Exercises for Section 5.2

Exercise 5.6 Suppose that $X_1, X_2, \ldots$ are independent and identically distributed Normal $(0, \sigma^2)$ random variables.

(a) Based on the result of Example 5.7, Give an approximate test at $\alpha = .05$ for $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_a : \sigma^2 \neq \sigma_0^2$.

(b) For $n = 25$, estimate the true level of the test in part (a) for $\sigma_0^2 = 1$ by simulating 5000 samples of size $n = 25$ from the null distribution. Report the proportion of cases in which you reject the null hypothesis according to your test (ideally, this proportion will be about .05).

5.3 Sample Correlation

Suppose that $(X_1, Y_1), (X_2, Y_2), \ldots$ are independent and identically distributed vectors with $\mathrm{E}\ X_i^4 < \infty$ and $\mathrm{E}\ Y_i^4 < \infty$. For the sake of simplicity, we will assume without loss of generality that $\mathrm{E}\ X_i = \mathrm{E}\ Y_i = 0$ (alternatively, we could base all of the following derivations on the centered versions of the random variables).

We wish to find the asymptotic distribution of the sample correlation coefficient, r. If we let

$$\begin{pmatrix} m_x \\ m_y \\ m_{xx} \\ m_{yy} \\ m_{xy} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i^2 \\ \sum_{i=1}^n Y_i^2 \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} \tag{5.5}$$

and

$$s_x^2 = m_{xx} - m_x^2, s_y^2 = m_{yy} - m_y^2, \text{ and } s_{xy} = m_{xy} - m_x m_y, \tag{5.6}$$

then $r = s_{xy}/(s_x s_y)$. According to the central limit theorem,

$$\sqrt{n} \left\{ \begin{pmatrix} m_x \\ m_y \\ m_{xx} \\ m_{yy} \\ m_{xy} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \sigma_x^2 \\ \sigma_y^2 \\ \sigma_{xy} \end{pmatrix} \right\} \xrightarrow{d} N_5 \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{Cov}\,(X_1, X_1) & \cdots & \mathrm{Cov}\,(X_1, X_1Y_1) \\ \mathrm{Cov}\,(Y_1, X_1) & \cdots & \mathrm{Cov}\,(Y_1, X_1Y_1) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}\,(X_1Y_1, X_1) & \cdots & \mathrm{Cov}\,(X_1Y_1, X_1Y_1) \end{pmatrix} \right\} \tag{5.7}$$

Let Σ denote the covariance matrix in expression (5.7). Define a function $\mathbf{g} : \mathbb{R}^5 \to \mathbb{R}^3$ such that $\mathbf{g}$ applied to the vector of moments in Equation (5.5) yields the vector (s_x^2, s_y^2, s_{xy}) as defined in expression (5.6). Then

$$\nabla \mathbf{g} \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} -2a & 0 & -b \\ 0 & -2b & -a \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Therefore, if we let

$$\begin{aligned} \Sigma^* &= \left[\nabla \mathbf{g} \begin{pmatrix} 0 \\ 0 \\ \sigma_x^2 \\ \sigma_y^2 \\ \sigma_{xy} \end{pmatrix}\right]^\top \Sigma \left[\nabla \mathbf{g} \begin{pmatrix} 0 \\ 0 \\ \sigma_x^2 \\ \sigma_y^2 \\ \sigma_{xy} \end{pmatrix}\right] \\ &= \begin{pmatrix} \operatorname{Cov}(X_1^2, X_1^2) & \operatorname{Cov}(X_1^2, Y_1^2) & \operatorname{Cov}(X_1^2, X_1Y_1) \\ \operatorname{Cov}(Y_1^2, X_1^2) & \operatorname{Cov}(Y_1^2, Y_1^2) & \operatorname{Cov}(Y_1^2, X_1Y_1) \\ \operatorname{Cov}(X_1Y_1, X_1^2) & \operatorname{Cov}(X_1Y_1, Y_1^2) & \operatorname{Cov}(X_1Y_1, X_1Y_1) \end{pmatrix}, \end{aligned}$$

then by the delta method,

$$\sqrt{n}\left\{\begin{pmatrix} s_x^2 \\ s_y^2 \\ s_{xy} \end{pmatrix} - \begin{pmatrix} \sigma_x^2 \\ \sigma_y^2 \\ \sigma_{xy} \end{pmatrix}\right\} \xrightarrow{d} N_3(\mathbf{0}, \Sigma^*). \tag{5.8}$$

As an aside, note that expression (5.8) gives the same marginal asymptotic distribution for $\sqrt{n}(s_x^2 - \sigma_x^2)$ as was derived using a different approach in Example 4.11, since $\operatorname{Cov}(X_1^2, X_1^2)$ is the same as τ^2 in that example.

