ΚΑΤΑΝΟΜΕΣ ΔΙΔΙΑΣΤΑΤΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Χαράλαμπος Α. Χαραλαμπίδης

21 Δεκεμβρίου 2009

ΑΝΕΞΑΡΤΗΣΙΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Ορισμός

(α) Έστω (X, Y) *διακριτή διδιάστατη τυχαία μεταβλητή με συνάρτηση πιθανότητας* $f_{X,Y}(x_i, y_j)$, $i, j = 0, 1, \ldots,$ *και περιθώριες συναρτήσεις πιθανότητας* $f_X(x_i)$, $i = 0, 1, \ldots,$ *και* $f_Y(y_j)$, $j = 0, 1, \ldots$.

Οι τυχαίες μεταβλητές X *και* Y *καλούνται (στοχαστικά) ανεξάρτητες αν και μόνο αν*

$$f_{X,Y}(x_i, y_j) = f_X(x_i) f_Y(y_j), \quad i, j = 0, 1, \ldots . \tag{1}$$

(β) Έστω (X, Y) *συνεχής διδιάστατη τυχαία μεταβλητή με συνάρτηση πυκνότητας* $f_{X,Y}(x, y)$, $x, y \in R$ *και περιθώριες συναρτήσεις πυκνότητας* $f_X(x)$, $x \in R$ *και* $f_Y(y)$, $y \in R$.

Οι τυχαίες μεταβλητές X *και* Y *καλούνται (στοχαστικά) ανεξάρτητες αν και μόνο αν*

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \quad x, y \in R. \tag{2}$$

ΑΝΕΞΑΡΤΗΣΙΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Επισημαίνουμε ότι, σύμφωνα με τον ορισμό, αν υπάρχει έστω και ένα ζεύγος τιμών (x_0, y_0) με $f_{X,Y}(x_0, y_0) \neq f_X(x_0) f_Y(y_0)$, τότε οι τυχαίες μεταβλητές X και Y είναι (στοχαστικά) εξαρτημένες.

Η έννοια της στοχαστικής ανεξαρτησίας δύο τυχαίων μεταβλητών επεκτείνεται άμεσα και σε ν τυχαίες μεταβλητές.

ΑΝΕΞΑΡΤΗΣΙΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Ας θεωρήσουμε την από κοινού συνάρτηση πιθανότητας υ τυχαίων μεταβλητών $X_1, X_2, \ldots, X_\upsilon$,

$$f(x_1, x_2, \ldots, x_\upsilon) = P(X_1 = x_1, X_2 = x_2, \ldots, X_\upsilon = x_\upsilon), \; x_i \in R_{X_i}, \; i = 1, 2, \ldots, \upsilon,$$

στην περίπτωση που αυτές είναι διακριτές ή την από κοινού συνάρτηση πυκνότητας,

$$f(x_1, x_2, \ldots, x_\upsilon) \geq 0, \quad x_i \in R_{X_i}, \quad i = 1, 2, \ldots, \upsilon,$$

στην περίπτωση που αυτές είναι συνεχείς. Τότε οι τυχαίες μεταβλητές $X_1, X_2, \ldots, X_\upsilon$ καλούνται (στοχαστικά) ανεξάρτητες αν και μόνο αν

$$f(x_1, x_2, \ldots, x_\upsilon) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_\upsilon}(x_\upsilon), \tag{3}$$

για $x_i \in R_{X_i}$, $i = 1, 2, \ldots, \upsilon$.

Δεσμευμένες κατανομές και στοχαστική ανεξαρτησία.
Έστω X και Y διακριτές τυχαίες μεταβλητές με από κοινού συνάρτηση πιθανότητας $f_{X,Y}(x_i, y_j)$, $i, j = 0, 1, \ldots,$ και περιθώριες συναρτήσεις πιθανότητας $f_X(x_i)$, $i = 0, 1, \ldots,$ και $f_Y(y_j)$, $j = 0, 1, \ldots$.
(α) Αν η δεσμευμένη συνάρτηση πιθανότητας της τ.μ. Y δεδομένης της $X = x_i$, $f_{Y|X}(y_j|x_i) = f_{X,Y}(x_i, y_j)/f_X(x_i)$, $j = 0, 1, \ldots,$ $i = 0, 1, \ldots,$ είναι

$$f_{Y|X}(y_j|x_i) = f_Y(y_j), \quad j = 0, 1, \ldots, \quad i = 0, 1, \ldots,$$

τότε, εξισώνοντας της δύο αυτές εκφράσεις, συμπεραίνουμε ότι

$$f_{X,Y}(x_i, y_j) = f_X(x_i) f_Y(y_j), \quad i = 0, 1, \ldots, \quad j = 0, 1, \ldots,$$

και επομένως οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες.
Ομοίως, αν $f_{X|Y}(x_i|y_j) = f_X(x_i)$, $i = 0, 1, \ldots,$ $j = 0, 1, \ldots,$ τότε οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες.

