Lection 8. Rank Estimators

α-trimmed mean

Problem 1. Prove that

$$\sigma^2(F) \leq \frac{\sigma^2}{(1-2\alpha)^2},$$

where $\sigma^2 = \mathbf{D}X$ and the inequality becomes an equality for some F.

So, the α-trimmered mean is a good robust alternative to the sample mean. However, as the asymptotic breakdown point of X_α equals $1/3$, $\sigma^2(F)$ can be 9 times larger than σ^2.

Rank-estimators

Definition 1. An estimator $f(F)$ is an *R-estimator* if it's a unique solution of the equation

$$\int_{\mathbb{R}} J\left(\frac{1}{2}\left(F(x)+1-F(2f(F)-x)\right)\right) dF(x) = 0, \tag{1}$$

where J is given function, $J(1-x) = -J(x)$.

Lemm 1. *If $F \in F_{symm}$, then $f(F) = \theta$*

Proof.

$$\int_{\mathbb{R}} J\left(\frac{1}{2}\left(F(x)+1-F(2\theta-x)\right)\right) dF(x) = \int_{-\infty}^{\theta} J\left(\frac{1}{2}\left(F(x)+1-F(2\theta-x)\right)\right) dF(x) + \int_{\theta}^{\infty} J\left(\frac{1}{2}\left(F(x)+1-F(2\theta-x)\right)\right) dF(x).$$

Let $u = 2\theta - x$, then

$$\int_{\theta}^{\infty} J\left(\frac{1}{2}\left(F(x)+1-F(2\theta-x)\right)\right) dF(x) = \int_{-\infty}^{\theta} J\left(\frac{1}{2}\left(F(2\theta-u)+1-F(u)\right)\right) dF(u) =$$
$$\int_{-\infty}^{\theta} J\left(1-\frac{1}{2}\left(1-F(2\theta-u)+F(u)\right)\right) dF(u) = -\int_{-\infty}^{\theta} J\left(\frac{1}{2}\left(1-F(2\theta-u)+F(u)\right)\right) dF(u)$$

Therefore,

$$\int_{\mathbb{R}} J\left(\frac{1}{2}\left(F(x)+1-F(2\theta-x)\right)\right) dF(x) = 0.$$

□

Lemm 2. *R-estimator $f(\widehat{F}_n)$ is asymptotically normal for θ with the asymptotical variance*

$$\sigma_F^2 = \frac{1}{\left(\int_{\mathbb{R}} J'(F(x))p(x)^2 dx\right)^2} \int_0^1 J^2(u)du.$$

Proof. Let's find the influence function of R-estimate. Let $F_\varepsilon(u) = F(u) + \varepsilon(\delta_x - F(u))$. Differentiating (1) with respect to ε at $\varepsilon = 0$, we get

$$\left(\int_{\mathbb{R}} J\left(\frac{1}{2}\left(F_\varepsilon(u)+1-F_\varepsilon(2f(F_\varepsilon)-u)\right)\right) dF_\varepsilon(u)\right)' = \int_{\mathbb{R}} J\left(\frac{1}{2}\left(F(u)+1-F(2f(F)-u)\right)\right) d(\delta_x(u)-F(u)) +$$
$$\int_{\mathbb{R}} J'\left(\frac{1}{2}\left(F(u)+1-F(2f(F)-u)\right)\right) \frac{1}{2}\left(\delta_x(u)-F(u)-(\delta_x-F)(2f(F)-u)-p(2f(F)-u)2f'(F)\right).$$

Since $F(2\theta - u) = 1 - F(u)$, we have

$$\int_{\mathbb{R}} J\left(\frac{1}{2}(F(u)+1-F(2f(F)-u))\right) d\delta_x(u) = J\left(\frac{1}{2}(F(x)+1-F(2f(F)-x))\right) = J(F(x)),$$

$$\int_{\mathbb{R}} J\left(\frac{1}{2}(F(u)+1-F(2f(F)-u))\right) dF(u)) = 0,$$

$$\int_{\mathbb{R}} J'\left(\frac{1}{2}(F(u)+1-F(2f(F)-u))\right) p(2f(F)-u)dF(u) = \int_{\mathbb{R}} J'(F(u))p(u)^2du.$$

Substituting $2f(F) - u$ for v, we get

$$\int_{\mathbb{R}} J'\left(\frac{1}{2}(F(u)+1-F(2f(F)-u))\right)((\delta_x - F)(u) - (\delta_x - F)(2f(F)-u))dF(u) =$$

$$\int_{\mathbb{R}} J'\left(\frac{1}{2}(F(2f(F)-v)+1-F(v))\right)((\delta_x - F)(2f(F)-v) - (\delta_x - F)(v))dF(2f(F)-v) =$$

$$-\int_{\mathbb{R}} J'\left(\frac{1}{2}(F(v)+1-F(2f(F)-v))\right)((\delta_x - F)(v) - (\delta_x - F)(2f(F)-v))dF(v) = 0.$$

Therefore,

$$I_F(f) = \frac{J(F(x))}{\int_{\mathbb{R}} J'(F(x))p(x)^2dx}.$$

Integrating $I_F(f)^2$ with respect to dF we prove the lemm. □

The Hodges-Lehmann estimator

Consider the functional $f(F) = med(F * F)/2$, where $*$ denotes the convolution. For $F \in \mathcal{F}_{Symm}$ $f(F) = med = \theta(F)$. Therefore, $f(\hat{F}_n)$ is a natural estimator for the $\theta(F)$.

