

## 12. Feltételes valószínűség és többdimenziós eloszlások

A mostani fejezetben kimondjuk a teljes valószínűség tételének azon verzióját, ahol a feltételben valószínűségi változó szerepel teljes eseményrendszer helyett. Ettől független témaként közelebbről megvizsgálunk néhány valószínűségi vektorváltozóhoz tartozó (ún. többdimenziós) nevezetes eloszlást, különös tekintettel a többdimenziós normális eloszlásra.

### 12.1. Teljes valószínűség tétele, folytonos eset

Az előző fejezet után maradhatott némi hiányérzet az olvasóban: míg a teljes várható érték **tételét** többféle formában is kimondtuk (teljes eseményrendszerrel és valószínűségi változó  $\{X = x\}$  színhalmazaival is diszkrét és folytonos esetben), addig a teljes valószínűség **tételét** csak teljes eseményrendszerre fogalmaztuk meg. Mi a helyzet az utóbbi tétellel akkor, ha a feltétel egy valószínűségi változó (színhalmaza)?

Ha  $X$  diszkrét valószínűségi változó, akkor a teljes valószínűség tétele nem újdonság:

$$\mathbb{P}(A) = \sum_{k \in S_X} \mathbb{P}(A \mid X = k) \cdot \mathbb{P}(X = k),$$

ahol  $S_X$  azon  $k$  értékek halmaza, ahol  $\mathbb{P}(X = k) > 0$ , és így van értelme a  $\mathbb{P}(A \mid X = k)$  feltételes valószínűségről beszélni. Ez az eredeti teljes valószínűség tételének speciális esete. Viszont ha  $X$  folytonos, akkor  $\mathbb{P}(A \mid X = k)$  értelmetlen. A probléma feloldása ugyanaz, mint a regressziós függvény esetében.

**12.1.1. Definíció.** Legyen  $X$  valószínűségi változó, és  $A$  esemény. Ekkor  $A$ -nak az  $X$ -re vett **feltételes valószínűsége** az

$$x \mapsto \mathbb{E}(\mathbf{1}_A \mid X = x)$$

regressziós függvény. Jelölése:  $\mathbb{P}(A \mid X = x)$ .

Itt a regressziós függvényt a 11.2 alfejezet definíciója szerint értjük, vagyis ez az a  $g$  függvény, amire  $g(X) = \mathbb{E}(\mathbf{1}_A \mid X)$ . Az  $\mathbb{E}(\mathbf{1}_A \mid X)$  regresszió pedig a (11) egyenlettel definiált. (Mivel  $\mathbb{E}(\mathbf{1}_A)$  azaz  $\mathbb{P}(A)$  véges, így  $g$  létezik a 11.2 alfejezet megjegyzése okán.) Szemléletesen,  $\mathbb{E}(\mathbf{1}_A \mid X = x)$  jelentése az  $A$  esemény valószínűsége (avagy precízebben, annak legjobb átlagos közelítése), tudván az  $X$  értékét.

Innen a teljes valószínűség tétele már megtippelhető:

**▲ 12.1.2. Tétel** (Teljes valószínűség tétele). *Legyen  $X$  folytonos valószínűségi változó, és  $A$  esemény. Ekkor*

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x) f_X(x) dx,$$

ahol  $f_X$  az  $X$  sűrűségfüggvénye.

**12.1.3. Példa.** Jelölje  $X$  egy hallgatónak a valószínűségszámítás vizsgára szánt felkészülési idejét. Tegyük fel, hogy  $X$  egyenletes eloszlású az  $[\varepsilon, 20]$  intervallumon (napban számolva, ahol  $\varepsilon$  20-nál kisebb, és őszintén remélem, hogy pozitív valós szám). Feltételezve, hogy  $x$  időt szán felkészülésre a hallgató,  $\left(\frac{x}{21}\right)^2$  a valószínűsége, hogy ötös érdemjegyet kap. Mi a valószínűsége az ötös vizsgának?

