Bios 661: 1 – 5; Bios 673: 2 – 6.

1. C&B 5.30

2. C&B 5.32

3. Two *sufficient* conditions for consistency, $T_n \to_p \theta$, are

   i. $\lim_{n\to\infty} E(T_n) = \theta$;

   ii. $\lim_{n\to\infty} Var(T_n) = 0$.

   Assume the distribution of income (in thousands of dollars) in a large U.S. city follows a Pareto density function

   $$f_Y(y) = \theta\gamma^\theta y^{-(\theta+1)}, \quad 0 < \gamma < y < \infty, 2 < \theta < \infty,$$

   where γ and θ are known parameters. Let $Y_1, \cdots, Y_n$ be a random sample from $f_Y(y)$. Show that the sample minimum $Y_{(1)}$ is a consistent estimator of γ, i.e., $Y_{(1)} \to_p \gamma$.

   Solution: First, we have

   $$\begin{aligned} F_Y(y) &= \int_\gamma^y \theta\gamma^\theta t^{-(\theta+1)} dt = -\gamma^\theta \; t^{-\theta}\Big|_\gamma^y \\ &= \gamma^\theta \left(\gamma^{-\theta} - y^{-\theta}\right) = 1 - \left(\frac{\gamma}{y}\right)^\theta, \quad 0 < \gamma < y < \infty. \end{aligned}$$

   Let $Z = Y_{(1)}$. One has $f_Z(z) = n\theta\gamma^{n\theta} z^{-n\theta+1}$, $0 < \gamma < z < \infty$. Using this density, one has

   $$E(Z^m) = \int_\gamma^\infty z^m n\theta\gamma^{n\theta} z^{-(n\theta+1)} dz = \frac{n\theta\gamma^m}{(n\theta - m)}, \quad n\theta > m.$$

   Therefore,

   $$E(Z) = \frac{n\theta\gamma}{n\theta - 1}, \quad \text{and} \quad Var(Z) = \frac{n\theta\gamma^2}{(n\theta-1)^2(n\theta-2)}.$$

   For (i) and (ii),

   $$\begin{aligned} \lim_{n\to\infty} E(Z) &= \lim_{n\to\infty} \frac{n\theta\gamma}{n\theta-1} = \lim_{n\to\infty} \frac{\theta\gamma}{\theta - 1/n} = \gamma, \\ \lim_{n\to\infty} Var(Z) &= \lim_{n\to\infty} \frac{n\theta\gamma^2}{(n\theta-1)^2(n\theta-2)} = \lim_{n\to\infty} \frac{\theta\gamma^2}{n^2(\theta-1/n)^2(\theta-2/n)} = 0. \end{aligned}$$

4. (midterm 1 in 2014) Suppose that $X_1, X_2, \ldots, X_n$ are iid random variables distributed as Poisson with mean $\mu > 0$. Denote $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. We are interested in constructing a confidence interval for μ.

(a) State the central limit theorem for $\bar{X}_n$.

Solution:

$$\sqrt{n}(\bar{X}_n - \mu)/\sqrt{\mu} \to_d N(0,1).$$

(b) What is the asymptotic variance of $T_n = \sqrt{n}(\bar{X}_n - \mu)$?

Solution:

$$\lim_{n\to\infty} Var(T_n/\sqrt{\mu}) = 1 \Rightarrow \lim_{n\to\infty} Var(T_n) = \mu.$$

(c) What is the appropriate function $h(\bar{X}_n)$ so that $h(\bar{X}_n)T_n \to_d N(0,1)$? What theorem(s) are needed to justify such claim?

Solution:

$$T_n/\sqrt{\mu} \to_d N(0,1) \Rightarrow T_n/\sqrt{\bar{X}_n} \to_d N(0,1),$$

by Slutsky Theorem. That implies $h(\bar{X}_n) = 1/\sqrt{\bar{X}_n}$.

