ΚΑΤΑΝΟΜΕΣ ΔΙΔΙΑΣΤΑΤΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ (Συνέχεια)

Χαράλαμπος Α. Χαραλαμπίδης

11 Ιανουαρίου 2010

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Η **δεσμευμένη μέση τιμή** μιας τυχαίας μεταβλητής Y σε δεδομένο σημείο μιας άλλης τυχαίας μεταβλητής $X = x$, συμβολιζόμενη με $E(Y|x)$ ή $m_{Y|X}(x)$ ή απλώς $m(x)$, αν δεν υπάρχει κίνδυνος σύγχυσης, δίνεται από τη σχέση

$$m_{Y|X}(x) \equiv E(Y|x) = \sum_{j=0}^{\infty} y_j \, f_{Y|X}(y_j|x), \tag{1}$$

αν η (X, Y) είναι διακριτή, και από τη σχέση

$$m_{Y|X}(x) \equiv E(Y|x) = \int_{-\infty}^{\infty} y \, f_{Y|X}(y|x) dy, \tag{2}$$

αν η (X, Y) είναι συνεχής.

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Σημειώνουμε ότι η δεσμευμένη μέση τιμή είναι μία πραγματική συνάρτηση $m_{Y|X}(x) \equiv E(Y|x)$, η οποία ορίζεται για κάθε $x = x_i$, $i = 0, 1, \ldots,$ αν η τ.μ. X είναι διακριτή, ή για κάθε $x \in R_X$, αν η τ.μ X είναι συνεχής.

Επισημαίνουμε ότι η δεσμευμένη μέση τιμή για συγκεκριμένο x είναι ένας πραγματικός αριθμός.

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Ορισμός

Έστω (X, Y) *μια διδιάστατη τυχαία μεταβλητή και* $m_{Y|X}(x) = E(Y|x)$ *η δεσμευμένη μέση τιμή της* Y *δεδομένης της* $X = x$. *Η καμπύλη με εξίσωση*

$$y = m_{Y|X}(x) \tag{3}$$

καλείται καμπύλη (μέσης) παλινδρόμησης της Y *στην* X. *Επίσης, η συνάρτηση* $m_{Y|X}(x)$ *καλείται συνάρτηση παλινδρόμησης της* Y *στη* X.

Κατά τον ίδιο τρόπο ορίζονται η συνάρτηση παλινδρόμησης της X στην Y: $m_{X|Y}(y) = E(X|y)$ και η καμπύλη (μέσης) παλινδρόμησης της X στην Y:

$$x = m_{X|Y}(y). \tag{4}$$

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Στην (ακραία) περίπτωση που ο συντελεστής συσχέτισης των τυχαίων μεταβλητών X και Y είναι κατ΄ απόλυτη τιμή ίσος με τη μονάδα, $\rho(X, Y) = \pm 1$, τότε, όπως έχουμε ήδη αποδείξει, υπάρχουν πραγματικοί αριθμοί $\alpha \neq 0$ και β τέτοιοι ώστε, με πιθανότητα 1, να ισχύει η σχέση $Y = \alpha X + \beta$. Επομένως

$$m_{Y|X}(x) = E(Y|x) = E(\alpha X + \beta|x) = \alpha x + \beta,$$

$$m_{X|Y}(y) = E(X|y) = E((Y - \beta)/\alpha|y) = (y - \beta)/\alpha,$$

και οι καμπύλες παλινδρόμησης $y = m_{Y|X}(x)$ και $x = m_{X|Y}(y)$ συμπίπτουν με την ευθεία $y = \alpha x + \beta$.

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Αν οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες, τότε

$$m_{Y|X}(x) = E(Y|x) = E(Y) = \mu_Y, \quad m_{X|Y}(y) = E(X|y) = E(X) = \mu_X$$

και οι καμπύλες παλινδρόμησης $y = m_{Y|X}(x)$ και $x = m_{X|Y}(y)$ είναι οι ευθείες $y = \mu_Y$ και $x = \mu_X$, οι οποίες είναι παράλληλες προς τους άξονες x και y, αντίστοιχα, και τέμνονται στο σημείο (μ_X, μ_Y).

