Large Sample Theory

Homework 2: Order Statistics

Due Date: October 27th

1. Show that if X has a beta $\beta(r, s)$ distribution,then

$$\begin{aligned} E(X^k) &= \frac{r\cdots(r+(k-1))}{(r+s)\cdots(r+s+(k-1))} \quad k=1,2,\cdots \\ Var(X) &= \frac{rs}{(r+s)^2(r+s+1)}. \end{aligned}$$

2. Let $X_1, \ldots, X_n$ be a sample from a uniform $U(0, 1)$ distribution. Deduce that $X_{(\ell)} - X_{(k)}$ has a $\beta(\ell - k, n - \ell + k + 1)$ distribution.

3. Let $X_1, \ldots, X_{2n+1}$ be a sample from a uniform $U(\theta - 0.5, \theta + 0.5)$ distribution.
   (a) Find the expectation and variance of $X_{(n+1)}$. Use the Chebyschev's inequality to show that $X_{(n+1)}$ is close to θ.
   (b) Derive the asymptotic distribution of the midrange, $\hat{\theta} = (X_{(2n+1)} + X_{(1)})/2$.

4. Let $X_1, \ldots, X_n$ be a sample from a population with density f and cdf F. Show that the conditional density of $(X_{(1)}, \ldots, X_{(r)})$ given $(X_{(r+1)}, \ldots, X_{(n)})$ is,

$$P(x_{(1)}, \ldots, x_{(r)}|x_{(r+1)}, \ldots, x_{(n)}) = \frac{r! \prod_{i=1}^{r} f(x_{(i)})}{F^r(x_{(r+1)})}$$

   if $x_{(1)} < \cdots < x_{(r)} < x_{(r+1)}$. Interpret this result.

5. Let the cdf F have a density f which is continuous and positive on an interval (a, b) such that $F(b) - F(a) = 1$, $-\infty \leq a < b \leq \infty$.
   (a) Show that if X has density f, then $Y = F(X)$ is uniformly distributed on $(0, 1)$.
   (b) Show that if $U \sim UNIF(0, 1)$, then $F^{-1}(U)$ has density f.
   (c) Let $U_{(1)} < \cdots < U_{(n)}$ be the order statistics of a sample of size n from a $UNIF(0, 1)$ population. Show that then, $F^{-1}(U_{(1)}) < \cdots < F^{-1}(U_{(n)})$ are distributed as the order statistics of a sample of size n from a population with density f.

6. Let $X_1, \ldots, X_n$ be iid according to the uniform distribution $U(0, \theta)$.
   (a) Derive the distribution of $n[X_{(n)}^{-1} - \theta^{-1}]$.
   (b) Find $n[EX_{(n)}^{-1} - \theta^{-1}]$ and $Var(X_{(n)}^{-1})$.

7. Let $X_1, X_2, \ldots$ be an infinite sequence of iid random variables with common cdf F. Let $u_1, u_2, \ldots$ be a sequence of nondecreasing real numbers. We say that an *exceedance* of the level u_n occurs at time i if $X_i > u_n$. Let A_n denote the number of exceedances of level u_n by $X_1, X_2, \ldots, X_n$.
   (a) Determine the distribution of A_n.
   (b) Suppose that u_n's are chosen such that $n[1 - F(u_n)] \to \lambda$ for some positive number λ as $n \to \infty$. Show that A_n converges in distribution and determine its

limiting distribution.
(c) Verify that the event $\{X_{(n:n-k+1)} \leq u_n\}$ is the same as the event $\{A_n < k\}$. Hence determine the limiting value of $P(X_{(n:n-k+1)} \leq u_n)$, Where k is held fixed and $n \to \infty$, if $n[1-F(u_n)] \to \lambda$.

8. (a) Determine $E|X|$ and $E\max(0,X)$, when X is $N(0,\sigma^2)$.
(b) Let $X_1,\ldots,X_n$ be a $N(\mu,\sigma^2)$ sample. Show that

$$E(\hat{\sigma}) = \frac{\sqrt{2}\Gamma(n/2)\sigma}{\sqrt{n}\Gamma(\frac{1}{2}(n-1))} \quad \text{and} \quad E(\tilde{\sigma}) = \sqrt{\frac{2(n-1)}{\pi n}}\sigma,$$

where $\hat{\sigma}$ and $\tilde{\sigma}$ are the sample standard and mean derivations, respectively. Namely,

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_i (X_i - \bar{X})^2} \quad \text{and} \quad \tilde{\sigma} = \frac{1}{n}\sum_i |X_i - \bar{X}|.$$

9. Let X be an absolutely continuous random variable with cdf F and having variance σ^2.
(a) Show that

$$\sigma^2 = \frac{1}{2}\int\int_{-\infty<x<y<\infty} F(x)[1-F(y)]dxdy.$$

(Hint: Try integration by parts.)
(b) Replace F by the sample distribution function F_n in (a). Is it closely related to the commonly used sample variance.

10. When U and V are iid with cdf F, $E|U-V|$ provides a measure of dispersion which we will denote by θ. Show that *Ginis mean difference*, given by

$$G_n = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}|X_i - X_j|,$$

is an unbiased estimator of θ. Also, prove that G_n can be expressed as

$$G_n = \frac{n+1}{n(n-1)}\sum_{i=1}^{n} 2\left(\frac{2i}{n+1} - 1\right)X_{(i)}.$$

11. Let X be an exponential random variable with distribution function $F(x) = 1-\exp(-\lambda t)$ and C be an independent random censoring time. It means that we can only observe $T = \min\{X, C\}$ and $\delta = 1_{\{X<C\}}$. Assume that we have a random sample of $(X_1,C_1),\ldots,(X_n,C_n)$ and we observe $(T_1,\delta_1),\ldots,(T_n,\delta_n)$.
(a) Derive $E(X_i - T_i|X_i > T_i)$.
(b) It is known that the method of moments leads to the estimate of θ as $\hat{\theta} = n/\sum_{i=1}^n X_i$. Since X_i cannot be observed, someone suggests to estimate unobservable X_i by $T_i + E(X_i - T_i|X_i > T_i)$ and solve the following equation

$$\theta = \frac{n}{\sum_{i=1}^n T_i\delta_i + \sum_{i=1}^n [T_i + E(X_i - T_i|X_i > T_i)](1-\delta_i)},$$

which leads to $\hat{\theta}_E$.
(a) Derive $\hat{\theta}_E$.
(b) Show that $\hat{\theta}_E$ is a consistent estimate of θ and derive its asymptotic distribution. (For solving (b), you can assume that C_i is exponentially distributed with parameter λ.)

12. The Weibull distribution with index α has c.d.f.

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - \exp(-(x/\lambda)^\alpha), & x \geq 0, \end{cases} \quad \lambda, \alpha > 0.$$

It is used in industrial reliability studies in situations where failure of a system comprising many similar components occurs when the weakest component fails. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed continuous non-negative random variables with p.d.f. $f(x)$ and c.d.f. $F(x)$ such that, as $x \to 0$, $f(x) \to kx^{\alpha-1}$, $F(x) \to kx^\alpha/\alpha$, where $k, \alpha > 0$. Let $W = (kn/\alpha)^{1/\alpha} X_{(1)}$.
(a) Show that, as $n \to \infty$, W has as its limiting distribution the Weibull distribution with index α.
(b) Explain why a probability plot for the Weibull distribution may be based upon plotting the logarithm of the rth order statistic against $\log\left[\log\left(1 - \frac{r}{n+1}\right)\right]$ and give the slope and intercept of such a plot.