Statistics 523, Problem Set 5 Solutions

Wellner; May 7, 2010

1. Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables with values in the measurable space $(S, \mathcal{S})$ be defined on $(\Omega, \mathcal{A}, P)$. Let $\mathcal{A}_n$ be the $\sigma-$field generated by all functions of $(X_1, X_2, \ldots)$ that are symmetric in their first n arguments. Prove that a sequence $\{U_n\}$ of $U-$statistics based on a fixed symmetric kernel h of order r (so that $h : S^r \to \mathbb{R}$) is a reverse martingale for $n \ge r$ with respect to the filtration $\mathcal{A}_r \supset \mathcal{A}_{r+1} \supset \cdots$.

   Solution: First, with $\underline{X} = (X_1, \ldots, X_n) \in S^n$,

$$U_n = \frac{1}{\binom{n}{r}} \sum_{1 \le i_1 < i_2 < \cdots < i_n \le n} h(X_{i_1}, \ldots, X_{i_r})$$

   and h symmetric, note that $U_n(\pi \underline{X}) = U_n(\underline{X})$ where $\pi \underline{X} = (X_{\pi(1)}, \ldots, X_{\pi(n)})$. Thus U_n is $\mathcal{A}_n$ measurable. Furthermore, with $\mathcal{B}_n$ the sigma-field generated by all symmetric functions of $X_1, \ldots, X_n$, for any $1 \le i_1 < \cdots < i_r \le n$,

$$\begin{aligned} E\{h(X_{i_1}, \ldots, X_{i_r})|\mathcal{A}_n\} &= E\{h(X_{i_1}, \ldots, X_{i_r})|\mathcal{B}_n\} \\ &= E\{h(X_1, \ldots, X_r)|\mathcal{B}_n\} = E\{h(X_1, \ldots, X_r)|\mathcal{A}_n\}, \end{aligned}$$

   Thus

$$\begin{aligned} E\{h(X_1, \ldots, X_r)|\mathcal{A}_n\} &= \frac{1}{\binom{n}{r}} \sum_{1 \le i_1 < i_2 < \cdots < i_n \le n} E\{h(X_{i_1}, \ldots, X_{i_r})|\mathcal{A}_n\} \\ &= E\left\{ \frac{1}{\binom{n}{r}} \sum_{1 \le i_1 < i_2 < \cdots < i_n \le n} h(X_{i_1}, \ldots, X_{i_r})|\mathcal{A}_n \right\} \\ &= E\{U_n|\mathcal{A}_n\} = U_n. \end{aligned}$$

   Hence $\{U_n, \mathcal{A}_n\}_{n \ge r}$ is a reverse martingale: for $r \le m < n$,

$$\begin{aligned} E\{U_m|\mathcal{A}_n\} &= E\left\{E\{h(X_1, \ldots, X_r)|\mathcal{A}_m\}|\mathcal{A}_n\right\} \\ &= E\{h(X_1, \ldots, X_r)|\mathcal{A}_n\} = U_n \quad \text{a.s..} \end{aligned}$$

2. Suppose that the kernel h in problem 1 satisfies $E|h(X_1, \ldots, X_r)| < \infty$. Show that the corresponding sequence of U statistics $\{U_n\}_{n\geq r}$ satisfies $U_n \rightarrow_{a.s.,1} Eh(X_1, \ldots, X_r)$. (For $r > 1$ the given condition is not necessary, but a simple necessary and sufficient condition appears to be unknown.) Use your result for $\{U_n\}$ to show a corresponding result for the sequence of $V-$statistics $\{V_n\}$ associated with $\{U_n\}$ under reasonable additional hypotheses.