Next, define the function $h(a, b, c) = c/\sqrt{ab}$, so that we have $h(s_x^2, s_y^2, s_{xy}) = r$. Then

$$[\nabla h(a, b, c)]^\top = \frac{1}{2}\left(\frac{-c}{\sqrt{a^3 b}}, \frac{-c}{\sqrt{ab^3}}, \frac{2}{\sqrt{ab}}\right),$$

so that

$$[\nabla h(\sigma_x^2, \sigma_y^2, \sigma_{xy})]^\top = \left(\frac{-\sigma_{xy}}{2\sigma_x^3\sigma_y}, \frac{-\sigma_{xy}}{2\sigma_x\sigma_y^3}, \frac{1}{\sigma_x\sigma_y}\right) = \left(\frac{-\rho}{2\sigma_x^2}, \frac{-\rho}{2\sigma_y^2}, \frac{1}{\sigma_x\sigma_y}\right). \tag{5.9}$$

Therefore, if A denotes the 1×3 matrix in Equation (5.9), using the delta method once again yields

$$\sqrt{n}(r - \rho) \xrightarrow{d} N(0, A\Sigma^* A^\top).$$

To recap, we have used the basic tools of the multivariate central limit theorem and the multivariate delta method to obtain a *univariate* result. This derivation of univariate facts via multivariate techniques is common practice in statistical large-sample theory.

Example 5.10 Consider the special case of bivariate normal (X_i, Y_i). In this case, we may derive

$$\Sigma^* = \begin{pmatrix} 2\sigma_x^4 & 2\rho^2\sigma_x^2\sigma_y^2 & 2\rho\sigma_x^3\sigma_y \\ 2\rho^2\sigma_x^2\sigma_y^2 & 2\sigma_y^4 & 2\rho\sigma_x\sigma_y^3 \\ 2\rho\sigma_x^3\sigma_y & 2\rho\sigma_x\sigma_y^3 & (1+\rho^2)\sigma_x^2\sigma_y^2 \end{pmatrix}. \tag{5.10}$$

In this case, $A\Sigma^* A^\top = (1-\rho^2)^2$, which implies that

$$\sqrt{n}(r-\rho) \xrightarrow{d} N\{0, (1-\rho^2)^2\}. \tag{5.11}$$

In the normal case, we may derive a variance-stabilizing transformation. According to Equation (5.11), we should find a function $f(x)$ satisfying $f'(x) = (1-x^2)^{-1}$. Since

$$\frac{1}{1-x^2} = \frac{1}{2(1-x)} + \frac{1}{2(1+x)},$$

we integrate to obtain

$$f(x) = \frac{1}{2}\log\frac{1+x}{1-x}.$$

This is called Fisher's transformation; we conclude that

$$\sqrt{n}\left(\frac{1}{2}\log\frac{1+r}{1-r} - \frac{1}{2}\log\frac{1+\rho}{1-\rho}\right) \xrightarrow{d} N(0,1).$$

Exercises for Section 5.3

Exercise 5.7 Verify expressions (5.10) and (5.11).

Exercise 5.8 Assume $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent and identically distributed from some bivariate normal distribution. Let ρ denote the population correlation coefficient and r the sample correlation coefficient.

(a) Describe a test of $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$ based on the fact that

$$\sqrt{n}[f(r) - f(\rho)] \quad \xrightarrow{d} \quad N(0,1),$$

where $f(x)$ is Fisher's transformation $f(x) = (1/2)\log[(1+x)/(1-x)]$. Use $\alpha = .05$.

(b) Based on 5000 repetitions each, estimate the actual level for this test in the case when $\mathrm{E}\,(X_i) = \mathrm{E}\,(Y_i) = 0$, $\mathrm{Var}\,(X_i) = \mathrm{Var}\,(Y_i) = 1$, and $n \in \{3, 5, 10, 20\}$.