Statistics 610 fall 2014
Homework # 3
Due: Thursday 25 September

[3.1] In class I discussed the Hodges estimator, $T_n := \overline{x}_n \mathbf{1}\{|\overline{x}_n| > \alpha_n\}$, where $\overline{x}_n = n^{-1}\sum_{i\le n} x_i$ and the x_i's are iid $N(\theta, 1)$ under the $\mathbb{P}_\theta$ model. Under squared-error loss, the risk function $R(\theta, T_n) := \mathbb{E}_\theta(T_n - \theta)^2$ beats the efficiency lower bound, in the sense that $\lim_{n\to\infty} nR(\theta, T_n) = 1$ for all $\theta \ne 0$ and $\lim_{n\to\infty} nR(0, T_n) = 0$. However, for finite n, the estimator T_n can do worse than the efficient estimator $\overline{x}_n$ at some values of θ near zero. For this Problem I want you to generate (maybe using R) some pictures that illustrate this phenomenon.

(i) Explain why $\overline{x}_n$ is efficient (as an estimator of θ) in the Fisherian sense.

(ii) Reparametrize by writing β for $\sqrt{n}\alpha_n$ and h for $\theta/\sqrt{n}$. (It also helps to note that $Z_n = \sqrt{n}(\overline{x}_n - \theta)$ has a $N(0, 1)$ distribution under each $\mathbb{P}_\theta$.) Express $nR(\theta, T_n) - nR(\theta, \overline{x}_n)$ as a function, $B(h, \beta)$, of h and β. (Note: there is no dependence on n.) Draw plots of $B(h, \beta)$ against h, for various (informative) values of β.

[3.2] In the vanTrees.pdf handout I stated the VT inequality for a family of densities $\{p(x, t) : t \in \Theta\}$, with $\Theta \subset \mathbb{R}$:

$$\int_\Theta q(t)\mathbb{E}_t |T(x) - \psi(t)|^2 \, dt \ge \frac{\left(\int \dot{\psi}(t)q(t)\, dt\right)^2}{\mathbb{I}_q + \int \mathbb{I}_p(t)q(t)\, dt}$$

but only proved it for the special case where $\psi(t) = t$. Show which parts of the argument on the handout need to be changed to get the more general result. (The LaTeX source file vanTrees.tex is in the handouts directory if you need it.)

[3.3] (Compare with Hall 1989) For each constant $K > 1$ define $\mathcal{F}_K$ to be the set of continuous probability densities on $(-1/2, +1/2)$ satisfying the constraints: $K^{-1} \le f(z) \le K$ and $|\dot{f}(z)| \le K$ and $|\ddot{f}(z)| \le K$ for $|z| < 1/2$. Consider the estimation of the value $f(0)$ based on independent observations $x_1, \dots, x_n$ from f. In this Problem you will show there is a constant $C = C_K$ for which

<1> $$\sup_{f\in\mathcal{F}_K} \mathbb{E}_f|T_n(x) - f(0)|^2 \ge Cn^{-4/5}$$

for every estimator $T_n(x) = T_n(x_1, \dots, x_n)$ of $f(0)$. Follow these steps.

(i) Suppose g is an infinitely differentiable function defined on the real line with the properties: (a) $g(z) = 0$ for $|z| \ge 1$; (b) $g(0) = 1$; and (c) $\int_{-1}^{+1} g(z)\, dz = 0$. Define

$$\zeta_h(z) = 1 + ct^2 g(z/t) \qquad \text{for } -1/2 < z < 1/2.$$

Show that $\{\zeta_t : |t| < \delta\} \subset \mathcal{F}_K$ if the constants c and δ are chosen small enough.

(ii) Show that $\mathbb{I}_\zeta(t) := \int \overset{\bullet}{\zeta}_t(z)^2/\zeta_t(z)\,dz \le C_1|t|^3$ if $|t| < \delta$, some constant C_1 (depending on K and δ and g).

(iii) Define $\gamma_n = n^{1/5}$. Let q be a smooth density on $\mathbb{R}$ with $q(t) = 0$ for $|t| \ge 1$. Define $q_n(t) = \gamma_n q(\gamma_n t)$ and $p_n(x,t) = \prod_{i \le n} \zeta_t(x_i)$ and $\psi(t) = \zeta_t(0)$. Write $\mathbb{E}_t$ for expectations with respect to the $p_n(\cdot,t)$ density.

Apply the version of the VT inequality from Problem [3.2] with q replaced by q_n and p replaced by p_n to show that

$$\int q_n(t)\mathbb{E}_t|T_n(x) - \zeta_t(0)|^2 \ge c/\left(\gamma_n^4 + n\gamma_n^{-1}\right)$$

for some constant $c > 0$.

(iv) Complete the proof of inequality <**1**>.

References

Hall, P. (1989). On convergence rates in non-parametric problems. *International Statistical Review 57*, 45–58.