Solution: First note that

$$\begin{aligned} E|U_n| &= E\{E\{h(X_1, \ldots, X_r)|\mathcal{A}_n\}|\} \leq EE\{|h(X_1, \ldots, X_n)||\mathcal{A}_n\} \\ &\qquad \text{by the conditional version of Jensen's inequality} \\ &= E|h(X_1, \ldots, X_r)| < \infty \ \text{ by hypothesis.} \end{aligned}$$

and $\{U_n, \mathcal{A}_n\}_{n\geq r}$ is a reversion martingale. By the reverse martingale convergence theorem 13.3.3,

$$U_n = E(h(X_1, \ldots, X_r)|\mathcal{A}_n) \rightarrow_{a.s.,1} E(h(X_1, \ldots, X_r)|\mathcal{A}_\infty)$$

where $\mathcal{A}_\infty = \cap_{n\geq r}\mathcal{A}_n$ is the *symmetric sigma field*. By the Hewitt - Savage $0-1$ law, $\mathcal{A}_\infty$ is trivial. Now $E\{E(h(x_1, \ldots, X_r)|\mathcal{A}_\infty)\} = Eh(X_1, \ldots, X_r)$, so Hewitt-Savage implies that $E\{h(X_1, \ldots, X_r)|\mathcal{A}_\infty\} = Eh(X_1, \ldots, X_r)$ a.s. and we conclude that $U_n \rightarrow_{a.s.,1} Eh(X_1, \ldots, X_r)$.

To relate this to the corresponding $V-$statistic, we note that

$$n^r V_n - n_{(r)} r! U_n) = \sum h(X_{i_1}, \ldots, X_{i_r}) \equiv (n^r - n_{(r)})W_n$$

where $n_{(r)} = n(n-1)\cdots(n-r+1) = n!/(n-r)!$ and the sum on the right side is over all terms $h(X_{i_1}, \ldots, X_{i_r})$ with at least one equality $i_a = i_b$ for $a \neq b$. Note that there are $n^r - n_{(r)}$ such terms and that $n^r - n_{(r)} = O(n^{r-1})$. Thus, if $E|h(X_{i_1}, \ldots, X_{i_r})| < \infty$ for all $1 \leq i_1 \leq \ldots \leq i_r \leq r$,

$$\begin{aligned} V_n &= \frac{r!}{n^r}\binom{n}{r}U_n + \frac{n^r - n_{(r)}}{n^r}W_n \\ &\rightarrow_{a.s.} \ 1 \cdot Eh(X_1, \ldots, X_r) + 0 \cdot w = Eh(X_1, \ldots, X_r) \end{aligned}$$

where w is the limit of several U statistics terms of order $r-1$ or smaller. See Serfling (1980), page 206 and de la Pena and Giné (1999), pages 235-237 for an alternative approach (with somewhat different hypotheses).

3. Williams, PwM, Problem E10.8, page 234. Suppose that $\Theta \sim \text{Uniform}(0,1)$, and that conditional on $\Theta = \theta$, $B_n \sim \text{Binomial}(n, \theta)$.
(a) Show that B_n has the same probability structure as in problem E10.1 on Polya's urn.
(b) Show that N_n^θ in that problem is a regular conditional density (with respect to Lebesgue measure) of Θ given $B_1, \ldots, B_n$.

Solution: (a) For any $0 \le k_1 \le \cdot \le k_n$ with $k_j \in \{0, \ldots, n\}$ and $0 \le k_j - k_{j-1} \le 1$ we have

$$\begin{aligned} P(B_1 = k_1, \ldots, B_n = k_n | \Theta = \theta) &= P(B_n - B_{n-1} = k_n - k_{n-1}, \ldots, B_1 = k_1 | \Theta = \theta) \\ &= \prod_{j=1}^n \theta^{k_j - k_{j-1}} (1-\theta)^{1-(k_j - k_{j-1})} = \theta^{k_n} (1-\theta)^{n-k_n}, \end{aligned}$$

and hence

$$\begin{aligned} P(B_1 = k_1, \ldots, B_n = k_n) &= \int_0^1 \theta^{k_n} (1-\theta)^{n-k_n} d\theta = \frac{\Gamma(k_n+1)\Gamma(n-k_n+1)}{\Gamma(n+2)} \\ &= \frac{k_n!(n-k_n)!}{(n+1)!}. \end{aligned}$$

