Metodološki zvezki, Vol. 1, No. 1, 2004, 99-108

The Distribution of the Ratio of Jointly Normal Variables

Anton Cedilnik[1], Katarina Košmelj[2], and Andrej Blejec[3]

Abstract

We derive the probability density of the ratio of components of the bivariate normal distribution with arbitrary parameters. The density is a product of two factors, the first is a Cauchy density, the second a very complicated function. We show that the distribution under study does not possess an expected value or other moments of higher order. Our particular interest is focused on the shape of the density. We introduce a shape parameter and show that according to its sign the densities are classified into three main groups. As an example, we derive the distribution of the ratio $Z = -B_{m-1}/(mB_m)$ for a polynomial regression of order *m*. For $m = 1$, Z is the estimator for the zero of a linear regression, for $m = 2$, an estimator for the abscissa of the extreme of a quadratic regression, and for $m = 3$, an estimator for the abscissa of the inflection point of a cubic regression.

1 Introduction

The ratio of two normally distributed random variables occurs frequently in statistical analysis. For example, in linear regression, $E(Y \mid x) = \beta_0 + \beta_1 x$, the value x_0 for which the expected response $E(Y)$ has a given value y_0 is often of interest. The estimator for x_0, the random variable $X_0 = (y_0 - B_0)/B_1$, is under the standard regression assumption expressed as the ratio of two normally distributed and dependent random variables B_0 and B_1, which are the estimators for β_0 and β_1 and whose distributions and dependence are known from regression theory.

[1] Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia; Anton.Cedilnik@bf.uni-lj.si

[2] Biotechnical Faculty, University of Ljubljana Jamnikarjeva 101, 1000 Ljubljana, Slovenia; Katarina.Kosmelj@bf.uni-lj.si

[3] National Institute of Biology, University of Ljubljana, Večna pot 111, 1000 Ljubljana, Slovenia; Andrej.Blejec@uni-lj.si

Similar to the example above is the situation of a quadratic regression, $E(Y\mid x)=\beta_0+\beta_1 x+\beta_2 x^2$, where the value sought is the x_0 for which $E(Y)$ reaches its extreme value. At this point, the first derivative must be zero. Hence, $X_0=-B_1/2B_2$ is expressed as the ratio of two normally distributed and dependent variables as well.

From the literature it is known that the distribution of the ratio $Z=X/Y$, when X and Y are independent, is Cauchy. The probability density function for a Cauchy variable $U{:}\,C(a,b)$ is $p_U(x)=\dfrac{b}{\pi\left((x-a)^2+b^2\right)}$, where the location parameter a is the median, while the quartiles are obtained from the location parameter a and the positive scale parameter b, $q_{1,3}=a\mp b$. This density function $p_U(x)$ has 'fat tails', hence U does not possess an expected value or moments of higher order (Johnson et al., 1994).

Some results about the ratio from the literature are:

(a) The ratio Z of two *centred* normal variables is a Cauchy variable (Jamnik, 1971: 149):

$$\begin{bmatrix}X\\Y\end{bmatrix}:\ N\left(\mu_X=\mu_Y=0,\sigma_X,\sigma_Y,\ \rho\neq\pm 1\right)\quad\Rightarrow$$

$$Z=\frac{X}{Y}:\ C\left(a=\rho\frac{\sigma_X}{\sigma_Y},b=\frac{\sigma_X}{\sigma_Y}\sqrt{1-\rho^2}\right)$$

The simplest case is the ratio of two *independent standardised* normal variables which is a 'standard' Cauchy variable $C(0,1)$.

(b) The ratio Z of two *non-centred independent* normal variables is a particular Cauchy-like distribution. This result is shown in Kamerud (1978).

(c) The ratio of two arbitrary normal variables is discussed in Marsaglia (1965) and leads again to a Cauchy-like distribution.

The case considered in (b) is not general and the result in the cited article is presented in a very implicit way. Marsaglia dealt with the ratio of two independent normal variables, having shown previously, however that any case could be transformed into this setting.

