Statistics 581, Problem Set 6 Solutions

Wellner; 11/9/2005

1. Chapter 2, Exercise 5.3, page 25. [Hint: use the fact that $\mathbb{S}_n(t_j) - \mathbb{S}_n(t_{j-1}) = n^{-1/2}\sum_{i=[nt_{j-1}]+1}^{[nt_j]} X_i$, $j = 1, \ldots, t_k$ with $t_0 \equiv 0$ are independent random variables.]

   Solution: Note that

$$V_{n,j} \equiv \mathbb{S}_n(t_j) - \mathbb{S}_n(t_{j-1}) = \frac{1}{\sqrt{n}} \sum_{i=[nt_{j-1}]+1}^{[nt_j]} X_i$$

   is the sum of $[nt_j] - [nt_{j-1}]$ i.i.d. random variable X_i with mean 0 and variance 1. Since $n^{-1}([nt_j] - [nt_{j-1}]) \to t_j - t_{j-1}$ it follows that

$$\begin{aligned} V_{n,j} &= \sqrt{n^{-1}([nt_j] - [nt_{j-1}])}\frac{1}{\sqrt{[nt_j] - [nt_{j-1}]}} \sum_{i=[nt_{j-1}]+1}^{[nt_j]} X_i \\ &\to_d \sqrt{t_j - t_{j-1}} Z_j \equiv V_j \qquad (0.1) \end{aligned}$$

   where $Z_j \sim N(0,1)$. Since the V_{nj}'s are independent for $j = 1, \ldots, k$, the convergence in (0.1) holds jointly for the vector $\underline{V}_n \equiv (V_{n,1}, \ldots, V_{n,k})^T$ (this can be seen either by arguing directly, or by characteristic functions) and we conclude that $\underline{V}_n = (V_{n,1}, \ldots, V_{n,k})^T \to_d (V_1, \ldots, V_k)^T \equiv \underline{V}$ where $\underline{V} \equiv (V_1, \ldots, V_k)^T \sim N_k(0, \text{diag}(t_j - t_{j-1}))$. Now $\mathbb{S}_n(t_j) = (1, 1, \ldots, 1, 0, \ldots, 0)^T V_n$ where there are j ones in the vector multiplying V_n (and $k - j$ zeros). Thus with M defined to be the matrix

$$M \equiv \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 0 & \cdot & \cdot & 0 \\ 1 & 1 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & 1 \end{pmatrix}$$

   we can write

$$\begin{pmatrix} \mathbb{S}_n(t_1) \\ \mathbb{S}_n(t_2) \\ \cdot \\ \cdot \\ \cdot \\ \mathbb{S}_n(t_k) \end{pmatrix} = M\underline{V}_n \to_d M\underline{V} \stackrel{d}{=} \begin{pmatrix} \mathbb{S}(t_1) \\ \mathbb{S}(t_2) \\ \cdot \\ \cdot \\ \cdot \\ \mathbb{S}(t_k) \end{pmatrix} \sim N_k(0, (t_i \wedge t_j)).$$

2. Consider a function $T : \mathcal{F} \to \mathbb{R}$ where $\mathcal{F}$ is some (sub) class of distribution functions F (examples include the mean, $T(F) = \mu(F) = \int x dF(x)$, the variance $T(F) = \sigma^2(F) = \int (x - \int y dF(y))^2 dF(x)$, the median $T(F) = F^{-1}(1/2)$, linear combinations of order statistics $T(F) = \int_0^1 F^{-1}(u)w(u)du$, the *Lorenz curve* (at $t \in (0,1)$), $T(F) =$

$\int_0^t F^{-1}(u)du / \int_0^1 F^{-1}(u)du \equiv L(t,F)$, and so forth). [The Lorenz curve gives the percentage of "income" received by the poorest fraction t of the income distribution.] The "principle of substitution" says that $T(F)$ can be estimated by $T(\widehat{F}_n)$. for some estimator $\widehat{F}_n$ of F. If T is sufficiently "smooth", then frequently the empirical distribution function $\mathbb{F}_n$ can be taken as the estimator $\hat{F}_n$ of F.

Give a treatment of consistency and asymptotic normality of the estimator $L(t,\mathbb{F}_n)$ of $L(t,F)$ based on our results from sections 2.4 and 2.6. You may assume that with $X \sim F$ we have $E_F|X| < \infty$ and $E_F X^2 < \infty$ (and any other additional assumptions you need).

