Stat210B: Theoretical Statistics **Lecture Date: March 07, 2007**

Asymptotics of Empirical Processes

Lecturer: Michael I. Jordan *Scribe: Wei-Chun Kao*

1

Lemma 1. *(van der Vaart, 1998, lemma 19.24) Let $\mathcal{F}$ be P-donsker, and $\hat{f}_n$ be a random sequence of functions taking values in $\mathcal{F}$ s.t.*

$$\int (f_n(x) - f_0(x))^2 dP(x) \xrightarrow{P} 0$$

for some f_0 in $\mathcal{L}_2(P)$, then we have

$$G_n(\hat{f}_n - f_0) \xrightarrow{P} 0,$$

and

$$G_n\hat{f}_n \xrightarrow{d} G_P f_0$$

with $G_n = \sqrt{n}(P_n - P)$.

Proof. Sketch: Uses uniform continuity of sample paths of G_p with CMT. □

2 An Example: Mean Absolute Deviation

Define mean absolute deviation

$$M_n = \frac{1}{n}\sum_i |X_i - \bar{X}_n|$$

Let F denote the unknown CDF. W.l.o.g, let $Fx = 0$. Let $F_n|X - \theta| = \frac{1}{n}\sum_i |X_i - \theta|$. If $Fx^2 < \infty$, and if $\theta \in \Theta$ for a compact Θ, then $\{|x - \theta|\}$ is F-Donsker (van der Vaart, 1998, example 19.7)

$$F(|x - \bar{X}_n| - |x|)^2 \leq |\bar{X}_n|^2 \xrightarrow{P} 0$$

By Lemma 1, we have

$$G_n|x - \bar{X}_n| - G_n|x| \xrightarrow{P} 0 \tag{1}$$

Consider

$$\sqrt{n}(M_n - F|x|) = \sqrt{n}(F|x - \bar{X}_n| - F|x|) + G_n|x| + o_P(1), \tag{2}$$

assume that $\theta \mapsto F|x - \theta|$ is differentiable at $\theta = 0$, differentiate $F|x - \theta|$ at $\theta = 0$ we have the derivative:

$$2F(0) - 1.$$

Apply Delta Method on $\sqrt{n}(F|x-\bar{X}_n| - F|x|)$, we have

$$\begin{aligned}\sqrt{n}(F|x-\bar{X}_n| - F|x|) &= -(2F(0)-1)\sqrt{n}((x-\bar{X}_n)-x) + o_P(1) & (3)\\ &= (2F(0)-1)\sqrt{n}\bar{X}_n & (4)\\ &= (2F(0)-1)\sqrt{n}(F_n - F)x + o_P(1) & (5)\\ &= (2F(0)-1)G_n x + o_P(1). & (6)\end{aligned}$$

Therefore, we have

$$\begin{aligned}\sqrt{n}(M_n - F|x|) &= ((2F(0)-1)x + |x|) + o_P(1) & (7)\\ &\xrightarrow{d} G_P((2F(0)-1)x + |x|). & (8)\end{aligned}$$

Thus, M_n is AN with mean 0 and variance equals to variance of $(2F(0)-1)X_1 + |X_1|$. We lose $(2F(0)-1)X$ term by not knowing the mean of X. When the mean and median are the same, $2F(0)-1=0$, in which case we don't incur any extra variance by having to estimate the location parameter.

3 AN of Z-estimators

Definition 2. A function $\psi_\theta(x)$ is Lipschitz if $\exists$ a function $\dot{\psi}(x)$ s.t.

$$||\psi_{\theta_1}(x) - \psi_{\theta_2}(x)|| \leq \dot{\psi}(x)||\theta_1 - \theta_2||$$

$\forall \theta_1, \theta_2$ in some neighborhood of θ_0 and $P\dot{\psi}^2 \leq \infty$.

Theorem 3. *(van der Vaart, 1998, Theorem 5.21) For each θ_0 in an open subset of Euclidean space, let $\psi_\theta(x)$ be Lipschitz. Assume $P||\psi_{\theta_0}||^2 \leq \infty$, $P\psi_\theta$ is differentiable at θ_0 with derivative V_{θ_0} (note that it is different from "ψ_θ is differentiable"). Let*

$$P_n \psi_{\hat{\theta}_n} = o_P(n^{-1/2})(\text{ a “near zero”}). \tag{9}$$

Assume $\hat{\theta}_n \xrightarrow{P} \theta_0$ (consistency). Then we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V_{\theta_0}^{-1} \frac{1}{\sqrt{n}} \sum_i \psi_{\theta_0}(x_i) + o_P(1),$$

and

$$\hat{\theta}_n - \theta_0 \text{ is AN with zero mean and covariance } V_{\theta_0}^{-1}(P\psi_{\theta_0}\psi_{\theta_0}')V_{\theta_0}^{-T}$$

Proof. (van der Vaart, 1998, example 19.7) shows that Lipschitz functions are P-Donsker. Apply Lemma 1 we have

$$G_n \psi_{\hat{\theta}_n} - G_n \psi_{\theta_0} \xrightarrow{P} 0$$

By assumption that $\sqrt{n}P_n\psi_{\hat{\theta}_n} = o_P(1)$, we have

$$\begin{aligned}G_n \psi_{\hat{\theta}_n} &= -\sqrt{n}P\psi_{\hat{\theta}_n} + o_P(1) & (10)\\ &= \sqrt{n}P(\psi_{\theta_0} - \psi_{\hat{\theta}_n}) + o_P(1), & (11)\end{aligned}$$

with $P\psi_{\theta_0} = 0$ by definition. Apply Delta Method, or (van der Vaart, 1998, Lemma 2.12), we have

$$\sqrt{n}V_{\theta_0}(\theta_0 - \hat{\theta}_n) + \sqrt{n}\ o_P(||\hat{\theta}_n - \theta_0||) = G_n\psi_{\theta_0} + o_P(1). \tag{12}$$

By invertability of V_{θ_o}, we have

$$\begin{aligned}\sqrt{n}||\hat{\theta}_n - \theta_o|| &\leq ||V_{\theta_0}^{-1}||\sqrt{n}||V_{\theta_0}(\hat{\theta}_n - \theta_0)|| \qquad (13)\\ &= O_P(1) + o_P(\sqrt{n}||\hat{\theta}_n - \theta_0||). \qquad (14)\end{aligned}$$

Inequality (14) is obtained by plugging (12) into (13) and using triangle inequality. Therefore, we have

$$\hat{\theta}_n \text{ is } \sqrt{n} - \text{consistent.} \qquad (15)$$

By (12) and (15), we have

$$\sqrt{n}V_{\theta_0}(\hat{\theta}_n - \theta_0) = -G_n\psi_{\theta_0} + o_P(1). \qquad (16)$$

Multiply both side by $V_{\theta_0}^{-1}$ to get the result. □

References

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.