ΑΝΕΞΑΡΤΗΣΙΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

(β) Αν οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες, οπότε ισχύει η σχέση

$$f_{X,Y}(x_i, y_j) = f_X(x_i) f_Y(y_j), \quad i = 0, 1, \ldots, \quad j = 0, 1, \ldots,$$

τότε, η δεσμευμένη συνάρτηση πιθανότητας, $f_{Y|X}(y_j|x_i) = f_{X,Y}(x_i, y_j)/f_X(x_i), j = 0, 1, \ldots, i = 0, 1, \ldots,$ είναι

$$f_{Y|X}(y_j|x_i) = f_Y(y_j), \; j = 0, 1, \ldots, \; i = 0, 1, \ldots .$$

Ομοίως, συμπεραίνουμε ότι $f_{X|Y}(x_i|y_j) = f_X(x_i), i = 0, 1, \ldots,$ $j = 0, 1, \ldots .$
Επομένως, οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες αν και μόνο αν οι δεσμευμένες συναρτήσεις πιθανότητας δεν εξαρτώνται από την τιμή της δεσμεύουσας τυχαίας μεταβλητής.

Ανάλογη παρατήρηση ισχύει και στην περίπτωση που οι X και Y είναι συνεχείς τυχαίες μεταβλητές.

Παράδειγμα

Ας θεωρήσουμε μία κληρωτίδα που περιέχει 5 αριθμημένα σφαιρίδια $\{1, 2, 3, 4, 5\}$ *από τα οποία το* $\{1\}$ *είναι άσπρο, τα* $\{2, 3\}$ *είναι μαύρα και τα* $\{4, 5\}$ *κόκκινα.*

Έστω ότι εξάγονται στην τύχη δύο σφαιρίδια

(α) χωρίς επανάθεση και

(β) με επανάθεση.

Αν X *είναι ο αριθμός των εξαγομένων άσπρων σφαιριδίων και* Y *ο αριθμός των εξαγομένων μαύρων σφαιριδίων, να υπολογισθούν η από κοινού και οι περιθώριες συναρτήσεις πιθανότητας των τυχαίων μεταβλητών* X *και* Y *και να εξετασθεί αν είναι ανεξάρτητες.*

(α) Ο δειγματικός χώρος, στην περίπτωση που η εξαγωγή των σφαιριδίων γίνεται χωρίς επανάθεση, είναι το σύνολο

$$\begin{aligned}\Omega = \{&(1,2),(1,3),(1,4),(1,5),(2,1),(2,3),(2,4),(2,5),(3,1),(3,2),\\&(3,4),(3,5),(4,1),(4,2),(4,3),(4,5),(5,1),(5,2),(5,3),(5,4)\}\end{aligned}$$

των 20 διατάξεων των 5 ανά 2. Τα 20 αυτά δειγματικά σημεία είναι ισοπίθανα. Η από κοινού συνάρτηση πιθανότητας $f(x,y)$ και οι περιθώριες συναρτήσεις πιθανότητας $f_X(x)$ και $f_Y(y)$ υπολογίζονται σύμφωνα με τον Ορισμό και δίνονται στον ακόλουθο Πίνακα.

ΑΝΕΞΑΡΤΗΣΙΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

$f_{X,Y}(x,y)$				
$x \backslash y$	0	1	2	$f_X(x)$
0	$1/10$	$4/10$	$1/10$	$3/5$
1	$2/10$	$2/10$	0	$2/5$
$f_Y(y)$	$3/10$	$6/10$	$1/10$	1

Πίνακας 5.4. Από κοινού και περιθώριες συναρτήσεις πιθανότητας

Παρατηρούμε ότι $f(x, y) \neq f_X(x) f_Y(y)$ και επομένως οι τυχαίες μεταβλητές X και Y δεν είναι ανεξάρτητες.

