Statistical appendix

Let U, V be random variables and $a \in \mathbb{R}$. Linear regression based on the use of the type $\mathbb{E}(U|V = a)$ of formulas with multinormal models are discussed, for example, in Section 4.3 of Weisberg, (2005), in Section 11.3 of Casella and Berger, (2002) and Chapters 4 and 7 in Johnson and Wichern, (2007). The conditioning with an event of type $\{V > a\}$ is more rare but it is used at least in Maddala, (1983) in the context of economical self-selection models. Similar mathematical formulae appear also in the context of *truncated* distributions Barr and Sherill, (1999), Johnson and Kotz, (1972) and Kotz, Balakrishnan and Johnson, (2005) since truncating a distribution is equivalent to conditioning on an interval.

Binormal distribution

A bivariate normal (*binormal*) distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is specified by the parameters $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and the positive definite covariance matrix $\boldsymbol{\Sigma}$, which can be written as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where $\sigma_1 > 0, \sigma_2 > 0$. The condition $-1 < \rho < 1$ is sufficient to guarantee that $\boldsymbol{\Sigma}$ is positive definite.

Assume first that $\sigma_1 = \sigma_2 = 1$, $\mu_1 = \mu_2 = 0$ and that $\rho \neq 0$. If (U, V) follows this normalized distribution then there exists a representation $V = Y_1$ and $U = \rho Y_1 + \sqrt{1-\rho^2}Y_2$, where Y_1 and $Y_2$are two *independent* $N(0,1)$ random variables. From this representation, it follows that the conditional distribution of U given $V = v$ is $N(\rho v, 1-\rho^2)$, in other words, $\mathbb{E}(U|V = v) = \rho v$. When the conditioning is done with, for example, the event of the type $\{V > \alpha\}$, the computations are done with the truncated distribution. The PDF of the truncated distribution, with the truncation $v > \alpha$ for the normalized pair (U, V) is

$$f_\alpha(u, v) = \begin{cases} \dfrac{exp\left(-\dfrac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right)}{2\pi\sqrt{1-\rho^2}(1-\Phi(\alpha))}, & u \in \mathbb{R}, \quad v > \alpha \\ 0, & u \in \mathbb{R}, \quad v \leq \alpha \end{cases}$$

where Φ is the CDF of the $N(0,1)$ distribution. Integration of $uf_\alpha(u, v)$ then gives

$$\mathbb{E}(U|V > \alpha) = \frac{\rho\varphi(\alpha)}{1-\Phi(\alpha)}$$

for the normalized pair (U, V). The φ above is the PDF of the $N(0,1)$ distribution. Applying this result to the general pair $(Z_1, Z_2) = (\sigma_1 U + \mu_1, \sigma_2 V + \mu_2) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $a \in \mathbb{R}$ gives

$$\mathbb{E}(Z_1|Z_2 > a) = \mathbb{E}(\sigma_1 U + \mu_1|\sigma_2 V + \mu_2 > a) = \sigma_1 \mathbb{E}\left(U\middle|V > \frac{a-\mu_2}{\sigma_2}\right) + \mu_1 = \mu_1 + \sigma_1\frac{\rho\varphi(\alpha)}{1-\Phi(\alpha)} \qquad (1)$$

where, in order to simplify the notation, define for all $a \epsilon \mathbb{R}$, $\alpha = \alpha(a) = (a - \mu_2)/\sigma_2$. The same approach gives

$$\mathbb{E}(Z_1|Z_2 \leq a) = \mu_1 - \sigma_1\frac{\rho\varphi(\alpha)}{\Phi(\alpha)}$$

Next, we show how to compute the variance of the conditioned variable. The variance is used to quantify the uncertainty of the binormal model. Using, for example, integration by parts rule the integration of $u^2 f_\alpha(u, v)$ gives

$$\mathbb{E}(U^2|V > \alpha) = 1 + \frac{\rho^2\alpha\varphi(\alpha)}{1-\Phi(\alpha)}$$

and, combining this with $\mathbb{E}(U|V > \alpha)$ gives

$$Var(U|V > \alpha) = 1 + \frac{\rho^2\alpha\varphi(\alpha)}{1-\Phi(\alpha)} - \left(\frac{\rho\varphi(\alpha)}{1-\Phi(\alpha)}\right)^2$$

for the normalized pair (U, V). The general formula is, with the shorthand notation $\alpha = \alpha(a)$,

$$Var(Z_1|Z_2 > a) = \sigma_1^2\left(1 + \frac{\rho^2\alpha\varphi(\alpha)}{1-\Phi(\alpha)} - \left(\frac{\rho\varphi(\alpha)}{1-\Phi(\alpha)}\right)^2\right) \tag{2}$$