Since there are $\binom{n}{k_n}$ ways of putting k_n ones in n positions, it follows that

$$P(B_n = k_n) = \frac{1}{n+1}, \quad \text{for} \quad k_n \in \{0, \ldots, n\}.$$

Arguing similarly, we find that

$$P(B_n = k_n, B_{n-1} = k_{n-1} | \Theta = \theta) = \frac{(n-1)!}{k_{n-1}!(n-1-k_{n-1})!} \theta^{k_n} (1-\theta)^{n-k_n},$$

and hence

$$\begin{aligned} P(B_n = k_n, B_{n-1} = k_{n-1}) &= \frac{(n-1)!}{k_{n-1}!(n-1-k_{n-1})!} \frac{k_n!(n-k_n)!}{(n+1)!} \\ &= \begin{cases} \frac{1}{(n+1)n}(k_{n-1}+1), & \text{if } k_n = k_{n-1}+1, \\ \frac{1}{(n+1)n}(n - k_{n-1}), & \text{if } k_n = k_{n-1}. \end{cases} \\ &= \begin{cases} \frac{1}{n} \frac{(k_{n-1}+1)}{n+1}, & \text{if } k_n = k_{n-1}+1, \\ \frac{1}{n}\left(1 - \frac{k_{n-1}+1}{n+1}\right), & \text{if } k_n = k_{n-1}. \end{cases} \end{aligned}$$

It follows immediately that

$$P(B_n = k_n | B_{n-1} = k_{n-1}) = \begin{cases} \frac{(k_{n-1}+1)}{n+1}, & \text{if } k_n = k_{n-1} + 1, \\ 1 - \frac{k_{n-1}+1}{n+1}, & \text{if } k_n = k_{n-1}. \end{cases}$$

Thus $\{B_n\}$ has the same probabilistic structure as in the Polya urn scheme.

(b) Now we want to show that for any bounded measurable function f

$$E\{f(\Theta)|B_1, \ldots, B_n\} = \int_0^1 f(\theta) \frac{(n+1)!}{B_n!(n-B_n)!} \theta^{B_n} (1-\theta)^{n-B_n} d\theta \quad \text{a.s..}$$

This will be proved if we show that

$$E\{1_A f(\Theta)\} = E\left\{1_A \int_0^1 f(\theta) \frac{(n+1)!}{B_n!(n-B_n)!} \theta^{B_n} (1-\theta)^{n-B_n} d\theta \right\} \qquad (1)$$

for any $A \in \mathcal{A}_n = \sigma\{B_1, \ldots, B_n\}$. Consider sets $A \in \mathcal{A}_n$ of the form

$$A = \{B_1 = k_1,\ B_2 = k_2, \ldots, B_n = k_n\} = \cap_{j=1}^n \{B_j - B_{j-1} = k_j - k_{j-1}\}$$

where $0 \le k_j - k_{j-1} \le 1$, $k_j - k_{j-1} \in \{0, 1\}$. Since all the sets of this form generate $\mathcal{A}_n$, it suffices to prove (1) for sets of this form. Now for sets A of this form,

$$\begin{aligned} E\{1_A f(\Theta)\} &= \int_0^1 P(B_1 = k_1, \ldots, B_n = k_n | \Theta = \theta) f(\theta) d\theta \\ &= \int_0^1 P(\cap_{j=1}^n \{B_j - B_{j-1} = k_j - k_{j-1} | \Theta = \theta) f(\theta) d\theta \\ &= \int_0^1 \theta^{k_n} (1-\theta)^{n-k_n} f(\theta) d\theta \end{aligned}$$

much as in the calculations in (a). Since this depends only on k_n, the value of B_n, and does not depend on $k_1, \ldots, k_{n-1}$, we can sum over the $\binom{n}{k_n}$ ways of putting k_n ones in n positions and use the uniform marginal distribution of B_n to find that (1) holds.