(β) Ο δειγματικός χώρος, στην περίπτωση που η εξαγωγή των σφαιριδίων γίνεται με επανάθεση, είναι το σύνολο

$$\begin{aligned}\Omega = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(2,1),(2,2),(2,3),(2,4),(2,5),\\ &(3,1),(3,2),(3,3),(3,4),(3,5),(4,1),(4,2),(4,3),(4,4),(4,5),\\ &(5,1),(5,2),(5,3),(5,4),(5,5)\}\end{aligned}$$

των 25 διατάξεων των 5 ανά 2 με επανάληψη. Τα 25 αυτά δειγματικά σημεία είναι ισοπίθανα. Η από κοινού συνάρτηση πιθανότητας $f(x,y)$ και οι περιθώριες συναρτήσεις πιθανότητας $f_X(x)$ και $f_Y(y)$ υπολογίζονται σύμφωνα με τον Ορισμό και δίνονται στον ακόλουθο Πίνακα.

ΑΝΕΞΑΡΤΗΣΙΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

$f_{X,Y}(x,y)$				
$x \backslash y$	0	1	2	$f_X(x)$
0	4/25	8/25	4/25	16/25
1	4/25	4/25	0	8/25
2	1/25	0	0	1/25
$f_Y(y)$	3/10	6/10	1/10	1

Πίνακας 5.5. Από κοινού και περιθώριες συναρτήσεις πιθανότητας

Παρατηρούμε ότι $f(x,y) \neq f_X(x) f_Y(y)$ και επομένως οι τυχαίες μεταβλητές X και Y δεν είναι ανεξάρτητες.

Παράδειγμα

Η κατανομή του ήχου σε διάφορα σημεία μιας αίθουσας διδασκαλίας είναι χρήσιμη στη μελέτη της ηχητικής της αίθουσας.
Έστω η συνάρτηση πυκνότητας του ήχου στο σημείο (X, Y) δίνεται από την

$$f_{X,Y}(x,y) = xye^{-(x^2+y^2)/2}, \quad 0 < x < \infty, \ \ 0 < y < \infty.$$

Να εξετασθεί κατά πόσον οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες.

ΑΝΕΞΑΡΤΗΣΙΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Η περιθώρια συνάρτηση πυκνότητας της X είναι η

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = xe^{-x^2/2} \int_0^{\infty} ye^{-y^2/2}dy = xe^{-x^2/2}, \ 0 < x < \infty.$$

Ομοίως

$$f_Y(y) = ye^{-y^2/2}, \ 0 < y < \infty.$$

Επομένως

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \quad 0 < x < \infty, \ 0 < y < \infty,$$

και οι X και Y είναι ανεξάρτητες.

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Ας θεωρήσουμε μία διδιάστατη τυχαία μεταβλητή (X, Y) και μία συνάρτηση αυτής $Z = g(X, Y)$.
Η Z είναι μία (μονοδιάστατη) τυχαία μεταβλητή και η συνάρτηση πιθανότητας $f_Z(z_\kappa)$, $\kappa = 0, 1, \ldots,$ ή πυκνότητας $f_Z(z)$, $z \in R$, αυτής προσδιορίζεται μέσω της συνάρτησης πιθανότητας $f_{X,Y}(x_i, y_j)$, $i, j = 0, 1, \ldots,$ ή πυκνότητας $f_{X,Y}(x, y)$, $x, y \in R$, της διδιάστατης τυχαίας μεταβλητής (X, Y).
Είναι επομένως ενδιαφέρον και έχει έννοια ο υπολογισμός της μέσης τιμής της $Z = g(X, Y)$.
Τούτο δεν είναι αναγκαίο να γίνεται σε κάθε μερική περίπτωση.

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Σχετικά ισχύει η ακόλουθη έκφραση

$$E(Z) \equiv E[g(X,Y)] = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} g(x_i, y_j) f_{X,Y}(x_i, y_j) \tag{4}$$

αν η διδιάστατη τυχαία μεταβλητή (X, Y) είναι διακριτή και η έκφραση

$$E(Z) \equiv E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dxdy \tag{5}$$

αν η (X, Y) είναι συνεχής.