The objective of our work is to derive the probability density for the ratio of components of the bivariate normal distribution for a general setting. Let the vector $\mathbf{W}=[X\ \ Y]^{\mathrm{T}}:\ N\left(\mu_X,\mu_Y,\sigma_X>0,\sigma_Y>0,\rho\right)$ be distributed normally, with the density (for $\rho\neq\pm 1$):

$$p_{\mathbf{W}}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\cdot\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X{}^2}-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y{}^2}\right]\right)$$

and with the expected value and the variance-covariance matrix

$$E(\mathbf{W}) = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \quad \operatorname{var}(\mathbf{W}) = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Our aim is to express the density function of the ratio $Z = X/Y$ explicitly, in terms of the parameters of the bivariate normal distribution. We shall also discuss the degenerate situation, $\rho = \pm 1$.

2 Probability density for the ratio

The following theorem is the basis for our derivation of the probability density for the ratio (Jamnik, 1971: 148).

Theorem 1. *Let* $\mathbf{W} = [X \;\; Y]^{\mathrm{T}}$ *be a continuously distributed random vector with a probability density function* $p_{\mathbf{W}}(x, y)$. *Then* $Z = X/Y$ *is a continuously distributed random variable with the probability density function*

$$p_Z(z) = \int_{-\infty}^{\infty} |y|\, p_{\mathbf{W}}(zy, y)dy = \int_0^{\infty} - \int_{-\infty}^{0} y\, p_{\mathbf{W}}(zy, y)dy\;. \tag{2.1}$$

For the derivation of $p_Z(z)$ for the ratio of the components of a bivariate normal vector we calculated the integral (2.1) using formulae in the Appendix. A long but straightforward calculation gives the next theorem.

Theorem 2. *The probability density for* $Z = X/Y$, *where* $[X \;\; Y]^{\mathrm{T}} : N(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho \neq \pm 1)$ *is expressed as a product of two terms:*

$$p_Z(z) = \frac{\sigma_X\sigma_Y\sqrt{1-\rho^2}}{\pi(\sigma_Y^2 z^2 - 2\rho\sigma_X\sigma_Y z + \sigma_X^2)} \cdot \left[\exp\left(-\frac{1}{2}\cdot \sup R^2\right)\cdot\left(1 + \frac{R\cdot\Phi(R)}{\varphi(R)}\right)\right] =$$

$$= \frac{\sigma_X\sigma_Y\sqrt{1-\rho^2}}{\pi(\sigma_Y^2 z^2 - 2\rho\sigma_X\sigma_Y z + \sigma_X^2)} \cdot \left[\exp\left(-\frac{1}{2}\cdot \sup R^2\right) + \sqrt{2\pi}\cdot R\cdot\Phi(R)\cdot\exp\left(-\frac{1}{2}\cdot\left[\sup R^2 - R^2\right]\right)\right] \tag{2.2}$$

where:

$$R = R(z) = \frac{(\sigma_Y^2\mu_X - \rho\sigma_X\sigma_Y\mu_Y)z - \rho\sigma_X\sigma_Y\mu_X + \sigma_X^2\mu_Y}{\sigma_X\sigma_Y\sqrt{1-\rho^2}\cdot\sqrt{\sigma_Y^2 z^2 - 2\rho\sigma_X\sigma_Y z + \sigma_X^2}} = \frac{\left(\frac{\mu_X}{\sigma_X} - \rho\frac{\mu_Y}{\sigma_Y}\right)z - \left(\rho\frac{\mu_X}{\sigma_X} - \frac{\mu_Y}{\sigma_Y}\right)\frac{\sigma_X}{\sigma_Y}}{\sqrt{1-\rho^2}\cdot\sqrt{z^2 - 2\rho\frac{\sigma_X}{\sigma_Y}z + \left(\frac{\sigma_X}{\sigma_Y}\right)^2}} \tag{2.2a}$$

