Statistics 583, Problem Set 7

Wellner; 5/11/2016

Reading: Chapter 8, sections 8.1- 8.4;
van der Vaart, *Asymptotic Statistics*, chapter 23, pages 326 - 340;
Wasserman, Chapters 2-3, pages 13-41.

Due: Wednesday, May 18, 2016

1. Let $T(P) \equiv \int\int h(x,y)dP(x)dP(y)$ for a fixed function $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with $\int\int |h(x,y)|dP(x)dP(y) < \infty$. The corresponding estimator $T(\mathbb{P}_n)$ is a $V-$statistic, and the natural unbiased estimator is

$$U_n = \frac{1}{\binom{n}{2}} \sum\sum_{1\le i<j\le n} h(X_i, X_j).$$

   (a) Show that $T(\mathbb{P}_n)$ is a biased estimator of $T(P)$ and compute the bias.
   (b) Find the influence function of $T(P)$.
   (c) What do you expect for the asymptotic variance of $\sqrt{n}(T(\mathbb{P}_n) - T(P))$?
   (d) What is the Hájek projection of U_n? How does it relate to the influence function you calculated in (b)?
   (e) How does the result in (c) compare with the limiting distribution of $\sqrt{n}(U_n - T(P))$?
   Hint: see van der Vaart, *Asymptotic Statistics*, section 12.1, pages 161 - 163.

2. Let $T(F) = \int(F - F_0)^2 dF_0$ Find the first and second order Gateaux derivatives of $T(F)$ at $F = F_0$. What limit distribution do you expect for $nT(\mathbb{F}_n)$?

3. Let F be a bivariate distribution function and define $T(F) = \int \varphi(F_1, F_2)dF_2$ if F_1 and F_2 are the (one-dimensional) marginal distribution functions of F and $\varphi : [0,1]^2 \to \mathbb{R}$ is a smooth fixed function.
   (a) Find the influence function of T.
   (b) Write out $T(\mathbb{F}_n$ where $\mathbb{F}_n$ is the bivariate empirical distribution function of $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d. as F.
   (c) What asymptotic variance do you expect for $\sqrt{n}(T(\mathbb{F}_n) - T(F))$?

4. (a) Given n distinct data items, show that the probability that a given data item does not appear in a bootstrap sample is $e_n = (1 - 1/n)^n$
   (b) Show that $e_n \to e^{-1} \approx .368$ as $n \to \infty$.
   (c) Hence show that the probability that each of B bootstrap samples contains an item i is $(1 - e_n)^B$. Evaluate this quantity for $n = 10, 20, 50, 100$ and $B =$

$10, 20, 50, 100$.

(d) Let $N_n \equiv \sum_{j=1}^{n} 1_{[M_j=0]}$ where $\underline{M} \equiv (M_1, \ldots, M_n) \sim \text{Mult}_n(n, \underline{1}/n)$. Show that $E(n^{-1}N_n) = e_n$ as computed in (a).

5. **Optional bonus problem 1:** Show that with N_n as defined in part (d) of the previous problem we have $\sqrt{n}(n^{-1}N_n - (1-1/n)^n) \to_d N(0, e^{-1}(1-2e^{-1}))$.

6. **Optional bonus problem 2:** Suppose that we observe $X_1, \ldots, X_n$ i.i.d. P on $\mathbb{R}^+ = [0, \infty)$ and assume that $P \in \mathcal{P}_0 \equiv \{P_\theta : (dP_\theta/d\lambda) = p_\theta, \ \theta \in \Theta\}$ where $\theta = (\alpha, \beta) \in (0, \infty)^2$ and $p_\theta = p_{\alpha,\beta}$ is the Weibull density given by $p_\theta(x) = (\beta/\alpha)(x/\alpha)^{\beta-1} \exp(-(x/\alpha)^\beta) 1_{(0,\infty)}(x)$. From Lehmann and Casella, TPE, Example 6.6.1 (page 468) and problems 6.6.1 - 6.6.3 (page 509), we know that the maximum likelihood estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ exists and is unique if $0 < X_{(1)} < X_{(n)}$.

(a) If, in fact, $P \notin \mathcal{P}_0$, to what function of P do you expect $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ converges (in probability)? [Hint: use the development in example 6.6.1 of Lehmann and Casella, TPE, page 468, to show that the solution of the population version of the score equations (with sampling from $P \notin \mathcal{P}$) leads to $\alpha(P) = \{E_P(X^\beta)\}^{1/\beta}$ where β is the solution of

$$\frac{E_P(X^\beta \log X)}{E_P X^\beta} - \frac{1}{\beta} = E_P(\log X),$$

assuming that $E_P(X^\beta |\log X|) < \infty$.]

(b) Show heuristically that $\theta(P) = (\alpha(P), \beta(P))$ minimizes $K(P, P_\theta)$ over Θ.

(c) In particular, if P has "half-normal" density $p(x) = 2\phi(x)1_{(0,\infty)}(x)$ where ϕ is the standard normal density, find $(\alpha, \beta) = (\alpha(P), \beta(P))$ corresponding to the "best-fitting" member of the Weibull family $P_{(\alpha(P),\beta(P))}$. Plot both p and $p_{(\alpha(P),\beta(P))}$ as functions of x.