Solution. First consistency. Since

$$L(t,F) = \int_0^t F^{-1}(u)du / \int_0^1 F^{-1}(u)du$$

and similarly for $L(t,\mathbb{F}_n)$, for consistency it suffices to show that both $\int_0^t \mathbb{F}_n^{-1}(u)du \to_p \int_0^t F^{-1}(u)du$ and $\int_0^1 \mathbb{F}_n^{-1}(u)du \to_p \int_0^1 F^{-1}(u)du$. Since $\int_0^1 F^{-1}(u)du = \int_{\mathbb{R}} x dF(x)$ and similarly with F^{-1}, F replaced by $\mathbb{F}_n^{-1}, \mathbb{F}_n$, if $E|X| = \int |x| dF(x) < \infty$, it follows from the weak (or strong) law of large numbers that

$$\begin{aligned}\int_0^1 \mathbb{F}_n^{-1}(u)du &= \int_{\mathbb{R}} x d\mathbb{F}_n(x) = \overline{X}_n \\ &\to_{p(a.s.)} EX = \int x dF(x) = \int_0^1 F^{-1}(u)du.\end{aligned}$$

To show that $\int_0^t \mathbb{F}_n^{-1}(u)du \to_p \int_0^t F^{-1}(u)du$, note that

$$\begin{aligned}\left|\int_0^t \mathbb{F}_n^{-1}(u)du - \int_0^t F^{-1}(u)du\right| &\le \int_0^t |\mathbb{F}_n^{-1}(u) - F^{-1}(u)|du \\ &\le \sup_{0\le u\le t} |\mathbb{F}_n^{-1}(u) - F^{-1}(u)| \times t \\ &\to_{a.s.} 0\end{aligned}$$

by (8) of Proposition 2.6.1, page 28 of the course notes if we assume in addition that F^{-1} is continuous on $[0,t]$ with $t \in (0,1)$.

To establish asymptotic normality By the delta-method applied to the function $g(x,y) = x/y$ it suffices to show that

$$\begin{pmatrix} \sqrt{n}(\int_0^t \mathbb{F}_n^{-1}(u)du - \int_0^t F^{-1}(u)du) \\ \sqrt{n}(\int_0^1 \mathbb{F}_n^{-1}(u)du - \int_0^1 F^{-1}(u)du) \end{pmatrix} \to_d \begin{pmatrix} Z_t \\ Z_1 \end{pmatrix}$$

where $(Z_t, Z_1)^T \sim N_2(0,\Sigma)$ for some covariance matrix Σ. As we have seen above $\int_0^1 \mathbb{F}_n^{-1}(u)du = \overline{X}_n$ and $\int_0^1 F^{-1}(u)du = E_F X \equiv \mu$, so we will assume $E_F X^2 < \infty$ to insure the convergence of the second coordinate in the last display. To establish the needed joint convergence we write, assuming that $Q = F^{-1}$ has derivative Q' on

$(0,1)$,

$$\begin{pmatrix} \sqrt{n}(\int_0^t \mathbb{F}_n^{-1}(u)du - \int_0^t F^{-1}(u)du) \\ \sqrt{n}(\int_0^1 \mathbb{F}_n^{-1}(u)du - \int_0^1 F^{-1}(u)du) \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \int_0^t \sqrt{n}\{F^{-1}(\mathbb{G}_n^{-1}(u)) - F^{-1}(u)\}du \\ \int_0^1 \sqrt{n}\{F^{-1}(\mathbb{G}_n^{-1}(u)) - F^{-1}(u)\}du \end{pmatrix}$$

$$= \begin{pmatrix} \int_0^t \sqrt{n}\{F^{-1}(\mathbb{G}_n^{-1}(u)) - F^{-1}(u)\}du \\ \int_0^1 \sqrt{n}\{F^{-1}(\mathbb{G}_n^{-1}(u)) - F^{-1}(u)\}du \end{pmatrix}$$

$$= \begin{pmatrix} \int_0^t Q'(u)\mathbb{V}_n(u)du \\ \int_0^1 Q'(u)\mathbb{V}_n(u)du \end{pmatrix} + o_p(1)$$

$$= \begin{pmatrix} \int_0^t \mathbb{V}_n(u)dF^{-1}(u) \\ \int_0^1 \mathbb{V}_n(u)dF^{-1}(u) \end{pmatrix} + o_p(1)$$

$$\to_d \begin{pmatrix} \int_0^t \mathbb{V}(u)dF^{-1}(u) \\ \int_0^1 \mathbb{V}(u)dF^{-1}(u) \end{pmatrix}$$

$$\sim N_2(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} \int_0^t\int_0^t (u\wedge v - uv)dF^{-1}(u)dF^{-1}(v) & \int_0^t\int_0^1 (u\wedge v - uv)dF^{-1}(u)dF^{-1}(v) \\ \int_0^t\int_0^1 (u\wedge v - uv)dF^{-1}(u)dF^{-1}(v) & \int_0^1\int_0^1 (u\wedge v - uv)dF^{-1}(u)dF^{-1}(v) \end{pmatrix}.$$

where, by formula (1.4.16) of the course notes, and by the inverse transformation,

$$\int_0^1\int_0^1 (u\wedge v - uv)dF^{-1}(u)dF^{-1}(v) = Var[F^{-1}(\xi)] = Var[X].$$

