U. Mortensen

Die n-dimensionale Normalverteilung

Es wird zunächst die 2-dimensionale Normalverteilung betrachtet. Die zufälligen Veränderlichen X und Y seien normalverteilt. Gesucht ist die gemeinsame Verteilung $f(x, y)$, wenn für die Korrelation $r_{xy} \neq 0$ gilt; der Fall unabhängiger, normalverteilter Variablen ergibt sich dann als Spezialfall für $r_{xy} = 0$.

Sind A und B zwei zufällige Ereignisse, so ist die bedingte Wahrscheinlichkeit $p(A|B)$ durch

$$p(A|B) = \frac{p(A \wedge B)}{p(B)}$$

gegeben ($\wedge$ steht für 'und'); es ist dann

$$p(A \wedge B) = p(A|B)p(B). \tag{0.1}$$

Weiter seien X und Y zwei zufällige Veränderliche mit $f(x, y)$ als gemeinsamer Dichte, und es sei $A = \{X = x\}$, $B = \{Y = y\}$. $g(y|x)$ sei die bedingte Dichte für Y, gegeben $X = x$, und $f_X(x) = \int f(x, y)dy$ sei die Randverteilung für X. Dann hat man, analog zu (0.1),

$$g(y|x) = \frac{f(x, y)}{f_X(x)}. \tag{0.2}$$

Die gemeinsame Dichte ergibt sich daraus als

$$f(x, y) = g(y|x)f_X(x). \tag{0.3}$$

Nach Voraussetzung ist

$$f_X(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left[-\frac{(x - \mu_x)^2}{2\mu_x}\right]. \tag{0.4}$$

Um g zu bestimmen, werde die Regressionsgleichung

$$y = ax + b + \varepsilon$$

betrachtet. Für festes x ist g die Dichte für Y, gegeben x (kurz: $Y|x$), und $\sigma^2_{y|x} = \sigma^2_\varepsilon$. Die Voraussetzung der Normalverteilung für Y bedeutet dann $Y \sim N(ax + b, \sigma^2_\varepsilon)$. Allgemein gilt für die unkonditionierte Varianz von Y

$$\sigma^2_y = a^2\sigma^2_x + \sigma^2_\varepsilon.$$

Es ist $r = r_{xy} = a\sigma_x/\sigma_y$, so dass $a = r\sigma_y/\sigma_x$. Dann ist

$$\sigma^2_y = \frac{r^2\sigma^2_x\sigma^2_y}{\sigma^2_x} + \sigma^2_\varepsilon = r^2\sigma^2_y + \sigma^2_\varepsilon,$$

woraus

$$\sigma_\varepsilon^2 = (1 - r^2)\sigma_y^2 \tag{0.5}$$

folgt, und für g findet man

$$g(y|x) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left[\frac{(y-(ax+b))^2}{2\sigma_\varepsilon^2}\right] = \frac{1}{\sigma_y\sqrt{(1-r^2)}\sqrt{2\pi}} \exp\left[\frac{(y-(ax+b))^2}{2\sigma_\varepsilon^2}\right]$$

Es ist $\mu_y = a\mu_x + b$ und also

$$y-(ax+b) = y-\mu_y-(ax+b)+\mu_y = y-\mu_y-(ax+b-a\mu_x-b) = y-\mu_y-a(x-\mu_x)$$

und

$$(y-(ax+b))^2 = (y-\mu_y)^2 + a^2(x-\mu_x)^2 - 2a(x-\mu_x)(y-\mu_y).$$

Substituiert man für a wieder $r\sigma_y/\sigma_x$, so erhält man den

Satz 0.1 *Sind X und Y gemeinsam normalverteilt, so ist die gemeinsame Dichte durch*

$$f(x,y) = A\exp\left[-\frac{1}{2(1-r^2)}\left(\frac{(y-\mu_y)^2}{\sigma_y^2} + \frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2r(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right], \tag{0.6}$$

mit

$$A = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}. \tag{0.7}$$

gegeben.

Ellipsen als Orte gleicher Wahrscheinlichkeit Es werde die Frage nach dem Ort von Punkten (x, y) betrachtet, für die die Dichte f denselben Wert hat. Dies ist die Menge der Punkte $\{(x, y)|f(x, y) = k_0\}$, k_0 eine Konstante. Es ist $f(x, y) = k_0$, wenn der Exponent der Exponentialfunktion in (0.6) einen bestimmten, konstanten Wert, etwa k_1, hat. Der Exponent ist durch

$$h(x,y) = \frac{1}{2(1-r^2)}\left(\frac{(y-\mu_y)^2}{\sigma_y^2} + \frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2r(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right) \tag{0.8}$$

gegeben. Es muß also $h(x, y) = k_1$ eine Konstante sein; (0.8) ist die Gleichung einer Ellipse mit dem Mittelpunkt (μ_x, μ_y) und einer durch den Wert von r bestimmten Orientierung, – für $r = 0$ ist die Ellipse achsenparallel. Ellipsen sind die geometrischen Orte für Punkte mit gleicher Wahrscheinlichkeit, wenn (X, Y) gemeinsam normalverteilt sind.