Az előző tétel jelölésével: tudjuk, hogy  $f_X(x) = \frac{1}{20-\varepsilon}$  ha  $\varepsilon \leq x \leq 20$ , illetve 0 egyébként. Továbbá  $\mathbb{P}(A \mid X = x) = \left(\frac{x}{21}\right)^2$ . Tehát

$$\mathbb{P}(A) = \int_{\varepsilon}^{20} \left(\frac{x}{21}\right)^2 \frac{1}{20-\varepsilon} dx = \left[ \frac{x^3}{3 \cdot 21^2 (20-\varepsilon)} \right]_{\varepsilon}^{20} = \frac{\varepsilon^2 + 20\varepsilon + 20^2}{3 \cdot 21^2}$$

Ha  $\varepsilon = 1$ , akkor ez kerekítve 0,3182.

*Megjegyzés.* A feltételes valószínűség speciális esete a feltételes eloszlásfüggvény:

$$F_{Y|X}(y \mid x) = \mathbb{P}(Y < y \mid X = x).$$

## 12.2. Többdimenziós eloszlások

Legyen  $\underline{X} = (X_1, \dots, X_m)$  valószínűségi vektorváltozó. Az egydimenziós esethez hasonlóan beszélhetünk az  $\underline{X}$  eloszlásáról (amit például az együttes eloszlásfüggvény ír le), ahogy ezt tettük is már az együttes eloszlás témakörénél. Nézzünk most néhány gyakrabban előkerülő többdimenziós eloszlást.

Nevezetes diszkrét eloszlás a binomiális. Hogyan általánosítható ez több változóra? Erre van egy kézenfekvő módszer: legyenek  $X_1, \dots, X_m$  együttesen függetlenek, és legyen  $X_i \sim B(n; p_i)$  valamilyen  $n \in \mathbb{N}$  és  $0 < p_i < 1$  számokra ( $i = 1, \dots, m$ ). Így értelmes többdimenziós eloszlást kapunk, de a binomiális eloszlás általánosításának nem ez az egyetlen módja.

**12.2.1. Példa.** Átcímkeztünk egy szabályos dobókockát: egy 1-es, két 2-es és három 3-mas számjegyet írtunk rá. Dobjunk 13-szor a kockával. Jelölje  $X_i$  a dobott  $i$  számjegyek számát. Mi a valószínűsége, hogy  $X_1 = 3$ ,  $X_2 = 4$  és  $X_3 = 6$ ?

A valószínűség kombinatorikus módon meghatározható:

$$\mathbb{P}(X_1 = 3, X_2 = 4, X_3 = 6) = \frac{13!}{3!4!6!} \left(\frac{1}{6}\right)^3 \left(\frac{1}{3}\right)^4 \left(\frac{1}{2}\right)^6 \approx 0,05364,$$

hiszen a 3 db 1-es, 4 db 2-es és 6 db 3-mas lehetséges elhelyezéseinek száma  $\frac{13!}{3!4!6!}$  (ismétléses permutáció), és az ilyen esetek valószínűsége  $p_1^3 p_2^4 p_3^6$ , ahol az  $i$  dobás valószínűsége  $p_i$ .

**12.2.2. Definíció.** Az  $\underline{X} = (X_1, \dots, X_m)$  valószínűségi vektorváltozó **polinomiális** (más néven: multinomiális) eloszlású,  $n \in \mathbb{N}$  és  $(p_1, p_2, \dots, p_m) \in [0, 1]^m$  paraméterekkel, ha  $p_1 + \dots + p_m = 1$  és

$$\mathbb{P}(X_1 = k_1, \dots, X_m = k_m) = \frac{n!}{k_1! k_2! \dots k_m!} p_1^{k_1} \dots p_m^{k_m}$$

minden  $0 \leq k_i \leq n$  ( $i = 1, \dots, m$ ),  $\sum_{i=1}^m k_i = n$  értékek esetén.

Ha  $m = 2$  és  $(p_1, p_2) = (p, 1 - p)$  valamilyen  $p \in [0, 1]$  esetén, akkor  $X_1$  eloszlása  $B(n; p)$  (az  $X_2$  pedig nem hordoz extra információt, hiszen  $X_2 = n - X_1$ ).