Χρησιμοποιώντας τις (4) και (5) προκύπτει άμεσα η σχέση

$$E[g(X,Y)+h(X,Y)] = E[g(X,Y)] + E[h(X,Y)] \tag{6}$$

και ως μερική περίπτωση αυτής συνάγουμε τη σχέση

$$E[g(X)+h(Y)] = E[g(X)] + E[h(Y)]. \tag{7}$$

Θέτοντας $g(X) = \alpha X$ και $h(Y) = \beta Y$, με α και β σταθερές, συμπεραίνουμε ότι

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y). \tag{8}$$

Επίσης, αν X και Y ανεξάρτητες τυχαίες μεταβλητές, τότε χρησιμοποιώντας (4) και (5), συνάγουμε άμεσα τη σχέση

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)], \tag{9}$$

και ειδικότερα τη σχέση

$$E(XY) = E(X)E(Y). \tag{10}$$

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Οι ροπές μιας διδιάστατης τυχαίας μεταβλητής ορίζονται ως μέση τιμή μιας συγκεκριμένης, σε κάθε περίπτωση, συνάρτησης αυτής.
Η από κοινού διακύμανση (διασπορά) δύο τυχαίων μεταβλητών εισάγεται στον ακόλουθο ορισμό.

Ορισμός

Έστω (X, Y) μία διδιάστατη τυχαία μεταβλητή με $\mu_X = E(X)$ και $\mu_Y = E(Y)$. Η συνδιακύμανση (ή συνδιασπορά) των τυχαίων μεταβλητών X και Y, συμβολιζόμενη με $C(X, Y)$ ή $\sigma_{X,Y}$, ορίζεται από τη σχέση

$$\sigma_{X,Y} \equiv C(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]. \tag{11}$$

Σύμφωνα με τον ορισμό, συμπεραίνουμε ως μερική περίπτωση τη σχέση $C(X, X) = V(X)$.

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Θεώρημα

(α) *Αν* X *και* Y *είναι τυχαίες μεταβλητές και* α, β, γ, *και* δ *είναι οποιεσδήποτε σταθερές, τότε*

$$C(\alpha X + \beta, \gamma Y + \delta) = \alpha\gamma\, C(X, Y). \tag{12}$$

(β) *Η συνδιακύμανση δύο τυχαίων μεταβλητών* X *και* Y *εκφράζεται ως*

$$C(X, Y) = E(XY) - E(X)E(Y). \tag{13}$$

Απόδειξη.

(α) Σύμφωνα με τον ορισμό και χρησιμοποιώντας τη γραμμικότητα της μέσης τιμής,

$$E(aX+\beta)=a\mu_X+\beta, \quad E(\gamma Y+\delta)=\gamma\mu_Y+\delta,$$

συνάγουμε διαδοχικά

$$\begin{aligned} C(aX+\beta\,,\ \gamma Y+\delta) &= E\{[aX+\beta-E(aX+\beta)]\cdot[\gamma Y+\delta-E(\gamma Y+\delta)]\} \\ &= E[(aX+\beta-a\mu_X-\beta)\cdot(\gamma Y+\delta-\gamma\mu_Y-\delta)] \\ &= a\gamma E[(X-\mu_X)(Y-\mu_Y)] = a\gamma\, C(X,Y). \end{aligned}$$

□

(β) Χρησιμοποιώντας τις (6) και (8) παίρνουμε

$$\begin{aligned} C(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] = E(XY-\mu_X Y-\mu_Y X+\mu_X\mu_Y) \\ &= E(XY)+E(-\mu_X Y-\mu_Y X+\mu_X\mu_Y) \\ &= E(XY)-\mu_X E(Y)-\mu_Y E(X)+\mu_X\mu_Y = E(XY)-E(X)E(Y). \end{aligned}$$

Παρατηρούμε ότι αν οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες, τότε $C(X,Y)=E(XY)-E(X)E(Y)=0$.
Το αντίστροφο δεν ισχύει γενικά.

Οι τυχαίες μεταβλητές με συνδιακύμανση μηδέν παρουσιάζουν ιδιαίτερο ενδιαφέρον και έτσι θεωρείται σκόπιμος ο ακόλουθος ορισμός.

