STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Apr 14

Lecturer: Xianyang Zhang

1 Donsker Class

A collection $\mathcal{F}$ of functions is called P-Donsker if the process $\{\sqrt{n}(P_n - P)f\}_{f\in\mathcal{F}}$ converges to a tight limit G indexed by $\mathcal{F}$ in $L^\infty(\mathcal{F})$. Here G is a gaussian process. In particular,

$$\left(\sqrt{n}(P_n - P)f_1, \cdots, \sqrt{n}(P_n - P)f_k\right) \to (G_{f_1}, \dots, G_{f_k})$$

and

$$\mathrm{cov}(G_{f_i}, G_{f_j}) = \mathrm{cov}(f_i(X), f_j(X)),$$

where $X \sim P$.

1.1 Example

Let $\Theta \subset \mathbb{R}^d$, where Θ is compact. Let

$$l_\theta(\cdot) : \Theta \times \mathcal{X} \to \mathbb{R}$$

with $l_\theta(\cdot)$ being $L(x)$-Lispschitz continuous in θ and $\mathbb{E}[L(X)^2] < \infty$. Then $\mathcal{F} = \{l_\theta(\cdot)\}_{\theta\in\Theta}$ is P-Donsker and

$$\left\{\sqrt{n}(P_n - P)l_\theta\right\}_{l_\theta\in\mathcal{F}} \xrightarrow{d} G_\theta.$$

where

$$\mathrm{cov}(G_{\theta_i}, G_{\theta_j}) = \mathrm{cov}(l_{\theta_i}(X), l_{\theta_j}(X))$$

for $X \sim P$.

1.2 Main theorem

Let $\mathcal{F}$ be a class of functions mapping from $\mathcal{X}$ to $\mathbb{R}$, and let F be an envelop function of $\mathcal{F}$, (i.e. for any $x \in \mathcal{X}$ and any $f \in \mathcal{F}$, $|f(x)| \le F(x)$). Suppose $PF^2 < \infty$ and

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\mathcal{F}, L_2(Q), \|F\|_{L_2(Q)}\epsilon\right)}\, d\epsilon < \infty,$$

where the sup is over all finitely supported measure Q. Then $\mathcal{F}$ is P-Donsker.

1.3 Idea of the proof

To prove the limit exists, we only need to check two conditions.

- For finite dimensional convergence, we only need to verify the Lindeberg's condition for multidimensional CLT. Here we will need to use the fact that for any $x \in \mathcal{X}$ and any $f \in \mathcal{F}$, $|f(x)| \le F(x)$.
- Below we sketch the proof for ASEC which is a more difficult part.

Define

$$\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \|f-g\|_{L_2(P)} \le \delta\}$$

and $G_n = \sqrt{n}(P_n - P)$. Note that

$$G_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(f(X_i) - \mathbb{E}_p[f(X_i)]\right).$$

The goal is to show

$$\lim_{\delta \to 0} \limsup_n P\left(\sup_{f,g\in\mathcal{F}, \|f-g\|_{L_2(P)} \le \delta} |G_n(f-g)| \ge \epsilon_0 \right) \to 0.$$

Verify yourself that this is equivalent to ASEC. Denote

$$\|G_n\|_{\mathcal{F}_\delta} = \sup_{f,g\in\mathcal{F}, \|f-g\|_{L_2(P)} \le \delta} |G_n(f-g)|.$$

