Advanced Models B 52805 (Midterm quiz, 2022)

P. Chigansky

- Exam duration is 90 min.
- All questions have the same weight.
- Any printed/handwritten material can be used.
- All means of communications are strictly prohibited, incl. calculators, etc.
- Answer the questions in a clear way, dubious solutions will not be given credit.

Problem

Consider the first centeral absolute moment functional

$$T(F)=\int_0^1 |x-\mu(F)|dF(x)$$

on the subset of continuous distributions on the interval $[0,1]$, where

$$\mu(F)=\int_0^1 xdF(x)$$

is the mean functional.

מכאן משתמע כי $1 \le x \le 0$ וגם $1 \le \mu(F) \le 0$

Note: In questions (3)-(5) you may find useful the integration by parts formula for the Lebesgue–Stieltjes integrals

$$\int_a^b h(x)dg(x)=h(x)g(x)\Big|_a^b-\int_a^b g(x)dh(x).$$

(1) Specify the plug-in estimator $T(\widehat{F}_n)$ and evaluate it at the sample $\{\frac{1}{2},\frac{1}{4},\frac{3}{4}\}$.

The plug-in estimator is

$$T(\widehat{F}_n)=\frac{1}{n}\sum_{j=1}^n |X_j-\overline{X}_n|.$$

For the given sample, $\overline{X}_n=\frac{1}{2}$ and hence the functional yields

$$\tfrac{1}{3}|\tfrac{1}{2}-\tfrac{1}{2}|+\tfrac{1}{3}|\tfrac{1}{4}-\tfrac{1}{2}|+\tfrac{1}{3}|\tfrac{3}{4}-\tfrac{1}{2}|=\tfrac{1}{6}.$$

(2) Prove that $T(\widehat{F}_n)$ is consistent.

הצעת פתרון בסוף העמוד

הצעה נוספת

Hint: use the following instance of the triangular inequality

$$\big||x-a|-|x-b|\big|\le |b-a|,\quad \forall x,a,b\in\mathbb{R}.$$

Note that

$$\big|T(\widehat{F}_n)-T(F)\big|=\Big|\frac{1}{n}\sum_{j=1}^n |X_j-\overline{X}_n|-T(F)\Big|\le$$

$$\Big|\frac{1}{n}\sum_{j=1}^n |X_j-\overline{X}_n|-\frac{1}{n}\sum_{j=1}^n |X_j-\mu(F)|\Big|+\Big|\frac{1}{n}\sum_{j=1}^n |X_j-\mu(F)|-T(F)\Big|.$$

The last term vanishes $\mathbb{P}_F$-a.s. as $n \to \infty$ by the LLN. It remains to argue that the first term vanishes as well. To this end, use the hinted inequality

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{j=1}^{n}|X_j-\overline{X}_n|-\frac{1}{n}\sum_{j=1}^{n}|X_j-\mu(F)|\right| \le \\
&\frac{1}{n}\sum_{j=1}^{n}\Big||X_j-\overline{X}_n|-|X_j-\mu(F)|\Big| \le \\
&\frac{1}{n}\sum_{j=1}^{n}|\overline{X}_n-\mu(F)| = |\overline{X}_n-\mu(F)| \xrightarrow[n\to\infty]{\mathbb{P}_F-a.s.} 0,
\end{aligned}$$

where the convergence is due to LLN.

(3) Show that the functional in question satisfies

$$T(F) = 2\int_0^{\mu(F)} F(x)dx.$$

כאן אנו מפצלים את הקטע $[0,1]$ לקטעים $[0,1]=[0,\mu(F)]\cup[\mu(F),1]$ לכן $[\mu(F),1]=[0,1]\backslash[0,\mu(F)]$

$$\begin{aligned}
&T(F) = \int_0^1 |x-\mu(F)|dF(x) = \\
&\int_0^{\mu(F)} (\mu(F)-x)dF(x) + \int_{\mu(F)}^1 (x-\mu(F))dF(x) = \\
&\int_0^{\mu(F)} (\mu(F)-x)dF(x) - \int_0^{\mu(F)} (x-\mu(F))dF(x) + \int_0^1 (x-\mu(F))dF(x) = \\
&2\int_0^{\mu(F)} (\mu(F)-x)dF(x) = 2(\mu(F)-x)F(x)\Big|_0^{\mu(F)} + 2\int_0^{\mu(F)} F(x)dx = \\
&2\int_0^{\mu(F)} F(x)dx
\end{aligned}$$

$|\int_0^1 x-\mu(F)dF = \mu(F)-\mu(F) = 0$

- באינטגרציה בחלקים אנו מביטים ב $\to 2\int_0^{\mu(F)} (\mu(F)-x)\cdot f_x dx$

אז $\begin{cases} v' = f_x & u = \mu(F)-x \\ v = F(x) & u' = -1 \end{cases}$

(4) Prove that T is Hadamard differentiable with the derivative

$$\dot{T}_F(G-F) = 2\int_0^{\mu(F)} \big(G(u)-F(u)\big)du + 2F(\mu(F))\int_0^1 ud\big(G(u)-F(u)\big)$$

for any distribution function G.