4. (a) Suppose that $X_1, X_2, \ldots$ are independent with probability distributions $(P_1, P_2, \ldots)$ where $P_k = N(0, 1)$ for each k or independent with

distributions $(Q_1, Q_2, \ldots)$ where $Q_k = N(c_k, 1)$ for some sequence of numbers c_k with $c_k \to 0$. Calculate $H(P_k, Q_k)$ in terms of c_k where $H(P, Q)$ is the Hellinger distance between P, Q given by

$$H^2(P,Q) = \frac{1}{2}\int \{\sqrt{p} - \sqrt{q}\}^2 d\mu$$

where $p = dP/d\mu$, $q = dQ/d\mu$ for some dominating measure μ (recall that $\mu = P + Q$ always works). If $Y_k \equiv dQ_k/dP_k$ for $k = 1, 2, \ldots$, and we define $M_n \equiv \prod_{k=1}^n Y_k$, for what sequences c_k does $M_n \to_{a.s.} M_\infty$ with $E(M_\infty) = 1$ under $P^\infty = \prod_{k=1}^\infty P_k$? For what sequences does $M_n \to_{a.s.} 0$ under P^∞?
(b) Consider a similar example when $P_k = \exp(1)$ for all k and $Q_k = \exp(c_k)$ with $c_k \to 1$.
Hint: compute in terms of the Hellinger affinity $\rho(P_k, Q_k) \equiv \int \sqrt{p_k q_k} d\mu$.

Solution: (a) Now

$$\begin{aligned} \rho(P_k, Q_k) &= \int_{-\infty}^{\infty} \sqrt{\phi(x)\phi(x - c_k)}dx \\ &= \int \sqrt{(2\pi)^{-1}\exp(-x^2/2)\exp(-(x - c_k)^2/2)}dx \\ &= \exp(-c_k^2/8). \end{aligned}$$

By Kakutani's theorem, $M_n = \prod_{k=1}^n \frac{dQ_k}{dP_k} \to_{a.s.} M_\infty$ with $E(M_\infty) = 1$ if and only if $\prod_{k=1}^\infty \rho(P_k, Q_k) > 0$. But

$$\prod_{k=1}^n \rho(P_k, Q_k) = \prod_{k=1}^n \exp(-c_k^2/8) = \exp(-\sum_{k=1}^n c_k^2/8),$$

so this converges to something positive if and only if $\sum_{k=1}^\infty c_k^2 < \infty$. For example, if $c_k = k^{-1/2}(\log k)^{-(1+\delta)/2}$ with $\delta > 0$, then $\sum_{k=1}^\infty c_k^2 < \infty$ and $E(M_\infty) = 1$. On the other hand if $\delta = 0$, then $\sum_{k=1}^\infty c_k^2 = \infty$, and $P(M_\infty = 0) = 1$.
(b) When P_k is the $\exp(1)$ distribution for all k and Q_k is $\exp(c_k)$, we

compute

$$
\begin{aligned}
\rho(P_, Q_k) &= \int_0^\infty \sqrt{\exp(-x) c_k \exp(-c_k x)} dx \\
&= \sqrt{c_k} \int_0^\infty \exp(-(1+c_k)x/2) dx = \sqrt{c_k} \frac{2}{1+c_k} \\
&= \frac{2\sqrt{c_k}}{1+c_k} = \frac{2\sqrt{1+b_k}}{2+b_k} \quad \text{with} \quad c_k \equiv 1+b_k, \\
&= \frac{2\sqrt{1+b_k}}{2(1+b_k/2)} \sim (1+b_k/2)(1-b_k/2) = 1 - b_k^2/4.
\end{aligned}
$$

Thus

$$
1 - \rho(P_k, Q_k) \sim \frac{1}{4} b_k^2,
$$

and

$$
\sum_{k=1}^{\infty} (1 - \rho(P_k, Q_k)) < \infty \quad \text{iff} \quad \sum_{k=1}^{\infty} b_k^2 < \infty.
$$