Die Ellipsengleichung läßt sich kompakter in Matrixschreibweise formulieren. Es sei

$$\Sigma = \begin{pmatrix} \sigma_x^2 & r\sigma_x\sigma_y \\ r\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}. \tag{0.9}$$

Σ ist die Varianz-Kovarianz-Matrix für die zufälligen Veränderlichen X und Y. Die zu Σ inverse Matrix Σ^{-1} ist durch

$$\Sigma^{-1} = \frac{1}{(1-r^2)} \left(\begin{array}{cc} 1/\sigma_x^2 & -r/\sigma_x\sigma_y \\ -r/\sigma_x\sigma_y & 1/\sigma_y^2 \end{array} \right) \tag{0.10}$$

gegeben. Dann kann die Ellipsengleichung (0.8) in der Form

$$(x-\mu_x, y-\mu_y)\Sigma^{-1} \left(\begin{array}{c} x-\mu_x \\ y-\mu_y \end{array} \right) = k \tag{0.11}$$

geschrieben werden. Setzt man

$$\vec{x} - \vec{\mu} = \left(\begin{array}{c} x-\mu_x \\ y-\mu_y \end{array} \right),$$

so kann man

$$(\vec{x}-\vec{\mu})'\Sigma^{-1}(\vec{x}-\vec{\mu}) = k \tag{0.12}$$

schreiben.

Ist P die Matrix der Eigenvektoren von Σ und Λ die Diagonalmatrix der zugehörigen Eigenwerte, so gilt

$$\Sigma = P\Lambda P'. \tag{0.13}$$

Die Inverse Σ^{-1} von Σ ist

$$\Sigma^{-1} = (P\Lambda P')^{-1} = (P')^{-1}\Lambda^{-1}P^{-1}.$$

Aber P ist orthonormal, so dass $P^{-1} = P'$ ist, und folglich $(P^{-1})^{-1} = P$, so dass

$$\Sigma^{-1} = P\Lambda^{-1}P'. \tag{0.14}$$

Σ^{-1} hat also die gleichen Eigenvektoren wie Σ, und die Eigenwerte von Σ^{-1} sind durch die reziproken Eigenwerte von Σ gegeben. Gleichung (0.12) definiert für konstanten Wert k eine Ellipse, deren Orientierung durch die Orientierung der Eigenvektoren der Varian-Kovarianz-Matrix Σ bzw. deren Inverse Σ^{-1} gegeben ist.

Gesucht ist eine Transformation der Vektoren $\vec{\xi} = \vec{x} - \vec{\mu}$ mit $\vec{\xi}'\Sigma^{-1}\vec{\xi}$ derart, dass die Ellipse in eine achsenparallele Ellipse übergeht. Es wird also eine Rotationsmatrix T gesucht derart, dass jeder Vektor ξ, der der Bedingung $\xi'\Sigma^{-1}\xi = k$ genügt, in einen Vektor $\vec{\eta} = T\xi$ übergeht, der der Bedigung $\vec{\eta}'N\vec{\eta} = k$ genügt, wobei N eine Diagonalmatrix ist. Also soll gelten

$$\vec{\eta}'N\vec{\eta} = \xi'T'NT\xi = k.$$

Dann muß aber $T'NT = \Sigma^{-1}$ gelten, woraus folgt, dass $T' = P$ die Matrix der Eigenvektoren von Σ^{-1} und $N = \Lambda^{-1}$ die Diagonalmatrix der Eigenwerte von Σ^{-1} ist.

Für den 2-dimensionalen Fall lassen sich die Ausdrücke für die Eigenwerte von Σ noch explizit angeben:

$$\lambda_1 = \frac{1}{2}\left(s_x^2 + s_y^2 + \sqrt{(s_x^2 - s_y^2)^2 - 4r^2 s_x^2 s_y^2}\right) \tag{0.15}$$

$$\lambda_2 = \frac{1}{2}\left(s_x^2 + s_y^2 - \sqrt{(s_x^2 - s_y^2)^2 - 4r^2 s_x^2 s_y^2}\right) \tag{0.16}$$

Offenbar ist $\lambda_1 \geq \lambda_2$, und λ_1 wird kleiner für $|r| \to 1$, während λ_2 für diesen Fall größer wird. Für $r = 0$ erhält man den Spezialfall $\lambda_1 = s_x^2$, $\lambda_2 = s_y^2$, so dass Σ^{-1} die Eigenwerte $1/\sigma_x^2$ und $1/\sigma_y^2$ hat. Die Längen der Halbachsen der achsenparallelen Ellipse sind gleich $\sqrt{k/\lambda_j}$, $j = 1, 2$, und wegen $\lambda_j = 1/\sigma_j^2$ folgt, dass die Längen gleich $\sigma_x\sqrt{k}$, $\sigma_y\sqrt{k}$, also proportional zu den Streuungen sind.