Definition 2. The estimator $W = med(\hat{F}_n * \hat{F}_n)/2 = MED((X_i + X_j)/2, i, j = 1, ..., n)$ is called the *Hodges-Lehmann estimator.*

Problem 2. Prove that W is R-estimate with $J(x) = x - 1/2$.

Therefore,

$$L_F(f) = \frac{F(x) - 1/2}{\int_{\mathbb{R}} p(x)^2dx}$$

and

$$\sigma^2(F) = \int_{\mathbb{R}} L_F(x; f)^2 dF(x) = \frac{\int_{\mathbb{R}}(2F(x)-1)^2dF(x)}{4\left(\int_{\mathbb{R}} p(x)^2dx\right)^2} = \frac{1}{4\left(\int_{\mathbb{R}} p(x)^2dx\right)^2}$$

In Hettmansperger, 1987 it was shown that

$$\frac{\sigma^2(F; \overline{X})}{\sigma^2(F; W)} = 12\left(\int_{\mathbb{R}} p(x)^2dx\right)^2 \int_{\mathbb{R}} x^2p(x)dx \geq \frac{108}{125}.$$

Therefore, the Hodges-Lehmann estimator is a nice alternative to the sample mean. Next time we prove that it's robust estimator.

Example 1. Let's compute the asymptotic breakdown point for W. Obviously, as $F(x) > \frac{1}{\sqrt{2}}$ we have $(F * F)(2x) > 1/2$. Since for every $G \in U_\delta(F)$ we have

$$G(x) \geq F(x - \delta) - \delta \geq F(x - \delta) - \delta > \frac{1}{\sqrt{2}}$$

as $x > \delta + F^{-1}\left(\delta + \frac{1}{\sqrt{2}}\right)$. Therefore, $f(G) < \infty$ for every $G \in U_{1-\frac{1}{\sqrt{2}}}(F)$. Similarly, $f(G) > -\infty$ for every $G \in U_{1-\frac{1}{\sqrt{2}}}(F)$. Since $\varepsilon^*(F) \geq 1 - \frac{1}{\sqrt{2}} \approx 0.29$.

Consider $U_\delta(F)$, where $\delta = 1 - \frac{1}{\sqrt{2}} + \varepsilon$ and consider

$$F_n(x) = \begin{cases} 0, & x < -n, \\ \delta, -n \le x \le F_0^{-1}(1 - \frac{1}{\sqrt{2}} + \varepsilon), \\ F_0(x), & x \ge F_0^{-1}(\delta). \end{cases}$$

Then, $\sup_{x\in\mathbb{R}} |F_n(x) - F_0(x)| \le \delta$, so $F_n \in U_\delta(F_0)$. If $X_1^{(n)}$, $X_2^{(n)} \sim F_n$ i.i.d, then

$$\mathbf{P}(X_1^{(n)} + X_2^{(n)} > -n) \le \mathbf{P}(X_1^{(n)} \le -n, X_2^{(n)} > n/2) + \mathbf{P}(X_1^{(n)} > n/2, X_2^{(n)} \le -n) + \mathbf{P}(X_1^{(n)} > -n, X_2^{(n)} > -n).$$

Since $\mathbf{P}(X_1^{(n)} > n/2) = 1 - F_0(n/2) \to 0$, $n \to \infty$ and $\mathbf{P}(X_1^{(n)} > -n, X_2^{(n)} > -n) = (1-\delta)^2$, we have

$$\liminf_{n\to\infty} (F_n * F_n)(-n/2) = \liminf_{n\to\infty} \mathbf{P}(X_1^{(n)} + X_2^{(n)} \le -n/2) \ge 1 - (1-\delta)^2 = 1 - \left(\frac{1}{\sqrt{2}} - \varepsilon\right)^2 > \frac{1}{2}.$$

Thus

$$\limsup_{n\to\infty} f(F_n) = \limsup_{n\to\infty} (F_n * F_n)^{-1}(1/2)/2 \le -\liminf_{n\to\infty} n/4 = -\infty.$$

Therefore, $\varepsilon^*(F) = 1 - \frac{1}{\sqrt{2}} \approx 0.29$.