

Mathematical Proof

Definition

We begin by defining notations that are needed for later sections. First, we define *moment* as the mean of a random variable or the mean of a given order of the variable:

$$\mu'_j = E(X^j), \quad (\text{A1})$$

where X is the target random variable, E is the expectation function, j is the order of moments, and μ'_j is the j order moments (mean) of random variable X . For example, μ'_1 or $E(X)$ is the population mean of the random variable X , which is the expected value of X . We will use μ to refer to μ'_1 in the following sections.

From this perspective, *central moments* are the expected value of the deviation of a random variable from a particular value. In this paper, we are only interested in the deviation from the population mean:

$$\mu_j = E[(X - \mu)^j], \quad (\text{A2})$$

where μ_j is the j order central moment of a random variable X from its population mean. The first central moment is 0. The second central moment (μ_2) is the population variance. The third central moment is simply the mean of the deviations cubed. The fourth central moment is simply the mean of the deviations to the fourth power. The sample estimates of the central moments are

$$m_j = E[(x - M)^j], \quad (\text{A3})$$

where m_j is the sample estimate of the j -order central moment of a random variable X from its population mean, and M is the sample mean. The sample central moment is only the central moment using the sample mean (rather than the population mean).

Another concept relating to skewness (γ_1), kurtosis (β_2), and excessive kurtosis (γ_2), and which is critical to understand directional dependency is the cumulant. The j -order *cumulant* (κ_j) is the value derived from taking the j^{th} -order derivative from the cumulant-generating function and substitute 0 for the argument variable of the function. The cumulant-generating function is the logarithm of the moment-generating function of a probability distribution (Gut, 2009). The first three orders of cumulants are equal to the first three orders of central moments: $\kappa_1 = \mu_1$, $\kappa_2 = \mu_2$, and $\kappa_3 = \mu_3$. If, however, $j > 3$, the j^{th} -order cumulant is not equal to the j^{th} -order central moment. The relation between the fourth-order cumulant and the fourth-order central moment is

$$\kappa_4 = \mu_4 - 3\mu_2^2. \quad (\text{A4})$$

where μ_2 is the second-order central moment, μ_4 is the fourth-order central moment, and κ_4 is the fourth-order cumulant of a random variable. Note that the cumulant of a normal distribution is 0 if $j > 2$ but the central moment of a normal distribution is not necessary equal to 0 if $j > 2$.

Skewness and excessive kurtosis can be written in terms of cumulants as

$$\gamma_1 = \frac{\kappa_3}{\sqrt{\kappa_2^3}} = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{E[(X - \mu)^3]}{\sqrt{(E[(X - \mu)^2])^3}}, \quad (\text{A5})$$

and,

$$\gamma_2 = \frac{\kappa_4}{\kappa_2^2} = \frac{\mu_4 - 3\mu_2^2}{\mu_2^2} = \frac{E[(X - \mu)^4]}{(E[(X - \mu)^2])^2} - 3. \quad (\text{A6})$$

where κ_2 , κ_3 , and κ_4 are the second-, third-, and the fourth-order cumulant of a random variable, and μ_2 , μ_3 , and μ_4 are the second-, third-, and the fourth-order central moment of a random variable.

The unbiased estimator of a cumulant takes a different functional form from its population cumulant. The unbiased estimator of a cumulant is also called a *k-statistic* (Cramér, 1946). The second-order *k-statistic* (k_2) is equal to the sample estimate of variance, which is

$$k_2 = \frac{n}{n-1} m_2, \quad (\text{A7})$$

where n is the sample size, and m_2 is the second-order sample central moment of a random variable.

The third-order *k-statistics* (k_3) and the fourth-order *k-statistics* (k_4) are

$$k_3 = \frac{n^2}{(n-1)(n-2)} m_3, \quad (\text{A8})$$

and,

$$k_4 = \frac{n^2[(n+1)m_4 - 3(n-1)m_2^2]}{(n-1)(n-2)(n-3)}, \quad (\text{A9})$$

where n is the sample size, and m_3 and m_4 are the third- and the fourth-order sample central moments of a random variable. The sample estimates of skewness (g_1) and excessive kurtosis (g_2) are (Cramér, 1946)

$$g_1 = \frac{k_3}{k_2^{3/2}} = \frac{\sqrt{n(n-1)}}{(n-2)} \frac{E[(X-M)^3]}{\sqrt{(E[(X-M)^2])^3}}, \quad (\text{A10})$$

and,

$$g_2 = \frac{k_4}{k_2^2} = \frac{(n+1)}{(n-2)(n-3)} \frac{E[(X-M)^4]}{(E[(X-M)^2])^2} - 3 \frac{(n-1)^2}{(n-2)(n-3)}. \quad (\text{A11})$$

where M is the sample mean. Other notations are defined in Equation A7-A9.

