάρτητες. Έστω οι τ.μ. X, Y που ορίζονται ως εξής:

$$P\{X=0\} = P\{X=1\} = P\{X=-1\} = \frac{1}{3} \text{ και } Y := \begin{cases} 0, & \text{αν } X \neq 0 \\ 1, & \text{αν } X = 0 \end{cases}$$

Προφανώς, οι X και Y είναι εξαρτημένες τ.μ. Από τον ορισμό των X, Y έχουμε ότι $X \cdot Y = 0$ και επομένως, $E(XY) = 0$.

$$\text{Επιπλέον, } E(X) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} - 1 \cdot \frac{1}{3} = 0. \text{ Άρα } \text{Cov}(X, Y) = 0.$$

$$\text{Επίσης, ισχύει ότι: } \text{Cov}(X+Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z).$$

Πρόταση: Για τις τ.μ. X, Y ισχύει ότι:

$$\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y) \pm 2\text{Cov}(X, Y)$$

Για τις τ.μ. $X_1, \dots, X_n$ η παραπάνω πρόταση επεκτείνεται ως εξής:

$$\text{Πρόταση: } \text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^n \text{Cov}(X_i, X_j)$$

Αν οι τ.μ. X, Y είναι ανεξάρτητες, τότε:

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) \text{ και, αν οι τ.μ. } X_1, \dots, X_n$$

είναι ανά δύο ανεξάρτητες τ.ρ. τότε:

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i).$$

#1 Για την #4, φροντιστήριο #3, υπολογίστε τη διασπορά του αριθμού των ανδρών που διαλέγουν τα διατάκτα καπέλα.

Λύση Ισχύει ότι:

$$\text{var}(X) = \text{var}(X_1 + \dots + X_N) = \sum_{i=1}^N \text{var}(X_i) + 2 \sum_{i=1}^N \sum_{j=i+1}^N \text{cov}(X_i, X_j)$$

$$\text{Όμως, } \text{var}(X_i) = E(X_i^2) - (E(X_i))^2 = P\{X_i=1\} \cdot 1^2 - (E(X_i))^2$$

$$= \frac{1}{N} - \left(\frac{1}{N}\right)^2 = \frac{N-1}{N^2}, \text{ για } i=1, \dots, N. \text{ Επίσης, για } i, j,$$

$$i < j, \text{ έχουμε: } \text{cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

Παρατηρούμε ότι: $X_i X_j = \begin{cases} 1, & \text{αν αμφότεροι οι άνδρες } i, j \text{ διαλέγουν τα καπέλα τους} \\ 0, & \text{σε οποιαδήποτε άλλη περίπτωση.} \end{cases}$

$$\text{Άρα, } E(X_i X_j) = P\{X_i=1, X_j=1\} = P\{X_i=1\} P\{X_j=1 | X_i=1\}$$

$$= \frac{1}{N} \cdot \frac{1}{N-1}. \text{ Συνεπώς, } \text{cov}(X_i, X_j) = \frac{1}{N(N-1)} - \left(\frac{1}{N}\right)^2$$

$$= \frac{1}{N^2(N-1)}.$$

Συνεπώς,

$$\begin{aligned} \text{var}(X_1 + \dots + X_N) &= N \frac{N-1}{N^2} + 2 \sum_{i=1}^N \sum_{j=i+1}^N \frac{1}{N^2(N-1)} \\ &= \frac{N-1}{N} + 2 \frac{1}{N^2(N-1)} \sum_{i=1}^N (N-i) = \frac{N-1}{N} + \frac{2}{N^2(N-1)} \left(N^2 - \frac{N(N+1)}{2} \right) \\ &= \frac{N-1}{N} + \frac{2}{N^2(N-1)} \left(\frac{2N^2 - N^2 - N}{2} \right) = \frac{N-1}{N} + \left(\frac{N^2 - N}{N^2(N-1)} \right) \end{aligned}$$

$$= \frac{N(N-1)}{N^2(N-1)} + \frac{N-1}{N} = \frac{1}{N} + \frac{N-1}{N} = 1. \quad \square$$

Η συνδιακύρυνση δύο τ.ρ. μεταβάλλεται αν μεταβληθούν οι μονάδες μέτρησης των τ.ρ. Ένα μέτρο εξάρτησης δύο τ.ρ. που δεν επηρεάζεται από πιθανές μεταβολές των μονάδων μέτρησης των μεταβλητών είναι ο συντελεστής συσχέτισης των τ.ρ.

$$X, Y \text{ που ορίζεται ως: } \rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \text{ όπου}$$

σ_X, σ_Y οι τυπικές αποκλίσεις των τ.ρ. X, Y τέτοιες ώστε $\sigma_X, \sigma_Y \neq 0$.

