Stat 710 11/13/02

Solutions to Problem Set 3

Ch. 5 #15. (X_i, Y_i) is *iid* with joint distribution given by $Y_i = f_{\vartheta_0}(X_i) + e_i$. Least squares says $\hat{\vartheta} \equiv \text{argmin}_\vartheta \; n^{-1} \sum_{i=1}^n (Y_i - f_\vartheta(X_i))^2$. What we are minimizing is asymptotically $E(Y_1 - f_\vartheta(X_1))^2$, so if this min is unique then it is given (subject to regularity conditions) as the solution to $0 =$

$$\nabla_\vartheta \, E(Y_1 - f_\vartheta(X_1))^2 \;=\; -2 \int \nabla_\vartheta f_\vartheta(x) \; \{E(e_1|X_1 = x) + f_{\vartheta_0}(x) - f_\vartheta(x)\} \, f_X(x) \, dx$$

If uniqueness holds (eg via a 2nd derivative condition) and $E(e|X = x) = 0$, then the condition for solution is met (and thus met uniquely) when $\vartheta_* = \vartheta_0$. But $E(e) = 0$ is not necessarily enough to make this work, even under regularity conditions! This says that the closest function $f_\vartheta(X_1)$ to $f_{\vartheta_0}(X_1) + E(e_1|X_1)$ may not be the one with $\vartheta = \vartheta_0$. For example, take $f_\vartheta(X_1) = \vartheta X_1$ and suppose $E(e_1|X_1) = X_1$. Then least-squares would give estimators converging not to ϑ_0 but to $\vartheta_0 + 1$, and in this example can have $E(X_1) = 0$, which means that also $E(e_1) = 0$. The important property (for consistency) is not expectation 0, but conditional expectation identically equal to 0.

#20. In this problem, there is a function $\ddot{\psi}_\vartheta(x)$ applying to individual x-coordinates, as well as a function $\ddot{\Psi}(\vartheta) = n^{-1} \sum_{i=1}^n \ddot{\psi}_\vartheta(X_i)$ applying to the whole data-sample, and these must be distinguished, because the domination-assumption by M applies directly only to the first one ! We are *assuming* that $\tilde{\vartheta} \xrightarrow{P} \vartheta_0$ and that there exist M, δ such that $|\ddot{\psi}_\vartheta(x)| \le M(x)$ for all x and ϑ such that $|\vartheta - \vartheta_0| < \delta$. Then $P(|\tilde{\vartheta} - \vartheta_0| \ge \delta) \to 0$ and $n^{-1} \sum_{i=1}^n M(X_i) \xrightarrow{P} EM(X_1) < \infty$ implies $P(|n^{-1} \sum_{i=1}^n M(X_i)| \ge 2EM(X_1)) \longrightarrow 0$, and for all $\epsilon > 0$,

$$P(|\tilde{\vartheta} - \vartheta_0| \, |\frac{1}{n} \sum_{i=1}^n \psi_{\ddot{\tilde{\vartheta}}}(X_i)| \;\ge\; \epsilon) \;\le\; P(|\tilde{\vartheta} - \vartheta_0| \ge \max(\delta, \; \frac{\epsilon}{2EM(X_1)}) +$$

$$+ \; P(|\tilde{\vartheta} - \vartheta_0 < \delta, \; |\frac{1}{n} \sum_{i=1}^n M(X_i)| \ge 2EM(X_1)) \;\longrightarrow\; 0$$

which implies $(\tilde{\vartheta} - \vartheta_0) \; \ddot{\Psi}_{\tilde{\vartheta}} = \; (\tilde{\vartheta} - \vartheta_0) \, n^{-1} \sum_{i=1}^n \; \ddot{\psi}_{\tilde{\vartheta}}(X_i) = o_P(1)$.

Ch. 6 #1. First, $(dQ_n/dP_n)(X) = \exp(\mu_n X - \mu_n^2/2)$, where $X \equiv Z \sim \mathcal{N}(0,1)$ under P_n and $X - \mu_n \equiv Z \sim \mathcal{N}(0,1)$ under Q_n. This likelihood-ratio sequence is convergent in distribution along subsequences whether or not μ_n is bounded, since it degenerates to 0 if $\mu_n \to \pm\infty$. Now by LeCam's First Lemma, we have mutual contiguity of P_n, Q_n iff for all subsequences of n along which μ_n converges to a finite or infinite limit, with $Z \sim \mathcal{N}(0,1)$, $\exp(\mu_n Z - \mu_n^2) \xrightarrow{\mathcal{D}} U$ with $P(U > 0) = 1$, $E(U) = 1$, i.e., $\mu_n \not\to \pm\infty$. Thus the criterion is that $\sup_n |\mu_n| < \infty$.

#2. Now $P_n \sim \mathcal{N}(0, 1/n)$, $Q_n \sim \mathcal{N}(\vartheta_n, 1/n)$, and $(dQ_n/dP_n)(T) = \exp(nT\vartheta_n - n\vartheta_n^2/2)$. Letting $Z \equiv T\sqrt{n} \sim \mathcal{N}(0,1)$ under P_n, we find again as in #1, via LeCam's First Lemma, that mutual contiguity holds iff for all sequences of n along which $\exp(\vartheta_n Z\sqrt{n} - n\vartheta_n^2/2) \xrightarrow{\mathcal{D}} U$, also $U > 0$ and $EU = 1$, so that $\mu_n = \sqrt{n}\vartheta_n = \mathcal{O}_P(1)$, as asserted. *A student in the class points out that in general, if for measures* P_n, Q_n *there is a one-to-one transformation* g_n *such that* $P_n^* \equiv P_n \circ g_n^{-1}$, $Q_n^* \equiv Q_n \circ g_n^{-1}$ *are mutually absolutely contiguous, then the same holds for* P_n, Q_n. *In this problem,* $g_n(x) = \sqrt{n}x$, *and the measures* P_n^*, Q_n^* *are exactly the same as the measures* P_n, Q_n *in #6.1.* (The general assertion holds because, if B_n are events with $P_n(B_n) \to 0$, then $A_n = \{g_n(x) : x \in B_n\}$ satisfies $P_n^*(A_n) \to 0$ which by contiguity implies $Q_n(B_n) = Q_n^*(A_n) \to 0$. The proof works the same way for the other direction of contiguity.)

#4. Suppose $\|P_n - Q_n\| \to 0$. Then $\forall A_n$, $P(A_n) \to 0$ implies

$$Q(A_n) \le |Q(A_n) - P(A_n)| + P(A_n) \le \|P_n - Q_n\| + P(A_n) \to 0$$

and similarly $P(A_n) \le \|P_n - Q_n\| + Q(A_n) \to 0$ if $Q(A_n) \to 0$. Thus convergence of variation distance to 0 implies mutual contiguity.

#6. The simplest example is $P_n \equiv P$ equal to twice Lebesgue measure on $[0, 1/2]$ and $Q_n \equiv Q$ Lebesgue on $[0, 1]$.