ECMA31000: Introduction to Empirical Analysis
Exam 1 2020 Solutions

Question 1 (10 points) Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and let $A, B \in \mathcal{F}$. Prove that if $\mathrm{P}(B) > 0$ and $\mathrm{P}(A|B) = \mathrm{P}(A|B^c)$, then A and B are independent events.

ANS: We have $\mathrm{P}(A) = \mathrm{P}(A|B)\,\mathrm{P}(B) + \mathrm{P}(A|B^c)\,\mathrm{P}(B^c)$ by the law of total probability, since $B \cup B^c$ forms a partition of Ω.

Upon substituting $\mathrm{P}(A|B) = \mathrm{P}(A|B^c)$, we obtain

$$\mathrm{P}(A) = \mathrm{P}(A|B)\left(\mathrm{P}(B) + 1 - \mathrm{P}(B)\right) = \mathrm{P}(A|B).$$

The definition of condtional probability then gives

$$\mathrm{P}(A) = \mathrm{P}(A|B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)},$$

from which it follows that $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$. Hence A and B are independent events.

Question 2 Let $\{X_n\}_{n \geq 1}$ and X be random variables defined on the same probability space.

a) (5 points) Suppose $X_n \overset{a.s.}{\to} X$. Is it true that $\mathrm{E}\left(|X_n - X|^2\right) \to 0$?

ANS: No, convergence almost surely does not imply convergence in 2−nd mean. Take

$$X_n = \begin{cases} 2^n & \text{with probability } \frac{1}{n^2}, \\ 0 & \text{with probability } 1 - \frac{1}{n^2}. \end{cases}$$

Then clearly

$$\mathrm{E}\left(X_n^2\right) = \frac{2^{2n}}{n^2} \to \infty,$$

but for any $\epsilon > 0$:

$$\begin{aligned} \mathrm{P}\left(\cup_{k \geq n} \{|X_k - 0| > \epsilon\}\right) &\leq \sum_{k=n}^{\infty} \mathrm{P}\left(|X_k - 0| > \epsilon\right) \\ &\leq \sum_{k=n}^{\infty} \mathrm{P}\left(X_k = 2^k\right) \end{aligned}$$

$$= \sum_{k=n}^{\infty} \frac{1}{k^2} \to 0$$

as $n \to \infty$. So $X_n \overset{a.s.}{\to} 0$ but $\mathrm{E}\left(|X_n - 0|^2\right) \to \infty$.

b) (10 points) Suppose $\mathrm{E}\left(|X_n - X|^2\right) = \frac{1}{n^2}$. Is it true that $X_n \overset{a.s.}{\to} X$?

ANS: Yes. For any $\epsilon > 0$,

$$\begin{aligned}
\mathrm{P}\left(\cup_{k\geq n}\{|X_k - X| > \epsilon\}\right) &\leq \sum_{k=n}^{\infty} \mathrm{P}\left(|X_k - X| > \epsilon\right) \\
&\leq \sum_{k=n}^{\infty} \frac{\mathrm{E}\left(|X_k - X|^2\right)}{\epsilon^2} \\
&= \sum_{k=n}^{\infty} \frac{1}{k^2\epsilon^2} \to 0
\end{aligned}$$

as $n \to \infty$.

c) (10 points) Let $\{X_i\}_{i\geq1}$ be an iid sequence with $\mathrm{E}(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Does $\mathrm{E}\left(\left|\bar{X}_n - \mu\right|\right) \to 0$ as $n \to \infty$?

ANS: Yes:

$$\begin{aligned}
\mathrm{E}\left(\left|\bar{X}_n - \mu\right|\right) &\leq \mathrm{E}\left(\left|\bar{X}_n - \mu\right|^2\right)^{1/2} \\
&= Var\left(\bar{X}_n\right)^{1/2} \\
&= \frac{\sigma}{\sqrt{n}} \to 0,
\end{aligned}$$

where the first inequality follows by Jensen's inequality.

Question 3 (15 points) Let $\{X_i\}_{i\geq1}$ be an iid sequence with $\mathrm{E}(X_i) = \mu$ and $Var(X_i) = 1$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Find constants $r \geq 0$ and c such that $n^r\left(\bar{X}_n^2 - c\right) \overset{d}{\to} Y$ for some non-degenerate random variable Y.

