Statistics 583, Problem Set 7

Wellner; 5/11/2011

Reading: Chapter 8, sections 8.4 - 8.5, pages 18 - 27;
Wasserman, Chapter 3, pages 13-41; begin reading Chapter 4
Due: Wednesday, May 18, 2011

1. The expression for the jackknife variance estimator for the median, in the display (1) on page 11 (3rd line from the bottom) in chapter 8 was derived under the assumption $n = 2m$ and that $T(\mathbb{F}_n) = X_{(m)}$ if $n = 2m - 1$, $T(\mathbb{F}_n) = (X_{(m)} + X_{(m+1)})/2$ if $n = 2m$.
(a) Derive the first equality in (1), page 11, using this definition of the sample median.
(b) Derive versions of the development in (1), page 11, using $T(F) = F^{-1}(1/2)$ (strictly). Does the asymptotic result in (1) still hold? Here is some further explanation of what I mean by "strictly" here: let $T_1(\mathbb{F}_n) = X_m$ if $n = 2m - 1$, $T_1(\mathbb{F}_n) = (X_{(m)} + X_{(m+1)})/2$ if $n = 2m$. This is one common definition of the median, and this is the definition used in (a). Let $T_2(\mathbb{F}_n) = \mathbb{F}_n^{-1}(1/2)$. This is my favorite definition of the median. Note that $T_2(\mathbb{F}_n) = T_1(\mathbb{F}_n)$ if $n = 2m - 1$, but $T_2(\mathbb{F}_n) \neq T_1(\mathbb{F}_n)$ if $n = 2m$. (What is the value of $T_2(\mathbb{F}_n)$ in this case?) T_2 is the definition of the median to be considered in 2(b)!

2. (a) Wasserman, problem 3.8.3, page 39, modified. Show that the claimed expression for v_{boot} given in the display for this problem is incorrect and find the correct expression. Here $v_{boot} = Var_{\mathbb{F}_n}(T_n)$ where $T_n = \overline{X}_n^2$. [Hint: see Dodd and Korn, *The American Statistician* **61** (2007), 127 - 131, and especially their appendix B, pages 130-131. Apparently the formula given by Wasserman in his problem is from Shao and Tu (1995), page 10; as noted by Dodd and Korn, the expression in Shao and Tu is incorrect.]
(b) Explain how the resulting formulas relate to how you would estimate the variance of $\overline{X}_n^2$ via the delta method.

3. Wasserman, problem 3.8.7, page 40.

4. Wasserman, problem 3.8.11, page 41.

5. **Optional bonus problem:** Suppose that $\{S_n\}$ and $\{T_n\}$ are arbitrary sequences of statistics (with $E(S_n^2) < \infty$ and $E(T_n^2) < \infty$ for each n) satisfying

$$\frac{Var(S_n - T_n)}{Var(T_n)} \to 0. \qquad (1)$$

(a) Show that (1) implies that $Var(S_n)/Var(T_n) \to 1$.
(b) Show that $2Cov(S_n, T_n) = Var(S_n) + Var(T_n) - Var(S_n - T_n)$.
(c) Let

$$R_n \equiv \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} - \frac{T_n - E(T_n)}{\sqrt{Var(T_n)}}.$$

Use the identity in (b) to show that

$$\begin{aligned} Var(R_n) &= 2 - 2\frac{Cov(S_n, T_n)}{\sqrt{Var(S_n)Var(T_n)}} \\ &= 2 - \sqrt{\frac{Var(S_n)}{Var(T_n)}} - \sqrt{\frac{Var(T_n)}{Var(S_n)}} + \sqrt{\frac{Var(T_n)}{Var(S_n)}} \cdot \frac{Var(S_n - T_n)}{Var(T_n)}. \end{aligned}$$

(d) Use the result of (a) together with the identity you proved in (c) to show that (1) implies $Var(R_n) \to 0$.
(e) Does the conclusion of (d) imply $R_n \to_p 0$?

6. **Optional bonus problem:** (Hard!) On page 12, line 4 of Chapter 8 of the lecture notes, it is claimed that if $E_F|X|^r < \infty$ for some $r > 0$, then for the median function $T(F) = F^{-1}(1/2)$ we have

$$nVar_F(T(\mathbb{F}_n)) \to \frac{1/4}{f^2(F^{-1}(1/2))}.$$

Prove (or disprove) this claim.