$$\sup R^2 = \frac{\sigma_Y^2\mu_X^2 - 2\rho\sigma_X\sigma_Y\mu_X\mu_Y + \sigma_X^2\mu_Y^2}{\sigma_X^2\sigma_Y^2(1-\rho^2)} = \frac{\left(\frac{\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{\mu_X}{\sigma_X}\frac{\mu_Y}{\sigma_Y} + \left(\frac{\mu_Y}{\sigma_Y}\right)^2}{1-\rho^2}, \tag{2.2b}$$

$$\sup R^2 - R^2 = \frac{(\mu_X - \mu_Y z)^2}{\sigma_Y^2 z^2 - 2\rho\sigma_X\sigma_Y z + \sigma_X^2} = \frac{\left(\frac{\mu_X}{\sigma_X}\frac{\sigma_X}{\sigma_Y} - \frac{\mu_Y}{\sigma_Y} z\right)^2}{z^2 - 2\rho\frac{\sigma_X}{\sigma_Y} z + \left(\frac{\sigma_X}{\sigma_Y}\right)^2}. \tag{2.2c}$$

The first factor in (2.2), the *standard part*, is the density for a non-centred Cauchy variable, $C\left(a = \rho\frac{\sigma_X}{\sigma_Y}, b = \frac{\sigma_X}{\sigma_Y}\sqrt{1-\rho^2}\right)$. We have to stress that this factor is independent of the expected values μ_X and μ_Y.

The second factor, the *deviant part*, is a complicated function of *z*, including also the error function $\Phi(.)$ (in Gauss form; see Appendix). We need four parameters: ρ, $\frac{\mu_X}{\sigma_X}$, $\frac{\mu_Y}{\sigma_Y}$ and $\frac{\sigma_X}{\sigma_Y}$, to fully describe the distribution. It is strictly positive and asymptotically constant – it has the same positive value for both $z = \pm\infty$, due to the fact that $R(\pm\infty) = \pm\frac{\sigma_Y\mu_X - \rho\sigma_X\mu_Y}{\sigma_X\sigma_Y\sqrt{1-\rho^2}}$. Therefore, the asymptotic behaviour of $p_Z(z)$ is the same as that of the Cauchy density, so $E(Z)$ and other moments do not exist.

We wrote the deviant part in (2.2) in two forms. The first form is nicer and can also be found in Marsaglia (1965), but the second form is better for numerical purposes.

A more detailed analysis of $p_Z(z)$ led us to the definition of the *shape parameter* ω:

$$\omega = \frac{\mu_Y}{\sigma_Y}\left(\frac{\mu_X}{\sigma_X} - \rho\frac{\mu_Y}{\sigma_Y}\right), \tag{2.3}$$

based on $R(\pm\infty)$ and $\frac{dR}{dz} = \frac{\sigma_X\sigma_Y\sqrt{1-\rho^2}(\mu_X - \mu_Y z)}{\left(\sigma_Y^2 z^2 - 2\rho\sigma_X\sigma_Y z + \sigma_X^2\right)^{3/2}}$. The sign of ω separates three different *types* of shape of $p_Z(z)$:

I. $\omega > 0$

II. $\omega < 0$

III. $\omega = 0$ which occurs in three variants:

- **a.** $\mu_Y \neq 0$,
- **b.** $\mu_Y = 0 \neq \mu_X$,
- **c.** $\mu_Y = 0 = \mu_X$.

The derivative of the deviant part led us to the definitions of two quantities for types I and II: $u = \dfrac{\mu_X}{\mu_Y} = \dfrac{\dfrac{\mu_X}{\sigma_X}}{\dfrac{\mu_Y}{\sigma_Y}} \cdot \dfrac{\sigma_X}{\sigma_Y}$ and $d = \dfrac{\rho \dfrac{\mu_X}{\sigma_X} - \dfrac{\mu_Y}{\sigma_Y}}{\dfrac{\mu_X}{\sigma_X} - \rho \dfrac{\mu_Y}{\sigma_Y}} \cdot \dfrac{\sigma_X}{\sigma_Y}$. u is the abscissa of the local maximum and d the abscissa of local minimum of the deviant part. For type I: $d < a < u$, and for type II: $u < a < d$; as previously, $a = \rho \dfrac{\sigma_X}{\sigma_Y}$, the centre of the standard part (see Figure 1).