Nun werde noch der Normalisierungsfaktor A für $f(x, y)$ (Gleichung (0.7)) betrachtet:

$$A = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}, \quad |r| < 1$$

Die Summe aller Wahrscheinlichkeiten muß bekanntlich stets gleich 1 sein. Bei stetigen Variablen entspricht die Summe einem Integral. Die eindimensionale Standardnormalverteilung ist durch die Dichte

$$f(z) = ae^{-z^2/2}, \quad -\infty < z < \infty$$

definiert, wobei a eine noch zu bestimmende Konstante ist. Es muß

$$\int_{-\infty}^{\infty} f(z)dz = \int_{-\infty}^{\infty} ae^{-z^2/2}dz = a\int_{-\infty}^{\infty} e^{-z^2/2}dz = 1$$

gelten. Aus der Analysis ist aber bekannt, dass

$$\int_{-\infty}^{\infty} e^{-z^2/2}dz = \sqrt{2\pi}$$

gilt, so dass $a\sqrt{2\pi} = 1$ folgt. Damit ergibt sich für die *Normierungskonstante* a

$$a = \frac{1}{2\pi}.$$

Für die 2-dimensionale Dichte verfährt man analog und gelangt so auf den Ausdruck für die Normierungskonstante A.

Natürlich will man auch 3-, 4- und allgemein n-dimensionale Dichten anwenden können. Während man bei einer 1-dimensionalen Verteilung mit dem Integral die Fläche unter der Dichtefunktion berechnet, um die Normierungskonstante zu bestimmen, muß man bei 2- und mehrdimensionalen Dichten ein Volumen unter einer Fläche berechnen. Es wäre mühsam und unökonomisch, die Konstante für jedes n separat zu berechnen. Man macht also von einem allgemeinen Resultat der Analysis Gebrauch, demzufolge das Volumen unter einer Fläche, die durch die n-dimensionale Gaußsche Dichte definiert ist, durch die *Determinante* der Matrix gegeben ist, die der inversen Varianz-Kovarianz-Matrix Σ^{-1} entspricht.

Die Determinante wird auf eine relativ komplizierte Weise aus den Elementen von Σ^{-1} berechnet, auf die hier nicht weiter eingegangen werden kann, zumal ein Modul zur Berechnung von Determinanten in jedem Statistikprogramm implementiert ist. Für den 2-dimensionalen Fall kann die Determinante allerdings leicht angegeben werden. Hat man etwa die Matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

so ist die Determinante durch

$$|M| = a \cdot d - b \cdot c$$

gegeben. Für Σ^{-1} erhält man demnach

$$|\Sigma^{-1}| = \frac{1}{(1-r^2)^2}\left(\frac{1}{\sigma_x^2\sigma_y^2} - \frac{r^2}{\sigma_x^2\sigma_y^2}\right) = \frac{1-r^2}{(1-r^2)^2}\left(\frac{1}{\sigma_x^2\sigma_y^2}\right) = \frac{1}{(1-r^2)\sigma_x^2\sigma_y^2}. \tag{0.17}$$

Der Vergleich mit der Definition von A zeigt, dass

$$A = \frac{1}{2\pi\sqrt{|\Sigma^{-1}|}} = \frac{1}{2\pi|\Sigma^{-1}|^{1/2}} \tag{0.18}$$

Für den allgemeinen, n-dimensionalen Fall erhält man

$$A = \frac{1}{(2\pi)^{(n/2)}|\Sigma^{-1}|^{1/2}}, \tag{0.19}$$

wobei Σ^{-1} natürlich eine $(n \times n)$-Matrix ist.

Der allgemeine Fall Jetzt sei ein n-dimensionaler Vektor $\vec{x} = (x_1, x_2, \dots, x_n)^t$ gegeben, und die Komponenten seien gemeinsam normalverteilt mit einer $n \times n$-Varianz-Kovarianz-Matrix Σ. Der Erwartungswert der j-ten Komponente sei μ_j, und der Vektor der Erwartungswerte sei $\vec{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^t$. Dann definiert

$$f(x_1, \dots, x_n) = A \exp\left[-(\vec{x} - \vec{\mu})^t\Sigma^{-1}(\vec{x} - \vec{\mu})\right] \tag{0.20}$$

mit der Normierungskonstanten

$$A = \frac{1}{(2\pi)^{1/2)n}|\Sigma^{-1}|^{1/2}} \tag{0.21}$$

den allgemeinen Fall. $|\Sigma^{-1}|^{1/2}$ entspricht dem σ im 1-dimensionalen Fall; man kann sagen, dass $|\Sigma^{-1}|^{1/2}$ die Gesamtvarianz des Vektors $\vec{x}$ repräsentiert.