BST 676 — Spring 2010 — Dr. Charnigo

Written Assignment 1

Written Assignment 1 is due on Wednesday 03 February at the end of class. You are encouraged to work in groups of two or three, but you may work individually if you prefer. In what follows, a generic sequence of random variables is denoted $Y_1, Y_2, \ldots$, and a generic random variable is denoted Y. Moreover, unless otherwise indicated, $X_1, X_2, \ldots$ are independent and identically distributed with mean $\mu \in (-\infty, \infty)$ and standard deviation $\sigma \in (0, \infty)$.

[25] 1. Suppose that

$$n^{\alpha}(Y_n - c) \xrightarrow{L} Y,$$

where $\alpha \in (0, \infty)$ and $c \in (-\infty, \infty)$ are constants. Prove that, for any $\beta \in (0, \infty)$, we have

$$n^{(\alpha-\beta)}(Y_n - c) \xrightarrow{P} 0.$$

Remark. You may also wonder about the large sample behavior of $W_n := n^{(\alpha+\beta)}(Y_n - c)$. (Here I have changed $\alpha - \beta$ to $\alpha + \beta$.) Interestingly, if Y is not equal to zero with probability one, then $n^{(\alpha+\beta)}(Y_n - c)$ does not converge at all. To see why, suppose without loss of generality that there exists a positive constant ϵ such that the cumulative distribution function of Y is continuous at ϵ and $\delta := P(Y > \epsilon) > 0$. Then we have $P(W_n > n^{\beta}\epsilon) = P(n^{\alpha}(Y_n - c) > \epsilon) \to P(Y > \epsilon) = \delta > 0$. Let y be an arbitrary real number. For large enough n, we have $y \leq n^{\beta}\epsilon$ so that $P(W_n \leq y) \leq P(W_n \leq n^{\beta}\epsilon)$. Thus, if $P(W_n \leq y)$ converges at all, the convergence is to a quantity no larger than $1 - \delta$. So, if we had $W_n \xrightarrow{L} W$ for some random variable W, we would also have $P(W \leq y) \leq 1 - \delta$ for every real number y. However, the last inequality is impossible because the cumulative distribution function for W should satisfy $\lim_{y\to\infty} P(W \leq y) = 1$.

[25] 2. Catalog the values of $\gamma \in (-\infty, \infty)$ for which $n^{\gamma}(\bar{X}_n - \mu)$:

(i) converges in probability to zero;

(ii) converges in law to a non-degenerate random variable;

(iii) does not converge in law.

Justify your answers by referencing one or more appropriate theorems from Unit I along with exercise 1 and the subsequent remark above.

[25] 3. Derive a large sample 95% confidence interval for the relative risk in a cohort study. (You may assume equal numbers of exposed subjects and non-exposed subjects.)

[25] 4. A multivariate generalization of the delta method (beyond the scope of BST 676) can be used to show that

$$\sqrt{n}(R_n - \rho) \xrightarrow{L} N(0, (1-\rho^2)^2),$$

where R_n is the sample (Pearson) correlation coefficient for bivariate normally distributed data and $\rho \in (-1, 1)$ is the corresponding population correlation coefficient. Yet, people usually don't construct 95% confidence intervals for ρ of the form

$$R_n \pm 1.96(1 - R_n^2)/\sqrt{n}.$$

Why do you suppose that people don't construct 95% confidence intervals of that form?

Instead, people often use Fisher's Z transformation to construct a 95% confidence interval for ρ. To understand where that comes from, let us employ the delta method with a yet-to-be-determined real-valued function h that is defined and whose first derivative is continuous on $(-1, 1)$. We have

$$\sqrt{n}(h(R_n) - h(\rho)) \xrightarrow{L} N(0, [h'(\rho)]^2(1-\rho^2)^2),$$

so choosing h with $h'(\rho) = 1/(1-\rho^2)$ results in convergence to $N(0, 1)$. (Statisticians refer to this process as "variance stabilization".) Use a partial fraction decomposition or an integral table to find h with $h'(\rho) = 1/(1-\rho^2)$. Then propose a 95% confidence interval for $h(\rho)$ and, finally, for ρ.