From the same symmetrization argument as before, we have

$$\begin{aligned}
P(\|G_n\|_{\mathcal{F}_\delta} \ge \epsilon_0) &\le \frac{2}{\epsilon_0} \mathbb{E}\left[\sup_{f\in\mathcal{F}_\delta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i f(X_i) \right| \right] \\
&\le \frac{2}{\epsilon_0} \mathbb{E}\left[\sup_{f\in\mathcal{F}_\delta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i (f(X_i) - \tilde{f}(X_i)) \right| \right] + \frac{2}{\epsilon_0} \mathbb{E}\left[\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i \tilde{f}(X_i) \right| \right] \\
&\le \frac{C}{\epsilon_0} \mathbb{E}\left[\int_0^{\theta_n} \sqrt{\log N(\mathcal{F}_\delta, \|\cdot\|_{L_2(P_n)}, \epsilon)}\, d\epsilon \right] + C\delta,
\end{aligned}$$

where $D_n = \sup_{f,g\in\mathcal{F}_\delta} \|f-g\|_{L_2(P_n)} \le 2\sup_{f\in\mathcal{F}_\delta} \|f\|_{L_2(P_n)}$ and $\theta_n = D_n/2 \le \sup_{f\in\mathcal{F}_\delta} \|f\|_{L_2(P_n)}$. Denote

$$I := \mathbb{E}\left[\int_0^{\theta_n} \sqrt{\log N(\mathcal{F}_\delta, \|\cdot\|_{L_2(P_n)}, \epsilon)}\, d\epsilon \right]$$

One can show that $N(\mathcal{F}_\delta, L_2(P_n), \epsilon) \le N(\mathcal{F}, L_2(P_n), \epsilon/2)^2$. To see this, note that

- Suppose $\{f_i\}_N$ is $\epsilon/2$-net of $\mathcal{F}$.
- Then $\{f_i - f_j : i \le N, j \le N\}$ has N^2 elements, which forms an ϵ-net of $\mathcal{F}_\delta$.

Replacing ϵ by $\|F\|_{L_2(P_n)}\epsilon$, we have

$$\begin{aligned}
I &\le C\mathbb{E}\left[\int_0^{\theta_n/\|F\|_{L_2(P_n)}} \|F\|_{L_2(P_n)} \sqrt{\log N(\mathcal{F}, L_2(P_n), \epsilon\|F\|_{L_2(P_n)})}\, d\epsilon \right] \\
&\le C\mathbb{E}\left[\int_0^{\theta_n/\|F\|_{L_2(P_n)}} \|F\|_{L_2(P_n)} \sup_Q \sqrt{\log N(\mathcal{F}, L_2(Q), \epsilon\|F\|_{L_2(Q)})}\, d\epsilon \right] \\
&\le C\sqrt{\mathbb{E}(\|F\|^2_{L_2(P_n)})} \sqrt{\mathbb{E}\left[\left(\int_0^{\theta_n/\|F\|_{L_2(P_n)}} \sup_Q \sqrt{\log N(\mathcal{F}, L_2(Q), \epsilon\|F\|_{L_2(Q)})}\, d\epsilon \right)^2 \right]}.
\end{aligned}$$

Recall that $\theta_n = D_n/2 \le \sup_{f\in\mathcal{F}_\delta} \|f\|_{L_2(P_n)}$. Note that

$$\begin{aligned}
\sup_{f\in\mathcal{F}_\delta} \|f\|^2_{L_2(P_n)} &= \sup_{f\in\mathcal{F}_\delta} P_n f^2 \\
&\le \sup_{f\in\mathcal{F}_\delta} |(P_n - P)f^2| + \sup_{f\in\mathcal{F}_\delta} |Pf^2|.
\end{aligned}$$

The first term goes to zero as $n \to +\infty$ because of the entropy condition. The second term goes to zero as $\delta \to 0$. Thus θ_n can be made small for large enough n and small enough δ. By DCT,

$$\mathbb{E}\left[\left(\int_0^{\theta_n/\|F\|_{L_2(P_n)}} \sup_Q \sqrt{\log N(\mathcal{F}, L_2(Q), \epsilon\|F\|_{L_2(Q)})}\, d\epsilon\right)^2\right],$$

will be small for large enough n and small enough δ. Thus I can be made arbitrarily small, which completes the proof.