(d) Use the last part to construct an approximate 95% confidence interval for μ. Give the upper and lower limits in explicit form.

Solution:

$$\begin{aligned} 0.95 &= P(-1.96 < h(\bar{X}_n)T_n < 1.96) \\ &= P(-1.96 < h(\bar{X}_n)\sqrt{n}(\bar{X}_n - \mu) < 1.96) \\ &= P\left(\bar{X}_n - \frac{1.96}{h(\bar{X}_n)\sqrt{n}} < \mu < \bar{X}_n + \frac{1.96}{h(\bar{X}_n)\sqrt{n}}\right) \end{aligned}$$

(e) Another approach to eliminate μ from the asymptotic variance is to find a function g such that $\sqrt{n}(g(\bar{X}_n) - g(\mu)) \to_d N(0,1)$. Find an explicit expression for $g(\mu)$.

Solution: According to the delta method,

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \to_d N(0, \{g'(\mu)\}^2\mu) \equiv N(0,1),$$

which means $\{g'(\mu)\}^2\mu = 1$ and $g(\mu) = 2\mu^{1/2}$.

(f) (For discussion, no need to return) Use the last part to construct an approximate 95% confidence interval for μ. Give the upper and lower limits in explicit form.

Solution:

$$\begin{aligned} 0.95 &= P(-1.96 < \sqrt{n}(g(\bar{X}_n) - g(\mu)) < 1.96) \\ &= P(g(\bar{X}_n) - 1.96/\sqrt{n} < g(\mu) < g(\bar{X}_n) + 1.96/\sqrt{n}) \\ &= P\left(\frac{1}{4}\{g(\bar{X}_n) - 1.96/\sqrt{n}\}^2 < \mu < \frac{1}{4}\{g(\bar{X}_n) + 1.96/\sqrt{n}\}^2\right) \end{aligned}$$

5. (midterm 1 in 2015) Let $X_1, \ldots, X_n$ be a random sample from a normal distribution $N(\mu, 1)$.

   (a) Find the limiting distribution of $U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$ by the central limit theorem.

   Solution The central limit theorem gives $\sqrt{n}(\sum_{i=1}^n X_i/n - \mu) \to_d N(0,1)$. That makes $U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \to_d N(0,1)$.

   (b) Show that $V_n = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \to 1$ in probability by the weak law of large numbers.

   Solution By the weak law of large numbers,

$$V_n = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \to_p E(X_1 - \mu)^2 = 1.$$

   We may also let $Y_i = (X_i - \mu)^2$ and Y_i follows χ_1^2. The weak law of large number also gives $V_n = \bar{Y} \to_p E(Y_1) = 1$.

   (c) Find the limiting distribution of $W_n = U_n / V_n$.

   Solution By Slutsky Theorem, $W_n \to_d N(0,1)$.

   (d) Find the limiting distribution of $\sqrt{n}(\bar{X}^2 - \mu^2)$, where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$.

   Solution We know that $\sqrt{n}(\bar{X} - \mu) \to_d N(0,1)$. Let $g(\mu) = \mu^2$. By the delta method, $\sqrt{n}(g(\bar{X}) - g(\mu)) \to_d N(0, \{g'(\mu)\}^2) \equiv N(0, 4\mu^2)$.

(e) Construct a 95% confidence interval for μ^2, either under a finite n (exact) or $n \to \infty$ (limiting).