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Θεώρημα

Έστω X *και* Y *οποιεσδήποτε τυχαίες μεταβλητές και* α *και* β *σταθερές. Τότε*

$$V(\alpha X + \beta Y) = \alpha^2 V(X) + \beta^2 V(Y) + 2\alpha\beta\, C(X, Y) \tag{14}$$

και ειδικότερα

$$V(X \pm Y) = V(X) + V(Y) \pm 2C(X, Y).$$

Αν οι τυχαίες μεταβλητές X *και* Y *είναι ανεξάρτητες* $\big($*ή απλώς ασυσχέτιστες*$\big)$*, τότε*

$$V(\alpha X + \beta Y) = \alpha^2 V(X) + \beta^2 V(Y) \tag{15}$$

και ειδικότερα

$$V(X \pm Y) = V(X) + V(Y).$$

Απόδειξη.

Χρησιμοποιώντας διαδοχικά την (6), παίρνουμε τη σχέση

$$
\begin{aligned}
V(\alpha X+\beta Y) &= E\{[(\alpha X+\beta Y)-E(\alpha X+\beta Y)]^2\}\\
&= E\{[(\alpha X+\beta Y)-(\alpha\mu_X+\beta\mu_Y)]^2\}\\
&= E\{[\alpha(X-\mu_X)+\beta(Y-\mu_Y)]^2\}\\
&= E[\alpha^2(X-\mu_X)^2+\beta^2(Y-\mu_Y)^2+2\alpha\beta(X-\mu_X)(Y-\mu_Y)]\\
&= \alpha^2E[(X-\mu_X)^2]+\beta^2E[(Y-\mu_Y)^2]+2\alpha\beta E[(X-\mu_X)(Y-\mu_Y)]
\end{aligned}
$$

από την οποία συμπεραίνουμε άμεσα την (14).
Στην περίπτωση που οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες (ή απλώς ασυσχέτιστες), $C(X,Y)=0$ και έτσι από την (14) προκύπτει η (15). $\square$

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Ορισμός

Έστω (X, Y) *μία διδιάστατη τυχαία μεταβλητή με συνδιακύμανση* $\sigma_{X,Y} = C(X, Y)$ *και διασπορές* $\sigma_X^2 = V(X)$ *και* $\sigma_Y^2 = V(Y)$.
Ο συντελεστής συσχέτισης των τυχαίων μεταβλητών X *και* Y*, συμβολιζόμενος με* $\rho(X, Y)$ *ή* $\rho_{X,Y}$ *ή, απλώς,* ρ*, ορίζεται από τη σχέση*

$$\rho \equiv \rho(X, Y) = \frac{C(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}. \tag{16}$$

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Θεώρημα

Αν X και Y είναι τυχαίες μεταβλητές και α, β, γ και δ είναι σταθερές, με $\alpha\gamma \neq 0$, τότε

$$\rho(\alpha X+\beta\,,\,\gamma Y+\delta) = \begin{cases} \rho(X,Y), & \text{αν } \alpha\gamma > 0, \\ -\rho(X,Y), & \text{αν } \alpha\gamma < 0. \end{cases} \tag{17}$$

Απόδειξη.

Χρησιμοποιώντας την (12) και τις σχέσεις $V(\alpha X+\beta) = \alpha^2 V(X)$, $V(\gamma Y+\delta) = \gamma^2 V(Y)$, συνάγουμε τη σχέση

$$\begin{aligned} \rho(\alpha X+\beta\,,\,\gamma Y+\delta) &= \frac{C(\alpha X+\beta\,,\,\gamma Y+\delta)}{\sqrt{V(\alpha X+\beta)}\,\sqrt{V(\gamma Y+\delta)}} \\ &= \frac{\alpha\gamma}{|\alpha|\cdot|\gamma|}\,\frac{C(X,Y)}{\sqrt{V(X)}\,\sqrt{V(Y)}} = \frac{\alpha\gamma}{|\alpha\gamma|}\rho(X,Y), \end{aligned}$$

η οποία είναι ισοδύναμη με την (17). $\square$

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Θεώρημα

Αν $\rho = \rho(X, Y)$ είναι ο συντελεστής συσχέτισης δύο τυχαίων μεταβλητών X και Y, τότε

$$-1 \leq \rho \leq 1. \tag{18}$$

Απόδειξη.