ANS: Note that by the SLLN and CMT, $\bar{X}_n^2 \overset{a.s.}{\to} \mu^2$, so there is not a non-degenerate limit if $r = 0$. If $r > 0$, then by Slutsky's theorem,

$$n^{-r}n^r\left(\bar{X}_n^2 - c\right) \overset{d}{\to} 0 \cdot Y = 0.$$

Since $\overset{d}{\to}$ to a constant implies $\overset{p}{\to}$ to the same constant, $\bar{X}_n^2 - c \overset{p}{\to} 0$, so c must be equal to μ^2. Now consider two cases: If $\mu = 0$, the CLT tells us that

$$\sqrt{n}\bar{X}_n \overset{d}{\to} \mathcal{N}(0,1),$$

so by the CMT,

$$n\bar{X}_n^2 \xrightarrow{d} \chi_1^2,$$

so $c = 0$ and $r = 1$. If $\mu \neq 0$, note that by the CLT

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0,1).$$

By the delta method, with $g(x) = x^2$, we obtain $g'(x) = 2x$, and

$$\sqrt{n}\left(g\left(\bar{X}_n\right) - g(\mu)\right) = \sqrt{n}\left(\bar{X}_n^2 - \mu^2\right) \xrightarrow{d} \mathcal{N}\left(0, g'(\mu)^2\right) \overset{d}{=} \mathcal{N}\left(0, 4\mu^2\right).$$

Question 4 a) (5 points) Is it true that $X_n = O_p(1)$ implies $X_n = o_p(1)$?

ANS: No. Let $X_n = 1$ for all n. Then for any $\epsilon > 0$, $\mathrm{P}(X_n \leq 1) = 1 \geq 1 - \epsilon$ for all n, so $X_n = O_p(1)$. On the other hand, $\mathrm{P}(|X_n - 0| > \epsilon) = 1$ for all n and $\epsilon < 1$, so $X_n \overset{p}{\not\to} 0$.

b) (10 points) Consider a sequence of random variables $\{X_n\}_{n\geq 1}$ such that $X_n > 0$ for all n. Prove that $X_n = O_p\left(\mathrm{E}(|X_n|^r)^{1/r}\right)$ for any $r > 0$ such that $\mathrm{E}(|X_n|^r)$ exists.

ANS: Need to show that for any $\delta > 0$, $\exists B_\delta > 0$ such that

$$\mathrm{P}\left(\left|\frac{X_n}{\mathrm{E}(|X_n|^r)^{1/r}}\right| > B_\delta\right) \leq \delta.$$

By Chebyshev's inequality,

$$\mathrm{P}\left(|X_n| > B_\delta \mathrm{E}(|X_n|^r)^{1/r}\right) \leq \frac{\mathrm{E}(|X_n|^r)}{B_\delta^r \mathrm{E}(|X_n|^r)} = \frac{1}{B_\delta^r} \leq \delta$$

if $B_\delta \geq \left(\frac{1}{\delta}\right)^{1/r}$.

c) (15 points) Show that if $\frac{X_n}{Y_n} \xrightarrow{p} 0$ and $Y_n = O_p(1)$, then $X_n \xrightarrow{p} 0$.

<u>Hint</u>: First show that $|X_n| \leq B_\delta \left|\frac{X_n}{Y_n}\right| \mathbf{1}(|Y_n| < B_\delta) + |X_n| \mathbf{1}(|Y_n| \geq B_\delta)$.

ANS: Note that $\left|\frac{X_n}{Y_n}\right| = o_p(1)$, so $|X_n| = \left|\frac{X_n}{Y_n}\right| \cdot |Y_n| = o_p(1) \cdot O_p(1) = o_p(1)$, which implies $X_n \xrightarrow{p} 0$. A more explicit argument using the hint follows.

$$\begin{aligned}
|X_n| &= |X_n| \mathbf{1}(|Y_n| < B_\delta) + |X_n| \mathbf{1}(|Y_n| \geq B_\delta) \\
&\leq B_\delta \left|\frac{X_n}{Y_n}\right| \mathbf{1}(|Y_n| < B_\delta) + |X_n| \mathbf{1}(|Y_n| \geq B_\delta),
\end{aligned}$$

where the inequality holds because when $|Y_n| < B_\delta$, we have $\frac{B_\delta}{|Y_n|} > 1$, so for any $\epsilon > 0$:

$$\mathrm{P}(|X_n| > \epsilon) \leq \mathrm{P}\left(\left[B_\delta \left|\frac{X_n}{Y_n}\right| \mathbf{1}(|Y_n| < B_\delta) + |X_n| \mathbf{1}(|Y_n| \geq B_\delta)\right] > \epsilon\right)$$

$$
\begin{aligned}
&\leq \mathrm{P}\left(B_\delta \left|\frac{X_n}{Y_n}\right| \mathbf{1}\left(|Y_n| < B_\delta\right) > \frac{\epsilon}{2}\right) + \mathrm{P}\left(|X_n| \, \mathbf{1}\left(|Y_n| \geq B_\delta\right) > \frac{\epsilon}{2}\right) \\
&\leq \mathrm{P}\left(\left|\frac{X_n}{Y_n}\right| > \frac{\epsilon}{2B_\delta}\right) + \mathrm{P}\left(|Y_n| \geq B_\delta\right) \\
&\leq \mathrm{P}\left(\left|\frac{X_n}{Y_n}\right| > \frac{\epsilon}{2B_\delta}\right) + \delta \text{ for all } \delta > 0,
\end{aligned}
$$

