Problem Set 5 Solutions

Problem 1 in lecture 5

(a) $\forall \epsilon > 0$, $\exists N = \frac{1}{\epsilon} + 1$, such that for all $n > N$, $|a_n - 0| = \frac{1}{n} < \frac{1}{N} < \frac{1}{1/\epsilon} = \epsilon$. Thus, $a_n \to 0$ as $n \to \infty$.

(b) $\forall \epsilon > 0$, $\exists N = \frac{1}{\epsilon} + 1$, such that for all $n > N$, $|a_n - 0| = \frac{1}{n}|\sin(\frac{\pi}{2}n)| \leq \frac{1}{n} < \frac{1}{N} < \frac{1}{1/\epsilon} = \epsilon$. Thus, $a_n \to 0$ as $n \to \infty$.

Problem 2 in lecture 5

(a) For $\epsilon > 0$ sufficiently small, we have $Pr(|X_n - 0| \leq \epsilon) = Pr(X_n = 0) = 1 - \frac{2}{n} \to 1$ as $n \to \infty$. Thus, $X_n \xrightarrow{p} 0$.

(b) $E(X_n) = -n \cdot \frac{1}{n} + 0 \cdot (1 - \frac{2}{n}) + n \cdot \frac{1}{n} = 0$.

(c) $Var(X_n) = E(X_n^2) - (E(X_n))^2 = (-n)^2 \cdot \frac{1}{n} + 0^2 \cdot (1 - \frac{2}{n}) + n^2 \cdot \frac{1}{n} - 0^2 = 2n$.

(d) $E(X_n) = 0 \cdot (1 - \frac{1}{n}) + n \cdot \frac{1}{n} = 1$.

(e) For the new distribution of X_n, we still have $X_n \xrightarrow{p} 0$ because for $\epsilon > 0$ sufficiently small, $Pr(|X_n - 0| \leq \epsilon) = Pr(X_n = 0) = 1 - \frac{1}{n} \to 1$ as $n \to \infty$. But $E(X_n) = 1 \nrightarrow 0$ as we show in the previous part. So we can conclude that $X_n \xrightarrow{p} 0$ is not sufficient for $E(X_n) \to 0$.

Problem 3 in lecture 5

(a) $E(\bar{Y}^*) = E(\frac{1}{n}\sum_{i=1}^n w_i Y_i) = \frac{1}{n}\sum_{i=1}^n w_i E(Y_i) = \frac{1}{n}\sum_{i=1}^n w_i \cdot \mu = \mu \cdot 1 = \mu$. Thus, $\bar{Y}^*$ is unbiased for μ.

(b) Since Y_i is i.i.d., $Var(\bar{Y}^*) = \frac{1}{n^2}\sum_{i=1}^n w_i^2 Var(Y_i) = \sigma^2(\frac{1}{n^2}\sum_{i=1}^n w_i^2)$, where $\sigma^2 = Var(Y_i)$ is the variance of Y_i.

(c) If $\frac{1}{n^2}\sum_{i=1}^n w_i^2 \to 0$, then $\forall \epsilon > 0$, $Pr(|\bar{Y}^* - \mu| > \epsilon) \leq \frac{Var(\bar{Y}^*)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}(\frac{1}{n^2}\sum_{i=1}^n w_i^2) \to 0$ by Chebyshev's Inequality. This tells us $Pr(|\bar{Y}^* - \mu| > \epsilon) \to 0$ as $n \to \infty$, or equivalently, $\bar{Y}^* \xrightarrow{p} \mu$. Thus, a sufficient condition for $\bar{Y}^* \xrightarrow{p} \mu$ is $\frac{1}{n^2}\sum_{i=1}^n w_i^2 \to 0$.

(d) Since $w_i \geq 0$,

$$\frac{1}{n^2}\sum_{i=1}^n w_i^2 \leq \frac{1}{n^2}\sum_{i=1}^n w_i \cdot (\max_{i \leq n} w_i) = \left(\frac{1}{n}\max_{i \leq n} w_i\right)\frac{1}{n}\sum_{i=1}^n w_i = \frac{1}{n}\max_{i \leq n} w_i.$$

Thus, $\frac{1}{n}\max_{i \leq n} w_i \to 0$ implies $\frac{1}{n^2}\sum_{i=1}^n w_i^2 \to 0$.