3. Suppose that $Z \sim N(0,1)$ and, for $\mu \in R$ and $\sigma > 0$, that $X = \mu + \sigma Z \sim P_{\mu,\sigma} = N(\mu, \sigma^2)$.
A. Compute the likelihood ratio

$$\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(x) = \frac{\sigma^{-1}\phi((x-\mu)/\sigma)}{\sigma^{-1}\phi(x/\sigma)} \qquad \text{and} \qquad Y \equiv \log\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X).$$

What is the distribution of Y under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?
B. Plot the function

$$l(\mu,\sigma;X) \equiv \log\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X)$$

as a function of μ.
C. Find the maximum value of the function $l(\mu;X)$ in B (as a function of μ) and the value of $\mu \equiv \hat{\mu}$ which achieves the maximum.
D. What is the distribution of $\hat{\mu}$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$? What is the distribution of $l(\hat{\mu};X)$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?

Solution: A. The likelihood ratio

$$\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(x) = \frac{\sigma^{-1}\phi((x-\mu)/\sigma)}{\sigma^{-1}\phi(x/\sigma)} = \frac{\exp(-(x-\mu)^2/(2\sigma^2))}{\exp(-x^2/(2\sigma^2))}$$

$$= \exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2}\frac{\mu^2}{\sigma^2}\right).$$

Hence

$$Y \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X) = \frac{\mu}{\sigma}\frac{X}{\sigma} - \frac{1}{2}\frac{\mu^2}{\sigma^2}.$$

Under $P_{0,\sigma}$ we find that $E(Y) = 0 - \frac{\mu^2}{2\sigma^2}$ and $Var(Y) = \mu^2/\sigma^2 \equiv V^2$ so that

$$Y \sim N(-\frac{1}{2}V^2, V^2) \qquad \text{under } \; P_{0,\sigma}.$$

Under $P_{\mu,\sigma}$ a similar computation gives $E(Y) = \mu^2/\sigma^2 - \mu^2/(2\sigma^2) = V^2/2$ and $Var(Y) = V^2$, so

$$Y \sim N(\frac{1}{2}V^2, V^2) \qquad \text{under } \; P_{\mu,\sigma}.$$

B and C. The function

$$l(\mu,\sigma;X) \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X) = \frac{\mu}{\sigma}\frac{X}{\sigma} - \frac{\mu^2}{2\sigma^2} = \frac{X^2}{2\sigma^2} - \frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}$$

is quadratic in μ with maximum value $X^2/(2\sigma^2)$ which is achieved at $\mu = \hat{\mu} \equiv X$.
D. Under $P_{0,\sigma}$, $\hat{\mu} = X \sim N(0,\sigma^2)$ and $l(\hat{\mu},\sigma;X) = X^2/(2\sigma^2) \sim \chi_1^2/2$. Under $P_{\mu,\sigma}$, $\hat{\mu} = X \sim N(\mu,\sigma^2)$ and $l(\hat{\mu},\sigma;X) = X^2/(2\sigma^2) \sim \chi_1^2(\delta)/2$ with $\delta = \mu^2/\sigma^2$.

4. Suppose that $X, X_1, X_2, \ldots, X_n$ are independent Poisson(λ) random variables:

$$P(X = k) \equiv p_k = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

Note that

$$\frac{p_k}{p_{k-1}} = \frac{\lambda}{k},$$

and hence whole family of alternative estimators $\{\tilde{\lambda}_n^{(k)}\}_{k\ge 1}$ is given by

$$\tilde{\lambda}_n^{(k)} = k\frac{\hat{p}_n(k)}{\hat{p}_n(k-1)}$$

where $\hat{p}_n(k) \equiv n^{-1}\sum_{i=1}^n 1_{[X_i = k]}$.
(a) Show that $\tilde{\lambda}_n \to_p \lambda$ for each $k = 1, 2, \ldots$.
(b) Show that

$$\sqrt{n}(\tilde{\lambda}_n^{(k)} - \lambda) \to_d N(0, \sigma_k^2(\lambda)) \quad \text{as} \quad n \to \infty$$

and compute $\sigma_k^2(\lambda)$ explicitly as a function of k and λ.
(c) What is the asymptotic relative efficiency of $\tilde{\lambda}_n^{(k)}$ to $\widehat{\lambda}_n = \overline{X}_n$ for $k > 1$?

