An Application of Empirical Process Theory

Jon A. Wellner

January 7, 2009

1. Two Theorems from Empirical Process Theory

Suppose that $X_1, \ldots, X_n$ are i.i.d. P on $(\mathcal{X}, \mathcal{A})$. We define the *empirical measure* $\mathbb{P}_n$ and the *empirical process* $\mathbb{G}_n$ by

$$\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}, \qquad \mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)\,.$$

Let $\mathcal{F}$ be a class of real-valued measurable functions on $\mathcal{X}$. Thus $f : \mathcal{X} \rightarrow R$. We define an *envelope function* for $\mathcal{F}$ to be an function $F_{\mathcal{F}}(x)$ satisfying

$$|f(x)| \le F_{\mathcal{F}}(x) \qquad \text{for all } \ x \in \mathcal{X} \ \text{ and all } \ \ f \in \mathcal{F}\,.$$

For $r \ge 1$ and a class of functions $\mathcal{F} \subset L_r(P)$, we define the $L_r(P)$ *covering numbers* $N(\epsilon, \mathcal{F}, L_r(P))$ to be the minimal number of $L_r(P)$-balls of radius ϵ needed to cover $\mathcal{F}$. The following analogues of the classical Glivenko-Cantelli and Donsker theorems are very useful in statistics:

Theorem 1. (Uniform covering numbers Glivenko-Cantelli theorem).
Suppose that $EF_{\mathcal{F}}(X) < \infty$ and $\sup_P N(\epsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\epsilon > 0$. Then

$$\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \rightarrow_{a.s.} 0 \qquad \text{as } \ n \rightarrow \infty\,.$$

Theorem 2. (Uniform covering numbers Donsker theorem).
Suppose that $EF_{\mathcal{F}}^2(X) < \infty$ and

$$\int_0^1 \sqrt{\sup_P \log N(\epsilon, \mathcal{F}, L_2(P))}\, d\epsilon < \infty$$

for every $\epsilon > 0$. Then

$$\mathbb{G}_n \Rightarrow \mathbb{G}_P \qquad \text{in } \ l^\infty(\mathcal{F})$$

where $\mathbb{G}_P$ is a $P-$Brownian bridge process; i.e. a mean zero Gaussian process with covariance

$$E(\mathbb{G}_P(f)\mathbb{G}_P(g)) = P(fg) - P(f)P(g)$$

and sample paths which are continuous with respect to the ρ_P pseudo-metric on $\mathcal{F}$ given by

$$\rho_P^2(f,g) = Var_P(f(X) - g(X)) .$$

Empirical process theory provides many classes of functions $\mathcal{F}$ for which the hypotheses of these two theorems hold, and other theorems of these basic types with somewhat different hypotheses. See Van der Vaart and Wellner (1996) and Dudley (1999).

The example in the following sections is taken from Pollard (1989).

2. A Consistency Proof

Suppose that $X, X_1, \ldots, X_n$ are i.i.d. real-valued random variables with distribution function F and corresponding probability measure P on $(R, \mathcal{B})$. We will assume that $E|X| < \infty$.

Consider the random variables

$$V_n = \frac{1}{n}\sum_{i=1}^n |X_i - \overline{X}_n| .$$

The question is, does $V_n \rightarrow_{a.s.} v$ for some constant v, and can we identify v? The strong law of large numbers does not apply immediately, because the random variables $|X_1 - \overline{X}_n|, \ldots, |X_n - \overline{X}_n|$ are dependent. We do know, from the SLLN, that $\overline{X}_n \rightarrow_{a.s.} \mu \equiv E(X)$, and, also by the SLLLN,

$$V_n^0 = \frac{1}{n}\sum_{i=1}^n |X_i - \mu| \rightarrow_{a.s.} E|X - \mu| .$$

Thus we naturally guess that

$$V_n \rightarrow_{a.s.} E|X - \mu| = v . \tag{2.1}$$

To prove that (2.1) holds, it is natural to introduce the functions

$$f_t(x) = |x - t| .$$

Note that if we set

$$\mathbb{H}_n(t) \equiv \mathbb{P}_n f_t = \frac{1}{n}\sum_{i=1}^n |X_i - t| ,$$

and

$$H(t) \equiv P f_t = P(|x - t|) = \int |x - t| dP(x) ,$$

then $V_n = \mathbb{H}_n(\overline{X}_n)$, and $v = H(\mu)$.