Világos, hogy az  $X_i$  változók nem függetlenek (hiszen például  $X_1, \dots, X_{m-1}$  egyértelműen meghatározza  $X_m$ -et), ugyanakkor az  $\underline{X}$  peremeloszlásai mind  $B(n; p_i)$  binomiális eloszlások. Ez a példa is mutatja, hogy a peremeloszlások nem határozzák meg az együttes eloszlást, továbbá, hogy nem mindig az együttesen független koordináták adják egy eloszlás természetes többváltozós általánosítását.

Egy másik érdekes többdimenziós eloszlás:

**12.2.3. Definíció.** Legyenek  $Y_1, Y_2, Y_3$  együttesen független valószínűségi változók, ahol  $Y_i \sim \text{Exp}(\lambda_i)$ , ( $i = 1, 2, 3$ ). Defináljuk az  $\underline{X} = (X_1, X_2)$  vektorváltozót:  $X_1 = \min(Y_1, Y_3)$  és  $X_2 = \min(Y_2, Y_3)$ . Az  $\underline{X}$  eloszlását **Marshall–Olkin-féle kétváltozós exponenciális eloszlásnak** (röviden Marshall–Olkin-eloszlásnak) hívják.<sup>55</sup>

A motiváció a következő: ha  $X$  exponenciális eloszlású, akkor teljesíti az örökifjúság feltételét, azaz  $\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t)$  minden  $s, t > 0$  esetén. Ennek lehetséges általánosítása a

$$\mathbb{P}(\underline{X} > \underline{t} + \underline{s} \mid \underline{X} > \underline{s}) = \mathbb{P}(\underline{X} > \underline{t})$$

feltétel, ahol  $\underline{t}, \underline{s} \in [0, \infty)^2$ , és a vektorok közti  $>$  reláció akkor teljesül, ha mindkét koordinátában külön-külön teljesül. Ez a fajta örökifjúság meghatároz egy értelmes kétdimenziós eloszlást: azt, aminek a koordinátái független, exponenciális eloszlású valószínűségi változók (vagyis ez nem a fenti Marshall–Olkin-eloszlás). Alternatív általánosítás viszont a következő feltétel:

$$(12) \quad \mathbb{P}(\underline{X} > \underline{t} \cdot \underline{1} + \underline{s} \mid \underline{X} > \underline{s}) = \mathbb{P}(\underline{X} > \underline{t} \cdot \underline{1}),$$

ahol  $\underline{s} \in [0, \infty)^2$ ,  $\underline{t} \geq 0$  és  $\underline{1} = (1, 1)$ . Ezt a tulajdonságot a független, exponenciális eloszlású koordinátákkal bíró valószínűségi vektorváltozón túl a fenti Marshall–Olkin-eloszlás is teljesíti.

<sup>55</sup>Érdekes, hogy a Marshall–Olkin-eloszlás nem folytonos, azaz nincs együttes sűrűségfüggvénye. Ennek az az oka, hogy a két koordináta pozitív eséllyel megegyezhet. Lásd A.W. Marshall, I. Olkin, A generalized bivariate exponential distribution, J. Appl. Probab. 4 (1967) 291–302.

**12.2.4. Példa.** Egy gépben két fontos alkatrész van. Jelölje  $X_1$  és  $X_2$  a két alkatrész (véletlen) élettartamát. Tegyük fel, hogy az alkatrészek kora nem befolyásolja, hogy elromlanak-e  $t$  idő alatt, vagyis ha az első alkatrész  $s_1$  idős, a második  $s_2$  idős, akkor annak a valószínűsége, hogy  $t$  ideig nem romlik el egyik alkatrész sem, ugyanaz, mintha mindkét alkatrész új lenne. Egyenlettel:

$$\mathbb{P}((X_1, X_2) > (t + s_1, t + s_2) \mid (X_1, X_2) > (s_1, s_2)) = \mathbb{P}((X_1, X_2) > (t, t))$$

tetszőleges  $s_1, s_2, t \in [0, \infty)$  esetén. Ez éppen az előző (12) egyenlet, azaz  $(X_1, X_2)$  Marshall–Olkin-eloszlású is lehet, valamilyen  $\lambda_1, \lambda_2, \lambda_3$  paraméterekkel. (Szemléletesen, az  $Y_3$  azt a közös hatást reprezentálja, ami mindkét alkatrészt egyszerre elronthatja.)

