

ΗΥ-217: Πιθανότητες - Χειμερινό Εξάμηνο 2017
Διδάσκων: Π. Τσακαλίδης

Φροντιστήριο 7
Συνδιασπορά και Συσχέτιση

Άσκηση 1.

(α) Αν X και Y είναι τυχαίες μεταβλητές, και α, β, γ και δ είναι σταθερές, να αποδείξετε ότι ισχύει η σχέση:

$$COV(\alpha X + \beta, \gamma Y + \delta) = \alpha\gamma COV(X, Y).$$

(β) Να αποδείξετε ότι η συνδιασπορά 2 τυχαίων μεταβλητών X και Y μπορεί να εκφραστεί ως εξής:

$$COV(X, Y) = E[XY] - E[X]E[Y].$$

Λύση.

(α) Σύμφωνα με τον ορισμό της συνδιασποράς, και χρησιμοποιώντας την ιδιότητα της γραμμικότητας της μέσης τιμής, έχουμε:

$$\begin{aligned} COV(\alpha X + \beta, \gamma Y + \delta) &= E\{[\alpha X + \beta - E(\alpha X + \beta)][\gamma Y + \delta - E(\gamma Y + \delta)]\} \\ &= E\{[\alpha X + \beta - \alpha E(X) - \beta][\gamma Y + \delta - \gamma E(Y) - \delta]\} \\ &= \alpha\gamma E\{[X - E(X)][Y - E(Y)]\} = \alpha\gamma COV(X, Y). \end{aligned}$$

(β) Χρησιμοποιώντας τον ορισμό της συνδιασποράς και την ιδιότητα της γραμμικότητας της μέσης τιμής, έχουμε:

$$\begin{aligned} COV(X, Y) &= E\{[X - E(X)][Y - E(Y)]\} = E\{XY - XE[Y] - YE[X] + E[X]E[Y]\} = \\ &= E[XY] - E\{XE[Y]\} - E\{YE[X]\} + E\{E[X]E[Y]\} \\ &= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] = E[XY] - E[X]E[Y]. \end{aligned}$$

Άσκηση 2.

Αν X και Y είναι 2 τυχαίες μεταβλητές, να αποδείξετε ότι ισχύει η ανισότητα *Caychy – Schwarz*:

$$(E[XY])^2 \leq E[X^2]E[Y^2].$$

Λύση.

Ορίζουμε την τυχαία μεταβλητή $X - \lambda Y$, όπου λ είναι πραγματικός αριθμός (σταθερά).

$\forall \lambda \in \mathbb{R}$ ισχύει ότι:

$$0 \leq E[(X - \lambda Y)^2] \Rightarrow 0 \leq \lambda^2 E[Y^2] - 2\lambda E[XY] + E[X^2].$$

Το παραπάνω τριώνυμο ως προς λ (έστω $f(\lambda)$) έχει ελάχιστο, διότι η 2η παράγωγός του ως προς λ ($f''(\lambda) = E[Y^2]$) είναι πάντα θετική. Η τιμή του λ που το ελαχιστοποιεί προκύπτει από την εξίσωση της 1ης παραγώγου ως προς λ ($f'(\lambda)$) με το 0:

$$f'(\lambda) = 0 \Rightarrow 2\lambda E[Y^2] - 2E[XY] = 0 \Rightarrow \lambda_{min} = \frac{E[XY]}{E[Y^2]}.$$

Η αντίστοιχη ελάχιστη τιμή του τριωνύμου είναι:

$$\begin{aligned} 0 &\leq E[(X - \lambda_{min} Y)^2] \Rightarrow 0 \leq E[(X - \frac{E[XY]}{E[Y^2]} Y)^2] \Rightarrow 0 \leq E[X^2 - 2\frac{XYE[XY]}{E[Y^2]} + Y^2(\frac{E[XY]}{E[Y^2]})^2] \\ &\Rightarrow 0 \leq E[X^2] - 2\frac{E[XY]E[XY]}{E[Y^2]} + E[Y^2]\frac{(E[XY])^2}{(E[Y^2])^2} \Rightarrow 0 \leq E[X^2] - 2\frac{(E[XY])^2}{E[Y^2]} + \frac{(E[XY])^2}{E[Y^2]} \\ &\Rightarrow 0 \leq E[X^2] - \frac{(E[XY])^2}{E[Y^2]} \Rightarrow (E[XY])^2 \leq E[X^2]E[Y^2]. \end{aligned}$$

Άσκηση 3.