Proofs of Equations 8 and 9

Dodge and Rousson (2001) relied on the fact that the cumulant (see Equation A4) of the sum of an independent random variable is equal to the sum of the cumulants of individual terms

(Gut, 2009). Thus, applying the cumulants on both sides of Equation 9 (X is the explanatory variable),

$$\kappa_j(Y) = \beta_{YX}^j \kappa_j(X) + \kappa_j(e). \quad (\text{A12})$$

where $\kappa_j(Y)$, $\kappa_j(X)$, and $\kappa_j(e)$ are the j -order cumulants of Y , X , and e , respectively, and β_{YX} is the regression coefficient of Y on X . If the regression error is normally distributed, then $\kappa_j(e) = 0$ when $j > 2$. As a result, Equation A12 can be transformed as

$$\beta_{YX}^j = \frac{\kappa_j(Y)}{\kappa_j(X)}. \quad (\text{A13})$$

In addition, the correlation coefficient between X and Y (ρ_{XY}) formula is

$$\rho_{XY} = \beta_{YX} \frac{\sigma_X}{\sigma_Y} = \beta_{YX} \frac{[\kappa_2(X)]^{1/2}}{[\kappa_2(Y)]^{1/2}}. \quad (\text{A14})$$

where σ_X and σ_Y are the standard deviations of X and Y , which are equal to the square root of the second-cumulants of X ($\kappa_2(X)$) and Y ($\kappa_2(Y)$), respectively. Combine Equations A13 and A14:

$$\rho_{XY}^j = \frac{\kappa_j(Y) [\kappa_2(X)]^{j/2}}{\kappa_j(X) [\kappa_2(Y)]^{j/2}}. \quad (\text{A15})$$

The notations are defined in Equations A12 and A14. Using Equations A5, A6, and A15, the directional dependency between two variables can be determined by skewness (Dodge & Rousson, 2000, 2001) and excessive kurtosis (Dodge & Yadegari, 2010) as

$$\rho_{XY}^3 = \frac{\gamma_1(Y)}{\gamma_1(X)}, \quad (\text{8, repeated})$$

and,

$$\rho_{XY}^4 = \frac{\gamma_2(Y)}{\gamma_2(X)}. \quad (\text{9, repeated})$$

Proof of Equation 15

The standardized regression coefficient formula is

$$\beta_{S_{YZ}} = \beta_{YZ} \frac{\sigma_Z}{\sigma_Y}, \quad (\text{A16})$$

where $\beta_{S_{YZ}}$ is the standardized regression coefficient of Z predicting Y , controlling for X . Other notations are defined in Equation A14.

Assuming that X and Z are independent and e_1 and e_2 are normally distributed, from Equation 13, 14, A5, and A16, the skewness of Y_1 and Y_2 will be

$$\gamma_1(Y_1) = \frac{\kappa_3(Y_1)}{[\kappa_2(Y_1)]^{3/2}} = \frac{\beta_{YX}^3 \kappa_3(X) + \beta_{YZ.1}^3 \kappa_3(Z_1)}{\sigma_{Y_1}^3} = \beta_{S_{YX}}^3 \gamma_1(X) + \beta_{S_{YZ.1}}^3 \gamma_1(Z_1), \quad (\text{A17})$$

$$\gamma_1(Y_2) = \frac{\kappa_3(Y_2)}{[\kappa_2(Y_2)]^{3/2}} = \frac{\beta_{YZ.2}^3 \kappa_3(Z_2)}{\sigma_{Y_2}^3} = \beta_{S_{YZ.2}}^3 \gamma_1(Z_2), \quad (\text{A18})$$

where $\beta_{S_{YZ.1}}$ is the standardized regression coefficients of Y on Z at Time 1 controlling for X , $\beta_{S_{YZ.2}}$ is the standardized regression coefficient of Y on Z at Time 2, and $\beta_{S_{YX}}$ is the standardized regression coefficient of Y on X controlling for Z_1 . Other notations are defined in Equations 13 and 14.

