# Κεφάλαιο 4

# Ειδικές πολυδιάστατες κατανομές

## 4.1 (6 Μαΐου 2009)

Θεωρούμε ένα τυχαίο πείραμα και $k$ ξένα ανά δύο ενδεχόμενα $A_1, \ldots, A_k$ στο δειγματικό χώρο $\Omega$. Αν ορίσουμε $A_0 = \Omega \backslash (A_1 \cup \cdots \cup A_k)$, τότε η οικογένεια $\{A_0, A_1, \ldots, A_k\}$ είναι διαμέριση του $\Omega$. Λέμε ότι το ενδεχόμενο $A_j$, $j = 1, \ldots, k$ αντιστοιχεί σε «επιτυχία $j$-είδους» και το ενδεχόμενο $A_0$ αντιστοιχεί σε αποτυχία. Αν ενδιαφερόμαστε να μελετήσουμε την πιθανότητα ενδεχομένων σχετικών με τις επιτυχίες $j$-είδους, τότε μπορούμε να υποθέσουμε ότι ο δειγματικός χώρος αποτελείται από $(k+1)$-σημεία: $\Omega = \{a, \epsilon_1, \ldots, \epsilon_k\}$. Λέμε ότι μελετάμε μια δοκιμή Bernoulli $k+1$ ενδεχομένων.

Θεωρούμε τώρα μια ακολουθία $n$ ανεξάρτητων δοκιμών Bernoulli $k+1$ ενδεχομένων. Υποθέτουμε ότι η πιθανότητα επιτυχίας $j$-είδους είναι $p_j$, $j = 1, \ldots, k$, και η πιθανότητα αποτυχίας είναι $p_0 = 1 - (p_1 + \cdots + p_k)$. Αν $X_j$ είναι το πλήθος των επιτυχιών $j$-είδους, τότε η συνάρτηση πιθανότητας του διακριτού τυχαίου διανύσματος $(X_1, \ldots, X_k)$ είναι η

$$f(x_1, \ldots, x_k) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k} p_0^{n-(x_1+\cdots+x_k)},$$

όπου

$$\binom{n}{x_1, \ldots, x_k} = \frac{n!}{x_1! \cdots x_k! (n - x_1 - \cdots - x_k)!}$$

και $x_j = 0, 1, \ldots, n$ (για $j = 1, \ldots, k$) και $x_1 + \cdots + x_k \leq n$. Η κατανομή του $(X_1, \ldots, X_k)$ είναι η **πολυωνυμική κατανομή** με παραμέτρους $p_1, \ldots, p_k$.

Συζητήσαμε πιο αναλυτικά την περίπτωση $k = 2$ (τριωνυμική κατανομή).

**Θεώρημα 4.1.1.** *Εστω ότι η διδιάστατη τυχαία μεταβλητή έχει την τριωνυμική κατανομή με συνάρτηση πιθανότητας την*

$$f(x, y) = \frac{n!}{x! y! (n - x - y)!} p_1^x p_2^y p_0^{n-(x+y)},$$

*όπου $x, y = 0, 1, \ldots, n$ και $x + y \leq n$. Τότε, η περιθώρια κατανομή της $X$ είναι διωνυμική, με συνάρτηση πιθανότητας την*

$$f_X(x) = \binom{n}{x} p_1^x (1 - p_1)^{n-x}, \quad x = 0, 1, \ldots, n$$

*και η περιθώρια κατανομή της $Y$ είναι διωνυμική, με συνάρτηση πιθανότητας την*

$$f_Y(y) = \binom{n}{y} p_2^y (1 - p_2)^{n-y}, \quad y = 0, 1, \ldots, n.$$

*Απόδειξη.* Γράφουμε

$$\begin{aligned} f_X(x) &= \sum_{y=0}^{n-x} f(x,y) \\ &= \sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y} \\ &= \frac{n!}{x!(n-x)!} p_1^x \sum_{y=0}^{n-x} \frac{(n-x)!}{y!(n-x-y)!} p_2^y (1-p_1-p_2)^{n-x-y} \\ &= \binom{n}{x} p_1^x \sum_{y=0}^{n-x} \binom{n-x}{y} p_2^y (1-p_1-p_2)^{n-x-y} \\ &= \binom{n}{x} p_1^x (p_2+1-p_1-p_2)^{n-x} \\ &= \binom{n}{x} p_1^x (1-p_1)^{n-x}. \end{aligned}$$

Ειδικότερα,

$$\mathbb{E}(X) = np_1 \text{ και } V(X) = np_1(1-p_1).$$

Εντελώς ανάλογα αποτελέσματα ισχύουν για την $Y$. □

**Θεώρημα 4.1.2.** *Έστω ότι η διδιάστατη τυχαία μεταβλητή έχει την τριωνυμική κατανομή με συνάρτηση πιθανότητας την*

$$f(x,y) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y p_0^{n-(x+y)},$$

*όπου $x, y = 0, 1, \ldots, n$ και $x + y \leq n$. Για κάθε $x = 0, 1, \ldots, n$, η δεσμευμένη κατανομή της $Y$ δεδομένου ότι $X = x$ είναι διωνυμική, με συνάρτηση πιθανότητας την*

$$f_{Y|X}(y|x) = \binom{n-x}{y} \left(\frac{p_2}{1-p_1}\right)^y \left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-x-y}, \quad y = 0, 1, \ldots, n-x$$

*και η δεσμευμένη κατανομή της $X$ δεδομένου ότι $Y = y$ είναι διωνυμική, με συνάρτηση πιθανότητας την*

$$f_{X|Y}(x|y) = \binom{n-y}{x} \left(\frac{p_1}{1-p_2}\right)^x \left(\frac{1-p_1-p_2}{1-p_2}\right)^{n-y-x}, \quad x = 0, 1, \ldots, n-y.$$