Ο συντελεστής συσχέτισης των τυχαίων μεταβλητών X και Y, σύμφωνα με τις (11) και (16), είναι

$$\rho \equiv \rho(X, Y) = \frac{C(X, Y)}{\sqrt{V(X)}\,\sqrt{V(Y)}} = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{\sigma_Y}\right)\right],$$

και εισάγοντας τις τυποποιημένες τυχαίες μεταβλητές

$$Z = \frac{X - \mu_X}{\sigma_X}, \quad W = \frac{Y - \mu_Y}{\sigma_Y},$$

δύναται να εκφρασθεί ως

$$\rho \equiv \rho(X, Y) = E(ZW). \tag{19}$$

Ας θεωρήσουμε την τυχαία μεταβλητή $U = tZ - W$, για $t \in R$. Τότε

$$E(U^2) = t^2 E(Z^2) - 2tE(ZW) + E(W^2) = t^2 - 2tE(ZW) + 1, \quad t \in R$$

εφ' όσον $E(Z^2) = V(Z) = 1$ και $E(W^2) = V(W) = 1$. Όμως $E(U^2) \geq 0$ και έτσι

$$t^2 - 2E(ZW)t + 1 \geq 0, \quad t \in R,$$

Το τριώνυμο αυτό δεν αλλάζει πρόσημο αν και μόνο αν η διακρίνουσά του δεν είναι θετική και επομένως

$$[E(ZW)]^2 - 1 \leq 0,$$

ή ισοδύναμα

$$-1 \leq E(ZW) \leq 1.$$

Η τελευταία σχέση, λόγω της (19), συνεπάγεται την (18).

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Η σημασία του συντελεστή συσχέτισης δύο τυχαίων μεταβλητών ως μέτρου συσχέτισης (γραμμικής εξάρτησης) τούτων απορρέει από το ακόλουθο θεώρημα.

Θεώρημα

Έστω X και Y δύο τυχαίες μεταβλητές με συντελεστή συσχέτισης $\rho = \rho(X, Y)$. Τότε $\rho = \pm 1$ αν και μόνο αν υπάρχουν πραγματικοί αριθμοί α και β, με $\alpha \neq 0$, τέτοιοι ώστε, με πιθανότητα 1, να ισχύει η γραμμική σχέση

$$Y = \alpha X + \beta. \tag{20}$$

Απόδειξη.

Αν ισχύει η γραμμική σχέση $Y = aX + \beta$, με πιθανότητα 1, τότε, σύμφωνα με την (12), $C(X,Y) = C(X, aX+\beta) = aC(X,X) = aV(X)$ και έτσι

$$\rho(X,Y) = \frac{C(X, aX+\beta)}{\sqrt{V(X)}\sqrt{V(aX+\beta)}} = \frac{aV(X)}{\sqrt{V(X)}\sqrt{a^2V(X)}} = \frac{a}{|a|} = \pm 1,$$

και μάλιστα $\rho = 1$ όταν $a > 0$, ενώ $\rho = -1$ όταν $a < 0$.

Αντίστροφα, ας θεωρήσουμε τις αντίστοιχες των X και Y τυποποιημένες τυχαίες μεταβλητές

$$Z = \frac{X - \mu_X}{\sigma_X} \quad \text{και} \quad W = \frac{Y - \mu_Y}{\sigma_Y},$$

και την $U = \rho Z - W$. □

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Τότε $E(U) = \rho E(Z) - E(W) = 0$ και

$V(U) = E(U^2) = \rho^2 E(Z^2) - 2\rho E(ZW) + E(W^2) = \rho^2 - 2\rho^2 + 1 = 1 - \rho^2$,

εφ' όσον $E(ZW) = \rho(X, Y) = \rho$. Έτσι, αν $\rho = \pm 1$, τότε $V(U) = 0$ και επειδή $E(U) = 0$, έπεται ότι $P(U = 0) = 1$.
Επομένως, αν $\rho = \pm 1$, τότε, με πιθανότητα 1, ισχύει η σχέση $W = \pm Z$ ή, ισοδύναμα, η σχέση

$$Y = \mu_Y \pm \frac{\sigma_Y}{\sigma_X}(X - \mu_X).$$

Συμπερασματικά, αν $\rho = 1$, η (20) ισχύει με $a = \sigma_Y / \sigma_X > 0$ και $\beta = \mu_Y - \mu_X \sigma_Y / \sigma_X$, ενώ αν $\rho = -1$, η (20) ισχύει με $a = -\sigma_Y / \sigma_X < 0$ και $\beta = \mu_Y + \mu_X \sigma_Y / \sigma_X$.