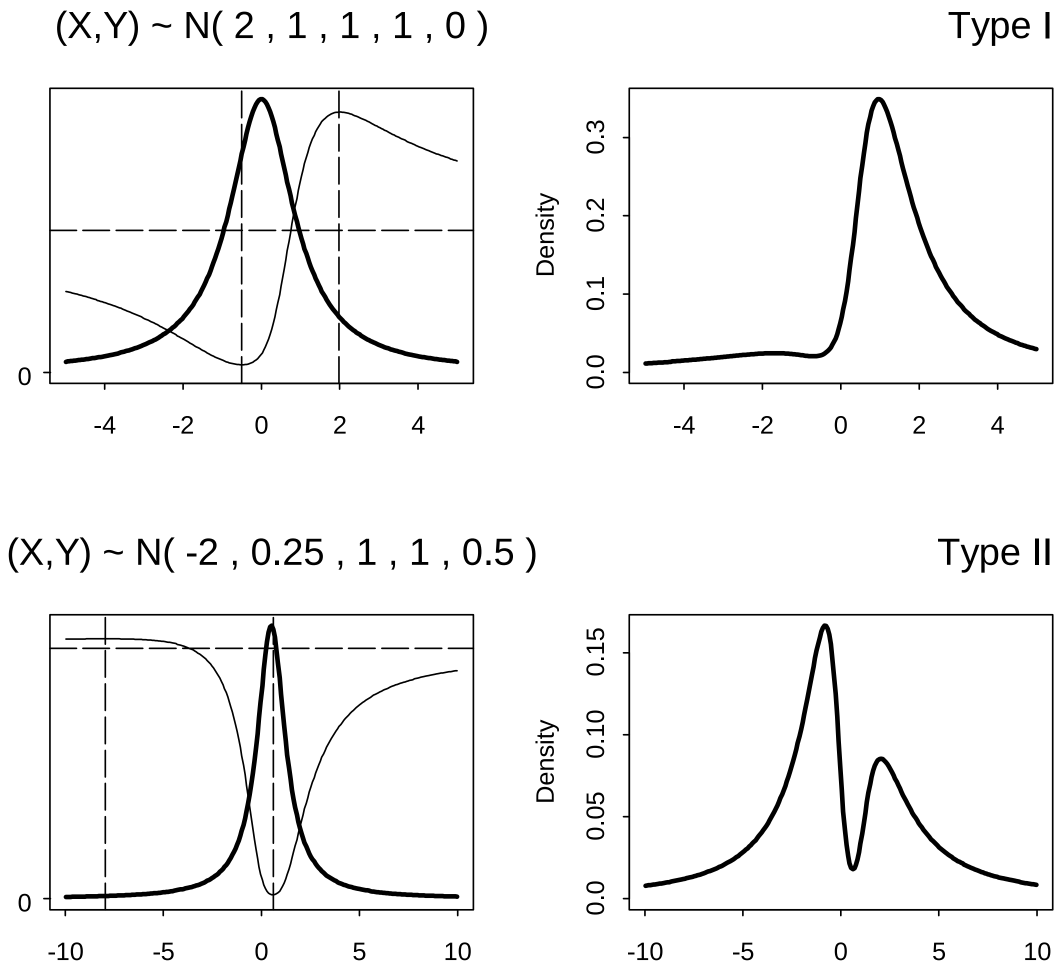

Figure 1: A case with a positive shape parameter (Type I) and with a negative shape parameter (Type II). On the left, the standard Cauchy part (thick line) and the deviant part (thin line) are presented; the functions are on different scales in order to depict the shapes of both functions on one plot. The vertical dashed lines indicate the abscissas of the local extremes of the deviant part, the horizontal dashed line is its asymptote. The right plot presents the graph of the density $p_Z(z)$.