*Απόδειξη.* Παρατηρούμε ότι, για κάθε $x = 0, 1, \ldots, n$,

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} \\ &= \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y} \frac{x!(n-x)!}{n!} \frac{1}{p_1^x (1-p_1)^{n-x}} \\ &= \frac{(n-x)!}{y!(n-x-y)!} \frac{p_2^y (1-p_1-p_2)^{n-x-y}}{(1-p_1)^{n-x}} \\ &= \binom{n-x}{y} \left(\frac{p_2}{1-p_1}\right)^y \left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-x-y}. \end{aligned}$$

**Πρόταση 4.1.3.** *Ισχύουν οι*

$$\mathbb{E}(XY) = n(n-1)p_1p_2 \text{ και } C(X,Y) = -np_1p_2.$$

*Σημείωση.* Η καμπύλη παλινδρόμησης της $Y$ στην $X$ είναι η ευθεία

$$y = m_{Y|X}(x) = \mathbb{E}(Y|x) = (n-x)\frac{p_1}{1-p_2}$$

και η καμπύλη παλινδρόμησης της $X$ στην $Y$ είναι η ευθεία

$$x = m_{X|Y}(y) = \mathbb{E}(X|y) = (n-y)\frac{p_2}{1-p_1}.$$

## 4.2 (11 Μαΐου 2009)

Θεωρούμε ανεξάρτητες κανονικές τυχαίες μεταβλητές $Z, W$ με μέση τιμή και διασπορά $\mu_Z, \sigma_Z^2$ και $\mu_W, \sigma_W^2$ αντίστοιχα. Αν θεωρήσουμε τον αντιστρέψιμο γραμμικό μετασχηματισμό

$$\begin{aligned} X &= aZ + bW \\ Y &= cZ + dW \end{aligned}$$

μπορούμε να υπολογίσουμε την από κοινού συνάρτηση πυκνότητας $f_{X,Y}$ των $X$ και $U$ χρησιμοποιώντας τη μέθοδο του Κεφαλαίου 2. Παρατηρήστε ότι

$$\mu_X = a\mu_Z + b\mu_W, \quad \mu_Y = c\mu_Z + d\mu_W$$

και

$$\sigma_X^2 = a^2\sigma_Z^2 + b^2\sigma_W^2, \quad \sigma_Y^2 = c^2\sigma_Z^2 + d^2\sigma_W^2.$$

Επίσης,

$$\rho := \rho_{X,Y} = \frac{ac\sigma_Z^2 + bd\sigma_W^2}{\sqrt{(a^2\sigma_Z^2 + b^2\sigma_W^2)(c^2\sigma_Z^2 + d^2\sigma_W^2)}}.$$

Μετά από πράξεις, βλέπουμε ότι

$$\begin{aligned} (1) \qquad f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \\ \times \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right). \end{aligned}$$

Λέμε ότι μια συνεχής διδιάστατη τυχαία μεταβλητή $(X, Y)$ με την παραπάνω πυκνότητα έχει τη **διδιάστατη κανονική κατανομή**.

**Θεώρημα 4.2.1.** *Έστω $(X, Y)$ διδιάστατη συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας την (1). Τότε, η κατανομή της $X$ και η κατανομή της $Y$ είναι κανονική. Πιο συγκεκριμένα:*

(i) *Η περιθώρια πυκνότητα της $X$ είναι η*

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right].$$

(ii) *Η περιθώρια πυκνότητα της $Y$ είναι η*

$$f_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right].$$

(iii) *Ισχύουν οι*

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

**Θεώρημα 4.2.2.** *Έστω $(X, Y)$ διδιάστατη συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας την (1). Τότε:*

(i) *Η δεσμευμένη κατανομή της $X$ δεδομένου ότι $Y = y$ είναι η κανονική κατανομή με μέση τιμή*

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

*και διασπορά*

$$\sigma_X^2(1-\rho^2).$$

(ii) *Η δεσμευμένη κατανομή της $Y$ δεδομένου ότι $X = x$ είναι η κανονική κατανομή με μέση τιμή*

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

*και διασπορά*

$$\sigma_Y^2(1 - \rho^2).$$

(iii) *Η καμπύλη παλινδρόμησης της $X$ στην $Y$ αλλά και της $Y$ στην $X$ είναι η ευθεία*

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

*Σημείωση.* Από την (1) ελέγχουμε άμεσα ότι αν οι $X$ και $Y$ είναι ασυσχέτιστες τότε

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

άρα οι $X$ και $Y$ είναι ανεξάρτητες.

**Παράδειγμα 4.2.3.** Θεωρούμε τον πληθυσμό των παντρεμένων ζευγαριών μιας περιοχής και υποθέτουμε ότι η διδιάστατη τυχαία μεταβλητή $(X,Y)$, όπου $X$ το ύψος του άντρα και $Y$ το ύψος της γυναίκας, ακολουθεί τη διδιάστατη κανονική κατανομή με $\mu_X = 175\,\mathrm{cm}$, $\mu_Y = 170\,\mathrm{cm}$, $\sigma_X = \sigma_Y = 5\,\mathrm{cm}$ και $\rho = 0.6$. Υπολογίστηκε η πιθανότητα, σε ένα ζευγάρι, ο άντρας να έχει ύψος $X$ μεγαλύτερο από $190\,\mathrm{cm}$ δεδομένου ότι η γυναίκα έχει ύψος $Y = 175\,\mathrm{cm}$.