### 12.3. Többdimenziós normális eloszlás

Bár számos többdimenziós eloszlásról lehetne beszélni, a legnevezetesebbet nem hagyhatjuk ki, ez a többváltozós normális eloszlás.

Hogyan tudnánk általánosítani a normális eloszlást kétdimenziós eloszlásként? Az egydimenziós normális eloszlás tipikusan egy fizikai mérés eredményének a tényleges érték körüli szóródását (hibáját) írja le. A kétdimenziós általánosítás meghatározásához tekintsünk egy kétdimenziós mérési eredményt, például egy olyan jeladó  $X$  szélességi és  $Y$  hosszúsági koordinátáit, aminek helyzetét nem ismerjük pontosan, de a jel alapján bemérjük. Idealizált esetben milyen tulajdonságot várnánk ettől az eloszlástól?

Egyrészt feltesszük, hogy az eloszlás *folytonos*, azaz létezik az  $f_{X,Y}$  együttes sűrűségfüggvény. Az egyszerűség kedvéért legyen a jeladó tényleges helye az origó. Természetes feltételezés, hogy az eloszlás *forgásszimmetrikus*, azaz  $f_{X,Y}$  értéke csak  $(x, y)$  hosszától függ. Egyenlettel:

$$(13) \quad f_{X,Y}(x, y) = h(x^2 + y^2)$$

valamilyen  $h$  valós függvényre. Másrészt, nem irreális feltétel az sem, hogy  $X$  és  $Y$  *függetlenek*, vagyis hogy az  $x$  és  $y$  koordinátában mért hibák nem befolyásolják egymást. Az  $X$  és  $Y$  függetlensége ekvivalensen:

$$(14) \quad f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \quad (\forall x, y \in \mathbb{R}).$$

Megmutatjuk, hogy ezek a feltételek meghatározzák az eloszlást.

**12.3.1. Állítás.** *Ha  $(X, Y)$  folytonos valószínűségi vektorváltozó, ami forgásszimmetrikus, és az  $X, Y$  koordináták függetlenek, akkor  $f_{X,Y}(x, y) = e^{a(x^2+y^2)-c}$  valamilyen  $a, c \in \mathbb{R}$  esetén, ahol  $a < 0$ .*

*Bizonyítás.* Helyettesítsünk  $y = 0$ -t a (13) és (14) egyenletekbe:

$$h(x^2 + 0^2) = f_{X,Y}(x, 0) = f_X(x) \cdot f_Y(0),$$

tehát  $f_X(x) = \frac{1}{f_Y(0)} h(x^2)$  ( $x \in \mathbb{R}$ ). Közben felhasználtuk, hogy ha  $f_Y(0) = 0$  lenne, akkor  $h$  azonosan nulla, ami lehetetlen. Hasonlóan,  $f_Y(y) = \frac{1}{f_X(0)} h(y^2)$  ( $y \in \mathbb{R}$ ). Visszahelyettesítve,

$$h(x^2 + y^2) = f_X(x) \cdot f_Y(y) = \frac{1}{f_Y(0)} h(x^2) \cdot \frac{1}{f_X(0)} h(y^2).$$