Problem 4 in lecture 5

(a) $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} E(X_i^2)$.

(b) $\frac{1}{n}\sum_{i=1}^n X_i^3 \xrightarrow{p} E(X_i^3)$.

(c) We cannot use WLLN and CMT to prove convergence in probability since there is no sample average in this expression. Actually, this statistic may not converge in probability.

The condition for it to converge is: the support of the population X is bounded from above. By saying bounded from above, we mean that there exists a $x_u \in \mathbb{R}$ such that $F_X(x_u) = 1$. x_u is called the upper bound of X. The least upper bound is defined as $\bar{x} = \arg\min_{x_u} F_X(x_u) = 1$. $\bar{x}$ must exist because $F_X(\cdot)$ is right-continuous. $F_X(x) = 1$ for $x \geq \bar{x}$ and $0 \leq F_X(x) < 1$ for $x < \bar{x}$.

The CDF of $\max_{i\leq n} X_i$ is $F(x) = Pr(\max_{i\leq n} X_i \leq x) = Pr(X_1 \leq x, \dots, X_n \leq x) = [Pr(X_i \leq x)]^n = [F_{X_i}(x)]^n$.

If the support of X is bounded from above, then we have a least upper bound $\bar{x}$. For $x \geq \bar{x}$, $F(x) = 1^n = 1$; for $x < \bar{x}$, $F(x) \to 0$ as $n \to \infty$. Thus, as $n \to \infty$, the limiting distribution is $\mathbb{1}(x \geq \bar{x})$, which is exactly the CDF of a fixed number $\bar{x}$. This implies $\max_{i\leq n} X_i \xrightarrow{d} \bar{x}$ and thus $\max_{i\leq n} X_i \xrightarrow{p} \bar{x}$. In words, the largest element in sample converges in probability to the least upper bound of the support of X.

If the support of X is not bounded from above, then $0 \leq F_X(x) < 1$ for all $x \in \mathbb{R}$. As $n \to \infty$, $F(x) \to 0$ for all $x \in \mathbb{R}$, which is not a CDF. In this case, $\max_{i\leq n} X_i$ does not converge in distribution and does not converge in probability, either.

(d) $Y_1 = \frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} E(X_i^2)$, $Y_2 = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{p} E(X_i)$. Since the function $g(Y_1, Y_2) = Y_1 - Y_2^2$ is continuous, $Y_1 - Y_2^2 \xrightarrow{p} E(X_i^2) - (E(X_i))^2 = Var(X_i)$.

(e) The function $g(Y_1, Y_2) = \frac{Y_1}{Y_2}$ is continuous for $Y_2 > 0$. Thus, $\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i} = \frac{Y_1}{Y_2} \xrightarrow{p} \frac{E(X_i^2)}{E(X_i)}$ assuming $E(X_i) > 0$.

(f) $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{p} E(X_i)$. Since the indicator function $\mathbb{1}(x > 0)$ is continuous on $(0, +\infty)$ and on $(-\infty, 0)$, we have: $\mathbb{1}(\frac{1}{n}\sum_{i=1}^n X_i > 0) \xrightarrow{p} 1$ if $E(X_i) > 0$; $\mathbb{1}(\frac{1}{n}\sum_{i=1}^n X_i > 0) \xrightarrow{p} 0$ if $E(X_i) < 0$; $\mathbb{1}(\frac{1}{n}\sum_{i=1}^n X_i > 0)$ does not converge in probability if $E(X_i) = 0$ (0 is a discontinuity point).

Problem 5 in lecture 5

Since X_i is i.i.d. and $E(\log(X))$ exists, by WLLN, we have $\log(\hat{\mu}) = \frac{1}{n}\sum_{i=1}^n \log(X_i) \xrightarrow{p} E(\log(X))$. Notice that the function $g(x) = e^x$ is continuous. By CMT, we have $\hat{\mu} = g(\log(\hat{\mu})) \xrightarrow{p} g(E(\log(X))) = \exp(E(\log(X))) = \mu$ as $n \to \infty$.

