Statistics 581, Solutions,Problem Set 6

Wellner; 11/9/2000

1. Consider the tests of $H : p = p_0$ versus $K : p \neq p_0$ based on Q_n and $4nH_n^2$ that we discussed in class and in problem set 3. These tests can be used to form confidence sets C_n and D_n for p as follows: define $Q_n(p_0) = \sum_{j=1}^k (N_j - np_{j0})^2/np_{j0}$ and $H_n^2(p_0) \equiv 4n\sum_{j=1}^k(\sqrt{\hat{p}_j} - \sqrt{p_{j0}})^2$, so that both $Q_n(p_0)$ and $H_n^2(p_0)$ are asymptotically χ_{k-1}^2 under the null hypothesis. Let

$$C_n = \{p_0 : Q_n(p_0) \leq \chi^2_{k-1,\alpha}\}$$

and

$$D_n = \{p_0 : H_n^2(p_0) \leq \chi^2_{k-1,\alpha}\}\,.$$

(a) Show that these both yield asymptotic $1 - \alpha$ confidence sets: e.g.

$$P_{p_0}(p_0 \in C_n) \to 1 - \alpha \qquad \text{as } \ n \to \infty\,.$$

(b) Can you use our "Theorem 2" about $Q_n(p_0)$ and $H_n^2(p_0)$ under $p \neq p_0$ to prove some other property of these confidence sets?

Solutions: (a) Note that

$$P_{p_0}(p_0 \in C_n) = P_{p_0}(Q_n(p_0) \leq \chi^2_{k-1,\alpha}) \to P(\chi^2_{k-1} \leq \chi^2_{k-1,\alpha}) = 1 - \alpha\,;$$

and, correspondingly,

$$P_{p_0}(p_0 \in D_n) = P_{p_0}(H_n^2(p_0) \leq \chi^2_{k-1,\alpha}) \to P(\chi^2_{k-1} \leq \chi^2_{k-1,\alpha}) = 1 - \alpha\,.$$

Thus C_n and D_n are both asymptotic $1 - \alpha$ confidence sets for p.
(b) When p_0 is true, and $p \neq p_0$, then

$$P_{p_0}(p \in C_n) = P_{p_0}(Q_n(p) \leq \chi^2_{k-1,\alpha}) \to 0$$

since $n^{-1}Q_n(p) \to_{p,a.s.} \sum_{j=1}^k (p_{0j} - p_j)^2/p_j > 0$ when $p_0 \neq p$ is true, and hence $Q_n(p) \to_p \infty$. Similarly,

$$P_{p_0}(p \in D_n) = P_{p_0}(H_n^2(p) \leq \chi^2_{k-1,\alpha}) \to 0$$

since $n^{-1}H_n^2(p) \to_{p,a.s.} 4H^2(p_0, p) > 0$ when $p_0 \neq p$ is true, and hence $H_n^2(p) \to_p \infty$.
(c) (Not assigned). When p_0 is true, and $p = p_n = p_0 + cn^{-1/2} \neq p_0$, with $1'c = 0$, then

$$P_{p_0}(p_n \in C_n) = P_{p_0}(Q_n(p_n) \leq \chi^2_{k-1,\alpha}) \to P(\chi^2_{k-1}(\delta) \leq \chi^2_{k-1,\alpha}) < 1 - \alpha$$

where the non-centrality parameter $\delta = \sum_{j=1}^k c_j^2/p_{0j}$. Thus the probability of "false coverage" for values of p of the (local) form p_n is related to the (local asymptotic) power of the test.

2. Suppose that $X_1, \ldots, X_n, \ldots$ are i.i.d. random vectors in R^k with common distribution function F and corresponding probability measure P on $(R^k, \mathcal{B}_k)$. Let $\mathbb{P}_n$ be the empirical measure defined by

$$\mathbb{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i} ,$$

and consider $\mathbb{P}_n$ and the empirical process $\mathbb{G}_n$ as indexed by a class of sets $\mathcal{C} \subset \mathcal{B}_k$:

$$\{\mathbb{P}_n(C) : \; C \in \mathcal{C}\}, \qquad \{\mathbb{G}_n(C) : \; C \in \mathcal{C}\} ,$$

