Econ 241A Probability, Statistics and Econometrics
Fall 2017

Problem Set 6

1. (Convergence in distribution does not imply convergence in probability). Define the following sequence of variables over the sample space generated by under a fair default coin toss, $S = \{H, T\}$:

$$X_n = \begin{cases} 1 \text{ if the coin toss is heads } s = H \\ 0 \text{ if the coin toss is tails } s = T \end{cases}$$

for all n (this means that $X_1 = X_2 = ... = X_n$). Now define the variable X such that

$$X = \begin{cases} 0 \text{ if the coin toss is heads } s = H \\ 1 \text{ if the coin toss is tails } s = T \end{cases}$$

(a) What is the cdf of X_1, X_2 and X_3? What is the cdf of X_n?

$$F_{X_i}(x_i) = \begin{cases} 1 \text{ if } 1 \leq x_i < \infty \\ \frac{1}{2} \text{ if } 0 \leq x_i < 1 \\ 0 \text{ if } -\infty < x_i < 0 \end{cases}$$

(b) What is the cdf of X?

$$F_X(x) = \begin{cases} 1 \text{ if } 1 \leq x < \infty \\ \frac{1}{2} \text{ if } 0 \leq x < 1 \\ 0 \text{ if } -\infty < x < 0 \end{cases}$$

(c) Does X_n converge in distribution to X?
Answer: Yes. The result is trivial since $F_{X_n}(x_n) = F_X(x), \forall n$.

(d) What is the cdf of $|X_n - X|$?
Answer: Let $Z = |X_n - X|$.
Note that because of the way the random variables are defined, it is never the case that $X_n = X$. Then:

$$F_Z(z) = \begin{cases} 1 \text{ if } 1 \leq z < \infty \\ 0 \text{ if } -\infty < z < 1 \end{cases}$$

(e) Using your answer to (d) find $P\left(|X_n - X| \geq \frac{1}{2}\right)$
Answer: $P\left(|X_n - X| \geq \frac{1}{2}\right) = P\left(Z \geq \frac{1}{2}\right) = 1$

(f) Does X_n converge in probability to X?
Answer: No, since

$$P\left(|X_n - X| \geq \frac{1}{2}\right) = P\left(Z \geq \frac{1}{2}\right) = 1, \forall n$$

and thus

$$\lim_{n\to\infty} P\left(|X_n - X| \geq \frac{1}{2}\right) = \lim_{n\to\infty} P\left(Z \geq \frac{1}{2}\right) = 1 > 0$$

2. Assume $X_1, ..., X_{20}$ is a random sample where $X_i \sim \text{exponential}(2)$ $\left(f_X(x) = \lambda e^{-\lambda x}\right)$. The first 4 uncentered moments of the exponential(2) distribution are given by: $\mathbb{E}(X_i) = \frac{1}{2}$, $\mathbb{E}(X_i^2) = \frac{1}{2}$, $\mathbb{E}(X_i^3) = \frac{3}{4}$ and $\mathbb{E}(X_i^4) = \frac{3}{2}$. Let $T = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

   (a) What is the mean of T?
   Answer: $\mathbb{E}(T) = \frac{1}{n} \cdot n \cdot \mathbb{E}(X_i^2) = \frac{1}{2}$

   (b) What is the variance of T?
   Answer: $\text{Var}(T) = \frac{1}{n^2} n\text{Var}(X_i^2) = \frac{1}{n}\left(\mathbb{E}(X_i^4) - \left(\mathbb{E}(X_i^2)\right)^2\right) = \frac{1}{n} \cdot (\frac{3}{2} - \frac{1}{4}) = \frac{5}{4n}$

   (c) What is the asymptotic distribution of T?
   Answer: Observe that T is an average of iid random variables, and hence we can apply the CLT. Thus we know $\sqrt{n}(T - \mu_{X^2}) \sim_A n(0, \sigma^2_{X^2})$, and where here $\mu_{X^2} = \mathbb{E}(X_i^2)$ and $\sigma^2_{X^2} = Var(X_i^2)$. We found these values in parts (a) and (b):

$$\sqrt{n}(T - \frac{1}{2}) \sim_A n(0, \frac{5}{4})$$

   This can also be expressed:

$$\frac{2\sqrt{n}}{\sqrt{5}}(T - \frac{1}{2}) \sim_A n(0, 1)$$

   (d) What is the approximate probability that $T \leq 1$?

$$P(T \leq 1) = P\left(\frac{2\sqrt{n}}{\sqrt{5}}(T - \frac{1}{2}) \leq \frac{2\sqrt{n}}{\sqrt{5}}(1 - \frac{1}{2})\right) \approx \Phi\left(\frac{2\sqrt{n}}{\sqrt{5}}(1 - \frac{1}{2})\right) = \Phi(2) \approx 0.98$$

   or so, where $\Phi(z)$ is the cdf for the standard normal distribution.

3. Let $\bar{X}$ denote the sample mean from a random sample of size n, from a population with exponential(λ) distribution. For convenience, let $\theta = \mathbb{E}(X) = \frac{1}{\lambda}$. So $\mathbb{E}(\bar{X}) = \theta$, $\text{Var}(\bar{X}) = \theta^2/n$, $\bar{X} \to_p \theta$ and $\frac{\sqrt{n}}{\theta}(\bar{X} - \theta) \to_d \text{n}(0, 1)$. Consider the sample statistic $U = \frac{1}{\bar{X}}$ (n.b. this is $1/\bar{X}$).

   (a) Use Slutsky theorem to show that $U \to_p \lambda$. Answer: Since $\bar{X} \to_p \theta$ and $\frac{1}{\bar{X}}$ (n.b. this is $1/\bar{X}$) is a continuous function of $\bar{X}$ the result follows directly from Thm. 5.5.4. (this is sometimes called the continuous mapping theorem).

