BST 676 — Spring 2010 — Dr. Charnigo

Written Assignment 1 Solutions

1. Clearly $n^{-\beta} \xrightarrow{a.s.} 0$, so $n^{-\beta} \xrightarrow{L} 0$ also. Applying Slutsky's Theorem #3, we obtain

$$n^{\alpha-\beta}(Y_n - c) = n^{\alpha}(Y_n - c) \times n^{-\beta} \xrightarrow{L} Y \times 0 = 0.$$

Finally, convergence in law to a constant is equivalent to convergence in probability to that same constant.

2. By the Central Limit Theorem,

$$n^{1/2}(\bar{X}_n - \mu) \xrightarrow{L} N(0, \sigma^2).$$

Exercise 1 indicates that for any $\beta > 0$ we have

$$n^{1/2-\beta}(\bar{X}_n - \mu) \xrightarrow{P} 0.$$

Since $\{1/2 - \beta : \beta \in (0, \infty)\} = (-\infty, 1/2)$, we conclude that

$$n^{\gamma}(\bar{X}_n - \mu) \xrightarrow{P} 0$$

whenever $\gamma \in (-\infty, 1/2)$. Since $N(0, \sigma^2)$ is not degenerate, the remark after exercise 1 indicates that for any $\beta > 0$ we have

$$n^{1/2+\beta}(\bar{X}_n - \mu) \text{ does not converge in law.}$$

Since $\{1/2 + \beta : \beta \in (0, \infty)\} = (1/2, \infty)$, we conclude that

$$n^{\gamma}(\bar{X}_n - \mu) \text{ does not converge in law}$$

whenever $\gamma \in (1/2, \infty)$. In summary:
- convergence in probability to zero occurs when $\gamma < 1/2$;
- convergence in law to a non-degenerate random variable occurs when $\gamma = 1/2$; and,
- convergence in law does not occur when $\gamma > 1/2$.

3. Suppose that we are performing a cohort study with n_1 exposed subjects and n_2 non-exposed subjects, with risk $p_1 \in (0, 1)$ for the exposed subjects and risk $p_2 \in (0, 1)$ for the non-exposed subjects. Let RR be a shorthand for the relative risk p_1/p_2.

Since a sample proportion is a sample mean of Bernoulli random variables, the Central Limit Theorem is applicable and yields

$$\sqrt{n_1}(a/n_1 - p_1) \xrightarrow{L} N(0, p_1(1 - p_1))$$

as $n_1 \to \infty$ and

$$\sqrt{n_2}(b/n_2 - p_2) \xrightarrow{L} N(0, p_2(1 - p_2))$$

as $n_2 \to \infty$. Moreover, a/n_1 may be regarded as independent of b/n_2 since there is no overlap between the exposed subjects and the non-exposed subjects.

Put

$$h(y) := \log y$$

for $y \in (0,1)$. Then, assuming that $\min\{a,b,c,d\} \geq 1$, we have $h(a/n_1) = \log[a/n_1]$ and $h(b/n_2) = \log[b/n_2]$. Moreover,

$$h'(y) = 1/y.$$

Thus, the delta method yields

$$\sqrt{n_1}(\log[a/n_1] - \log[p_1]) \xrightarrow{L} N(0, (1-p_1)/p_1) \tag{1}$$

and

$$\sqrt{n_2}(\log[b/n_2] - \log[p_2]) \xrightarrow{L} N(0, (1-p_2)/p_2). \tag{2}$$

Suppose that n_1 is constrained to equal n_2. Let n without subscript denote their common value. Then, applying Slutsky's Theorem #4 to (1) and (2), we obtain

$$\sqrt{n}\left(\log\left[\frac{a/n}{b/n}\right] - \log[RR]\right) \xrightarrow{L} N(0, (1-p_1)/p_1 + (1-p_2)/p_2)$$

or, equivalently,

$$\frac{\sqrt{n}}{\sqrt{(1-p_1)/p_1 + (1-p_2)/p_2}}\left(\log\left[\frac{a/n}{b/n}\right] - \log[RR]\right) \xrightarrow{L} N(0,1). \tag{3}$$