Solution According to the limiting distribution in (d),

$$0.95 \approx P\left(-1.96 < \frac{\sqrt{n}(\bar{X}^2 - \mu^2)}{2\mu} < 1.96\right).$$

We can search μ in a range such that the inequality within the bracket can be satisfied. However, it is quite complicate and uneasy. What can be much easier to obtain the confidence interval is that we replace μ in the denominator by $\bar{X}$. By Slutsky Theorem,

$$\frac{\sqrt{n}(\bar{X}^2 - \mu^2)}{2\bar{X}} \to_d N(0,1).$$

Hence,

$$0.95 \approx P\left(-1.96 < \frac{\sqrt{n}(\bar{X}^2 - \mu^2)}{2\bar{X}} < 1.96\right),$$

and

$$0.95 \approx P\left(\bar{X}^2 - \frac{1}{\sqrt{n}}1.96(2\bar{X}) < \mu^2 < \bar{X}^2 + \frac{1}{\sqrt{n}}1.96(2\bar{X})\right).$$

Since μ^2 is always positive, one may consider use

$$\left(\max\left\{\bar{X}^2 - \frac{1}{\sqrt{n}}1.96(2\bar{X}), 0\right\}, \bar{X}^2 + \frac{1}{\sqrt{n}}1.96(2\bar{X})\right)$$

as the 95% confidence interval for μ^2.

One may also construct an exact confidence interval using the fact that $\bar{X}$ follows Normal(μ,$1/n$). Hence, one can write

$$\begin{aligned} 0.95 &= P\left(-1.96 < \frac{\bar{X} - \mu}{\sqrt{1/n}} < 1.96\right) \\ &= P\left(\bar{X} - 1.96\sqrt{1/n} < \mu < \bar{X} + 1.96\sqrt{1/n}\right), \end{aligned}$$

which is an exact confidence interval for μ since the distribution is exactly normal. The confidence interval for μ^2 needs to consider both ends of $\bar{x} - 1.96\sqrt{1/n}$ and $\bar{x} + 1.96\sqrt{1/n}$ to make sure the coverage is correct.

6. Suppose that X_n is a random variable following a binomial distribution $B(n, \theta)$, where $\theta \in (0, 1)$. Let

$$Y_n = \begin{cases} \log(X_n/n), & X_n \geq 1, \\ 1, & X_n = 0. \end{cases}$$

Show that $\lim_{n\to\infty} Y_n = \log\theta$ a.s (or in probability) and $\sqrt{n}(Y_n - \log\theta) \to_d N(0, (1-\theta)/\theta)$.

Solution: We can assume $X_n = \sum_{i=1}^n Z_i$, where Z_i, $i = 1, \ldots, n$, is a binary random variable that follows a Bernoulli distribution with mean θ. One can use the Strong Law of Large Numbers (SSLN) to claim $X_n/n = n^{-1}\sum_{i=1}^n Z_i$ converges almost surely to θ. One also can use a sufficient condition

$$\sum_{n=1}^{\infty} P\left(\left|\frac{X_n}{n} - \theta\right| \geq \epsilon\right) < \infty$$

to check for the convergence almost surely. In fact

$$P\left(\left|\frac{X_n}{n} - \theta\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^4} E\left(\frac{X_n}{n} - \theta\right)^4 = \frac{c_1}{\epsilon^4 n^3} + \frac{c_2(n-1)}{\epsilon^4 n^3},$$

where $c_1 = \theta^4(1-\theta) + (1-\theta)^4\theta$ and $c^2 = \theta^2(1-\theta)^2$. Hence,

$$\sum_{n=1}^{\infty} P\left(\left|\frac{X_n}{n} - \theta\right|\right) = \sum_{n=1}^{\infty} \frac{c_1}{\epsilon^4 n^3} + \frac{c_2(n-1)}{\epsilon^4 n^3} < \infty.$$

For the convergence of Y_n, one can write $Y_n = \log(X_n/n)I(X_n \geq 1) + I(X_n = 0)$. If one can show $I(X_n \geq 1)$ converges to 1 almost surely, we can conclude Y_n converges to $\log\theta$ almost surely. To show that $I(X_n \geq 1)$ converges to 1 almost surely, we can use the sufficient condition mentioned above to show

$$\sum_{n=1}^{\infty} P\left(I(X_n = 0) \geq \epsilon\right) = \sum_{n=1}^{\infty} P(X_n = 0) = \sum_{n=1}^{\infty} (1-\theta)^n = \frac{1}{\theta} - 1 < \infty.$$