   (b) Use the Delta method to find the limiting distribution of $\sqrt{n}(U - \lambda)$.
   Answer: Because $U = \frac{1}{\bar{X}}$ (n.b. this is $1/\bar{X}$), $\lambda = \frac{1}{\theta}$, and $\sqrt{n}(\bar{X} - \theta) \to_d \text{n}(0, \theta^2)$, we have $\sqrt{n}(U - \lambda) \to_d n(0, \theta^2 g'(\theta)^2)$. Note that $g(\theta) = \frac{1}{\theta}$, so $g'(\theta) = -\frac{1}{\theta^2}$. Thus:

$$\sqrt{n}(U - \lambda) \to_d n(0, \lambda^2)$$

   (c) Use your result to approximate $P(U \leq 5/2)$ with a random sample of size 16, from an exponential population with $\lambda = 2$.
   Answer: Because of (c), we know U is approximately distributed normal with mean and variance both equalling two. So,

$$P(U \leq 5/2) = P(\frac{\sqrt{16}(U - 2)}{2} \leq \frac{\sqrt{16}(5/2 - 2)}{2}) = \Phi(1) = 0.84$$

(d) Use the following result to find the exact value for $P(U \leq 5/2)$ with a random sample of size 16, from an exponential population with $\lambda = 2$. Let $X_1, X_2, ..., X_n$ be a random sample from an exponential(2) distribution. Define $\bar{X} = \frac{1}{n}\sum_i X_i$, then:

$$2n\lambda\bar{X} \sim \chi^2_{2n}$$

Answer: Define $W = 2n\lambda\bar{X}$, so that $U = \frac{2n\lambda}{W}$. Then $W = \frac{2n\lambda}{U}$, and finally with evaluating at $u = 5/2$ (because the map is monotonic decreasing):

$$F_U(u) = 1 - F_W\left(\frac{2n\lambda}{u}\right) = 0.78$$

Notice that this value is fairly different from the approximation. The sample size is relatively small.

4. In a population, the random variable X = length of unemployment (in months) has the exponential distribution with parameter $\lambda = 2$. Consider a random sample of unemployment lengths where the sample size is $n = 21$. Let T be the proportion of the sampled persons who have been unemployed between 0.4158 and 1 months.

   Approximate the probability that T lies between 0.4 and 0.5. Hint: define the random variable

$$U_i = \begin{cases} 1 \text{ if } 0.4158 \leq X_i \leq 1 \\ 0 \text{ otherwise} \end{cases}$$

   Answer: Let's define the pdf of U_i:

$$f_{U_i}(u) = \begin{cases} \exp(-2(0.4158)) - \exp(-2) \text{ if } u = 1 \\ 1 - \exp(-2(0.4158)) - \exp(-2) \text{ if } u = 0 \end{cases}$$

$$f_{U_i}(u) = \begin{cases} 0.3 \text{ if } u = 1 \\ 0.7 \text{ if } u = 0 \end{cases}$$

   First, we know that $T = \frac{1}{n}\sum_i U_i$, and because T is the average of Bernoulli r.v.s, it is Binomial, and has $\mathbb{E}(T) = 0.3$ and $\text{Var}(T) = 0.21$. Applying the CLT, we can assert: $\frac{\sqrt{21}}{\sqrt{0.21}}(T - 0.3) = 10(T - 0.3) \to_d \text{n}(0,1)$. Now, we apply th assympotic results and transform to standard normal: $P(0.4 < T < 0.5) = P(1 < 10(T - 0.3) < 2)$ so that $\approx P(1 < Z < 2) = 0.136$ is the approximation.

In addition, solve the following problems from Casella and Berger: 5.21 and 5.31.

5.21 (see text for Q): Answer: (from solutions): Let m denote the median. For the case with general n, we are interested in:

$$P(\max(X_1, \dots, X_n) > m) = 1 - P(X_i \leq m, \forall i) = 1 - [P(X_i \leq m)]^n = 1 - (\frac{1}{2})^n$$

5.31 (see text for Q):
We know that $\sigma^2_{\bar{X}} = 9/100$. By Chebyshev's inequaliy, we have that

$$P(|X - \mu_x| \geq k\sigma) \leq 1/k^2$$
$$P(|X - \mu_x| \geq 3k/10) \leq 1/k^2$$
$$P(-3k/10 < X - \mu_x < 3k/10) \geq 1 - 1/k^2$$

We need $1 - 1/k^2 \geq 0.9$ which implies $k \geq \sqrt{10} = 3.16$ and $3k/10 = 0.9487$. Then,

$$P(-0.9487 < X - \mu_x < 0.9487) \geq 0.9$$

Now, $\sqrt{n}(\bar{X} - \mu) \xrightarrow[n\to\infty]{d} n(0, \sigma_x^2)$ by the Central Limit Theorem (CLT). So $\sqrt{n}\frac{(\bar{X}-\mu)}{3/10} \sim_A n(0,1)$ (here $\sim_A$ denotes "asymptotically distributed"). So

$$P\left(-1.645 < \frac{\bar{X} - \mu}{3/10} < 1.645\right) = P(-0.4935 < \bar{X} - \mu < 0.4935) = 0.9$$

The CLT guarantees convergence in distribution of the average of an large enough i.i.d. sample. In our case, $n = 100$ which is large enough (in other cases it might be too low to suffice convergence to a normal distribution or the limit distribution we are working with).

Chevyshev's gives a more conservative estimate than the limits provided by the CLT. In other words, they are too wide relative to the asymptotic distribution, which is normal.