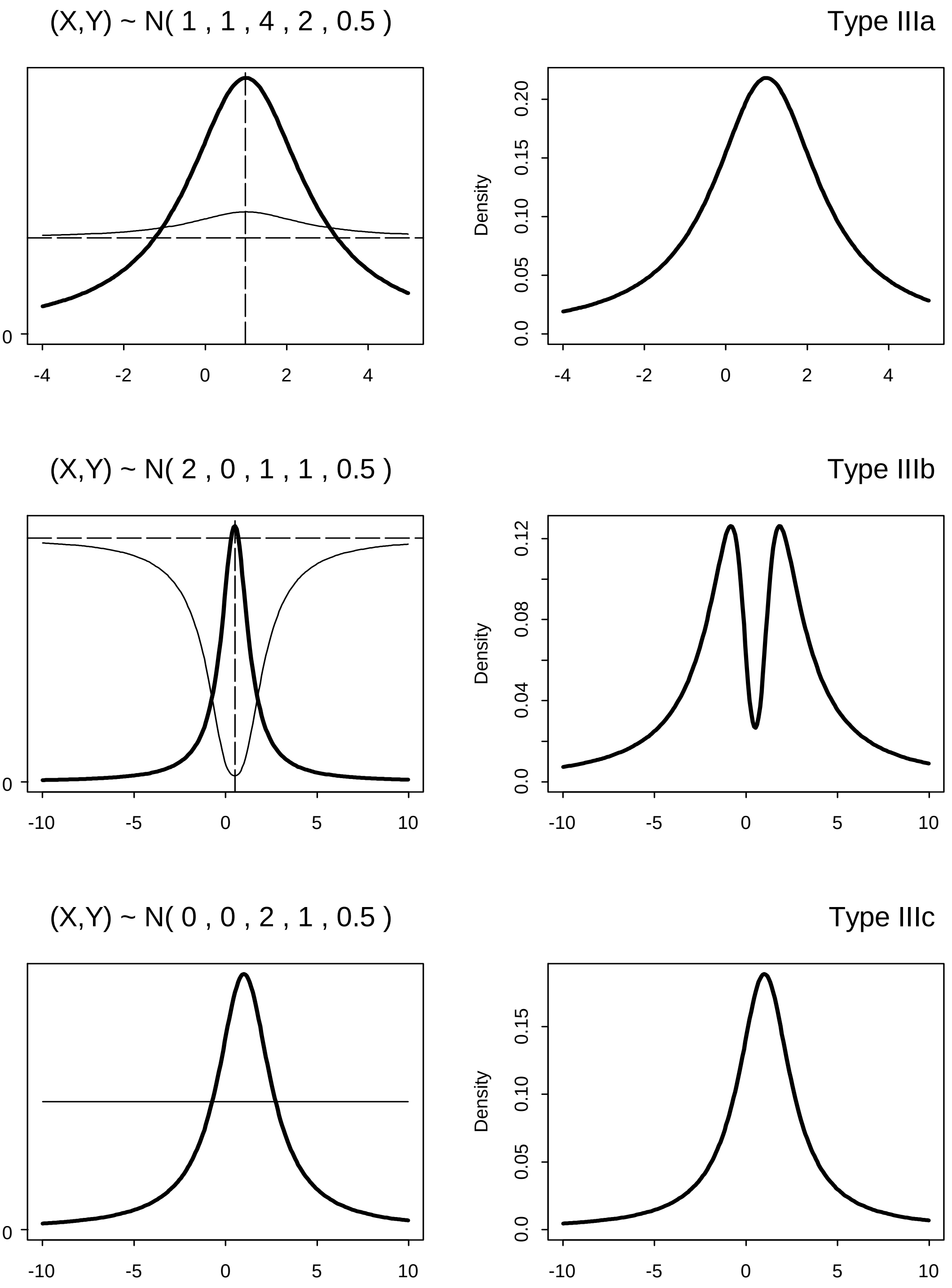

Figure 2: Three cases having zero value of the shape parameter (Type III). On the left, the standard Cauchy part (thick line) and the deviant part (thin line) are presented; the functions are on different scales in order to depict the shapes of both functions on one plot. The vertical dashed line indicates the abscissa of the local extreme of the deviant part, the horizontal dashed line is its asymptote. The right plot presents the graph of the density $p_Z(z)$.

