# ECO 519. Spring 2006

# Homework # 1 (due: March 8th)

1. (a) Generalize the result in Example 24 in Pollard (1984) to an arbitrary quantile of $Y$. Make any additional assumptions you need.

   (b) Suppose $X_n \Rightarrow X$ and let $\{x_n\}$ be a sequence of real numbers such that $x_n \to x$ and $\Pr\{X = x\} = 0$, show that $\Pr\{X_n \geq x_n\} \to \Pr\{X \geq x\}$ (Hint: Remember the continuous mapping theorem).

2. Read and replicate on your own the statements and proofs of Theorem 1 (and Lemma 1) and Theorem 2 (and Lemma 2) in Huber (1967).

   **Definition (this is relevant for the next problem):** Take a sequence of random vectors $X_n \in \mathbb{R}^k$. We say that $X_n = O_p(n^r)$ if $\forall\ \epsilon > 0$, there exists $M_\epsilon$ and $n_\epsilon$ such that $\Pr\big\{\|n^{-r}X_n\| > M_\epsilon\big\} < \epsilon$ for all $n \geq n_\epsilon$. We say that $X_n = o_p(n^r)$ if $n^{-r}X_n \xrightarrow{p} 0$.

3. Let $X_n$ and $Y_n$ be two real-valued random variables. Show the following:

   (a) If $X_n = O_p(n^{r_1})$ and $Y_n = O_p(n^{r_2})$, then $X_nY_n = O_p(n^{r_1+r_2})$. Let $\overline{r} = \max\{r_1, r_2\}$, then $X_n + Y_n = O_p(n^{\overline{r}})$.

   (b) If $X_n = o_p(n^{r_1})$ and $Y_n = o_p(n^{r_2})$, then $X_nY_n = o_p(n^{r_1+r_2})$. Let $\overline{r} = \max\{r_1, r_2\}$, then $X_n + Y_n = o_p(n^{\overline{r}})$.

   (c) If $X_n = O_p(n^{r_1})$ and $Y_n = o_p(n^{r_2})$, then $X_nY_n = o_p(n^{r_1+r_2})$ Let $\overline{r} = \max\{r_1, r_2\}$, then $X_n + Y_n = O_p(n^{\overline{r}})$.

   *Hint:* The following probability inequalities (which come from the Axioms of Probability): $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$ and $\Pr(A \cap B) \leq \Pr(A) + \Pr(B)$ might be useful.

4. Suppose $\theta_n$ is a minimizer of $G_n(\theta)$, and $\theta_0$ is a minimizer of $G(\theta)$ (you may think of $G(\theta)$ as the probability limit of $G_n(\theta)$, although that is not crucial here). Suppose $\theta_n \xrightarrow{p} \theta_0$ and also that

(a) There exists a neighborhood $\mathcal{N}$ of $\theta_0$ and a constant $\kappa > 0$ for which

$$G(\theta) \geq G(\theta_0) + \kappa\|\theta - \theta_0\|^2 \text{ for all } \theta \in \mathcal{N}.$$

(b) Uniformly over $o_p(1)$ neighborhoods of $\theta_0$ (that is, uniformly over any neighborhood of the form $\{\theta : \|\theta - \theta_0\| = o_p(1)\}$),

$$G_n(\theta) - G_n(\theta_0) = G(\theta) - G(\theta_0) + O_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\right) + o_p\big(\|\theta - \theta_0\|^2\big) + O_p(b_n),$$

where $\{b_n\}$ is a sequence of nonnegative real numbers that satisfies $b_n \to 0$ as $n \to \infty$.

Show that if these conditions are satisfied, we must have

$$\|\theta_n - \theta_0\| = O_p\big(\max\{b_n^{1/2}, n^{-1/2}\}\big).$$

Given this, provide sufficient conditions that guarantee $\|\theta_n - \theta_0\| = O_p(n^{-1/2})$ (in this case, we say that $\theta_n$ is a "root-n consistent" estimator of $\theta_0$).

5. (This problem can be seen as a follow-up of the previous problem, which provided sufficient conditions for root-n consistency. Now we provide sufficient conditions for root-n asymptotic normality.) Let $\Theta$ be a subset of $\mathbb{R}^k$. Suppose $\theta_n$ minimizes $G_n(\theta)$ over $\Theta$ and suppose $\theta_n$ is root-n consistent for $\theta_0$, where $\theta_0$ is in the interior of $\Theta$. Suppose also that uniformly over $O_p(n^{-1/2})$ neighborhoods of $\theta_0$ (that is, uniformly over any set of the form $\big\{\theta : \|\theta - \theta_0\| = O_p(n^{-1/2})\big\}$),

$$G_n(\theta) - G_n(\theta_0) = \frac{1}{2}(\theta - \theta_0)'V(\theta - \theta_0) + \frac{1}{\sqrt{n}}(\theta - \theta_0)'W_n + o_p(1/n)$$

where $V$ is a positive definite matrix and $W_n$ converges in distribution to a $\mathcal{N}(0, \Delta)$ random vector. Show that in this case, $\sqrt{n}(\theta_n - \theta_0) \Rightarrow \mathcal{N}(0, V^{-1}\Delta V^{-1})$. Explain why problems 1 and 2 are a generalization of the condition $\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(x_i, \theta_0) + \sqrt{n}\lambda(\theta_n) = o_p(1)$, which is the starting point to prove asymptotic normality in Huber's framework.