Έστω (X, Y) μία διακριτή δισδιάστατη τυχαία μεταβλητή, με από κοινού συνάρτηση πυκνότητας πιθανότητας που δίνεται από τον παρακάτω τύπο:

$$f_{X,Y}(x, y) = \frac{x+y}{21}, \quad x = 1, 2, \quad y = 1, 2, 3.$$

Να υπολογιστεί ο συντελεστής συσχέτισης των τυχαίων μεταβλητών X και Y .

Λύση.

Με βάση τον ορισμό του συντελεστή συσχέτισης των X και Y , έχουμε:

$$\rho(X, Y) = \frac{COV(X, Y)}{\sqrt{VAR[X]VAR[Y]}}.$$

Για τις τυχαίες μεταβλητές X και Y , έχουμε:

$$E[XY] = \sum_{y=1}^3 \sum_{x=1}^2 xy f_{X,Y}(x, y) = \sum_{y=1}^3 \sum_{x=1}^2 xy \frac{x+y}{21} = \sum_{y=1}^3 \frac{y(y+1)+2y(y+2)}{21} = \sum_{y=1}^3 \frac{y(3y+5)}{21} = \frac{24}{7}.$$

Για τον υπολογισμό της περιθώριας συνάρτησης πυκνότητας πιθανότητας της τυχαίας μεταβλητής X , αθροίζουμε την από κοινού συνάρτηση πυκνότητας πιθανότητας ως προς y :

$$f_X(x) = \sum_{y=1}^3 f_{X,Y}(x, y) = \sum_{y=1}^3 \frac{x+y}{21} = \frac{(x+1)+(x+2)+(x+3)}{21} = \frac{3x+6}{21} = \frac{x+2}{7}, \quad x = 1, 2.$$

Για τον υπολογισμό της περιθώριας συνάρτησης πυκνότητας πιθανότητας της τυχαίας μεταβλητής Y , αθροίζουμε την από κοινού συνάρτηση πυκνότητας πιθανότητας ως προς x :

$$f_Y(y) = \sum_{x=1}^2 f_{X,Y}(x, y) = \sum_{x=1}^2 \frac{x+y}{21} = \frac{(y+1)+(y+2)}{21} = \frac{2y+3}{21}, \quad y = 1, 2, 3.$$

Συνεπώς, για την τυχαία μεταβλητή X θα έχουμε:

$$\begin{aligned} E[X] &= \sum_{x=1}^2 x f_X(x) = \sum_{x=1}^2 x \frac{x+2}{7} = \frac{1(1+2)+2(2+2)}{7} = \frac{3+8}{7} = \frac{11}{7}. \\ E[X^2] &= \sum_{x=1}^2 x^2 f_X(x) = \sum_{x=1}^2 x^2 \frac{x+2}{7} = \frac{1^2(1+2)+2^2(2+2)}{7} = \frac{3+16}{7} = \frac{19}{7}. \\ VAR[X] &= E[X^2] - (E[X])^2 = \frac{19}{7} - \left(\frac{11}{7}\right)^2 = \frac{19}{7} - \frac{121}{49} = \frac{133-121}{49} = \frac{12}{49}. \end{aligned}$$

Αντίστοιχα, για την τυχαία μεταβλητή Y θα έχουμε:

$$\begin{aligned} E[Y] &= \sum_{y=1}^3 y f_Y(y) = \sum_{y=1}^3 y \frac{2y+3}{21} = \frac{1(2 \times 1 + 3) + 2(2 \times 2 + 3) + 3(2 \times 3 + 3)}{21} = \frac{5+14+27}{21} = \frac{46}{21}. \\ E[Y^2] &= \sum_{y=1}^3 y^2 f_Y(y) = \sum_{y=1}^3 y^2 \frac{2y+3}{21} = \frac{1^2(2 \times 1 + 3) + 2^2(2 \times 2 + 3) + 3^2(2 \times 3 + 3)}{21} = \frac{5+28+81}{21} = \frac{114}{21}. \\ VAR[Y] &= E[Y^2] - (E[Y])^2 = \frac{114}{21} - \left(\frac{46}{21}\right)^2 = \frac{114}{21} - \frac{2116}{441} = \frac{2394-2116}{441} = \frac{278}{441}. \end{aligned}$$

Για τον υπολογισμό της συνδιασποράς των X και Y , έχουμε:

$$COV(X, Y) = E[XY] - E[X]E[Y] = \frac{24}{7} - \frac{11}{7} \times \frac{46}{21} = \frac{24}{7} - \frac{506}{147} = \frac{504-506}{147} = -\frac{2}{147}.$$

Ως εκ τούτου, ο συντελεστής συσχέτισης των X και Y θα είναι τελικά:

$$\rho(X, Y) = \frac{COV(X, Y)}{\sqrt{VAR[X]VAR[Y]}} = \frac{-\frac{2}{147}}{\sqrt{\frac{12}{49} \times \frac{278}{441}}} = -\frac{1}{\sqrt{834}}.$$

Άσκηση 4.