Since $\overline{X}_n$ is within $\delta > 0$ of μ with high probability (or even a.s. for $n > N_\omega$), it is natural to consider the collection of such functions f_t for values of t with $|t-\mu| \le \delta$:

$$\mathcal{F} = \{f_t : \ |t-\mu| \le \delta\}.$$

Now it is easily seen that an envelope function $F_{\mathcal{F}}(x)$ for the class $\mathcal{F}$ is just $F_{\mathcal{F}}(x) = |x-\mu| + \delta$, which satisfies

$$E(F_{\mathcal{F}}(X)) = \delta + E|X-\mu| < \infty.$$

Moreover, the class $\mathcal{F}$ is a *VC-subgraph class* of functions (see e.g. Van der Vaart and Wellner (1996), section 2.6), and hence it satisfies

$$\sup_P N(\epsilon, \mathcal{F}, L_1(P)) \le \frac{K}{\epsilon^\gamma}$$

for a number γ related to the VC-dimension of $\mathcal{F}$. In particular these covering numbers are finite for every $\epsilon > 0$. Hence, by the Glivenko-Cantelli Theorem 1,

$$\sup_{f_t \in \mathcal{F}} |\mathbb{P}_n(f_t) - P(f_t)| = \sup_{t:|t-\mu|\le\delta} |\mathbb{P}_n(f_t) - P(f_t)| \to_{a.s.} 0. \tag{2.2}$$

Alternatively, the functions $f_t(x) = |x-t| \equiv f(x,t)$ are continuous in t for every x, $[\mu-\delta, \mu+\delta]$ is compact, and the argument above provides an integrable envelope function F for the class of functions $\mathcal{F} \equiv \{f(\cdot,t) : \ |t-\mu| \le \delta\}$, namely $F(x) = |x-\mu| + \delta$. Thus by Theorem 4.4.1 of Wellner's 580's class notes (or Theorem 16(a), page 108, Ferguson, ACILST), (2.2) holds.

This allows us to write

$$\begin{aligned}
|V_n - v| &= |\mathbb{H}_n(\overline{X}_n) - H(\overline{X}_n) + H(\overline{X}_n) - H(\mu)| \\
&\le |\mathbb{H}_n(\overline{X}_n) - H(\overline{X}_n)| + |H(\overline{X}_n) - H(\mu)| \\
&= |\mathbb{P}_n(f_{\overline{X}_n}) - P(f_{\overline{X}_n})| + |H(\overline{X}_n) - H(\mu)| \\
&\le \sup_{|t-\mu|\le\delta} |\mathbb{P}_n(f_t) - P(f_t)| + |\overline{X}_n - \mu| \\
&\to_{a.s.} 0 + 0 = 0
\end{aligned}$$

by (2.2) together with the observation that $|f_t(x) - f_\mu(x)| \le |t-\mu|$ for all $x \in R$.

3. An Asymptotic Normality Proof

Suppose the same basic set-up as in the previous section, but now assume that $E(X^2) < \infty$. Our question now is, does it hold that

$$\sqrt{n}(V_n - v) \to_d \mathbb{Z}$$

for some random variable $\mathbb{Z}$, and what is the distribution of $\mathbb{Z}$? Since V_n^0 is the sum of i.i.d. random varibles, each with expected value v, and $E|X-v|^2 < \infty$, it follows from the CLT that

$$\sqrt{n}(V_n^0 - v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{|X_i - \mu| - v\} = \mathbb{G}_n(f_\mu) \to_d \mathbb{G}_P(f_\mu) \sim N(0, Var(f_\mu(X))),$$

so we suspect that the random variable $\mathbb{G}_P(f_\mu)$ will enter in the limiting distribution.