Type III describes the marginal case, not likely to occur in practice. In variant IIIa (resp. IIIb), the deviant part has only a maximum (resp. a minimum) at $z = a$. In variant IIIc, the deviant part is equal to constant 1 (see Figure 2).

The median $M(Z)$ and mode(s) can not be obtained analytically for the general case; further numerical calculations have to be done for each particular case. But we have derived some partial results. For type I : $M(Z) > \rho \frac{\sigma_X}{\sigma_Y}$, for type II: $M(Z) < \rho \frac{\sigma_X}{\sigma_Y}$, for type III: $p_Z(z)$ is symmetric and $M(Z) = \rho \frac{\sigma_X}{\sigma_Y}$. Variants IIIa and IIIc are unimodal; generally, $p_Z(z)$ may be uni- or bimodal.

The distribution function and quantiles require numerical integration.

3 Degenerate situation

Now, let us consider the case $\rho = \pm 1$, but still with $\sigma_X > 0$, $\sigma_Y > 0$. Then, the distribution of $\mathbf{W} = [X \ \ Y]^{\mathrm{T}}$ is degenerate, and with probability 1, it holds $\frac{Y - \mu_Y}{\sigma_Y} = \rho \cdot \frac{X - \mu_X}{\sigma_X}$; hence: $Z = \frac{X}{Y} = \rho \frac{\sigma_X}{\sigma_Y} + \frac{\mu_X - \rho \frac{\sigma_X}{\sigma_Y} \mu_Y}{Y}$. Since the marginal distribution $Y : N(\mu_Y, \sigma_Y)$ is the usual normal distribution, it is easy to find the probability density for *Z* from the following theorem.

Theorem 3. *If* $Y : N(\mu_Y, \sigma_Y)$ *and* $Z = a + \frac{c}{Y}$, $c \neq 0$, *then Z has the density given by*

$$p_Z(z) = \frac{|c|}{\sigma_Y \sqrt{2\pi}} \cdot (z-a)^{-2} \cdot \exp\left(-\frac{1}{2\sigma_Y{}^2} \left[\frac{c}{z-a} - \mu_Y \right]^2 \right) .$$

The function $p_Z(z)$ from this theorem is much simpler than (2) and it is rather easy to find its characteristics, including quantiles and distribution function. Also, there are two modes that can be found explicitly, and between them there is a removable singularity $p_Z(a) = 0$. The expected value, as in non-degenerate cases, does not exist.

It is worth noting that in the degenerate case the shape parameter (3) is zero precisely when $p_Z(z)$ is symmetric, as in the non-degenerate case. According to the sign of the shape parameter, the relations between the median $M(Z)$ and the quantity $a = \rho \frac{\sigma_X}{\sigma_Y}$ remain the same, as well.