By the Weak Law of Large Numbers, we have

$$a/n \xrightarrow{P} p_1 \quad \text{and} \quad b/n \xrightarrow{P} p_2. \tag{4}$$

Applying the Continuous Mapping Theorem and Slutsky's Theorem #2 to (4), we obtain

$$\frac{1-a/n}{a/n} + \frac{1-b/n}{b/n} = c/a + d/b \xrightarrow{P} (1-p_1)/p_1 + (1-p_2)/p_2. \tag{5}$$

Then, applying the Continuous Mapping Theorem and Slutsky's Theorem #2 to (5), we obtain

$$\frac{\sqrt{(1-p_1)/p_1 + (1-p_2)/p_2}}{\sqrt{c/a + d/b}} \xrightarrow{P} 1. \tag{6}$$

Finally, applying Slutsky's Theorem #3 to (3) and (6) yields

$$\frac{1}{\sqrt{(c/a + d/b)/n}}\left(\log\left[\frac{a/n}{b/n}\right] - \log[RR]\right) \xrightarrow{L} N(0,1),$$

from which

$$\log\left[\frac{a/n}{b/n}\right] \pm 1.96\sqrt{(c/a + d/b)/n}$$

is justified as an approximate 95% confidence interval for $\log[RR]$ and, hence,

$$\frac{a/n}{b/n}\exp\left[\pm 1.96\sqrt{(c/a + d/b)/n}\right]$$

is justified as an approximate 95% confidence interval for RR.

4. A confidence interval of the form

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

is not constrained to lie inside $(-1, 1)$ when $n < 16$.

A solution to

$$h'(\rho) = 1/(1 - \rho^2) = (1/2)/(1 - \rho) + (1/2)/(1 + \rho)$$

is

$$h(\rho) = -(1/2)\log[1 - \rho] + (1/2)\log[1 + \rho] = (1/2)\log[(1 + \rho)/(1 - \rho)],$$

which mathematicians call the inverse hyperbolic tangent function.

An approximate 95% confidence interval for $h(\rho)$ is

$$h(R_n) \pm 1.96/\sqrt{n},$$

and so an approximate 95% confidence interval for ρ is

$$\frac{(1 + R_n)\exp[\pm 2 \times 1.96/\sqrt{n}] - (1 - R_n)}{(1 + R_n)\exp[\pm 2 \times 1.96/\sqrt{n}] + (1 - R_n)}.$$

Remark 1. What is special about $n = 16$? When $n = 16$, a confidence interval of the form

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

is approximately

$$R_n \pm (1 - R_n^2)/2.$$

The quadratic equation

$$R_n^2/2 + R_n - 1/2 = 1$$

has a double root at $R_n = 1$, while the quadratic equation

$$R_n^2/2 - R_n - 1/2 = -1$$

has a double root at $R_n = -1$. So $R_n \in (-1, 1)$ implies that the confidence interval also lies inside $(-1, 1)$ when $n = 16$. Moreover, for a fixed $R_n \in (-1, 1)$, the confidence interval becomes narrower as n increases, so $R_n \in (-1, 1)$ implies that the confidence interval also lies inside $(-1, 1)$ when $n > 16$.

The above demonstration that $n = 16$ is special relies on the fact that the confidence level is 95%. For a generic $\alpha \in (0, 1)$, can you figure out at which n a confidence interval of the form

$$R_n \pm z_{1-\alpha/2}(1 - R_n^2)/\sqrt{n}$$

is guaranteed to lie inside $(-1, 1)$?

Remark 2. Some textbooks propose a 95% confidence interval of the form

$$\frac{(1 + R_n)\exp[\pm 2 \times 1.96/\sqrt{n-3}] - (1 - R_n)}{(1 + R_n)\exp[\pm 2 \times 1.96/\sqrt{n-3}] + (1 - R_n)}.$$

Can you convince yourself that this is justified by applying Slutsky's Theorem #3 to $\sqrt{n}(h(R_n) - h(\rho))$ and $\sqrt{(n-3)/n}$?