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Στη γενική περίπτωση η καμπύλη παλινδρόμησης $y = m_{Y|X}(x)$ δεν είναι κατ' ανάγκη ευθεία και, από άποψη εφαρμογών, είναι συχνά χρήσιμη η προσέγγιση αυτής από μία ευθεία γραμμή $y = \alpha x + \beta$.
Ο υπολογισμός των σταθερών α και β γίνεται με την ελαχιστοποίηση (ως προς α και β) της μέσης τετραγωνικής απόκλισης της Y από την $\alpha X + \beta$,

$$E\{[Y - (\alpha X + \beta)]^2\}. \tag{5}$$

Σημειώνεται ότι η ευθεία αυτή αποτελεί την βέλτιστη προσέγγιση της καμπύλης παλινδρόμησης $y = m_{Y|X}(x)$ με την έννοια της ελαχιστοποίησης της μέσης τετραγωνικής απόκλισης

$$E\{[m_{Y|X}(X) - (\alpha X + \beta)]^2\}. \tag{6}$$

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Θεώρημα

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

$$E(X) = \mu_X, \quad E(Y) = \mu_Y, \quad V(X) = \sigma_X^2, \quad V(Y) = \sigma_Y^2, \quad \rho(X, Y) = \rho.$$

Τότε η ευθεία $y = ax + \beta$ *για την οποία ελαχιστοποιείται η μέση τετραγωνική απόκλιση* (5) *έχει κλίση*

$$a = \rho \frac{\sigma_X}{\sigma_Y} \tag{7}$$

και τέμνει τον άξονα των y στο σημείο

$$\beta = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X. \tag{8}$$

Επιπλέον

$$E\{[Y - (aX + \beta)]^2\} = \sigma_Y^2(1 - \rho^2). \tag{9}$$

Απόδειξη.

Η μέση τετραγωνική απόκλιση (5) θεωρούμενη ως συνάρτηση των a και β, έστω $Q(a,\beta)$, δύναται να γραφεί στη μορφή

$$\begin{aligned}
Q(a,\beta) =& E\{[(Y-\mu_Y) - a(X-\mu_X) + (\mu_Y - a\mu_X - \beta)]^2\} \\
=& a^2 E[(X-\mu_X)^2] + E[(Y-\mu_Y)^2] \\
& - 2aE[(X-\mu_X)(Y-\mu_Y)] + (\mu_Y - a\mu_X - \beta)^2 \\
=& a^2\sigma_X^2 + \sigma_Y^2 - 2a\rho\sigma_X\sigma_Y + (\mu_Y - a\mu_X - \beta)^2,
\end{aligned}$$

Λαμβάνοντας τις μερικές παραγώγους της συνάρτησης $Q(a,\beta)$ ως προς a και β και θέτοντας αυτές ίσες με μηδέν, συνάγουμε το σύστημα των εξισώσεων:

$$a\sigma_X^2 - \rho\sigma_X\sigma_Y - \mu_X(\mu_Y - a\mu_X - \beta) = 0, \quad \mu_Y - a\mu_X - \beta = 0.$$

$\square$

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Η λύση του συστήματος αυτού είναι η

$$a = \rho \frac{\sigma_X}{\sigma_Y}, \quad \beta = \mu_X - \rho \frac{\sigma_X}{\sigma_Y} \mu_Y$$

και ελαχιστοποιεί την συνάρτηση $Q(a, \beta)$.
Εισάγοντας τις τιμές αυτές στη συνάρτηση $Q(a, \beta)$ λαμβάνουμε την ελάχιστη μέση τετραγωνική απόκλιση:

$$\begin{aligned} Q(a,\beta) =& E\left\{\left[(Y - \mu_Y) - \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X)\right]^2\right\} \\ =& E[(Y - \mu_Y)^2] + \rho^2 \frac{\sigma_Y^2}{\sigma_X^2} E[(X - \mu_X)^2] - 2\rho \frac{\sigma_Y}{\sigma_X} E[(X - \mu_X)(Y - \mu_Y)] \\ =& \sigma_Y^2(1 - \rho^2). \end{aligned}$$

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Ορισμός

Έστω (X, Y) μία διδιάστατη τυχαία μεταβλητή με

$$E(X) = \mu_X, \quad E(Y) = \mu_Y, \quad V(X) = \sigma_X^2, \quad V(Y) = \sigma_Y^2, \quad \rho(X, Y) = \rho.$$

Η ευθεία

$$y = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

καλείται ευθεία γραμμικής παλινδρόμησης της Y στη X.
Η ελάχιστη μέση τετραγωνική απόκλιση

$$E\left\{\left[(Y - \mu_Y) - \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X)\right]^2\right\} = \sigma_Y^2(1 - \rho^2)$$

καλείται μέσο τετραγωνικό σφάλμα της γραμμικής παλινδρόμησης της Y στη X ή υπόλοιπο διασποράς.

