Statistics 583, Problem Set 3
Wellner; 4/13/2016

Reading: Chapter 2, sections 2.1- 2.7;
Ferguson, Chapters 1-10 + 13.
Van der Vaart, Chapters 2-4 + 21.
Chapter 3, Sections 3.1 - 3.5.

Due: Wednesday, April 20, 2016

1. Suppose that U is a random variable with a Uniform$(0,1)$ distribution. For each integer $n \ge 1$ define $X_n \equiv c_n 1_{[0,1/n]}(U) + 1$ for some sequence of real numbers c_n.
(a) Compute $E(X_n)$ (for an arbitrary sequence c_n).
(b) Show that $X_n \rightarrow_{a.s.} 1 \equiv X$ for any real sequence c_n.
(c) Find a sequence c_n so that $E\{\underline{\lim} X_n\} < \underline{\lim} E\{X_n\} < \infty$, showing that strict inequality can occur in in Fatou's inequality.
(d) For $c_n = n/\log(n+1)$, show that $E(\lim X_n) = \lim E(X_n)$ and that $\{X_n\}$ is uniformly integrable, but that the domination hypothesis of the dominated convergence theorem fails.

2. Suppose that $\underline{N}_n \sim \text{Mult}_k(n, \underline{p}_0)$ where $\sum_{j=1}^k p_{0,j} = 1$. Let $\underline{\widehat{p}}_n \equiv \underline{N}_n/n$.
(a) Use the multivariate CLT to show that
$$\sqrt{n}(\underline{\widehat{p}}_n - \underline{p}_0) \rightarrow_d \underline{Y} \sim N_k(0, \text{diag}(\underline{p}_0) - \underline{p}_0 \underline{p}_0^T).$$
(b) Use the result in (a) to prove that $C_n \equiv \sum_{j=1}^k (N_j - np_{0,j})^2/(np_{0,j}) \rightarrow_d \chi^2_{k-1}$.
(c) Use the result in (a) to prove that $H_n^2 \equiv 4n \sum_{j=1}^k (\widehat{p}_{n,j}^{1/2} - p_{0,j}^{1/2})^2 \rightarrow_d \chi^2_{k-1}$.

3. Let $\{X_n\}_{n=0}^\infty$ be a sequence of random variables with distribution functions $\{F_n\}_{n=0}^\infty$ and that $X_n \rightarrow_d X_0$; i.e. $F_n(x) \rightarrow F_0(x)$ for all $x \in C(F_0) \equiv \{x \in \mathbb{R} :$ F_0 is continuous at $x\}$. Show that this holds if and only if $Eg(X_n) = \int_{\mathbb{R}} g dF_n \rightarrow \int_{\mathbb{R}} g dF_0 = Eg(X_0)$ for all bounded and continuous functions g. [Hint: bracket $1_{(-\infty,x]}(y)$ above and below by continuous functions.]

4. (a) Show that the key condition of the Liapunov CLT implies the Lindeberg condition.
(b) Use the Cramér - Wold device together with the Lindeberg-Feller CLT to prove that for the linear regression model discussed in class on 4/11-13, the estimators $\widehat{\alpha}_n = \overline{Y}_n$ and $\widehat{\beta}_n = \sum_{i=1}^n Y_i(x_i - \overline{x}_n)/\sum_1^n (x_i - \overline{x}_n)^2$ are jointly asymptotically normal.

5. Suppose that Y is a random variable with $E(Y^2) < \infty$.
(a) Show that

$$Var(Y) = E\{Var(Y|X)\} + Var\{E(Y|X)\}\,;$$

i.e.

$$E(Y - EY)^2 = E\{E[(Y - E(Y|X))^2|X]\} + E\{[E(Y|X) - E(Y)]^2\}\,.$$

(b) Interpret (a) geometrically. [Hint: the space of all (centered) square integrable random variables is a Hilbert space.]
(c) Suppose that $Y \sim \chi^2_n(\delta)$. Compute $E(Y)$ and $Var(Y)$. [Hint: here it helps to use the following conditional representation of the random variable Y having the noncentral chisquare distribution with n degrees of freedom and noncentrality parameter δ: $(Y|K = k) \sim \chi^2_{2k+n}$ where $K \sim$ Poisson$(\delta/2)$. See e.g. JonW Chapter 1 notes, pages 15-16. Then use $E(Y) = E\{E(Y|K)\}$ and (a).]
(d) Show that

$$\frac{\chi^2_n(\delta) - (n+\delta)}{\sqrt{2n + 4\delta}} \to_d N(0,1)$$

as either $n \to \infty$ or $\delta \to \infty$.

6. (Bonus problem 1). Suppose that $X_1, \ldots, X_n$ are i.i.d. with distribution function F which has positive density f at its quartiles $F^{-1}(1/4)$ and $F^{-1}(3/4)$ and at its median $F^{-1}(1/2)$.
(a) Let $Q_n = (X_{(3n/4)} + X_{(n/4)})/2 = (\mathbb{F}_n^{-1}(3/4) + \mathbb{F}_n^{-1}(1/4))$, the mid-quartile range. Find the asymptotic distribution of Q_n as an estimator of the population mid-quartile range $Q = Q(F) = (F^{-1}(3/4) + F^{-1}(1/4))/2$. That is, prove that

$$\sqrt{n}(Q_n - Q) \to_d \text{“}\textit{something}\text{”}$$

and find “something”.
(b) Find the joint limiting distribution of $(\sqrt{n}(M_n - m), \sqrt{n}(Q_n - Q))$ where $m = F^{-1}(1/2)$.
(c) Assuming that the underlying distribution is Cauchy(θ, σ) ($X = \sigma Y + \theta$ where $Y \sim$ Cauchy$(0,1)$), compare the variances of the mid-quartile range Q_n and the median M_n as estimators of θ.

7. (Bonus problem 2). (a) Suppose that $\underline{Y}_n = (Y_{n,1}, \ldots, Y_{n,d})$ where the components $\{Y_{n,j}\}_{j=1}^d$ of $\underline{Y}_n$ are independent and $Y_{n,j} \to_d Y_j$ for each $j = 1, \ldots, d$. Show that $\underline{Y}_n \to_d \underline{Y}$ in $\mathbb{R}^d$.
(b) Chapter 2, JW notes, Exercise 5.3, page 25. Show that if $X_1, X_2, \ldots$ are i.i.d. with $E(X_1) = 0$, $Var(X_1) < \infty$, then $\mathbb{S}_n \to_{f.d.} \mathbb{S}$ where $\mathbb{S}$ is standard Brownian motion on $[0,1]$. [Hint: One approach uses the fact that $\mathbb{S}_n(t_j) - \mathbb{S}_n(t_{j-1}) = n^{-1/2}\sum_{i=[nt_{j-1}]+1}^{[nt_j]} X_i$, $j = 1, \ldots, k$ with $t_0 \equiv 0$ are independent random variables, and then uses (a) together with continuous mappiing.]