Problem Set #3

1. Let $X_i, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution P on S and let $\mathbf{F}$ be a class of measureable functions on S. Suppose that for each $\epsilon > 0$ there exists a finite number of pairs of functions $(f_j^l, f_j^u), j = 1, \ldots, k$ satisfying

   (i) $E_P[|f_j^l(X_i)|] < \infty$ and $E_P[|f_j^u(X_i)|] < \infty$ for all $1 \leq j \leq k$;

   (ii) $E_P[f_j^u(X_i) - f_j^l(X_i)] < \epsilon$ for all $1 \leq j \leq k$;

   (iii) for each $f \in \mathbf{F}$ there exists j such that $f_j^l \leq f \leq f_j^u$.

   Prove that

   $$\sup_{f \in \mathbf{F}} |\frac{1}{n} \sum_{1 \leq i \leq n} f(X_i) - E_P[f(X_i)]| \to 0 \text{ a.s.}$$

   (Hint: Generalize the proof of the classical Glivenko-Cantelli Theorem given in class.)

2. Let $\mathbf{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$ be a class of measureable functions on S where Θ is compact metric space. Suppose

   (i) $f(x, \theta)$ is a continuous function of θ for all $x \in S$;

   (ii) there exists F on S such that $|f(\cdot, \theta)| \leq F$ for all $\theta \in \Theta$ and $x \in S$ and $E[F(X_i)] < \infty$.

   This exercise steps you through verifying that this class of functions satisfies the three conditions of the preceding exercise.

   (a) Let $U_n(\theta^*)$ be an open ball centered at $\theta^* \in \Theta$ with radius $1/n$. Use the Dominated Convergence Theorem to argue that

   $$E_P[\sup_{\theta \in U_n(\theta^*)} f(X_i, \theta) - \inf_{\theta \in U_n(\theta^*)} f(X_i, \theta)] \to 0$$

   as $n \to \infty$.

(b) Let $\epsilon > 0$ be given. Conclude from (a) that for each $\theta^* \in \Theta$ there exists an open ball $U(\theta^*)$ centered at θ^* for which

$$E_P[\sup_{\theta \in U(\theta^*)} f(X_i, \theta) - \inf_{\theta \in U(\theta^*)} f(X_i, \theta)] < \epsilon \ .$$

(c) Use compactness of Θ to construct a finite subcover $\{U_j : 1 \leq j \leq k\}$ of Θ from the open cover $\{U(\theta) : \theta \in \Theta\}$. For $1 \leq j \leq k$, define

$$\begin{aligned} f_j^l(x) &= \inf_{\theta \in U_j} f(x, \theta) \\ f_j^u(x) &= \sup_{\theta \in U_j} f(x, \theta) \ . \end{aligned}$$

Verify that these functions satisfy the desired properties.

3. Let $X_i, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with continuous distribution F on $\mathbf{R}$. Define

$$F^{-1}(u) = \inf\{x \in \mathbf{R} : F(x) \geq u\}$$

and let $\hat{F}_n(x)$ be the empirical distribution function of the $X_i, i = 1, \ldots, n$. The following exercise steps you through proving that

$$\sup_{x \in \mathbf{R}} \sqrt{n} |\hat{F}_n(x) - F(x)| \tag{1}$$

is a pivot; that is, its distribution does not depend on F.

(a) Show that $F^{-1}(u) \leq x$ if and only if $u \leq F(x)$.

(b) Let $U \sim \text{Unif}(0, 1)$. Use (a) to show that $F^{-1}(U) \sim F$. (This useful trick is known as the quantile transformation.)

(c) Use (a) and (b) to show that the distribution of (1) is the same as the distribution of

$$\sup_{x \in \mathbf{R}} \sqrt{n} |\frac{1}{n} \sum_{1 \leq i \leq n} \mathbf{1}\{U_i \leq F(x)\} - F(x)| \ . \tag{2}$$

(d) Complete the proof by using the fact that F is continuous to show that the distribution of (2) is the same as the distribution of

$$\sup_{u \in [0,1]} \sqrt{n} \left| \frac{1}{n} \sum_{1 \le i \le n} \mathbf{1}\{U_i \le u\} - u \right| . \tag{3}$$

(e) We only used continuity of F in part (d). How would the distributions of (1) and (3) be related if F were not continuous?

4. Let F_n and F be nonrandom distribution functions on $\mathbf{R}$. Suppose F_n converges in distribution to F and that F is continuous. Show that

$$\sup_{x \in \mathbf{R}} |F_n(x) - F(x)| \to 0 .$$

(This result holds more generally on $\mathbf{R}^k$ with $k \ge 1$. It is sometimes referred to as Polya's Theorem.)

5. Let $X_i, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution P on $\mathbf{R}$ with finite, nonzero variance $\sigma^2(P)$. Let $J_n(x, P)$ denote the distribution of $\sqrt{n}(\bar{X}_n - \theta(P))$. Let $J(x, P) = \Phi(x/\sigma(P))$. Show that

$$\sup_{x \in \mathbf{R}} |J_n(x, \hat{P}_n) - J(x, P)| \to 0$$

a.s., where $\hat{P}_n$ is the empirical distribution of the $X_i, i = 1, \ldots, n$. (Hint: Use the above exercise.)

6. Let $X_i, i = 1, \ldots, n_X$ be a i.i.d. sequence of random variables with distribution P_X on $\mathbf{R}$ with finite, nonzero variance $\sigma^2(P_X)$. Independently, let $Y_i, i = 1, \ldots, n_X$ be a i.i.d. sequence of random variables with distribution P_Y on $\mathbf{R}$ with finite, nonzero variance $\sigma^2(P_Y)$. Suppose n_X and n_Y tend to infinity in a way such that $n_X/(n_X + n_Y) \to \rho \in (0, 1)$. Construct a bootstrap confidence interval for $\theta(P_X, P_Y) = \mu(P_X) - \mu(P_Y)$ of nominal level $1 - \alpha$ and show that it asymptotically has the correct coverage probability.