2 Goodness of fit statistics

2.1 Kolmogorov-Smirnoff test

Suppose we observe $X_1, \ldots, X_n \sim^{i.i.d} F$. We aim to test the null hypothesis that

$$H_0 : F = F_0.$$

Let F_n be the empirical cdf. The Kolmogorov-Smirnoff test statistic is defined as

$$KS_n = \sqrt{n} \sup_{t\in\mathbb{R}} |F_n(t) - F_0(t)|,$$

Under the null hypothesis,

$$KS_n = \sqrt{n} \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| = \sup_{f\in\mathcal{F}} |\sqrt{n}(P_n - P)f|$$

where $\mathcal{F} = \{1[\cdot \le t], t \in \mathbb{R}\}$.

2.2 Claim

$\mathcal{F}$ is a Donsker class. To see this, we first note that the envelop function can be taken as $F \equiv 1$. Second, one can show

$$\int_0^{\infty} \sup_Q \sqrt{\log N(\mathcal{F}, L_2(Q), \epsilon\|F\|_{L_2(Q)})}\, d\epsilon < \infty.$$

Thus we have $\sqrt{n}(P_n - P)f \to^d G_P(f)$ or equivalently

$$\sqrt{n}(F_n(t) - F(t)) \xrightarrow{d} G_F(t),$$

where G_F is a Brownian Bridge with

$$\mathrm{Cov}(G_F(t), G_F(s)) = F(t \wedge s) - F(t)F(s).$$

Note that the map $f \mapsto \sup_{t\in\mathbb{R}} |f(t)|$ is continuous in $\|\cdot\|_\infty$ as$|\|f\|_\infty - \|g\|_\infty| \le \|f - g\|_\infty$. By the continuous mapping theorem

$$KS_n = \sup_{t\in\mathbb{R}} |\sqrt{n}(F_n(t) - F(t))| \to \sup_{t\in\mathbb{R}} |G_F(t)| = \sup_{t\in\mathbb{R}} |G_\lambda(F(t))| = \sup_{u\in(0,1)} |G_\lambda(u)|,$$

where λ is the uniform distribution/measure on $(0, 1)$. We can see that

$$\mathrm{Cov}(G_\lambda(F(t)), G_\lambda(F(s))) = \lambda(F(t) \wedge F(s)) - \lambda(F(t))\lambda(F(s)) = F(t \wedge s) - F(t)F(s) = \mathrm{Cov}(G_F(t), G_F(s)).$$

2.3 Cramer-Von Mises Statistics

The Cramer-Von Mises Statistic is defined as

$$\mathrm{CV}_n = n \int (F_n(t) - F_0(t))^2 dF_0(t).$$

Under the null,

$$\mathrm{CV}_n = n \int (F_n(t) - F(t))^2 dF(t) = \int \{\sqrt{n}(F_n(t) - F(t))\}^2 dF(t).$$

The map $f \mapsto \int f^2(t)dF(t)$ is continuous w.r.t. $\|\cdot\|_\infty$. By the continuous mapping theorem

$$CV_n \xrightarrow{d} \int G_F(t)^2 dF(t) = \int G_\lambda(F(t))^2 dF(t) = \int G_\lambda(u)^2 du.$$

2.4 Simulate the limiting distributions

Let $X_1, \ldots, X_n \sim^{i.i.d} N(0,1)$. Then we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \bar{X}_n) \xrightarrow{t} G_\lambda(t), \quad t \in (0,1),$$

where $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$.

2.5 Summary of Empirical Process

	Consistency	Inference
Classical Probability Theory	LLN	CLT
Empirical Process Theory	ULLN (Glivenko-Cantelli class)	Uniform CLT (Donsker class)

Some key techniques:

- covering number, bracketing number
- discretization, approximation of an infinite class by finite/countable class
- VC-dimnesion
- Concerntration inequality
- Rademacher complexity
- Symmetrization
- Chaining argument
- Peeling device...