Solution: (a) First, $(\hat{p}_n(k-1), \hat{p}_n(k)) \to_p (p_k, p_{k-1})$ by the WLLN. The function $g: R^2 \to R$ defined by $g(u,v) = v/u$ is continuous at (u,v) with $u \ne 0$. Hence it follows from the Mann-Wald (or continuous mapping) theorem that

$$g(\hat{p}_n(k-1), \hat{p}_n(k)) \to_p g(p_{k-1}, p_k) = \frac{\lambda^k e^{-\lambda}/k!}{\lambda^{k-1}e^{-\lambda}/(k-1)!} = \frac{\lambda}{k},$$

and hence

$$\tilde{\lambda}_n = kg(\hat{p}_n(k-1), \hat{p}_n(k)) \to_p kg(p_{k-1}, p_k) = \lambda.$$

(b) Note that $g(u,v)$ is differentiable at (u,v) with $u \neq 0$ and $\nabla g(u,v) = (-v/u, 1)/u$. Hence $\nabla g(p_{k-1}, p_k) = (-\lambda/k, 1)/p_{k-1}$. From the Multivariate CLT we have

$$\sqrt{n}\begin{pmatrix} \hat{p}_n(k-1) - p_{k-1} \\ \hat{p}_n(k) - p_k \end{pmatrix} \to_d N_2(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} p_{k-1}(1-p_{k-1} & -p_{k-1}p_k \\ -p_{k-1}p_k & p_k(1-p_k) \end{pmatrix}.$$

Hence it follows from the delta method (or g'-theorem) that

$$\sqrt{n}(g(\hat{p}_n(k-1), \hat{p}_n(k)) - g(p_{k-1}, p_k)) \to_d N(0, \nabla g \Sigma (\nabla g)^T) = N(0, A^2)$$

and we can easily calculate

$$A^2 \equiv A^2(k, \lambda) = \nabla g \Sigma (\nabla g)^T = \frac{\lambda}{k}\left(1 + \frac{\lambda}{k}\right)\frac{1}{p_{k-1}}.$$

Thus it follows that

$$\begin{aligned} \sqrt{n}(\tilde{\lambda}_n - \lambda) &= k\sqrt{n}(g(\hat{p}_n(k-1), \hat{p}_n(k)) - g(p_{k-1}, p_k)) \\ &\to kN(0, A^2) = N(0, k^2A^2). \end{aligned}$$

(c) It follows from (b) that the ARE of $\tilde{\lambda}_n^{(k)}$ with respect to $\widehat{\lambda}_n = \overline{X}_n$ is

$$ARE(\tilde{\lambda}_n^{(k)}, \widehat{\lambda}_n) = \frac{\lambda}{k\lambda(1+\lambda/k)/p_{k-1}} = \frac{p_{k-1}}{\lambda + k}.$$

See Figure 1 below for a set of plots showing that these estimators get worse as k grows; $k = 1$ is the only one which has an ARE anywhere close to 1.

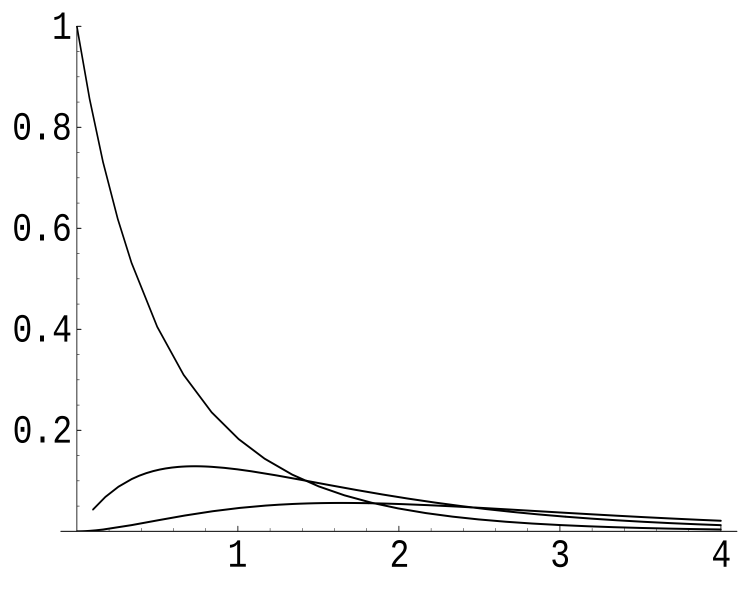

Figure 1: ARE for $k = 1, 2, 3$.