Let G_t be a curve of distributions such that $\|G_t - G\|_\infty \to 0$ for some distribution G. Denote $F_t := F + t(G-F)$, then

$$\begin{aligned}
&T(F_t)-T(F) = 2\int_0^{\mu(F_t)} F_t(x)dx - 2\int_0^{\mu(F)} F(x)dx = \\
&2\Big(\int_0^{\mu(F_t)} F_t(x)dx - \int_0^{\mu(F_t)} F(x)dx\Big) + 2\Big(\int_0^{\mu(F_t)} F(x)dx - \int_0^{\mu(F)} F(x)dx\Big) = \\
&2\int_0^{\mu(F_t)} \big(F_t(x)-F(x)\big)dx + 2\int_{\mu(F)}^{\mu(F_t)} F(x)dx.
\end{aligned}$$

נשים לב כי $0 \le \mu(F) \le \mu(F_t)$

מכיון ש $\mu(F_t) = \int_0^1 xd[F+t(F-G)] = \int_0^1 xdF + \int_0^1 txd(F-G)] = \mu(F) + some\_positive\_\_number$

גרסא אינטגרלית של משפט לגרנז

$$\int_{\mu(F)}^{\mu(F_t)} F(x)dx = F(\zeta)\cdot[\mu(F_t)-\mu(F)],\ \mu(F)\le\zeta\le\mu(F_t)$$

By continuity of the integral with respect to its limit

$$\frac{1}{t}\int_0^{\mu(F_t)}\big(F_t(x)-F(x)\big)dx = \int_0^{\mu(F_t)}\big(G(x)-F(x)\big)dx \xrightarrow[t\to 0]{} \int_0^{\mu(F)}\big(G(x)-F(x)\big)dx.$$

By continuity of F,

$$\frac{1}{t}\int_{\mu(F)}^{\mu(F_t)} F(x)dx = \frac{\mu(F_t)-\mu(F)}{t}\boxed{\frac{1}{\mu(F_t)-\mu(F)}\int_{\mu(F)}^{\mu(F_t)} F(x)dx} \xrightarrow[t\to 0]{} \int_0^1 xd\big(G(x)-F(x)\big)\boxed{F\big(\mu(F)\big)}.$$

Thus we get

$$\dot{T}_F(G-F) = 2\int_0^{\mu(F)}\big(G(u)-F(u)\big)du + 2F\big(\mu(F)\big)\int_0^1 ud\big(G(u)-F(u)\big).$$

(5) Show that the influence function is $L_F(x) = \psi_F(x) - \mathbb{E}_F\psi_F$ with

$$\psi_F(x) = 2\big(\mu(F)-x\big)\mathbf{1}_{\{x\le\mu(F)\}} + 2F\big(\mu(F)\big)x.$$

The first integral in the expression for $\dot{T}_F(G-F)$ can be written as

$$\int_0^{\mu(F)}\big(G(x)-F(x)\big)dx = \big(G(x)-F(x)\big)x\Big|_0^{\mu(F)} - \int_0^{\mu(F)} xd\big(G(x)-F(x)\big) =$$

$$G(\mu(F)) - F(\mu(F))\big)\mu(F) - \int_0^{\mu(F)} xd\big(G(x)-F(x)\big) =$$

$$\mu(F)\int_0^1 \mathbf{1}_{\{x\le\mu(F)\}}d\big(G(x)-F(x)\big) - \int_0^1 \mathbf{1}_{\{x\le\mu(F)\}}xd\big(G(x)-F(x)\big)$$

and hence

$$\dot{T}_F(G-F) = \int_0^1 2\Big(\big(\mu(F)-u\big)\mathbf{1}_{\{u\le\mu(F)\}} + F\big(\mu(F)\big)u\Big)d\big(G(u)-F(u)\big).$$

The claim follows by evaluating this expression at $\Delta_x(u) = \mathbf{1}_{\{u\ge x\}}$.

פתרון שאלה 2

$$\Big|\frac{1}{n}\sum_{i=1}^n|X_i-\overline{X_n}| - T(F)\Big| = \Big|\frac{1}{n}\sum_{i=1}^n|X_i-\overline{X_n}-\mu(F)+\mu(F)| - T(F)\Big|$$

$$\overset{\triangle}{\le}\Big|\frac{1}{n}\sum_{i=1}^n|X_i-\mu(F)| + \sum_{i=1}^n|\mu(F)-\overline{X_n}| - T(F)\Big| = \Big|\frac{1}{n}\sum_{i=1}^n|X_i-\mu(F)| - T(F) + \frac{1}{n}\sum_{i=1}^n|\mu(F)-\overline{X_n}|\Big|$$

$$\overset{\triangle}{\le}\Big|\frac{1}{n}\sum_{i=1}^n|X_i-\mu(F)| - T(F)\Big| + \frac{1}{n}\sum_{i=1}^n|\mu(F)-\overline{X_n}|$$

$$\Big|\frac{1}{n}\sum_{i=1}^n|X_i-\mu(F)| - T(F)\Big| \xrightarrow{L.L.N} 0$$

$$\frac{1}{n}\sum_{i=1}^n|\mu(F)-\overline{X_n}| = |\mu(F)-\overline{X_n}| \xrightarrow{L.L.N} 0$$