Ισχύει ότι, αν X_1, X_2 τ.ρ. και $Y_1 = a_1 X_1 + b_1, Y_2 = a_2 X_2 + b_2$ όπου a_1, a_2, b_1, b_2 σταθερές τότε $\rho(Y_1, Y_2) = \rho(X_1, X_2)$.

Επίσης $|\rho(X, Y)| \leq 1$. Έστω X, Y δύο τ.ρ. γραμμικά εξαρτημένες δηλαδή $Y = aX + b$ με a, b σταθερές και $a \neq 0$. Τότε

$$\rho(X, Y) = \begin{cases} 1, & \text{αν } a > 0 \\ -1, & \text{αν } a < 0 \end{cases}$$

Αντίστροφα, αν $|\rho(X, Y)| = 1$, υπάρχουν σταθερές $a \neq 0$ και b τέτοιες ώστε $Y = aX + b$. Ο συντελεστής συσχέτισης δύο τ.ρ. X, Y είναι ένα μέτρο του βαθμού της γραμμικής εξάρτησης των X, Y . Μία τιμή του συντελεστή συσχέτισης κοντά στο $+1$ ή στο -1 είναι ένδειξη υψηλού βαθμού γραμμικής εξάρτησης μεταξύ των X και Y ενώ μία τιμή του συντελεστή συσχέτισης κοντά στο μηδέν είναι ένδειξη ότι δεν υπάρχει γραμμική εξάρτηση χωρίς αυτό να αποκλείει εξάρτηση κάποιας άλλης μορφής. Αν $\rho(X, Y) = 0$, οι τ.ρ. X, Y καλούνται ασυσχέτιστες (η *uncorrelated*).

#2 Ένας αριθμός X επιλέγεται τυχαία στο διάστημα $(0, 1)$. Έστω R η απόκλιση του αριθμού που επιλέχτηκε

από ένα σταθερό αριθμό a , όπου $0 \leq a \leq 1$. (α) Να βρεθεί η συντελεστής συσχέτισης των τ.ρ. X και R . (β) Για ποια τιμή του a οι τ.ρ. X και R είναι ασυσχέτιστες;

Λύση (α) Η τ.ρ. X ακολουθεί την ομοιόμορφη κατανομή στο διάστημα $(0,1)$ ενώ για την τ.ρ. R ισχύει ότι:

$$R = |a - X|.$$

Αφού $f_X(x) = 1$, $0 < x < 1$ έχουμε:

$$E[R] = E[|a - X|] = \int_{-\infty}^{\infty} |a - x| f_X(x) dx = \int_0^a (a - x) dx + \int_a^1 (x - a) dx$$

$$= a^2 - \left[\frac{x^2}{2} \right]_0^a + \left[\frac{x^2}{2} \right]_a^1 - a(1 - a) = a^2 - \left(\frac{a^2}{2} \right) + \left(\frac{1}{2} - \frac{a^2}{2} \right) - a + a^2$$

$$= a^2 - \frac{a^2}{2} + \frac{1}{2} - \frac{a^2}{2} - a + a^2 = a^2 - a + \frac{1}{2}.$$

$$E[R^2] = E[|a - X|^2] = \int_{-\infty}^{\infty} (a - x)^2 f_X(x) dx = \int_0^1 (a - x)^2 dx$$

$$= \int_0^1 (a^2 + x^2 - 2ax) dx = a^2 + \left[\frac{x^3}{3} \right]_0^1 - 2a \left[\frac{x^2}{2} \right]_0^1$$