Παράδειγμα

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

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

Να υπολογισθούν (α) οι καμπύλες (μέσης) παλινδρόμησης, (β) οι ευθείες γραμμικής παλινδρόμησης και (γ) να συγκριθούν.

(α) Οι δεσμευμένες συναρτήσεις πυκνότητας της Y δεδομένης της $X = x$ και της X δεδομένης της $Y = y$ δίνονται από τις

$$f_{Y|X}(y|x) = \frac{1}{1-x}, \;\; x \leq y \leq 1, \;\; (0 \leq x < 1),$$

$$f_{X|Y}(x|y) = \frac{1}{y}, \;\; 0 \leq x \leq y, \;\; (0 < y \leq 1).$$

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Επομένως

$$m_{Y|X}(x) = E(Y|x) = \frac{1}{1-x}\int_x^1 y dy = \frac{1-x^2}{2(1-x)} = \frac{1+x}{2},$$

$$m_{X|Y}(y) = E(X|y) = \frac{1}{y}\int_0^y x dx = \frac{y^2}{2y} = \frac{y}{2}$$

και οι καμπύλες παλινδρόμησης της Y στη X και της X στην Y είναι οι

$$y = \frac{x+1}{2}, \;\; 0 \leq x < 1$$

και

$$x = \frac{y}{2}, \;\; 0 \leq y < 1.$$

αντίστοιχα.

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

(β) Οι μέσες τιμές, οι διασπορές και ο συντελεστής συσχέτισης των τυχαίων μεταβλητών X και Y δίνονται από τις (βλ. Παράδειγμα ;;)

$$\mu_X = \frac{1}{3}, \quad \mu_Y = \frac{2}{3}, \quad \sigma_X^2 = \sigma_Y^2 = \frac{1}{18}, \quad \rho = \frac{1}{2}.$$

Η ευθεία γραμμικής παλινδρόμησης της Y στη X, $y = ax + \beta$, σύμφωνα με τις (7) και (8), έχει κλίση

$$a = \rho \frac{\sigma_Y}{\sigma_X} = \frac{1}{2},$$

και τέμνει τον άξονα των y στο σημείο

$$\beta = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X = \frac{2}{3} - \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2}.$$

Επομένως η ευθεία αυτή είναι η

$$y = \frac{x+1}{2}, \quad 0 \le x < 1$$

ΚΑΜΠΥΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Ομοίως, για την ευθεία γραμμικής παλινδρόμησης της X στην Y, έστω $x = \gamma y + \delta$, παίρνουμε

$$\gamma = \rho \frac{\sigma_X}{\sigma_Y} = \frac{1}{2}, \quad \delta = \mu_X - \rho \frac{\sigma_X}{\sigma_Y} \mu_Y = \frac{1}{3} - \frac{1}{2} \cdot \frac{2}{3} = 0.$$

και έτσι

$$x = \frac{y}{2}, \quad 0 \leq y < 1.$$

(γ) Παρατηρούμε ότι οι καμπύλες παλινδρόμησης της Y στη X και της X στην Y είναι ευθείες και (αναγκαστικά) συμπίπτουν με τις αντίστοιχες ευθείες γραμμικής παλινδρόμησης της Y στη X και της X στην Y.

5.11 Η από κοινού συνάρτηση πιθανότητας $f(x, y)$ των τυχαίων μεταβλητών X και Y δίνεται από τις σχέσεις

$$f(1,1) = f(2,1) = \frac{1}{8}, \quad f(1,2) = \frac{1}{4}, \quad f(2,2) = \frac{1}{2}.$$

(α) Να βρεθούν οι περιθώριες συναρτήσεις πιθανότητας $f_X(x)$ και $f_Y(y)$, οι δεσμευμένες συναρτήσεις πιθανότητας $f_{X|Y}(x|y)$ και $f_{Y|X}(y|x)$ και να εξετασθεί κατά πόσον οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες.
(β) Να υπολογισθούν οι μέσες τιμές $E(X)$ και $E(Y)$ και οι δεσμευμένες μέσες τιμές $E(X|y)$, $y = 1, 2$, και $E(Y|x)$, $x = 1, 2$.

