Statistics 581, Problem Set 6

Wellner; 11/1/2017

Reminder: Midterm exam: Monday, November 3.
Reading: Lecture Notes Chapter 3, sections 1-2;
Ferguson, ACILST, chapters 19-20, pages 126 - 139;
vdVaart, Asym. Stat., sections 8.1-8.3, pages 108-112.

Due: Wednesday, November 8, 2017.

1. Suppose that $X_1, \ldots, X_n$ are i.i.d. with distribution function F which has positive density f at its quartiles $F^{-1}(1/4)$ and $F^{-1}(3/4)$ and at its median $F^{-1}(1/2)$.
(a) Let $M_n = \mathbb{F}_n^{-1}(1/2)$ be the median and $Q_n = \mathbb{F}_n^{-1}(3/4) - \mathbb{F}_n^{-1}(1/4)$, the interquartile range. Find the joint asymptotic distribution of M_n and Q_n as estimators of the population median $m = M(F) = F^{-1}(1/2)$ and interquartile-quartile range $q = Q(F) = F^{-1}(3/4) - F^{-1}(1/4)$. That is, prove that
$$\sqrt{n}(M_n - m, Q_n - q)^T \to_d \text{“something”}$$
and find “something”.
(b) Assuming that the underlying distribution F is Cauchy(μ, σ) ($X = \sigma Y + \mu$ where $Y \sim$ Cauchy$(0, 1)$), express μ and σ in terms of $m = M(F)$ and $q = Q(F)$.
(c) Use the relations you derived in (b) to propose estimators of μ and σ based on M_n and Q_n. Show that your estimators $\hat{\mu}_n$ and $\hat{\sigma}_n$ satisfy
$$\sqrt{n}(\hat{\mu}_n - \mu, \hat{\sigma}_n - \sigma) \to_d N_2(0, \Sigma)$$
and compute Σ.
Hint: A standard Cauchy random variable has d.f. $G(x) = 1/2 + (1/\pi)\text{arctan}(x)$ and inverse distribution function $G^{-1}(u) = \tan(\pi(u - 1/2))$.

2. (a) Let $\mathbb{U}_X$ and $\mathbb{U}_Y$ be two independent Brownian bridge processes on $[0, 1]$, and let $\lambda \in [0, 1]$. Show that the process $\mathbb{U}$ defined by $\mathbb{U} = \sqrt{1-\lambda}\,\mathbb{U}_X - \sqrt{\lambda}\,\mathbb{U}_Y$ is also a Brownian bridge process.
(b) Suppose that $X_1, \ldots, X_m$ are i.i.d. F and $Y_1, \ldots, Y_n$ are i.i.d. G, with the X's and Y's independent. Let $\mathbb{F}_m$ and $\mathbb{G}_n$ denote the empirical df's of the X's and Y's respectively. Suppose that $\lambda_N \equiv m/N \to \lambda \in (0, 1)$ where $N \equiv m + n$. Show that
$$\begin{aligned} \mathbb{X}_{m,n} &\equiv \sqrt{\frac{mn}{N}}(\mathbb{F}_m - F) - \sqrt{\frac{mn}{N}}(\mathbb{G}_n - G) \\ &\Rightarrow \sqrt{1-\lambda}\,\mathbb{U}_X(F) - \sqrt{\lambda}\,\mathbb{U}_Y(G) \equiv \mathbb{X}. \end{aligned}$$
(c) A distribution function F is said to be *stochastically smaller than another distribution function* G, and we write $F <_s G$, if $F(x) \ge G(x)$ for all $x \in \mathbb{R}$ with strict inequality for some $x \in \mathbb{R}$. Note that this means $F^{-1}(u) \le G^{-1}(u)$ for all $0 < u < 1$ so that the random variables resulting from a Skorokhod construction with one uniform random variable ξ satisfy satisfy $X^* \equiv F^{-1}(\xi) \le G^{-1}(\xi) \equiv Y^*$.

Consider testing $H_0: \ F = G$ continuous versus $H_1: \ F <_s G$ based on the one-sided Kolmogorov-Smirnov statistic

$$D^+_{m,n} \equiv \sqrt{\frac{mn}{N}} \|(\mathbb{F}_m - \mathbb{G}_n)^+\|_\infty = \sqrt{\frac{mn}{N}} \sup_{x \in \mathbb{R}} (\mathbb{F}_m(x) - \mathbb{G}_n(x));$$

here the notation f^+ is the *positive part* of the function f: $f^+(x) \equiv \max\{f(x), 0\}$. Use the result of B to show that under H_0 it follows that

$$D^+_{m,n} \to_d \|\mathbb{U}^+\|_\infty = \sup_{0 \le t \le 1} \mathbb{U}(t).$$

(d) To test the effectiveness of vitamin B_1 in stimulating growth in mushrooms, vitamin B_1 was applied to 13 mushrooms selected at random from a group of 24, while the remaining 11 did not receive this treatment. The weights of the mushrooms at the end of the period of observation were:

$\underline{X} = (18, 14.5, 13.5, 12.5, 23, 24, 21, 17, 18.5, 9.5, 14)$, $m = 11$;
$\underline{Y} = (27, 34, 20.5, 29.5, 20, 28, 20, 26.5, 22, 24.5, 34, 35.5, 19)$, $n = 13$.

Plot the two empirical df's and compute $D^+_{m,n}$. What is the approximate P - value for testing H_0 versus $H_1: \ F <_s G$? You may use your favorite tables of the distribution of $D^+_{m,n}$, or the asymptotic distribution.

3. A. Compute and plot the score for location $-f'/f(x)$ when:
(a) $f = \phi$, the standard normal density;
(b) $f(x) = \exp(-x)/(1 + \exp(-x))^2$ (logistic);
(c) $f(x) = (1/2)\exp(-|x|)$ (double exponential);
(d) $f(x) = t_k$, the $t-$density with $k-$degrees of freedom;
(e) $f(x) = \exp(-x)\exp(-\exp(-x))$; (f) $f(x) = \phi(x)\Phi(ax)$ where $\Phi(x)$ is the standard normal distribution function and $a > 0$.
B. A density f is called *log-concave* if $\log f$ is a concave function. Which of the densities in (a) - (e) are log-concave?

4. **Optional bonus problem 1:** Chapter 2, Exercise 6.3, page 35. Show that the partial sum process $\mathbb{S}_n$ defined in terms of i.i.d. X_i's with $E(X_1) = 0$ and $E(X_1^2) = 1$ satisfies $\mathbb{S}_n \to_{f.d.} \mathbb{S}$ where $\mathbb{S}$ is standard Brownian motion on $[0, \infty)$.
Hint: One approach uses 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.

5. **Optional bonus problem 2:** For the one-sample empirical distribution function we know that $P(\sup_{0 \le t \le 1} \mathbb{U}_n(t) > \lambda) \le \exp(-2\lambda^2)$; This refinement of the Dvoretzky-Kiefer-Wolfowitz inequality is due to Massart (1990). What is known about inequalities of this same type in the case of the two-sample Kolmogorov statistics in problem 3 above?