$$= a^2 + \frac{1}{3} - 2a \frac{1}{2} = a^2 - a + \frac{1}{3}$$

$$\text{Var}(R) = E[R^2] - (E[R])^2 = a^2 - a + \frac{1}{3} - \left(a^2 - a + \frac{1}{2} \right)^2$$

$$= a^2 - a + \frac{1}{3} - \left((a^2 - a)^2 + \frac{1}{4} + 2(a^2 - a) \frac{1}{2} \right)$$

$$= a^2 - a + \frac{1}{3} - \left(a^4 - 2a^3 + a^2 + \frac{1}{4} + a^2 - a \right)$$

$$= \cancel{a^2} - \cancel{a} + \frac{1}{3} - a^4 + 2a^3 - \cancel{a^2} - \frac{1}{4} - a^2 + \cancel{a}$$

$$= 2a^3 - a^4 - a^2 + \frac{1}{12}.$$

Επίσης, $E[XR] = E[X|a - X|] = \int_{-\infty}^{\infty} x|a - x| f_X(x) dx =$

$$\begin{aligned}
&= \int_0^a x(a-x) dx + \int_a^1 x(x-a) dx = \\
&= \int_0^a (ax - x^2) dx + \int_a^1 (x^2 - ax) dx = \\
&= \left[a \frac{x^2}{2} - \frac{x^3}{3} \right]_0^a + \left[\frac{x^3}{3} - a \frac{x^2}{2} \right]_a^1 \\
&= \frac{a^3}{2} - \frac{a^3}{3} + \frac{1}{3} - \frac{a}{2} - \frac{a^3}{3} + \frac{a^3}{2} = a^3 - \frac{2a^3}{3} - \frac{a}{2} + \frac{1}{3} \\
&= \frac{a^3}{3} - \frac{a}{2} + \frac{1}{3} \quad \text{Όπως } E[X] = \frac{1}{2}, \text{ var}(X) = \frac{1}{12} \text{ οπότε}
\end{aligned}$$

$$\begin{aligned}
\rho(X, R) &= \frac{E(XR) - E(X)E(R)}{\sqrt{\text{var}(X)} \sqrt{\text{var}(R)}} \\
&= \frac{\frac{a^3}{3} - \frac{a}{2} + \frac{1}{3} - \frac{1}{2} \left(a^2 - a + \frac{1}{2} \right)}{\sqrt{\frac{1}{12}} \sqrt{2a^3 - a^4 - a^2 + \frac{1}{12}}} \\
&= \frac{\frac{a^3}{3} - \cancel{\frac{a}{2}} + \frac{1}{3} - \frac{a^2}{2} + \cancel{\frac{a}{2}} - \frac{1}{4}}{\frac{1}{\sqrt{12}} \sqrt{2a^3 - a^4 - a^2 + \frac{1}{12}}} \cdot \frac{12}{12} = \frac{4a^3 - 6a^2 + 1}{\sqrt{12} \sqrt{2a^3 - a^4 - a^2 + \frac{1}{12}}}
\end{aligned}$$

(β) Για να είναι οι τ.μ. X, R ασυσχέτιστες θα πρέπει ο συντελεστής συσχέτισης $\rho(X, R)$ να είναι ίσος με μηδέν ή ισοδύναμα $4a^3 - 6a^2 + 1 = 0$. Με παραγοντοποίηση (με σχήμα Horner) έχουμε:

$$\begin{array}{ccc|c} 4 & -6 & 0 & 1 \\ & 2 & -2 & -1 \\ \hline 4 & -4 & -2 & 0 \end{array} \quad \frac{1}{2}$$