Now we write, assuming that $H(t) \equiv P(f_t)$ has a derivative $H'(\mu)$ at $t = \mu$,

$$\begin{aligned}
\sqrt{n}(V_n - v) &= \sqrt{n}(\mathbb{P}_n(f_{\overline{X}_n}) - Pf_\mu) \\
&= \sqrt{n}(\mathbb{P}_n(f_{\overline{X}_n}) - Pf_{\overline{X}_n}) + \sqrt{n}(Pf_{\overline{X}_n} - Pf_\mu) \\
&= \mathbb{G}_n(f_{\overline{X}_n}) - \mathbb{G}_n(f_\mu) + \mathbb{G}_n(f_\mu) \\
&\qquad + \sqrt{n}(H(\overline{X}_n) - H(\mu)) \\
&= \mathbb{G}_n(f_\mu) + H'(\mu)\sqrt{n}(\overline{X}_n - \mu) \\
&\qquad + \left\{ \frac{H(\overline{X}_n) - H(\mu)}{\overline{X}_n - \mu} - H'(\mu) \right\} \sqrt{n}(\overline{X}_n - \mu) \\
&\qquad + \mathbb{G}_n(f_{\overline{X}_n}) - \mathbb{G}_n(f_\mu) \\
&\equiv I_n + II_n + III_n\,.
\end{aligned}$$

Now

$$I_n = \mathbb{G}_n(f_\mu + H'(\mu)g_\mu) \rightarrow_d \mathbb{G}_P(f_\mu + H'(\mu)g_\mu) \sim N(0, V^2)$$

where $g_\mu(x) \equiv x - \mu$ and

$$V^2 = Var(f_\mu(X) + H'(\mu)g_\mu(X)) = Var(|X - \mu| + H'(\mu)(X - \mu))\,.$$

If we show that $II_n = o_p(1)$ and $III_n = o_p(1)$, then we have proved the desired convergence in distribution and have identified $\mathbb{Z}$ as $N(0, V^2)$. Now $II_n = o_p(1)$ since $\sqrt{n}(\overline{X}_n - \mu) = O_p(1)$ and the difference quotient involving H converges to $H'(\mu)$ since $H(t)$ is differentiable at μ. Finally, since $\overline{X}_n$ is in the interval $[\mu - \delta, \mu + \delta]$ with high probability, the third term is bounded above by

$$\sup_{|t-\mu| \le \delta} |\mathbb{G}_n(f_t) - \mathbb{G}_n(f_\mu)|\,,$$

which can be made arbitrarily small by choosing δ sufficiently small in view of the Donsker property of the class of functions $\mathcal{F}$. The Donsker property entails that, for every $\epsilon > 0$,

$$\limsup_{n\to\infty} Pr\left(\sup_{|t-\mu| \le \delta} |\mathbb{G}_n(f_t) - \mathbb{G}_n(f_\mu)| > \epsilon \right) \to 0 \qquad \text{as} \qquad \delta \to 0$$

and hence we can make III_n arbitrarily small by choice of δ.

Now we examine the hypothesis H is differentiable at $t = \mu$, and compute $H'(\mu)$. When does this differentiability hold? The difference quotient is

$$\frac{H(t) - H(\mu)}{t - \mu} = \frac{P(f_t) - P(f_\mu)}{(t - \mu)} = P\left(\frac{|x - t| - |x - \mu|}{t - \mu} \right) \equiv P(D_t(x))$$

where the functions D_t satisfy $|D_t(x)| \le 1$ for all t and x, and

$$D_t(x) \to 2 \cdot 1_{(-\infty,\mu)}(x) - 1 = \text{sign}(\mu - x)$$

for $x \neq \mu$. Thus $D_t(x) \to 2 \cdot 1_{(-\infty,\mu)}(x) - 1$ a.s. P if $P(\{\mu\}) = 0$. Hence if $P(\{\mu\}) = 0$, the dominated convergence theorem yields

$$\frac{H(t) - H(\mu)}{t - \mu} \to H'(\mu) = P(2 \cdot 1_{(-\infty,\mu)} - 1) = F(\mu) - (1 - F(\mu)) = 2F(\mu) - 1\,.$$

Thus we see that $H'(\mu) = 0$ for any distribution with $F(\mu) = 1/2$, and in this case the second term in the limit does not contribute.

REFERENCES

Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge.

Pollard, D. (1989). Asymptotics via empirical processes (with discussion). *Statist. Sci.* **4**, 341-366.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Department of Statistics
University of Washington
P.O. Box 354322
Seattle, Washington 98195-4322
U.S.A.
e-mail: jaw@stat.washington.edu