4 Examples

Now, let us discuss the two problems presented in the Introduction. First, we will consider a linear regression $E(Y \mid x) = \beta_0 + \beta_1 x$. We shall be interested in the x-axis intercept: $X_0 = -B_0 / B_1$, where B_0 and B_1 denote the estimators for β_0 and β_1. Under the assumption that $Y \mid x : N(\beta_0 + \beta_1 x,\ \sigma_{reg})$, the variable X_0 is expressed as the ratio of two normally distributed and dependent random variables $-B_0$ and B_1. Given the data $\{(x_i, y_i),\, i = 1,\ldots,n\}$ $(x_1 < x_n)$, we denote: $\bar{x} = \frac{1}{n}\sum x_i$, $w = \sqrt{\frac{1}{n}\sum x_i^{\,2}}$ and $q = \frac{\sigma_{reg}}{\sqrt{n(w^2 - \bar{x}^2)}}$. Then: $\begin{bmatrix} B_0 \\ B_1 \end{bmatrix} : N\left(\beta_0, \beta_1, qw, q, -\frac{\bar{x}}{w}\right)$, $\begin{bmatrix} -B_0 \\ B_1 \end{bmatrix} : N\left(-\beta_0, \beta_1, qw, q, \frac{\bar{x}}{w}\right)$. Hence, X_0 has a distribution with density function (2) on making the substitution: $\mu_X \to -\beta_0$, $\mu_Y \to \beta_1$, $\sigma_X \to q \cdot \bar{s}$, $\sigma_Y \to q$, $\rho \to \bar{x}/w$.

Now we shall be concerned with a general polynomial regression $E(Y \mid x) = \beta_0 + \beta_1 x + \ldots + \beta_m x^m$, $m \geq 1$. Let us define $Z = -B_{m-1}/(mB_m)$. For $m = 1$, $Z = X_0$ from the first example, the estimator for the zero of a linear regression. For $m = 2$, Z is an estimator for the abscissa of the extreme of a quadratic regression, and for $m = 3$, Z is an estimator for the abscissa of the inflection point of a cubic regression.

Introduce the following two data matrices: $\mathbf{v} = \left[x_i^k \ \begin{pmatrix} i = 1,\ldots,n \\ k = 0,\ldots,m \end{pmatrix} \right]_{n\times(m+1)}$, the matrix of powers of x-s , and $\mathbf{Y} = [y_i \quad (i = 1,\ldots,n)]_{n\times 1}$. The regularity condition, that there are at least $m+1$ distinct x-s, implies that the rank of $\mathbf{v}$ is precisely $m+1$. Hence, $\mathbf{v}^{\mathrm{T}} \cdot \mathbf{v}$ is invertible and $\mathbf{d} = (\mathbf{v}^{\mathrm{T}} \cdot \mathbf{v})^{-1} = [d_{jk}\ (j,k = 0,\ldots,m)]_{(m+1)\times(m+1)}$. Let $\boldsymbol{\beta} = [\beta_k \ \ (k = 0,\ldots,m)]_{(m+1)\times 1}$ be the column of the regression coefficients, and $\mathbf{B} = [B_k\, (k = 0,\ldots,m)]_{(m+1)\times 1}$ the column of their estimators. The normal system of equations in matrix form is then $\mathbf{v}^{\mathrm{T}} \cdot \mathbf{v} \ \cdot \mathbf{B} = \mathbf{v}^{\mathrm{T}} \cdot \mathbf{Y}$, and its solution is

$$\mathbf{B} = \mathbf{d} \cdot \mathbf{v}^{\mathrm{T}} \cdot \mathbf{Y} \tag{4.1}$$

As usual, we shall suppose that the y-s are independent normally distributed random variables with $E(y_i) = \beta_0 + \beta_1 x_i + \ldots + \beta_m x_i^{\,m}$, $\mathrm{var}(y_i) = \sigma_{reg}^2$. Hence, the vector $\mathbf{Y}$ is normally distributed with $E(\mathbf{Y}) = \mathbf{v} \cdot \boldsymbol{\beta}$, $\operatorname{var}(\mathbf{Y}) = \sigma_{reg}^2 \mathbf{I}$. According to (4), $\mathbf{B}$ is also normally distributed, $E(\mathbf{B}) = (\mathbf{d} \cdot \mathbf{v}^{\mathrm{T}}) \cdot E(\mathbf{Y}) = \boldsymbol{\beta}$, $\operatorname{var}(\mathbf{B}) = (\mathbf{d} \cdot \mathbf{v}^{\mathrm{T}}) \cdot \operatorname{var}(\mathbf{Y}) \cdot (\mathbf{d} \cdot \mathbf{v}^{\mathrm{T}})^{\mathrm{T}} = \sigma_{reg}^2 \mathbf{d}$.