Jelöljük ezt át a következőképp:  $u = x^2$ ,  $v = y^2$  és  $c = \ln(f_X(0)f_Y(0))$ . Ekkor a fenti egyenlet logaritmus:

$$\ln h(u + v) = \ln h(u) + \ln h(v) - c.$$

Legyen  $G(u) = \ln h(u) - c$ . Az utolsó egyenletből  $c$ -t levonva mindkét oldalról  $G(u + v) = G(u) + G(v)$  adódik. Ez ugyanaz a Cauchy-egyenlet, amiről **korábban** már beszéltünk. Integrálható megoldása ennek csak a  $G(u) = a \cdot u$  függvény, valamilyen  $a \in \mathbb{R}$  esetén. Tehát  $h(u) = e^{au-c}$ , vagyis  $f_{X,Y}(x, y) = e^{a(x^2+y^2)-c}$  valamilyen  $a, c \in \mathbb{R}$  esetén. Ha  $a \geq 0$  lenne, akkor nem lehetne  $f_{X,Y}$  integrálja 1.  $\square$

A paraméterek alkalmas megválasztásával adódik a standard normális eloszlás. Általánosan,  $n$ -dimenziós esetben ez a következő.

**12.3.2. Definíció.** Az  $\underline{X} = (X_1, \dots, X_n)$  valószínűségi vektorváltozó  **$n$ -dimenziós standard normális eloszlású**, ha folytonos, és együttes sűrűségfüggvénye:

$$f_{\underline{X}}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2} \quad (x_1, \dots, x_n \in \mathbb{R}).$$

Hogyan kapjuk a nem feltétlenül standard, többdimenziós normális eloszlásokat?

**12.3.3. Definíció.** Az  $\underline{Y} = (Y_1, \dots, Y_n)$  valószínűségi vektorváltozó **többdimenziós normális eloszlású**, ha létezik  $\underline{A} \in \mathbb{R}^{n \times n}$ ,  $\underline{\mu} \in \mathbb{R}^n$  és  $\underline{X}$   $n$ -dimenziós standard normális eloszlású valószínűségi vektorváltozó, amire

$$\underline{Y} = \underline{A} \cdot \underline{X} + \underline{\mu},$$

$\underline{X}$ -et oszlopvektorként kezelve. Az  $\underline{Y}$  eloszlása **nemelfajuló**, ha  $\underline{A}$  választható nemelfajuló mátrixnak (azaz  $\det(\underline{A}) \neq 0$ ).

Ez a leírásmód eltér az egydimenziós esetben alkalmazott paraméterezéstől, ahol egy (nem feltétlenül standard) normális eloszlást a várható értékével és a szórásnégyzetével adtunk meg. Vizsgáljuk meg a többdimenziós normális eloszlás hasonló paramétereit.

**12.3.4. Definíció.** Egy  $\underline{Y} = (Y_1, \dots, Y_n)$  valószínűségi vektorváltozó **várható érték vektora** az  $(\mathbb{E}Y_1, \dots, \mathbb{E}Y_n)$   $\mathbb{R}^n$ -beli vektor. Jelölés  $\mathbb{E}\underline{Y}$ .

A kovarianciamátrix szintén kifejezhető a várható érték vektor segítségével. Ha oszlopvektorokként kezeljük az  $\underline{Y}$  és  $\mathbb{E}\underline{Y}$  vektorokat, akkor

$$\text{cov}(\underline{Y}) = \mathbb{E}((\underline{Y} - \mathbb{E}\underline{Y}) \cdot (\underline{Y} - \mathbb{E}\underline{Y})^T) \in \mathbb{R}^{n \times n},$$

ahol a szorzás az  $n \times 1$  és  $1 \times n$  alakú mátrixok mátrixszorzatát jelöli, illetve a kapott mátrix várható értékét koordinátáinként értelmezzük.

**▲ 12.3.5. Állítás.** Legyen  $\underline{X} = (X_1, \dots, X_n)$  standard normális eloszlású valószínűségi vektorváltozó, és  $\underline{Y} = \underline{A} \cdot \underline{X} + \underline{\mu}$ . Ekkor  $\mathbb{E}\underline{Y} = \underline{\mu}$  és  $\text{cov}(\underline{Y}) = \underline{A} \cdot \underline{A}^T$ .

Ezekkel a paraméterekkel felírható a többdimenziós normális eloszlás sűrűségfüggvénye is.