Μπορούμε να γράψουμε:

$$\left(a - \frac{1}{2}\right)(4a^2 - 4a - 2) = 0$$

-6-

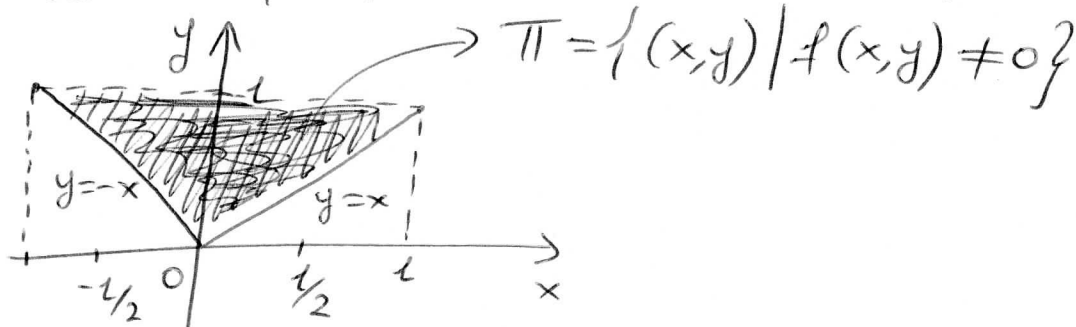
$\Rightarrow (2a - 1)(2a^2 - 2a - 1) = 0$. Η τελευταία εξίσωση έχει μοναδική λύση στο διάστημα $(0, 1)$ την $a = \frac{1}{2}$.

#3 Η από κοινού συνάρτηση πυκνότητας της διδιάστατης συνεχούς τυχάιας μεταβλητής (X, Y) δίνεται από τον τύπο:

$$f(x, y) = \begin{cases} 1, & |x| < y < 1 \\ 0, & \text{διαφορετικά} \end{cases}$$

Να δειχθεί ότι, ενώ οι τ.ρ. X, Y είναι ασυσχέτιστες, δεν είναι ανεξάρτητες.

Λύση Για τις τ.ρ. X, Y έχουμε $R_X = (-1, 1)$, $R_Y = [0, 1]$



$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_{|x|}^1 dy = 1 - |x|, \quad -1 < x < 1$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_{-y}^y dx = 2y, \quad 0 \leq y < 1$$

Οι X, Y δεν είναι ανεξάρτητες διότι δεν ισχύει η σχέση

$$f(x, y) = f_X(x) f_Y(y) \quad \forall x, y \quad (\text{π.χ. αν πάρουμε } x = y = \frac{1}{2})$$

$$\text{Θα έχουμε } 1 = f\left(\frac{1}{2}, \frac{1}{2}\right) \neq f_X\left(\frac{1}{2}\right) f_Y\left(\frac{1}{2}\right) = \frac{1}{2}.$$

Για τις μέσες τιμές των τ.ρ. X, Y και XY έχουμε:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-1}^1 x(1-|x|) dx = \int_{-1}^0 x(1+x) dx + \int_0^1 x(1-x) dx$$

$$= \left[\frac{x^2}{2} + \frac{x^3}{3} \right]_{-1}^0 + \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_{-1}^0 = 0$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 2y^2 dy = 1$$

και

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x,y) dx dy = \int_0^1 \int_{-y}^y xy dx dy$$

$$= \int_0^1 y \left[\frac{x^2}{2} \right]_{-y}^y dy = \int_0^1 y \cdot 0 dy = 0$$

$$\text{Αφού } \text{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0 - 0 \cdot 1 = 0$$

έπεται ότι οι τ.ρ. X, Y είναι ασυσχέτιστες.

#4 Ο χρόνος ζωής ενός ηλεκτρονίου εξαρτήματος που είναι τοποθετημένο σε μία συσκευή περιγράφεται από μία τ.ρ. Y , ενώ ο χρόνος ζωής της συσκευής περιγράφεται από μία άλλη τ.ρ. X . Αν το εξάρτημα καταστραφεί, παύει να λειτουργεί και η συσκευή χωρίς να συμβαίνει το αντίστροφο. Υποθέτουμε ότι η από κοινού συνάρτηση πυκνότητας των τ.ρ. X και Y σε χρόνια δίνεται από τον τύπο: $f(x,y) = \frac{1}{4y} e^{-y/4}$, $0 < x < y < \infty$ και

$f(x,y) = 0$, αλλιώς. (α) Ναδειχτεί ότι ο αναφερόμενος επιπλέον χρόνος που ζει το εξάρτημα σε σχέση με τη συσκευή είναι ίσος με 7 χρόνια. (β) Ναδειχτεί ότι οι τ.ρ. X και Y είναι θετικά συσχετισμένες και να εξηγηθεί διαισθητικά το αποτέλεσμα.