Introduce two matrices: $\mathbf{u} = \begin{bmatrix} 0 & \cdots & 0 & -1 & 0 \\ 0 & \cdots & 0 & 0 & m \end{bmatrix}_{2\times(m+1)}$ and $\mathbf{W} = \mathbf{u} \cdot \mathbf{B} = \begin{bmatrix} -B_{m-1} \\ mB_m \end{bmatrix}_{2\times 1}$.

$\mathbf{W}$ is also a normal variable with $E(\mathbf{W}) = \mathbf{u} \cdot E(\mathbf{B}) = \mathbf{u} \cdot \boldsymbol{\beta} = \begin{bmatrix} -\beta_{m-1} \\ m\beta_m \end{bmatrix}$ and $\operatorname{var}(\mathbf{W}) = \mathbf{u} \cdot \operatorname{var}(\mathbf{B}) \cdot \mathbf{u}^{\mathrm{T}} = \sigma_{reg}^2 \mathbf{u} \cdot \mathbf{d} \cdot \mathbf{u}^{\mathrm{T}} = \sigma_{reg}^2 \begin{bmatrix} d_{m-1,m-1} & -md_{m-1,m} \\ -md_{m-1,m} & m^2 d_{m,m} \end{bmatrix}$. Therefore, the distribution of $\mathbf{W}$ is $N\left(-\beta_{m-1}, m\beta_m, \sigma_{reg}\sqrt{d_{m-1,m-1}}, m\sigma_{reg}\sqrt{d_{m,m}}, -\frac{d_{m-1,m}}{\sqrt{d_{m-1,m-1}d_{m,m}}} \right)$.

Hence, Z has a distribution with density function (2) with the exchange: $\mu_X \to -\beta_{m-1}$, $\mu_Y \to m\beta_m$, $\sigma_X \to \sigma_{reg}\sqrt{d_{m-1,m-1}}$, $\sigma_Y \to m\sigma_{reg}\sqrt{d_{m,m}}$, $\rho \to -\frac{d_{m-1,m}}{\sqrt{d_{m-1,m-1}d_{m,m}}}$.

References

[1] Jamnik, R. (1971): *Verjetnostni račun*. Mladinska knjiga, Ljubljana.

[2] Johnson N.L, Kotz, S., and Balakrishnan, N. (1994): *Continuous Univariate Distributions*. **1**. John Wiley and Sons.

[3] Kamerud D. (1978): The random variable *X*/*Y*, *X*, *Y* normal. *The American Mathematical Monthly*, **85**, 207.

[4] Marsaglia, G. (1965): Ratios of normal variables and ratios of sums of uniforms variables. *JASA*, **60**, 163-204.

Appendix

$$a > 0 \quad \Rightarrow \quad \int_0^m t \cdot \exp(-at^2 + bt + c)dt =$$

$$= \frac{e^c}{2a}\left[1 - \exp(bm - am^2)\right] \; + \; b\sqrt{\frac{\pi}{a}} \cdot \frac{e^c}{2a} \cdot \exp\left(\frac{b^2}{4a}\right) \cdot \left[\Phi\left(\frac{b}{\sqrt{2a}}\right) \; + \; \Phi\left(m\sqrt{2a} - \frac{b}{\sqrt{2a}}\right)\right]$$

$$\int_{-\infty}^{\infty} |t| \cdot \exp(-at^2 + bt + c)dt = \int_0^{\infty} \; + \; \int_0^{-\infty} t \cdot \exp(-at^2 + bt + c)dt \; = \; \frac{e^c}{a}\left[1 + r \cdot \frac{\Phi(r)}{\varphi(r)}\right],$$

where: $r = \dfrac{b}{\sqrt{2a}}$ $\qquad \varphi(r) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}}$ $\qquad \Phi(r) = \displaystyle\int_0^r \varphi(x)\,dx = \frac{1}{2}\operatorname{erf}\left(\frac{r}{\sqrt{2}}\right).$