**12.3.6. Állítás.** Legyen  $\underline{Y}$  nemelfajuló  $n$ -dimenziós normális eloszlású vektorváltozó. Jelölje a várható érték vektorát  $\underline{\mu}$ , a kovarianciamátrixát  $\underline{\Sigma}$ . Ekkor  $\underline{Y}$  sűrűségfüggvénye

$$f_{\underline{Y}}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\underline{\Sigma})^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{x} - \underline{\mu})^T \underline{\Sigma}^{-1}(\underline{x} - \underline{\mu})},$$

ahol  $\det(\underline{\Sigma})$  a  $\underline{\Sigma}$  determinánsa,  $\underline{\Sigma}^{-1}$  pedig az inverz mátrixa.

A kitevőben a szorzat egy hármasszorzat (vektor, mátrix és megint vektor tagokkal), ami valós számot eredményez. A mátrix tag kétdimenziós esetben:

$$\underline{\Sigma} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad \Rightarrow \quad \underline{\Sigma}^{-1} = \frac{1}{\det(\underline{\Sigma})} \begin{pmatrix} c & -b \\ -b & a \end{pmatrix} \quad \det(\underline{\Sigma}) = ac - b^2,$$

ahol  $a = \mathbb{D}^2(Y_1)$ ,  $b = \text{cov}(Y_1, Y_2)$  és  $c = \mathbb{D}^2(Y_2)$ .

Az állítás fontos következménye, hogy egy nemelfajuló normális eloszlást meghatároz a  $\underline{\mu}$  várható érték vektora és a  $\underline{\Sigma}$  kovarianciamátrixa. (Vegyük észre, hogy adott  $\underline{\Sigma}$  többféle  $\underline{A}$  mátrixból is előállhat, ezért ez nem nyilvánvaló állítás.) Valójában az elfajuló esettel is ez a helyzet, de ekkor nincs sűrűségfüggvényünk, de ezzel itt részletesebben nem foglalkozunk.

A fentiek miatt értelmes a következő jelölés:

**Jelölés.** Az  $n$ -dimenziós normális eloszlást  $N(\underline{\mu}, \underline{\Sigma})$  jelöli, ahol  $\underline{Y} = \underline{A} \cdot \underline{X} + \underline{\mu}$ ,  $\underline{X}$   $n$ -dimenziós standard normális, és  $\underline{\Sigma} = \underline{A} \cdot \underline{A}^T$ . Speciálisan, a standard normális eloszlás jelölése  $N(\underline{0}, \underline{I})$ , ahol  $\underline{0}$  az  $n$ -dimenziós nullvektor, és  $\underline{I}$  az  $n$ -dimenziós egységmátrix.

Vegyük észre, hogy sem a standard, sem az általános esetben nem beszéltünk még az  $Y_i$  koordináták eloszlásáról, sőt szóba sem került az egydimenziós normális eloszlás. Kérdés tehát, hogy mik a normális eloszlás marginálisai? A válasz mérsékeltlen meglepő:

**12.3.7. Állítás.** Legyen  $\underline{Y} \sim N(\underline{\mu}, \underline{\Sigma})$ , ahol  $\underline{\mu} \in \mathbb{R}^n$  és  $\underline{\Sigma} \in \mathbb{R}^{n \times n}$ . Ekkor  $Y_i \sim N(\mu_i, \Sigma_{i,i})$ .

A standard esetben ennél többet is tudunk: mivel a sűrűségfüggvény szorzattá bomlik (hiszen  $\frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x_i^2}$ ), így az  $X_i$  koordináták együttesen független, egydimenziós standard normális eloszlásúak. Vagyis a normális eloszlásnál teljesül az a szép tulajdonság, ami a polinomiálisnál vagy a Marshall–Olkin-eloszlásnál nem: a természetes többdimenziós általánosítás az egydimenziós eloszlások együttesen független példányai, vektorba rendezve.