6. Let $X_1, \ldots, X_n$ be an iid sample with probability measure $P \in \mathcal{P}$ (the probability measure belongs to a family of probability measures denoted by $\mathcal{P}$). Throughout, we will let $P_n$ denote the empirical distribution of our sample. Take a real-valued function $\psi : \mathbb{R} \to \mathbb{R}$ that satisfies the following conditions:

(i) $\psi$ is monotone nondecreasing.

(ii) $\psi(-\infty) < 0 < \psi(\infty)$

(iii) $\big|\psi(x)\big| \leq M < \infty$ for all $x$

We will denote the parameter of interest $\theta$ by $\theta(P)$ just to remind you that it depends on the underlying true distribution.

(a) Show that any $\theta(P)$ that satisfies

$$E_P\big[\psi(X_i - \theta(P))\big] \geq 0 \geq E_P\big[\psi(X_i - \theta')\big] \text{ for all } \theta' > \theta(P)$$

must be finite.

(b) Suppose that for all $P \in \mathcal{P}$, $\theta(P)$ is the unique solution to $E_P\big[\psi(X_i - \theta)\big] = 0$. Let $\widehat{\theta}_n = \theta(P_n)$ (i.e, the solution to $0 = E_{P_n}\big[\psi(X_i - \theta)\big] = \frac{1}{n}\sum_{i=1}^n \psi(X_i - \theta)$). Show that for any $P \in \mathcal{P}$, $\widehat{\theta}_n \xrightarrow{p} \theta(P)$ (Hint: Show first that $E_P\big[\psi(X_i - \theta)\big]$ is nonincreasing in $\theta$, then use a dominated convergence argument and the fact that $\psi(X, \theta) \to \psi(-\infty) < 0$ as $\theta \to \infty$ and $\psi(X, \theta) \to \psi(\infty) > 0$ as $\theta \to -\infty$. To see how to use these results to establish consistency formally, see for example Lemma 2 and Theorem 2 in Huber, pp. 225-226.)

(c) Assume the conditions of the previous part hold. Let $\lambda(\theta) = E_P\psi(X_i - \theta)$ and $\tau^2(\theta) = \text{Var}_P\psi(X_i - \theta)$. Suppose $\lambda'(\theta) < 0$ exists and suppose that for any sequence $\theta_n \to \theta$ that satisfies $\theta_n = \theta + t/\sqrt{N}$ for some $t$ the following is true:

$$\frac{1}{\sqrt{n}\tau(\theta)} \sum_{i=1}^n \big[\psi(X_i - \theta_n) - \lambda(\theta_n)\big] \Rightarrow \mathcal{N}(0, 1).$$

Given this, let $\widehat{\theta}_n$ be the estimator defined in the previous part. Show that

$$\sqrt{n}(\widehat{\theta}_n - \theta) \Rightarrow \mathcal{N}\left(0, \frac{\tau^2(\theta)}{[\lambda'(\theta)]^2}\right).$$

Hint: The key is to note that for any $t$ and any $\theta_n$ such that $\theta_n = \theta + t/\sqrt{N}$, we have: $\Pr\big(\sqrt{n}(\widehat{\theta}_n - \theta) < t\big) = \Pr(\widehat{\theta}_n < \theta_n) = \Pr\big(\sum_{i=1}^n \psi(X_i - \theta_n) > 0\big)$.

(d) Suppose the cdf of $X$, $F(x)$ is continuous. Let $\theta$ denote the median of $X$ and suppose that $F'(\theta) = f(\theta) > 0$. Use the previous two parts to find the asymptotic distribution of the sample median, $\widehat{\theta}_n$. Compare your result with Example 24 in Pollard (1984). (Hint: Use $\psi(x) = \text{sign}(x)$).

7. Huber (1967) claims that we can relax Assumption (N-3)(iii) to $E[u(x,\theta,d)^2] < o\big(|\log d|^{-1}\big)$. Verify if the claim is true. (Hint: You don't have to replicate the entire proof, focus on the steps in which Chebushev's inequality is invoked and how things change afterwards if the assumption is relaxed in the proposed way.)