Appendix A. Construction of asymptotic confidence intervals

To construct $1-\alpha$ asymptotic confidence interval, one has to show that the plug-in estimator for the asymptotic variance $V_F = \mathrm{Var}_F(L_F)$ is consistent. This, in turn, amounts to showing that the plug-in estimators for the first two moments of ψ_F are consistent, i.e., $\mathbb{E}_{\widehat{F}_n} \psi_{\widehat{F}_n} \xrightarrow[n\to\infty]{\mathbb{P}_F} \mathbb{E}_F \psi_F$ and $\mathbb{E}_{\widehat{F}_n} \big(\psi_{\widehat{F}_n}\big)^2 \xrightarrow[n\to\infty]{\mathbb{P}_F} \mathbb{E}_F (\psi_F)^2$. Since the expression for ψ_F involves discontinuity, this takes some work. To demonstrate the point, consider convergence of the first moment,

$$
\begin{aligned}
\mathbb{E}_F \psi_F =& 2\int_0^1 \mu(F)\mathbf{1}_{\{x\le\mu(F)\}} dF(x) - 2\int_0^1 x\mathbf{1}_{\{x\le \mu(F)\}} dF(x) + 2F\big(\mu(F)\big)\mu(F) = \\
& -2\int_0^{\mu(F)} x dF(x) + 4F\big(\mu(F)\big)\mu(F) = 2\int_0^{\mu(F)} F(x)dx + 2F\big(\mu(F)\big)\mu(F).
\end{aligned}
$$

Obviously, $\mu(\widehat{F}_n) = \overline{X}_n \xrightarrow{\mathbb{P}_F-a.s.} \mu(F)$ and

$$
\begin{aligned}
|\widehat{F}_n(\mu(\widehat{F}_n)) - F(\mu(F))| \le& |\widehat{F}_n(\mu(\widehat{F}_n)) - F(\mu(\widehat{F}_n))| + |F(\mu(\widehat{F}_n)) - F(\mu(F))| \le \\
& \|\widehat{F}_n - F\|_\infty + |F(\mu(\widehat{F}_n)) - F(\mu(F))| \xrightarrow[n\to\infty]{\mathbb{P}_F-a.s.} 0,
\end{aligned}
$$

where the first term converges by G-C theorem and the second - by continuity of F. It remains to argue that the integral term converges as well,

$$
\begin{aligned}
& \Big| \int_0^{\mu(\widehat{F}_n)} \widehat{F}_n(x)dx - \int_0^{\mu(F)} F(x)dx \Big| \le \\
& \Big| \int_0^{\mu(\widehat{F}_n)} \widehat{F}_n(x)dx - \int_0^{\mu(\widehat{F}_n)} F(x)dx \Big| + \Big| \int_0^{\mu(\widehat{F}_n)} F(x)dx - \int_0^{\mu(F)} F(x)dx \Big| \le \\
& \int_0^{\mu(\widehat{F}_n)} |\widehat{F}_n(x) - F(x)|dx + \Big| \int_{\mu(F)}^{\mu(\widehat{F}_n)} F(x)dx \Big| \le \\
& \|\widehat{F}_n - F\|_\infty + \Big| \int_{\mu(F)}^{\mu(\widehat{F}_n)} F(x)dx \Big| \xrightarrow[n\to\infty]{\mathbb{P}_F-a.s.} 0.
\end{aligned}
$$

Convergence of the second moment can be argued along the same lines (try!) and thus

$$
V_{\widehat{F}_n} \xrightarrow[n\to\infty]{\mathbb{P}_F-a.s.} V_F.
$$

Then by Slutsky's lemma and the nonparametric Delta method theorem, the interval with the endpoints at

$$
T(\widehat{F}_n) \pm \frac{1}{\sqrt{n}} \sqrt{V_{\widehat{F}_n}} z_{1-\alpha/2}
$$

is an asymptotic $1-\alpha$ confidence interval.

הצעת פתרון שאלה 2

$$
\frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{X_n}| = \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{X_n} + \mu(F) - \mu(F)| \overset{\triangle}{\le} \frac{1}{n}\sum_{i=1}^{n} |x_i - \mu(F)| + \frac{1}{n}\sum_{i=1}^{n} |\overline{X_n} - \mu(F)|
$$

$$
= \frac{1}{n}\sum_{i=1}^{n} |x_i - \mu(F)| + \overline{X_n} - \mu_F \overset{L.L.N}{\rightarrow} E(|X_1 - \mu(F)|) = \int_0^1 |u - \mu(F)| du
$$