Problem 6 in lecture 5

(a) $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$

(b) By CLT

$$\begin{aligned}\sqrt{n}(\hat{\mu}_k - \mu_k) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i^k - \mu_k) \\ &\xrightarrow{d} N(0, V_{X^k})\end{aligned}$$

where $V_{X^k} = Var(X^k) = E(X^{2k}) - [E(X^k)]^2 = E(X^{2k}) - \mu_k^2$

Problem 7 in lecture 5

(a) $\hat{m}_k = \left\{\frac{1}{n}\sum_{i=1}^n X_i^k\right\}^{1/k}$

(b) Note that $\hat{m}_k = \left\{\frac{1}{n}\sum_{i=1}^n X_i^k\right\}^{1/k} = \hat{\mu}_k^{1/k} = g(\hat{\mu}_k)$ where $g(\cdot) = (\cdot)^{1/k}$, and similarly $m_k = \mu_k^{1/k} = g(\mu_k)$.

Since we show in the previous question that

$$\sqrt{n}(\hat{\mu}_k - \mu_k) \xrightarrow{d} N(0, V_{X^k}),\ V_{X^k} = E(X^{2k}) - \mu_k^2$$

Apply delta method

$$\begin{aligned}\sqrt{n}(\hat{m}_k - m_k) &= \sqrt{n}\,[g(\hat{\mu}_k) - g(\mu_k)] \\ &\xrightarrow{d} N(0, V)\end{aligned}$$

where

$$V = (g'(\mu_k))^2 V_{X^k} = \left(\frac{1}{k}(\mu_k)^{\frac{1}{k}-1}\right)^2 V_{X^k} = \frac{1}{k^2}(\mu_k)^{\frac{2}{k}-2}(E(X^{2k}) - \mu_k^2)$$

Problem 8 in lecture 5

(a) Delta method implies

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &= \sqrt{n}(\hat{\mu}^2 - \mu^2) \\ &\approx 2\mu\sqrt{n}(\hat{\mu} - \mu) \\ &\xrightarrow{d} 2\mu N(0, v^2) \sim N(0, 4\mu^2 v^2).\end{aligned}$$

(b) The asymptotic distribution degenerate to 0

$$0N(0, v^2) \sim 0.$$

(c) Under the assumption $\mu = 0$, we have

$$\begin{aligned}n\hat{\beta} &= v^2\left(\sqrt{n}\frac{\hat{\mu} - \mu}{v}\right)^2 \\ &\xrightarrow{d} v^2\{N(0,1)\}^2 \\ &\sim v^2\chi_1^2\end{aligned}$$

(d) Delta method relies on the first order Taylor approximation of the function. Let g be a infinitely many times differentiable function. If I take Taylor expansion, I can write

$$\begin{aligned}g(\hat{\theta}) - g(\theta) &= g'(\theta)(\hat{\theta} - \theta) + \frac{1}{2}g''(\theta)(\hat{\theta} - \theta)^2 \\ &+ \sum_{k=3}^{\infty}\frac{1}{k!}\frac{d^k g(\theta)}{d\theta^k}(\hat{\theta} - \theta)^k.\end{aligned}$$

Because $(\hat{\theta}-\theta) = O_p(n^{-1/2})$, $(\hat{\theta}-\theta)^k = O_p(n^{-k/2})$. That means, higher order term converge faster. As a result, when $g'(\theta) \neq 0$, the first order term dominates and the above expansion implies

$$\sqrt{n}[g(\hat{\theta}) - g(\theta)] = g'(\theta)\sqrt{n}(\hat{\theta}-\theta) + o_p(1).$$

When $g'(\theta) = 0$, the first order term degenerate to zero. For this case, the highest order term is the second. Thus, the rate of convergence and asymptotic behavior of the estimator be determined by the second order term, which is $O_p(n^{-1})$. Indeed, the above expansion now implies

$$n[g(\hat{\theta}) - g(\theta)] = \frac{1}{2}g''(\theta)n(\hat{\theta}-\theta)^2 + o_p(1).$$

When second order term is also zero, then the third order term dominates.
You do not need to know the definition of $O_p(n^{-1})$ and $o_p(n^{-1})$; Here, just think that $o_p(n^{-1})$ means that the speed of convergence in probability is as fast as the speed that n^{-1} converges to 0, and $O_p(n^{-1})$ is the speed of convergence in distribution, for example, if an estimator $\hat{\mu}$ has an asymptotic distribution $n(\hat{\mu}-\mu) \xrightarrow{d} N(0,\sigma^2)$, then $\hat{\mu} = O_p(n^{-1})$.