To show convergence in probability, one can use the definition to show

$$\lim_{n\to\infty} P\left(I(X_n = 0) > \epsilon\right) = \lim_{n\to\infty} P\left(X_n = 0\right) = \lim_{n\to\infty} (1-\theta)^n = 0.$$

One hence can claim $I(X_n = 0) \to_p 0$ and $I(X_n \geq 1) = 1 - I(X_n = 0) \to_p 1$. In fact, the sequence $\sqrt{n}I(X_n = 0)$ also converges to 0 in probability since

$$\lim_{n\to\infty} P\left(\sqrt{n}I(X_n = 0) > \epsilon\right) = \lim_{n\to\infty} P\left(X_n = 0\right) = 0.$$

The asymptotic normality of $\log(X_n/n)$ can be easily shown using the delta method. That is, we can have

$$\sqrt{n}\{\log(X_n/n) - \log(\theta)\} \to_d N(0, \{g'(\theta)\}^2\theta(1-\theta)) \equiv N(0, (1-\theta)/\theta),$$

where $g(\theta) = \log(\theta)$ and $g'(\theta) = \theta^{-1}$. The asymptotic normality of Y_n can then be proved using Slutsky Theorem, and one can claim Y_n has the same limiting distribution as $\log(X_n/n)$. Specifically,

$$\begin{aligned}\sqrt{n}(Y_n - \log(\theta)) &= \sqrt{n}\{\log(X_n/n)I(X_n \geq 1) + I(X_n = 0) - \log(\theta)\} \\ &= \sqrt{n}\{\log(X_n/n) - \log(\theta)\}I(X_n \geq 1) + \sqrt{n}I(X_n \geq 1)\log(\theta) \\ &\quad + \sqrt{n}I(X_n = 0) - \sqrt{n}\log(\theta) \\ &\to_d N(0, (1-\theta)/\theta),\end{aligned}$$

by the fact that $\sqrt{n}I(X_n = 0) \to_p 0$ and Slutsky Theorem.

7. (Bios 673 class material) Let $X_1, \ldots, X_n$ be independent random variables. Suppose that $\sum_{i=1}^n (X_i - EX_i)/\sigma_n \to_d N(0,1)$, where $\sigma_n^2 = \text{Var}(\sum_{i=1}^n X_i)$. Show that $n^{-1}\sum_{i=1}^n (X_i - EX_i) \to_p 0$ if and only if $\lim_{n\to\infty} \sigma_n/n = 0$.

Solution (Sufficient) If $\lim_{n\to\infty} \sigma_n/n = 0$, we can claim

$$\frac{1}{n}\sum_{i=1}^n (X_i - EX_i) = \underbrace{\frac{\sigma_n}{n}}_{\to 0} \underbrace{\frac{1}{\sigma_n}\sum_{i=1}^n (X_i - EX_i)}_{\to_d N(0,1)} \to_d 0,$$

by Slutsky's theorem. Since the variable converges in distribution to a degenerate point mass, we can also claim it converges in probability to 0.

(Necessary) If $n^{-1}\sum_{i=1}^n (X_i - EX_i) \to_p 0$ but $\lim_{n\to\infty} \sigma_n/n = c \in (0, \infty]$. We have

$$\frac{1}{\sigma_n}\sum_{i=1}^n (X_i - EX_i) = \underbrace{\frac{n}{\sigma_n}}_{\to c} \underbrace{\frac{1}{n}\sum_{i=1}^n (X_i - EX_i)}_{\to_p 0} \to_p 0,$$

which is apparently contradicted to the given condition that $\sum_{i=1}^n (X_i - EX_i)/\sigma_n \to_d N(0,1)$. We hence can claim $\lim_{n\to\infty} \sigma_n/n = 0$.