Έστω (X, Y) μία συνεχής δισδιάστατη τυχαία μεταβλητή, με από κοινού συνάρτηση πυκνότητας πιθανότητας που δίνεται από τον παρακάτω τύπο:

$$f_{X,Y}(x, y) = 8xy, 0 \leq x \leq y \leq 1.$$

Να υπολογιστεί ο συντελεστής συσχέτισης των τυχαίων μεταβλητών X και Y .

Λύση.

Με βάση τον ορισμό του συντελεστή συσχέτισης των X και Y , έχουμε:

$$\rho(X, Y) = \frac{COV(X, Y)}{\sqrt{VAR[X]VAR[Y]}}.$$

Για τις τυχαίες μεταβλητές X και Y , έχουμε:

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy = \int_0^1 \int_0^y xy \times 8xy dx dy = 8 \int_0^1 \int_0^y x^2 y^2 dx dy = \frac{8}{3} \int_0^1 y^5 dy = \frac{4}{9}.$$

Για τον υπολογισμό της περιθώριας συνάρτησης πυκνότητας πιθανότητας της τυχαίας μεταβλητής X , ολοκληρώνουμε την από κοινού συνάρτηση πυκνότητας πιθανότητας ως προς y :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_x^1 8xy dy = 8x \int_x^1 y dy = 4x(1 - x^2), 0 \leq x \leq 1.$$

Για τον υπολογισμό της περιθώριας συνάρτησης πυκνότητας πιθανότητας της τυχαίας μεταβλητής Y , ολοκληρώνουμε την από κοινού συνάρτηση πυκνότητας πιθανότητας ως προς x :

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \int_0^y 8xy dx = 8y \int_0^y x dx = 4y^3, 0 \leq y \leq 1.$$

Συνεπώς, για την τυχαία μεταβλητή X θα έχουμε:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \times 4x(1 - x^2) dx = 4 \int_0^1 x^2(1 - x^2) dx = \frac{8}{15}. \\ E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 \times 4x(1 - x^2) dx = 4 \int_0^1 x^3(1 - x^2) dx = \frac{1}{3}. \\ VAR[X] &= E[X^2] - (E[X])^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{1}{3} - \frac{64}{225} = \frac{75-64}{225} = \frac{11}{225}. \end{aligned}$$

Αντίστοιχα, για την τυχαία μεταβλητή Y θα έχουμε:

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \times 4y^3 dy = 4 \int_0^1 y^4 dy = \frac{4}{5}. \\ E[Y^2] &= \int_{-\infty}^{\infty} y^2 f_Y(y) dy = \int_0^1 y^2 \times 4y^3 dy = 4 \int_0^1 y^5 dy = \frac{2}{3}. \\ VAR[Y] &= E[Y^2] - (E[Y])^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{3} - \frac{16}{25} = \frac{50-48}{75} = \frac{2}{75}. \end{aligned}$$

Για τον υπολογισμό της συνδιασποράς των X και Y , έχουμε:

$$COV(X, Y) = E[XY] - E[X]E[Y] = \frac{4}{9} - \frac{8}{15} \times \frac{4}{5} = \frac{4}{9} - \frac{32}{75} = \frac{300-288}{675} = \frac{12}{675} = \frac{4}{225}.$$

Ως εκ τούτου, ο συντελεστής συσχέτισης των X και Y θα είναι τελικά:

$$\rho(X, Y) = \frac{COV(X, Y)}{\sqrt{VAR[X]VAR[Y]}} = \frac{\frac{4}{225}}{\sqrt{\frac{11}{225} \times \frac{2}{75}}} = \frac{4}{\sqrt{66}}.$$

Άσκηση 5.

Υποθέστε ότι οι X και Y είναι τυχαίες μεταβλητές με την ίδια διασπορά. Δείξτε ότι οι $X - Y$ και $X + Y$ είναι ασυσχέτιστες.

Λύση.