where

$$\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P) .$$

(a) Show that $\mathbb{G}_n \rightarrow_{f.d.} \mathbb{G}_P$ where $\mathbb{G}_P$ is a P-Brownian bridge process indexed by $\mathcal{C}$: i.e. show that for any integer m and sets $C_1, \ldots, C_m \in \mathcal{C}$,

$$(\mathbb{G}_n(C_1), \ldots, \mathbb{G}_n(C_m)) \rightarrow_d (\mathbb{G}_P(C_1), \ldots, \mathbb{G}_P(C_m)) \sim N_m(0, \Sigma)$$

where $\Sigma = (\sigma_{jj'})$ is given by

$$\sigma_{jj'} = P(C_j \cap C_{j'}) - P(C_j)P(C_{j'}) .$$

(b) When $\mathcal{C} = \mathcal{O} \equiv \{(-\infty, x] : \; x \in R^k\}$ specialize the result in (a) and show that it gives the finite-dimensional convergence of the empirical distribution function $\mathbb{F}_n$: i.e.
(i) show that $\mathbb{P}_n((-\infty, x]) = \mathbb{F}_n(x)$;
(ii) show that $P((-\infty, x]) = F(x)$;
(iii) show that $\mathbb{Y}(x) \equiv \mathbb{G}_P((-\infty, x])$ has mean zero and covariance

$$E\{\mathbb{Y}(x)\mathbb{Y}(y)\} = F(x \wedge y) - F(x)F(y), \qquad x, y \in R^k .$$

Solution: (a) This follows directly from the multivariate central limit theorem: for any integer m and sets $C_1, \ldots, C_m \in \mathcal{C}$,

$$\begin{pmatrix} \mathbb{G}_n(C_1) \\ \cdot \\ \cdot \\ \cdot \\ \mathbb{G}_n(C_m) \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} 1_{C_1}(X_i) - P(C_1) \\ \cdot \\ \cdot \\ \cdot \\ 1_{C_m}(X_i) - P(C_m) \end{pmatrix} \equiv \sqrt{n}\, \underline{Y}_n$$

where $\underline{Y}_1, \ldots, \underline{Y}_n$ are i.i.d. with $E(\underline{Y}_i) = 0$,

$$E(\underline{Y}_i \underline{Y}_i') = (P(C_j \cap C_{j'}) - P(C_j)P(C_{j'})) \equiv \Sigma .$$

Hence by the multivariate central limit theorem

$$\begin{pmatrix} \mathbb{G}_n(C_1) \\ \cdot \\ \cdot \\ \cdot \\ \mathbb{G}_n(C_m) \end{pmatrix} \rightarrow_d \begin{pmatrix} \mathbb{G}_P(C_1) \\ \cdot \\ \cdot \\ \cdot \\ \mathbb{G}_P(C_m) \end{pmatrix} \sim N_m(0, \Sigma) . \tag{0.1}$$

In other words, $\mathbb{G}_n \rightarrow_{f.d.} \mathbb{G}_P$ as indexed by the class $\mathcal{C}$.
(b) When $\mathcal{C} = \mathcal{O} \equiv \{(-\infty, x] : \ x \in R^k\}$, we compute
(i) $\mathbb{P}_n((-\infty, x]) = n^{-1} \sum_{i=1}^{n} 1_{(-\infty,x]}(X_i) = \mathbb{F}_n(x)$;
(ii) $P((-\infty, x]) = P(X \le x) = F(x)$;
(iii) and we have, with $\mathbb{Y}(x) \equiv \mathbb{G}_P((-\infty, x])$,

$$\begin{aligned} E\{\mathbb{Y}(x)\mathbb{Y}(y)\} &= E\{\mathbb{G}_P((-\infty, x])\mathbb{G}_P((-\infty, y]) \\ &= P((-\infty, x] \cap (-\infty, y]) - P((-\infty, x])P((-\infty, y]) \\ &= P((-\infty, x \wedge y]) - P((-\infty, x])P((-\infty, y]) \\ &= F(x \wedge y) - F(x)F(y) \quad \text{by three applications of } \ (ii). \end{aligned}$$