

# Statistics 581/582, Winter Quarter 2008

## Problem Set 9

**Reading:** Ferguson, Sections 7–12.

**Problem 31 (asymptotic distribution of the sample correlation coefficient, 6 points).** Let  $(X_1, Y_1)', (X_2, Y_2)', \dots$  be a sample from a bivariate distribution with finite fourth moments  $EX^4$  and  $EY^4$ . The sample correlation coefficient  $r = s_{XY}/(s_X s_Y)$  is a natural estimator of the correlation coefficient  $\rho = \sigma_{XY}/(\sigma_X \sigma_Y)$ . This problem discusses the large sample behavior of this estimator and reviews Section 8 of Ferguson. Please consult the textbook for notation, hints and further results.

- (a) Show that

$$\sqrt{n}(r - \rho) \longrightarrow_d \mathcal{N}(0, \gamma^2)$$

is asymptotically normal, and express the limit variance  $\gamma^2$  in terms of the moments of the random vector  $(X, Y)'$ .

- (b) Compute the correlation coefficient  $\rho$  and the limit variance  $\gamma^2$  in the case in which  $(X, Y)'$  is uniform on the union of the two squares  $(0, 1) \times (0, 1)$  and  $(-1, 0) \times (-1, 0)$  in the Euclidean plane.

- (c) Show that if  $(X, Y)'$  is bivariate normal with correlation coefficient  $\rho \in (-1, 1)$ , then the limit variance is  $\gamma^2 = (1 - \rho^2)^2$ . Find a variance-stabilizing transformation  $g : (-1, 1) \rightarrow \mathbb{R}$  such that

$$\sqrt{n}(g(r) - g(\rho)) \longrightarrow_d \mathcal{N}(0, 1)$$

is asymptotically standard normal, independently of the value of the correlation coefficient  $\rho \in (-1, 1)$ .

- (d) Design and implement an interesting simulation study in the situation of part (b) or part (c).

*Hint:* The use of computer algebra systems might prove useful. In part (a), use computer algebra systems to facilitate matrix and vector calculations. In parts (b) and (c), use computer algebra systems to find higher moments by integration, or by evaluating derivatives of characteristic functions. I will e-mail a Maple worksheet to get you started.

**Problem 32 (difference-sign test, 2 points).** Let  $X_0, X_1, X_2, \dots$  be a sequence of independent, identically distributed random variables from a continuous distribution. For  $t = 1, 2, \dots$  let the random variable  $Z_t$  be one if  $X_t > X_{t-1}$  and zero otherwise. Then  $S_n = Z_1 + \dots + Z_n$  represents the number of times that the sequence  $X_0, X_1, \dots, X_n$  increases. Find an asymptotic distribution for  $S_n$  and discuss its uses.

**Problem 33 (asymptotic distribution of a sample autocorrelation coefficient, 4 points).** A *time series* is a sequence  $X_1, X_2, \dots$  of observations, where the subscript refers to time  $t$ . One often aims to check whether a time series is compatible with the null hypothesis of an independent identically distributed sequence of random variables with finite variance  $\sigma^2$  and finite third and fourth moment  $\mu_3$  and  $\mu_4$ , respectively.

A commonly used test is based on the *sample autocorrelation function*. Specifically, the *sample autocorrelation coefficient* at lag 1 based on the first  $n + 1$  observations is defined as

$$\hat{\rho}_n(1) = \frac{\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)(X_{t+1} - \bar{X}_n)}{\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2}.$$

Show that under the null hypothesis

$$\sqrt{n} \hat{\rho}_n(1) \longrightarrow_d \mathcal{N}(0, 1) \quad \text{as} \quad n \longrightarrow \infty.$$

*Hint.* The sample autocorrelation coefficient  $\hat{\rho}_n(1)$  and the estimator

$$r_n = \frac{\frac{1}{n} \sum_{t=1}^n X_t X_{t+1} - \bar{X}_n^2}{\frac{1}{n} \sum_{t=1}^n X_t^2 - \bar{X}_n^2}$$

are asymptotically equivalent under the null hypothesis. Hence, it suffices to show that

$$\sqrt{n} r_n \longrightarrow_d \mathcal{N}(0, 1) \quad \text{as} \quad n \longrightarrow \infty.$$

Without loss of generality, we may assume that  $X_t$  is centered. Consider the random vector  $Z_t = (Z_{t1}, Z_{t2}, Z_{t3})' = (X_t, X_t^2, X_t X_{t+1})'$  for  $t = 1, 2, \dots$ . For real constants  $a, b$  and  $c$ , let

$$Y_t = aZ_{t1} + bZ_{t2} + cZ_{t3} \quad \text{where} \quad t = 1, 2, \dots$$

Apply the Central Limit Theorem for  $m$ -dependent sequences to show that

$$\sqrt{n} (\bar{Y}_n - b\sigma^2) \longrightarrow_d \mathcal{N}(0, a^2\sigma^2 + b^2(\mu_4 - \sigma^4) + c^2\sigma^4 + 2ab\mu_3).$$

Conclude that

$$\sqrt{n} \left( \bar{Z}_n - \begin{pmatrix} 0 \\ \sigma^2 \\ 0 \end{pmatrix} \right) \longrightarrow_d \mathcal{N}_3 \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \mu_3 & 0 \\ \mu_3 & \mu_4 - \sigma^4 & 0 \\ 0 & 0 & \sigma^4 \end{pmatrix} \right),$$

and apply the Delta Method to complete the proof.

**Problem 34 (asymptotic equivalence of normalized sequences, 2 points).**

- (a) Let  $U_1, U_2, \dots$  and  $V_1, V_2, \dots$  be sequences of normalized random variables. Show that if  $\text{corr}(U_n, V_n) \rightarrow 1$  then  $U_n - V_n \rightarrow_P 0$ .
- (b) Let  $X_1, X_2, \dots$  and  $Y_1, Y_2, \dots$  be sequences of random variables with finite variances. Use the result in part (a) to show that if

$$\frac{X_n - EX_n}{\sqrt{\text{var}(X_n)}} \rightarrow_d X \quad \text{and} \quad \text{corr}(X_n, Y_n) \rightarrow 1,$$

then

$$\frac{Y_n - EY_n}{\sqrt{\text{var}(Y_n)}} \rightarrow_d X.$$

Tilmann Gneiting, December 7, 2007. Solutions are due Friday, January 11, 2008 at the beginning of the class session.