A normális eloszlás több egyéb tulajdonsága okán is a „túl szép, hogy igaz legyen” díjas eloszlás első számú jelöltje; ezeket a tulajdonságokat a következő állításban foglaljuk össze:

**12.3.8. Következmény.** Legyen  $(Y_1, Y_2) \sim N(\underline{\mu}, \underline{\Sigma})$  kétdimenziós normális eloszlású valószínűségi vektorváltozó. Ekkor

- (1) tetszőleges  $c_1, c_2 \in \mathbb{R}$  esetén,  $c_1 Y_1 + c_2 Y_2$  egydimenziós normális eloszlású, vagy konstans,
- (2) ha  $\text{corr}(Y_1, Y_2) = 0$ , akkor  $Y_1$  és  $Y_2$  függetlenek,
- (3) az  $\mathbb{E}(Y_2 | Y_1)$  regresszió megegyezik az  $Y_2$ -nek az  $Y_1$ -re vett lineáris regressziójával, azaz

$$\mathbb{E}(Y_2 | Y_1) = \frac{b}{a} Y_1 + \left( \mu_2 - \frac{b}{a} \mu_1 \right), \quad \text{ahol } \underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \underline{\Sigma} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Az eloszlás vizualizációjáról még érdemes szót ejteni: hogyan is néz ki egy normális eloszlás sűrűségfüggvénye, például kétdimenziós esetben?

A standard esetben egy „domb” az origó körül (ahogy egydimenziós esetben is), ami forgásszimmetrikus, azaz a szintvonalai körök. Nem standard esetben a szintvonalak ellipszisek lesznek. Tehát a nem standard normális eloszlás nem feltétlenül forgásszimmetrikus, de továbbra is tengelyesen szimmetrikus az ellipszis(ek) főtengelyeire. Tekintsük az egyik ilyen ellipszist.

Az egyszerűség kedvéért tegyük fel, hogy  $\underline{\mu} = \underline{0}$ , vagyis az ellipszis középpontja az origó. Az ellipszis főtengelyei egymásra merőlegesek, így létezik olyan  $\underline{U} \in \mathbb{R}^{2 \times 2}$  ortogonális transzformáció, ami a főtengelyeket átviszi a koordinátatengelyekbe. Kiszámolható, hogy ekkor  $\underline{U} \cdot \underline{Y} \sim N(\underline{0}, \underline{D})$ , ahol  $\underline{D}$  diagonális mátrix. A következmény második pontja szerint ekkor  $\underline{U} \cdot \underline{Y}$  két koordinátája független. Összefoglalva, megfelelő koordináta-rendszert választva minden normális eloszlás független, egydimenziós, normális eloszlású valószínűségi változókból áll.

A diagonalizálással kapott független valószínűségi változók szórásai implicit módon korábban is megjelentek a normális eloszlás felírásában: ha  $\underline{D} = \text{diag}(\sigma_1^2, \sigma_2^2)$ , akkor a sűrűségfüggvényben megjelenő  $\det(\underline{\Sigma})^{\frac{1}{2}}$  éppen  $\sigma_1 \cdot \sigma_2$ , azaz a szórások szorzata. A kovarianciamátrix determinánsa nem változik ortogonális transzformáció alkalmazása esetén, így mindegy, hogy az eredeti  $\underline{Y}$  vagy a transzformált  $\underline{U} \cdot \underline{Y}$  kovarianciamátrixáról beszélünk. Vizuálisabban, ez méri az ellipszis területének az egységkör területéhez viszonyított arányát.

Többdimenziós eloszlások esetén a (teljes) variancia mérésére a kovarianciamátrix determinánsa mellett a  $\text{Tr}(\underline{\Sigma})$  nyoma is használatos mennyiség. A diagonalizált változó szórásaival kifejezve  $\text{Tr}(\underline{\Sigma}) = \sigma_1^2 + \sigma_2^2$ . Szemléletesen, ez az  $\underline{Y}$ -nak a  $\underline{\mu}$ -tól való eltéréseinek az átlagos hossz négyzetét méri.

A többdimenziós normális eloszlás további mélységeiért lásd:

- J.K. Patel, C.B. Read, Handbook of the Normal Distribution, CRC Press, 1982.
- Y.L. Tong, The Multivariate Normal Distribution, Springer, 1990.