Combine Equations A17 and A18,

$$\frac{\gamma_1(Y_1)}{\gamma_1(Y_2)} = \frac{\beta_{S_{YX}}^3 \gamma_1(X) + \beta_{S_{YZ.1}}^3 \gamma_1(Z_1)}{\beta_{S_{YZ.2}}^3 \gamma_1(Z_2)}. \quad (15, \text{repeated})$$

The Mathematical Proof of the Effect of Unobserved Explanatory Variable

We will prove that Equations 8–11 will not apply if a covariate is not observed in the model. We will only focus on the directional dependency using skewness. The proof for excessive kurtosis can be derived similarly.

Assume that X is an explanatory variable, Y is a response variable, Z is an unobserved covariate, and e is the regression error predicting Y by X and Z . Then, the regression equation is

$$Y = \beta_{YX}X + \beta_{YZ}Z + e, \quad (\text{A19})$$

where β_{YX} is the regression coefficient of Y on X controlling for Z , β_{YZ} is the regression coefficient of Y on Z controlling for X , and e is the regression error. We will consider two cases

separately: X and Z are independent and X and Z are correlated. First, if X , Z , and e are independent, from Equation A19, the third-order cumulant of Y , $\kappa_3(Y)$ will be

$$\kappa_3(Y) = \beta_{YX}^3 \kappa_3(X) + \beta_{YZ}^3 \kappa_3(Z) + \kappa_3(e). \quad (\text{A20})$$

where $\kappa_3(Y)$, $\kappa_3(X)$, and $\kappa_3(e)$ are the third-order cumulants of Y , X , and e , respectively, β_{YX} is the regression coefficient of Y on X controlling Z , and β_{YZ} is the regression coefficient of Y on Z , controlling X . Assuming that the regression error is normally distributed ($\kappa_3(e) = 0$), the regression coefficient will be

$$\beta_{YX} = \sqrt[3]{\frac{\kappa_3(Y) - \beta_{YZ}^3 \kappa_3(Z)}{\kappa_3(X)}}. \quad (\text{A21})$$

All notations are defined in Equation A20. Thus, from Equations A5, A14, A16, and A21, the correlation coefficient will be

$$\rho_{XY} = \frac{\sigma_X}{\sigma_Y} \sqrt[3]{\frac{\kappa_3(Y) - \beta_{YZ}^3 \kappa_3(Z)}{\kappa_3(X)}} = \sqrt[3]{\frac{\gamma_1(Y) - \beta_{S_{YZ}}^3 \gamma_1(Z)}{\gamma_1(X)}}. \quad (\text{A22})$$

where σ_X and σ_Y are the standard deviation of X and Y , respectively, and $\gamma_1(X)$, $\gamma_1(Y)$, and $\gamma_1(Z)$ are the skewness of X , Y , and Z , respectively. Other notations are defined in Equation A20. Because X and Z are not correlated in this model, $\beta_{S_{YZ}}$ is equivalent to the correlation between Y and Z (ρ_{YZ}). That is,

$$\rho_{XY} = \sqrt[3]{\frac{\gamma_1(Y) - \rho_{YZ}^3 \gamma_1(Z)}{\gamma_1(X)}}. \quad (\text{A23})$$

where $\gamma_1(X)$, $\gamma_1(Y)$, and $\gamma_1(Z)$ are the skewness of X , Y , and Z , respectively, and ρ_{YZ} is the correlation between Y and Z . The correlation, ρ_{YZ} , is cubed and the cube-root of the full equation is taken to get the correlation of X and Y . The key consequence of Equation A23 is that the accuracy of the directional dependency test is dependent on the covariate, Z . Thus, the ratio of the skewness can be either underestimated or overestimated by the effect of the skewness of any unobserved explanatory variable. Unfortunately, one never knows whether the effect of an

unobserved explanatory variable underestimates or overestimates the ratio of the skewness. Thus, this unfortunate fact makes the ratio of skewness unreliable in determining directional dependency.

If the covariate correlates with the explanatory variable, a similar proof can be used with the help of residual centering (Lance, 1988; Little, Bovaird, & Widaman, 2006). We will use residual centering in the covariate first. Then, the explanatory variable, X , will be independent to the residual of Z . Next, the sum-of-independent-cumulants rule will be used.