Similar computations give

$$Var(Z_1|Z_2 \le a) = \sigma_1^2\left(1 - \frac{\rho^2\alpha\varphi(\alpha)}{\Phi(\alpha)} - \left(\frac{\rho\varphi(\alpha)}{\Phi(\alpha)}\right)^2\right)$$

Trinormal distribution

A trivariate normal (*trinormal*) distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is specified by the parameters $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ and the positive definite covariance matrix $\boldsymbol{\Sigma}$ which can be written as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$$

where σ_1^2, σ_2^2 and σ_3^2 are univariate marginal variances and ρ_{12}, ρ_{13} and ρ_{23} are pairwise correlations. If $\rho_{ij} = \rho_{ik} = 0$ for some i, then there is no reason to include the i:th variable in the trinormal distribution model since it cannot provide any information about the j:th or k:th variable via the model.

Assume now that Y_1, Y_2 and Y_3 are independent $N(0,1)$ random variables and let the matrix $\boldsymbol{A}$ be defined as

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{12} & \sqrt{1-\rho_{12}^2} & 0 \\ \rho_{13} & \dfrac{\rho_{23}-\rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}^2}} & \sqrt{1-\rho_{13}^2 - \dfrac{(\rho_{23}-\rho_{12}\rho_{13})^2}{1-\rho_{12}^2}} \end{pmatrix}$$

Then the vector of random variables defined by $(X_1, X_2, X_3)^T = \boldsymbol{A}(Y_1, Y_2, Y_3)^T$ is trinormal with covariance matrix

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$$

and location vector $(0,0,0)$. The conditional distribution $(X_1, X_2)^T|X_3 = x_3$ is binormal with location vector $(\rho_{13}x_3, \rho_{23}x_3)^T$ and covariance matrix

$$\begin{pmatrix} 1-\rho_{13}^2 & \rho_{12}-\rho_{13}\rho_{23} \\ \rho_{12}-\rho_{13}\rho_{23} & 1-\rho_{23}^2 \end{pmatrix}$$

see, for example, Result 4.6 in Johnson and Wichern, (2007). The conditional distribution $X_1|X_2 = x_2, X_3 = x_3$ is univariate normal with expected value

$$\left(\frac{\rho_{12}-\rho_{13}\rho_{23}}{1-\rho_{23}^2}\right)x_2 + \left(\frac{\rho_{13}-\rho_{12}\rho_{23}}{1-\rho_{23}^2}\right)x_3$$

and variance

$$1 - \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{23}^2}$$

see Result 4.6 in Johnson and Wichern, (2007). The general formula for the conditional expectation is then found by applying the above to the random variables $Z_i = \sigma_i X_i + \mu_i$, $i = 1,2,3$. The conditional variance is

$$Var(Z_1 | Z_2 = z_2, Z_3 = z_3) = \sigma_1^2 \left(1 - \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{23}^2} \right) \quad (3)$$

The conditional variance is smaller than the marginal variance; the uncertainty decreases since correlation means that the two conditioning variables contain information about the conditioned variable.

References

Barr, D. R., & Sherill, E. T. (1999). Mean and Variance of Truncated Normal Distributions. *The American Statistician*, *53*(4), 357–361.

Casella, G., & Berger, R. L. (2002). *Statistical Inference* (2nd ed.). Duxbury.

Johnson, N. L., & Kotz, S. (1972). Bivariate and Trivariate Normal Distributions. In *Distributions in Statistics: Continuous Multivariate Distributions* (pp. 84–131). John Wiley & Sons, Inc.

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). Pearson Prentice Hall.

Kotz, S., Balakrishnan, N., & Johnson, N. L. (2005). Bivariate and Trivariate Normal Distributions. In *Continuous Multivariate Distributions* (pp. 251–348). https://doi.org/10.1002/0471722065.ch46

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Economics*. Cambridge University Press.

Weisberg, S. (2005). *Applied Linear Regression* (3rd ed.). John Wiley & Sons.