Λύση

(α) Θα υπολογίσουμε τις περιθώριες συναρτήσεις πυκνότητας και τις μέσες τιμές των τ.ρ. X και Y . Ισχύει ότι το σύνολο τιμών των X και Y είναι $R_X = R_Y = (0, \infty)$.

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_x^{\infty} \frac{1}{49} e^{-y/7} dy = -\frac{1}{7} \left[e^{-y/7} \right]_x^{\infty} \\ = \frac{1}{7} e^{-x/7}, \quad x > 0$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^y \frac{1}{49} e^{-y/7} dx = \frac{1}{49} y e^{-y/7}, \quad y > 0$$

Αφού η τ.ρ. $X \sim \text{Ευθετική} (\lambda = 1/7)$ και η $Y \sim \text{Gamma}(\lambda = 1/7, a=2)$ έχουμε ότι: $E[X] = \frac{1}{\lambda} = 7$, $E[Y] = \frac{a}{\lambda} = 14 = \frac{2}{1/7}$. Άρα, ο

αναμενόμενος χρόνος σε έτη που βίει το εξάρτημα σε σχέση με τη συσκευή, επιπλέον, είναι:

$$E[Y-X] = E[Y] - E[X] = 14 - 7 = 7.$$

(β) Έχουμε:

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dy dx = \int_0^{\infty} \frac{1}{49} x \left(\int_x^{\infty} y e^{-y/7} dy \right) dx$$

Ομως

$$\int_x^{\infty} y e^{-y/7} dy = -7 \left\{ \left[y e^{-y/7} \right]_x^{\infty} - \int_x^{\infty} e^{-y/7} dy \right\} \\ = -7 \left\{ -x e^{-x/7} + 7 \left[e^{-\frac{1}{7}y} \right]_x^{\infty} \right\} = e^{-x/7} \left(1 + \frac{1}{7}x \right) \cdot 49$$

Άρα

$$E(XY) = \frac{1}{49} \int_0^{\infty} x e^{-x/7} \left(1 + \frac{x}{7} \right) 49 dx =$$

$$= \int_0^{\infty} \left(x e^{-\frac{x}{7}} + \frac{1}{7} x^2 e^{-\frac{x}{7}} \right) dx = \int_0^{\infty} x e^{-\frac{x}{7}} dx + \frac{1}{7} \int_0^{\infty} x^2 e^{-\frac{x}{7}} dx$$

Το πρώτο ολοκλήρωμα το έχουμε υπολογίσει. Είναι της γενικής μορφής $I_1(a, b) = e^{-ab} \frac{(1+ab)}{b^2}$ με $a=0, b=1/7$

$$\text{άρα } \int_0^{\infty} x e^{-\frac{x}{7}} dx = 49$$

Το δεύτερο ολοκλήρωμα είναι της γενικής μορφής

$$I_2(a, b) = \int_a^{\infty} t^2 e^{-bt} dt = -\frac{1}{b} \left\{ \left[t^2 e^{-bt} \right]_a^{\infty} - 2 \int_a^{\infty} t e^{-bt} dt \right\}$$

$$= -\frac{1}{b} \left\{ -a^2 e^{-ab} - 2 I_1(a, b) \right\} = e^{-ab} \frac{(ab)^2 + 2(1+ab)}{b^3}$$

$$\text{Είναι } \frac{1}{7} \int_0^{\infty} x^2 e^{-\frac{x}{7}} dx = \frac{1}{7} I_2(0, 1/7) = \frac{1}{7} \frac{2}{(1/7)^3} = 98$$

$$\text{Συνεπώς, } E(XY) = 49 + 98 = 147 \text{ και}$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 147 - 7 \cdot 14 = 49 > 0$$

$\Rightarrow \text{Cov}(X, Y) > 0$ αναμενόμενο αποτέλεσμα διότι όσο αυξάνει ο χρόνος ζωής της συζύγου αναμένεται να αυξάνει και ο χρόνος ζωής του εξαρτήματος.