where the final inequality follows because $B_\delta < \infty$ may be chosen such that $\mathrm{P}\left(|Y_n| \geq B_\delta\right) \leq \delta$, because $Y_n = O_p(1)$. The first term on the RHS converges to 0 for any $\delta > 0$, and δ can be chosen arbitrarily small, so the RHS of the final inequality is arbitrarily close to 0 for n large enough. So $\mathrm{P}\left(|X_n| > \epsilon\right) \to 0$, which means $X_n \xrightarrow{p} 0$.

Question 5 Consider the following example from Lecture 7. We wish to learn the population mean wage (or earnings potential), $\mathrm{E}(W)$. For each individual i in our sample, we observe whether or not they are employed. That is, we observe

$$
D_i = \begin{cases} 1 & \text{if individual } i \text{ is employed,} \\ 0 & \text{if individual } i \text{ is unemployed.} \end{cases}
$$

Assume that $0 < \mathrm{P}(D_i = 1) < 1$. We observe the wage of individual i, W_i, if and only if they are employed. The wage of unemployed individuals is not observed. That is, we observe

$$
Z_i = \begin{cases} W_i & \text{if individual } i \text{ is employed,} \\ 0 & \text{if individual } i \text{ is unemployed.} \end{cases}
$$

We assume $\mathrm{E}(W)$ exists, and we observe an iid sample $\{Z_i, D_i\}_{i=1}^n$.

a) (10 points) Is $\mathrm{E}(W)$ point identified? Explain using the definition of point identification.

ANS: No. Using the definition of point identification, we need to show that changing the value of $\mathrm{E}(W)$ does not necessarily change the joint distribution of observables (Z, D). We can do this by increasing the wages of all unemployed individuals by 1 dollar, since

$$
\begin{aligned}
\mathrm{E}(W_i) &= \mathrm{E}(W_i | D_i = 1)\, \mathrm{P}(D_i = 1) + \mathrm{E}(W_i | D_i = 0)\, \mathrm{P}(D_i = 0) \\
&< \mathrm{E}(W_i | D_i = 1)\, \mathrm{P}(D_i = 1) + \mathrm{E}(W_i + 1 | D_i = 0)\, \mathrm{P}(D_i = 0),
\end{aligned}
$$

because $\mathrm{P}(D_i = 0) > 0$. This does not change the joint distribution of Z, D, since the distribution of wages for employed individuals remains unchanged, while it still holds that $Z_i = 0$ for all unemployed individuals. Finally, the proportion of unemployed individuals remains the same also. In other words, $F_{Z,D} = F_{Z|D} \times F_D$ is unchanged.

b) (10 points) We try to estimate $\mu_W = \mathrm{E}(W)$ using the following estimator:

$$\hat{\mu}_W = \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n D_i}.$$

Find the almost sure limit of $\hat{\mu}_W$ and justify your answer. Write your answer as a feature of the joint distribution of (W, D). Derive a condition under which this almost sure limit equals $\mathrm{E}(W)$.

ANS: Since the random vectors $\{Z_i, D_i\}$ are iid, by the SLLN:

$$\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n Z_i \\ \frac{1}{n}\sum_{i=1}^n D_i \end{pmatrix} = \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} Z_i \\ D_i \end{pmatrix} \xrightarrow{a.s.} \mathrm{E}\left(\begin{pmatrix} Z_i \\ D_i \end{pmatrix}\right) = \begin{pmatrix} \mathrm{E}(Z_i) \\ \mathrm{E}(D_i) \end{pmatrix} = \begin{pmatrix} \mathrm{E}(W_i \mathbf{1}(D_i = 1)) \\ \mathrm{P}(D_i = 1) \end{pmatrix}.$$

By the continuous mapping theorem, therefore, it follows that

$$\hat{\mu}_W = \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n D_i} = \frac{\frac{1}{n}\sum_{i=1}^n Z_i}{\frac{1}{n}\sum_{i=1}^n D_i} \xrightarrow{a.s.} \frac{\mathrm{E}(W_i \mathbf{1}(D_i = 1))}{\mathrm{P}(D_i = 1)}.$$

Note that this last expression equals $\mathrm{E}(W_i | D_i = 1)$. If wages are mean independent of employment status, then $\mathrm{E}(W_i | D_i = 1) = \mathrm{E}(W_i)$.