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Παράδειγμα

Έστω (X, Y) *μία διακριτή διδιάστατη τυχαία μεταβλητή με συνάρτηση πιθανότητας*

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

Να υπολογισθεί ο συντελεστής συσχέτισης των τυχαίων μεταβλητών X *και* Y.

Η μέση τιμή του γινομένου των τυχαίων μεταβλητών X *και* Y*, σύμφωνα με την (4), είναι*

$$E(XY) = \sum_{y=1}^{3}\sum_{x=1}^{2} xy\frac{x+y}{21} = \sum_{y=1}^{3}\frac{y(1+y)+2y(2+y)}{21}$$

$$= \sum_{y=1}^{3}\frac{y(3y+5)}{21} = \frac{24}{7}.$$

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Οι περιθώριες συναρτήσεις πιθανότητας των X και Y δίνονται από τις

$$f_X(x) = \frac{x+2}{7}, \quad x = 1, 2, \quad f_Y(y) = \frac{2y+3}{21}, \quad y = 1, 2, 3.$$

Συνεπώς,

$$E(X) = \sum_{x=1}^{2} x \, \frac{x+2}{7} = \frac{11}{7}, \quad E(X^2) = \sum_{x=1}^{2} x^2 \, \frac{x+2}{7} = \frac{19}{7},$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{19}{7} - \left(\frac{11}{7}\right)^2 = \frac{12}{49}$$

και

$$E(Y) = \sum_{y=1}^{3} y \, \frac{2y+3}{21} = \frac{46}{21}, \quad E(Y^2) = \sum_{y=1}^{3} y^2 \, \frac{2y+3}{21} = \frac{114}{21},$$

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{114}{21} - \left(\frac{46}{21}\right)^2 = \frac{278}{441}.$$

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Η συνδιακύμανση των τυχαίων μεταβλητών X και Y είναι ίση με

$$C(X,Y) = E(XY) - E(X)E(Y) = \frac{24}{7} - \frac{11}{7} \cdot \frac{46}{21} = -\frac{2}{147}$$

και επομένως

$$\rho(X,Y) = \frac{C(X,Y)}{\sqrt{V(X)}\,\sqrt{V(Y)}} = \frac{-2/147}{\sqrt{12/49}\,\sqrt{278/441}} = -\frac{1}{\sqrt{834}} \cong -0{,}03.$$

Η μικρή (κοντά στο 0) αυτή τιμή του ρ δείχνει ότι υπάρχει πολύ ασθενής (αμελητέα) αρνητική γραμμική εξάρτηση των X και Y.

Παράδειγμα

Έστω (X, Y) μια συνεχής διδιάστατη τυχαία μεταβλητή με συνάρτηση πυκνότητας

$$f_{X,Y}(x, y) = 2, \quad 0 \leq x \leq y \leq 1.$$

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

Η μέση τιμή του γινομένου των τυχαίων μεταβλητών X και Y, σύμφωνα με την (5), είναι

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{X,Y}(x, y) dxdy = 2\int_0^1 \int_0^y xydxdy = \int_0^1 y^3 dy = \frac{1}{4}.$$

Οι περιθώριες συναρτήσεις πυκνότητας των X και Y δίνονται από τις

$$f_X(x) = 2(1-x),\ \ 0 \leq x \leq 1,\ \ f_Y(y) = 2y,\ \ 0 \leq y \leq 1,$$

και έτσι

$$E(X^r) = \int_{-\infty}^{\infty} x^r f_X(x)dx = 2\int_0^1 x^r(1-x)dx = \frac{2}{(r+1)(r+2)},\ \ r = 1, 2, \ldots,$$

$$E(Y^r) = \int_{-\infty}^{\infty} y^r f_Y(y)dy = 2\int_0^1 y^{r+1}dy = \frac{2}{r+2},\ \ r = 1, 2, \ldots,$$

οπότε

$$E(X) = \frac{1}{3},\ \ V(X) = E(X^2) - [E(X)]^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18},$$

$$E(Y) = \frac{2}{3},\ \ V(Y) = E(Y^2) - [E(Y)]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Η συνδιακύμανση των X και Y, σύμφωνα με την (13), είναι ίση με

$$C(X, Y) = E(XY) - E(X)E(Y) = \frac{1}{4} - \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{36}$$

και επομένως

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{1/36}{\sqrt{1/18}\sqrt{1/18}} = \frac{1}{2}.$$