Καθώς η συνδιασπορά παραμένει αναλλοίωτη όταν προσθέτουμε μία σταθερά σε μία τυχαία μεταβλητή, μπορούμε να υποθέσουμε χωρίς βλάβη της γενικότητας ότι οι X και Y έχουν μηδενική μέση τιμή ($E[X] = E[Y] = 0$). Ως εκ τούτου, έχουμε:

$$\begin{aligned}
COV(X - Y, X + Y) &= E[(X - Y)(X + Y)] = E[X^2 + XY - YX - Y^2] \\
&= E[X^2 - Y^2] = E[X^2] - E[Y^2] \\
&= [VAR(X) + (E[X])^2] - [VAR(Y) + (E[Y])^2] \\
&= VAR[X] - VAR[Y] = 0
\end{aligned}$$

, καθώς οι X και Y έχουν την ίδια διασπορά ($VAR[X] = VAR[Y]$).

Άσκηση 6.

Θεωρήστε 4 τυχαίες μεταβλητές W, X, Y, Z , για τις οποίες ισχύουν:

$$\begin{aligned}
E[W] &= E[X] = E[Y] = E[Z] = 0 \\
VAR[W] &= VAR[X] = VAR[Y] = VAR[Z] = 1.
\end{aligned}$$

Επιπλέον, υποθέστε ότι οι W, X, Y, Z είναι ασυσχέτιστες ανά ζεύγη. Να υπολογιστούν οι συντελεστές συσχέτισης $\rho(A, B)$ και $\rho(A, C)$, όπου:

$$A = W + X, B = X + Y, C = Y + Z.$$

Λύση.

Εφόσον οι W, X, Y, Z είναι ασυσχέτιστες ανά ζεύγη, έχουμε:

$$\begin{aligned}
COV(W, X) &= 0 \Rightarrow E[WX] - E[W]E[X] = 0 \Rightarrow E[WX] = E[W]E[X]. \\
COV(W, Y) &= 0 \Rightarrow E[WY] - E[W]E[Y] = 0 \Rightarrow E[WY] = E[W]E[Y]. \\
COV(W, Z) &= 0 \Rightarrow E[WZ] - E[W]E[Z] = 0 \Rightarrow E[WZ] = E[W]E[Z]. \\
COV(X, Y) &= 0 \Rightarrow E[XY] - E[X]E[Y] = 0 \Rightarrow E[XY] = E[X]E[Y]. \\
COV(X, Z) &= 0 \Rightarrow E[XZ] - E[X]E[Z] = 0 \Rightarrow E[XZ] = E[X]E[Z]. \\
COV(Y, Z) &= 0 \Rightarrow E[YZ] - E[Y]E[Z] = 0 \Rightarrow E[YZ] = E[Y]E[Z].
\end{aligned}$$

Με βάση τον ορισμό του συντελεστή συσχέτισης των A και B , έχουμε:

$$\rho(A, B) = \frac{COV(A, B)}{\sqrt{VAR[A]VAR[B]}}.$$

Για τον υπολογισμό της συνδιασποράς των A και B , έχουμε:

$$\begin{aligned}
COV(A, B) &= E[AB] - E[A]E[B] = E[(W + X)(X + Y)] - E[W + X]E[X + Y] \\
&= E[WX + WY + X^2 + XY] - (E[W] + E[X])(E[X] + E[Y]) = \\
&\quad E[WX] + E[WY] + E[X^2] + E[XY] - (0 + 0)(0 + 0) \\
&= E[W]E[X] + E[W]E[Y] + E[X^2] + E[X]E[Y] \\
&= 0 \times 0 + 0 \times 0 + E[X^2] + 0 \times 0 \\
&= E[X^2] = VAR[X] + (E[X])^2 = 1 + 0^2 = 1.
\end{aligned}$$

Για τον υπολογισμό της διασποράς των A και B , έχουμε:

$$\begin{aligned}
VAR[A] &= VAR[W + X] = VAR[W] - 2COV(W, X) + VAR[X] = 1 - 2 \times 0 + 1 = 2. \\
VAR[B] &= VAR[X + Y] = VAR[X] - 2COV(X, Y) + VAR[Y] = 1 - 2 \times 0 + 1 = 2.
\end{aligned}$$

Ως εκ τούτου, ο συντελεστής συσχέτισης των A και B θα είναι τελικά:

$$\rho(A, B) = \frac{COV(A, B)}{\sqrt{VAR[A]VAR[B]}} = \frac{1}{\sqrt{2 \times 2}} = \frac{1}{\sqrt{4}} = \frac{1}{2}.$$