ΑΣΚΗΣΕΙΣ

(α)

$f(x,y)$			
$x \backslash y$	1	2	$f_X(x)$
1	1/8	1/4	3/8
2	1/8	1/2	5/8
$f_Y(y)$	1/4	3/4	1

$f_{X\mid Y}(x\mid y)$		
$x \backslash y$	1	2
1	1/2	1/3
2	1/2	2/3

$f_{Y\mid X}(y\mid x)$		
$x \backslash y$	1	2
1	1/3	2/3
2	1/5	4/5

Επειδή $f(1,1) \neq f_X(1)f_Y(1)$ οι τ.μ. X και Y δεν είναι ανεξάρτητες.

ΑΣΚΗΣΕΙΣ

$$E(X) = 1 \cdot \frac{3}{8} + 2 \cdot \frac{5}{8} = \frac{13}{8}, \;\; E(Y) = 1 \cdot \frac{1}{4} + 2 \cdot \frac{3}{4} = \frac{7}{4},$$

$$E(X|Y=1) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}, \;\; E(X|Y=2) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{3} = \frac{5}{3},$$

$$E(Y|X=1) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{3} = \frac{5}{3}, \;\; E(Y|X=2) = 1 \cdot \frac{1}{5} + 2 \cdot \frac{4}{5} = \frac{9}{5}.$$

5.13 Έστω ότι η διδιάστατη τυχαία μεταβλητή (X, Y) κατανέμεται ομοιόμορφα στο τρίγωνο $0 < x < y < \vartheta$, όπου $\vartheta > 0$ είναι παράμετρος.
(α) Δείξτε ότι η συνάρτηση πυκνότητας αυτής δίνεται από την

$$f(x, y) = \frac{2}{\vartheta^2}, \quad 0 < x < y < \vartheta.$$

(β) Υπολογίστε το συντελεστή συσχέτισης $\rho(X, Y)$ και τις δεσμευμένες μέσες τιμές $E(X|y), 0 < y < \vartheta$ και $E(Y|x), 0 < x < \vartheta$.

ΑΣΚΗΣΕΙΣ

(α) Το ορθογώνιο ισοσκελές τρίγωνο

$$\Omega = \{(x, y) \in R^2 : 0 < x < y < \partial\}$$

έχει εμβαδό $\mu(\Omega) = \partial^2/2$ και έτσι

$$f(x, y) = \frac{1}{\mu(\Omega)} = \frac{2}{\partial^2}, \quad 0 < x < y < \partial.$$

(β)

$$f_X(x) = \frac{2}{\partial^2}\int_x^{\partial} dy = \frac{2(\partial - x)}{\partial^2}, \quad 0 < x < \partial,$$

$$f_Y(y) = \frac{2}{\partial^2}\int_0^{y} dx = \frac{2y}{\partial^2}, \quad 0 < y < \partial,$$

ΑΣΚΗΣΕΙΣ

$$E(XY) = \frac{2}{\partial^2}\int_0^{\partial} y\Big\{\int_0^{y} x dx\Big\}dy = \frac{2}{\partial^2}\int_0^{\partial} y\Big[\frac{x^2}{2}\Big]_0^y dy$$

$$= \frac{1}{\partial^2}\int_0^{\partial} y^3 dy = \frac{1}{\partial^2}\Big[\frac{y^4}{4}\Big]_0^{\partial} = \frac{\partial^2}{4},$$

$$E(X) = \frac{2}{\partial^2}\int_0^{\partial} x(\partial - x)dx = \frac{2}{\partial^2}\Big[\frac{\partial x^2}{2} - \frac{x^3}{3}\Big]_0^{\partial} = \frac{2}{\partial^2}\Big(\frac{\partial^3}{2} - \frac{\partial^3}{3}\Big) = \frac{\partial}{3},$$

$$E(Y) = \frac{2}{\partial^2}\int_0^{\partial} y^2 dy = \frac{2}{\partial^2}\Big[\frac{y^3}{3}\Big]_0^{\partial} = \frac{2\partial}{3},$$

$$C(X,Y) = E(XY) - E(X)E(Y) = \frac{\partial^2}{4} - \frac{2\partial^2}{9} = \frac{\partial^2}{36}.$$