First, we use residual centering on Z as

$$Z = \beta_{ZX}X + u, \quad (\text{A24})$$

where u is the residual-centered covariate, which is not necessary to be normally distributed, and β_{ZX} is the regression coefficient predicting Z from X . Because X and u are independent, the third-order cumulant of u ($\kappa_3(u)$) will be

$$\kappa_3(u) = \kappa_3(Z) - \beta_{ZX}^3 \kappa_3(X). \quad (\text{A25})$$

where $\kappa_3(X)$ and $\kappa_3(Z)$ are the third-order cumulants of X and Z , respectively. Combining Equations 12 and A24,

$$Y = (\beta_{YX} + \beta_{YZ}\beta_{ZX})X + \beta_{YZ}u + e. \quad (\text{A26})$$

where β_{YX} is the regression coefficient of Y on X controlling for Z , β_{YZ} is the regression coefficient of Y on Z controlling for X , e is the regression error of Y controlling for the effects of X and Z . Assuming that X , u , and e are independent,

$$\kappa_3(Y) = (\beta_{YX} + \beta_{YZ}\beta_{ZX})^3 \kappa_3(X) + \beta_{YZ}^3 \kappa_3(u) + \kappa_3(e). \quad (\text{A27})$$

where $\kappa_3(Y)$ and $\kappa_3(e)$ are the third-order cumulants of Y and e . Combining Equations A25 and A27,

$$\kappa_3(Y) = (\beta_{YX} + \beta_{YZ}\beta_{ZX})^3 \kappa_3(X) + \beta_{YZ}^3 (\kappa_3(Z) - \beta_{ZX}^3 \kappa_3(X)) + \kappa_3(e). \quad (\text{A28})$$

Because population regression error is normally distributed ($\kappa_3(e) = 0$), the regression coefficient will be

$$\beta_{YX} = \left[\sqrt[3]{\frac{\kappa_3(Y) - \beta_{YZ}^3 \kappa_3(Z)}{\kappa_3(X)} + \beta_{YZ}^3 \beta_{ZX}^3} \right] - \beta_{YZ} \beta_{ZX}. \quad (\text{A29})$$

All notations are defined in Equation A24-A27. Therefore, from Equations, A5, A14, A20, and A29, the correlation coefficient between the explanatory variable and the response variable will be

$$\rho_{XY} = \left[\sqrt[3]{\frac{\gamma_1(Y) - \beta_{S_{YZ}}^3 \gamma_1(Z)}{\gamma_1(X)} + \beta_{S_{YZ}}^3 \beta_{S_{ZX}}^3} \right] - \beta_{S_{YZ}} \beta_{S_{ZX}}, \quad (\text{A30})$$

where $\beta_{S_{ZX}}$ is the standardized regression coefficient of X predicting Z , $\beta_{S_{YZ}}^3$ is the standardized regression coefficient of Z predicting Y controlling for X , $\gamma_1(X)$, $\gamma_1(Y)$, and $\gamma_1(Z)$ are the skewness of X , Y , and Z , respectively. The standardized regression coefficient of X predicting Z is equal to the correlation between X and Z (ρ_{XZ}) because the regression equation (Equation A24) involves only one predictor, then

$$\rho_{XY} = \left[\sqrt[3]{\frac{\gamma_1(Y) - \beta_{S_{YZ}}^3 \gamma_1(Z)}{\gamma_1(X)} + \beta_{S_{YZ}}^3 \rho_{XZ}^3} \right] - \beta_{S_{YZ}} \rho_{XZ}, \quad (\text{A31})$$

where $\beta_{S_{YZ}}$ is the standardized regression coefficient of Z predicting Y , controlling for X , ρ_{XZ} is the correlation between X and Z . Other notations are defined in Equation 13 and the proof of this equation is given in the Appendix A. In the special case where X and Z are uncorrelated, Equation A31 reduces to Equation A23. When a covariate predicting the response variable exists, the correlation between X and Y depends on (a) the skewness of Y , (b) the skewness of X , (c) the skewness of Z , (d) the standardized regression coefficient of Y on Z controlling for X , and (e) the correlation between X and Z . When the covariate is unobserved, the skewness of Z , the standardized regression coefficient of Y on Z controlling for X , and the correlation between X and

Z are not known. Hence, inferring directional dependency based on the skewness of Y and X might lead to an erroneous solution if unobserved covariates exist.

Supplementary Materials for

Pornprasertmanit, S., & Little, T. D. (in press). Determining directional dependency in causal associations. *International Journal of Behavioral Development*.