Ακολουθώντας το ίδιο σκεπτικό, με βάση τον ορισμό του συντελεστή συσχέτισης των A και C , έχουμε:

$$\rho(A, C) = \frac{COV(A, C)}{\sqrt{VAR[A]VAR[C]}}.$$

Για τον υπολογισμό της συνδιασποράς των A και C , έχουμε:

$$\begin{aligned} COV(A, C) &= E[AC] - E[A]E[C] = E[(W + X)(Y + Z)] - E[W + X]E[Y + Z] \\ &= E[WY + WZ + XY + XZ] - (E[W] + E[X])(E[Y] + E[Z]) = \\ &= E[WY] + E[WZ] + E[XY] + E[XZ] - (0 + 0)(0 + 0) \\ &= E[W]E[Y] + E[W]E[Z] + E[X]E[Y] + E[X]E[Z] \\ &= 0 \times 0 + 0 \times 0 + 0 \times 0 + 0 \times 0 = 0. \end{aligned}$$

Για τον υπολογισμό της διασποράς των A και C , έχουμε:

$$\begin{aligned} VAR[A] &= VAR[W + X] = VAR[W] - 2COV(W, X) + VAR[X] = 1 - 2 \times 0 + 1 = 2. \\ VAR[C] &= VAR[Y + Z] = VAR[Y] - 2COV(Y, Z) + VAR[Z] = 1 - 2 \times 0 + 1 = 2. \end{aligned}$$

Ως εκ τούτου, ο συντελεστής συσχέτισης των A και C θα είναι τελικά:

$$\rho(A, C) = \frac{COV(A, C)}{\sqrt{VAR[A]VAR[C]}} = \frac{0}{\sqrt{2 \times 2}} = \frac{0}{\sqrt{4}} = \frac{0}{2} = 0.$$

Άσκηση 7.

Υποθέστε ότι η X είναι μία τυχαία μεταβλητή για την οποία ισχύουν οι εξής σχέσεις:

$$E[X] = 0, E[X^2] = 1, E[X^3] = 0, E[X^4] = 3.$$

Έστω μία νέα τυχαία μεταβλητή Y , η οποία ορίζεται ως εξής:

$$Y = a + bX + cX^2.$$

Υπολογίστε τον συντελεστή συσχέτισης $\rho(X, Y)$.

Λύση.

Με βάση τον ορισμό του συντελεστή συσχέτισης των X και Y , έχουμε:

$$\rho(X, Y) = \frac{COV(X, Y)}{\sqrt{VAR[X]VAR[Y]}}.$$

Για τον υπολογισμό της συνδιασποράς των X και Y , έχουμε:

$$\begin{aligned} COV(X, Y) &= E[XY] - E[X]E[Y] = E[X(a + bX + cX^2)] - E[X]E[a + bX + cX^2] \\ &= E[aX + bX^2 + cX^3] - E[X](a + bE[X] + cE[X^2]) \\ &= aE[X] + bE[X^2] + cE[X^3] - aE[X] - b(E[X])^2 - cE[X]E[X^2] \\ &= a \times 0 + b \times 1 + c \times 0 - a \times 0 - b \times 0^2 - c \times 0 \times 1 = b. \end{aligned}$$

Για τον υπολογισμό της διασποράς των X και Y , έχουμε:

$$\begin{aligned} VAR[X] &= E[X^2] - (E[X])^2 = 1 - 0^2 = 1. \\ VAR[Y] &= VAR[a + bX + cX^2] = E[(a + bX + cX^2)^2] - (E[a + bX + cX^2])^2 \\ &= E[(a + bX)^2 + c^2X^4 + 2cX^2(a + bX)] - (E[a] + bE[X] + cE[X^2])^2 \\ &= E[(a + bX)^2] + c^2E[X^4] + 2cE[X^2(a + bX)] - (a + b \times 0 + c \times 1)^2 \\ &= E[a^2 + 2abX + b^2X^2] + c^2E[X^4] + 2cE[aX^2 + bX^3] - (a + c)^2 \\ &= E[a^2] + 2abE[X] + b^2E[X^2] + c^2E[X^4] + 2acE[X^2] + 2bcE[X^3] - (a + c)^2 \\ &= a^2 + 2ab \times 0 + b^2 \times 1 + c^2 \times 3 + 2ac \times 1 + 2bc \times 0 - (a^2 + 2ac + c^2) \\ &= a^2 + b^2 + 3c^2 + 2ac - a^2 - 2ac - c^2 = b^2 + 2c^2. \end{aligned}$$

Ως εκ τούτου, ο συντελεστής συσχέτισης των X και Y θα είναι τελικά:

$$\rho(X, Y) = \frac{COV(X, Y)}{\sqrt{VAR[X]VAR[Y]}} = \frac{b}{\sqrt{1 \times (b^2 + 2c^2)}} = \frac{b}{\sqrt{b^2 + 2c^2}}.$$