ECMA 31000: Problem Set 2

Joe Hardwick

Due October 14 by 11:59PM

Question 1 a) Prove, using the definition of convergence in probability, that if $\{X_n\}_{n\geq 1}$ is a sequence of random variables such that $\mathrm{E}(X_n) = 0$ and $Var(X_n) = \frac{1}{n}$, then $X_n \xrightarrow{p} 0$.

b) Fix $(\Omega, \mathcal{F}, \mathrm{P})$. Show that if $A_1, \ldots, A_n$ is any sequence of events in $\mathcal{F}$, then

$$\mathrm{P}\left(\cup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

c) Prove that if $\{X_n\}_{n\geq 1}$ satisifes $\mathrm{E}(X_n) = 0$ and $Var(X_n) = \frac{1}{n^2}$ for all n, then $X_n \xrightarrow{a.s.} 0$. (Hint: Use part (b), and the 2nd definition of $\xrightarrow{a.s.}$ discussed in class).

Question 2 Let $X_n : [0,1] \to \mathbb{R}$ be a sequence of random variables defined on $([0,1], \mathcal{B}([0,1]), \lambda)$ where $\mathcal{B}([0,1])$ is the Borel sigma algebra on $[0,1]$. The only property of this probability space you will need is that λ satisfies

$$\lambda([a,b]) = \lambda((a,b]) = \lambda((a,b)) = \lambda([a,b)) = b - a$$

for $0 \leq a \leq b \leq 1$. Consider

$$X_n(\omega) = 2^n \mathbf{1}\left(\omega \leq \frac{1}{n}\right).$$

a) Show that $X_n \xrightarrow{a.s.} 0$.

b) Is it true that $\mathrm{E}(X_n) \to 0$?

c) Now consider another sequence Y_n of non-negative random variables such that $Y_n \leq K$ for some constant $K > 0$. If $Y_n \xrightarrow{a.s.} 0$, does $\mathrm{E}(Y_n) \to 0$? Prove it or provide a counterexample.

Question 3 Suppose that X_n is a sequence of random variables such that $\mathrm{E}(X_n) \to \mu$ and $Var(X_n) \to 0$. Show that $X_n \xrightarrow{p} \mu$.

Question 4 Suppose $\{X_i\}_{\geq 1}$ is a sequence of independent random variables with $\mathrm{E}(X_i) = \mu$ for all $i \geq 1$, and $\max_i \mathrm{E}(X_i^4) = K < \infty$. Show that $\bar{X}_n \xrightarrow{a.s.} \mu$.

Hint: Assume $\mu = 0$ for brevity.

Question 5 Show that if $X_n \overset{a.s.}{\to} X$ and $X_n \overset{p}{\to} Y$ then $X_n \overset{a.s.}{\to} Y$.

Hint: If $X_n \overset{a.s.}{\to} X$ but $X_n \overset{a.s.}{\nrightarrow} Y$, then $\mathrm{P}(X \neq Y) > 0$. This means there are constants $c, \delta > 0$ such that $\mathrm{P}(|X - Y| > c) > \delta$.

Question 6 Show that for $(K \times 1)$ random vectors $\{X_n\}_{n \geq 1}, X$:

$$X_n \overset{a.s.}{\to} X \iff X_{n,i} \overset{a.s.}{\to} X_i \text{ for all } i = 1, \ldots, K;$$
$$\mathrm{E}(\|X_n - X\|^r) \to 0 \iff \mathrm{E}(|X_{n,i} - X_i|^r) \to 0 \text{ for all } i = 1, \ldots, K.$$

Question 7 Show that if $\mathrm{E}(\|X_n - X\|^r) \to 0$, then $\mathrm{E}(\|X_n - X\|^s) \to 0$ for $0 < s < r$.

(Hint: Jensen!)

Question 8 a) Let $\{X_i\}_{i \geq 1}$ be an iid sequence of $U[0, \theta]$ random variables. For each n, derive the distribution of $\max_{i \leq n} X_i$,and show that $\max_{i \leq n} X_i \overset{p}{\to} \theta$.

b) Show that $n(\theta - \max_{i \leq n} X_i) \overset{d}{\to} X$ where X has an exponential distribution with CDF

$$F_X(x) = \begin{cases} 0 & x < 0, \\ 1 - \exp\left(-\frac{x}{\theta}\right) & x \geq 0. \end{cases}$$

Question 9 (Computational Question) Let $\{X_i\}_{i \geq 1}$ be an iid sequence such that $X_i \sim Bernoulli(p)$. That is:

$$X_i = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1 - p. \end{cases}$$

a) Show that $\mathrm{E}(X_i) = p$ and $Var(X_i) = p(1-p)$, and that $\mathrm{P}\left(\bar{X}_n - \epsilon < p < \bar{X}_n + \epsilon\right) \geq 1 - \frac{1}{4n\epsilon^2}$.
(Hint: Chebyshev!)

b) We call $\left[\bar{X}_n - \epsilon, \bar{X}_n + \epsilon\right]$ a confidence interval for p. Suppose you want to use your bound to ensure that the probability the true parameter p lies inside your confidence interval is at least 0.95. How large a sample must you take if $\epsilon = 0.1$?

c) Simulate n iid draws from this distribution with $p = 0.4$, for each of $n = 25, 50, 100$. Let $\epsilon = 0.1$ and compute the confidence intervals for each n based on your simulated data. Does the true value of p lie inside the confidence interval? Repeat this exercise 250 times for each value of n, (though you don't need to display the results of each replication). For each value n, report the proportion of your replications for which the true value of p lies in your confidence interval. Are these proportions generally greater or lower than the bound derived in a)? Why?