ΑΣΚΗΣΕΙΣ

$$E(X^2) = \frac{2}{\vartheta^2}\int_0^{\vartheta} x^2(\vartheta - x)dx = \frac{2}{\vartheta^2}\Big[\frac{\vartheta x^3}{3} - \frac{x^4}{4}\Big]_0^{\vartheta} = \frac{2}{\vartheta^2}\frac{\vartheta^4}{12} = \frac{\vartheta^2}{6},$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{\vartheta^2}{6} - \frac{\vartheta^2}{9} = \frac{\vartheta^2}{18},$$

$$E(Y^2) = \frac{2}{\vartheta^2}\int_0^{\vartheta} y^3 dy = \frac{2}{\vartheta^2}\Big[\frac{y^4}{4}\Big]_0^{\vartheta} = \frac{\vartheta^2}{2},$$

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{\vartheta^2}{2} - \frac{4\vartheta^2}{9} = \frac{\vartheta^2}{18},$$

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

ΑΣΚΗΣΕΙΣ

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{y}, \quad 0 < x < y,$$

$$E(X|y) = \frac{1}{y}\int_0^y x dx = \frac{1}{y}\left[\frac{x^2}{2}\right]_0^y = \frac{y}{2}, \quad 0 < y < \partial,$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{\partial - x}, \quad x < y < \partial,$$

$$E(Y|x) = \frac{1}{\partial - x}\int_x^{\partial} y dy = \frac{1}{\partial - x}\left[\frac{y^2}{2}\right]_x^{\partial} = \frac{\partial^2 - x^2}{2(\partial - x)} = \frac{\partial + x}{2}, \quad 0 < x < \partial.$$

ΑΣΚΗΣΕΙΣ

5.14 Έστω (X, Y) μια διακριτή διδιάστατη τυχαία μεταβλητή με συνάρτηση πιθανότητας

$$f(x,y) = \frac{x+y-2}{18}, \quad x = 1,2,3, \quad y = 1,2,3.$$

Να υπολογισθούν
(α) ο συντελεστής συσχέτισης $\rho(X, Y)$ και
(β) η καμπύλη (μέσης) παλινδρόμησης της X στην Y, $x = m_{X|Y}(y)$, για $y = 1, 2, 3$.

(a)

$$f_X(x) = \sum_{y=1}^{3} \frac{x+y-2}{18} = \frac{(x-1)+x+(x+1)}{18} = \frac{x}{6}, \quad x = 1,2,3,$$

$$f_Y(y) = \frac{y}{6}, \quad y = 1,2,3.$$

$$E(XY) = \sum_{y=1}^{3}\sum_{x=1}^{3} xy\frac{x+y-2}{18} = \frac{1}{18}\sum_{y=1}^{3}[y(y-1)+2y^2+3y(y+1)]$$

$$= \frac{1}{9}\sum_{y=1}^{3} y(3y+1) = \frac{48}{9} = \frac{16}{3},$$

ΑΣΚΗΣΕΙΣ

$$E(X) = E(Y) = \frac{1}{6}\sum_{x=1}^{3} x^2 = \frac{14}{6} = \frac{7}{3},$$

$$E(X^2) = E(Y^2) = \frac{1}{6}\sum_{x=1}^{3} x^3 = \frac{36}{6} = 6,$$

$$V(X) = V(Y) = 6 - \left(\frac{7}{3}\right)^2 = 6 - \frac{49}{9} = \frac{5}{9},$$

$$C(X,Y) = E(XY) - E(X)E(Y) = \frac{16}{3} - \frac{49}{9} = -\frac{1}{9},$$

$$\rho(X,Y) = \frac{C(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{-1/9}{5/9} = -\frac{1}{5}.$$

ΑΣΚΗΣΕΙΣ

(β)

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{x+y-2}{3y}, \quad x = 1,2,3,$$

$$m_{X|Y}(y) = E(X|y) = \sum_{x=1}^{3} x \cdot \frac{x+y-2}{3y} = \frac{(y-1)+2y+3(y+1)}{3y}$$
$$= \frac{2(3y+1)}{y},$$

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