Preservation Theorems for Glivenko-Cantelli classes and Consistency of the Nonparametric MLE for Generalized Interval Censoring

Jon A. Wellner and Aad van der Vaart

University of Washington
and Free University, Amsterdam

Talk to be given at IMS-BS Meeting, Guanajuato, May 16, 2000.

Email: jaw@stat.washington.edu

http: //www.stat.washington.edu/ jaw/jaw.research.html

Outline.

1. **Introduction:**
2. **Glivenko-Cantelli Theorems:**
3. **Preservation of Glivenko-Cantelli Classes:**
4. **Application to Interval Censoring:**
5. **Problems:**

Introduction.

Questions:

A. What operations preserve Glivenko-Cantelli classes of functions $\mathcal{F}$?

B. Can we simplify the consistency theorems of Schick and Yu (2000) for "mixed case" interval censoring?

C. Can we extend the consistency theorems of Schick and Yu (2000) to more complicated settings?

Glivenko-Cantelli theorems.

Theorem 1. (Giné and Zinn, 1984). Suppose that $\mathcal{F}$ is $L_1(P)-$bounded and NLSM(P). Then the following are equivalent:
A. $\mathcal{F}$ is a strong Glivenko-Cantelli class for P:

$$\left(\sup_{f\in\mathcal{F}}|\mathbb{P}_n(f)-P(f)|\right)^* \to_{a.s.} 0 \qquad (1)$$

B. $\mathcal{F}$ has an envelope function $F \in L_1(P)$ and the classes $\mathcal{F}_M \equiv \{f 1_{[F\le M]}: \ f\in\mathcal{F}\}$ satisfy

$$\frac{1}{n}E^* \log N(\epsilon, \mathcal{F}_M, L_r(\mathbb{P}_n)) \to 0 \qquad (2)$$

for every $\epsilon > 0$ and for some (all) $r \in (0,\infty]$ where $\|f\|_{L_r(P)} \equiv (P|f|^r)^{r^{-1}\wedge 1}$.

3. Preservation of Glivenko-Cantelli Classes

For classes $\mathcal{F}_1, \ldots, \mathcal{F}_k$ of functions $f_i : \mathcal{X} \to R$ and a function $\varphi : R^k \to R$, let $\varphi(\mathcal{F}_1, \ldots, \mathcal{F}_k)$ be the class of functions

$$x \mapsto \varphi(f_1(x), \ldots, f_k(x))$$

where $f_i \in \mathcal{F}_i$ for $i = 1, \ldots, k$.

Theorem 2. (van der Vaart and Wellner). Suppose that $\mathcal{F}_1, \ldots, \mathcal{F}_k$ are $P-$Glivenko-Cantelli classes of functions and that $\varphi : R^k \to R$ is continuous. Then $\mathcal{H} \equiv \varphi(\mathcal{F}_1, \ldots, \mathcal{F}_k)$ is $P-$Glivenko-Cantelli provided that it has an integrable envelope.

Corollary 1. (Dudley, 1998). Suppose that $\mathcal{F}$ is a strong Glivenko-Cantelli class for P with $PF < \infty$, J is a possibly unbounded interval including the ranges of all $f \in \mathcal{F}$, φ is continuous and monotone on J, and for some finite constants c, d, $|\varphi(y)| \le c|y| + d$ for all $y \in J$. Then $\varphi(\mathcal{F})$ is also a strong Glivenko-Cantelli class for P.

Corollary 2. (Dudley, 1998). Suppose that $\mathcal{F}$ is a strong Glivenko-Cantelli class for P with $PF < \infty$, and g is a fixed bounded function ($\|g\|_\infty < \infty$). Then the class of functions

$$g \cdot \mathcal{F} \equiv \{g \cdot f : \ f \in \mathcal{F}\}$$

is a strong Glivenko-Cantelli class for P.

Corollary 3. (Giné and Zinn, 1984). Suppose that $\mathcal{F}$ is a uniformly bounded strong Glivenko-Cantelli class for P, and $g \in L_1(P)$ is a fixed function. Then the class of functions

$$g \cdot \mathcal{F} \equiv \{g \cdot f : \ f \in \mathcal{F}\}$$

is a strong Glivenko-Cantelli class for P.

Lemma. Suppose that $\varphi : K \to R$ is continuous and $K \subset R^k$ is compact. Then for every $\epsilon > 0$ there exists a $\delta > 0$ such that for all n and for all $a_1, \ldots, a_n, b_1, \ldots, b_n \in K \subset R^k$

$$\frac{1}{n}\sum_{i=1}^n \|a_i - b_i\| < \delta$$

implies

$$\frac{1}{n}\sum_{i=1}^n |\varphi(a_i) - \varphi(b_i)| < \epsilon\,.$$

Here $\|\cdot\|$ can be any norm on R^k; in particular $\|x\|_r = (\sum_1^k |x_i|^r)^{1/r}$ for $r \in [1, \infty)$ or $\|x\|_\infty = \max_{1 \le i \le k} |x_i|$ for $x = (x_1, \ldots, x_k) \in R^k$.

Proof of Theorem 2 (sketch): Let H an envelope function for $\mathcal{H} = \varphi(\underline{\mathcal{F}})$. Using the lemma with $\|\cdot\|$ the L_1- norm $\|\cdot\|_1$, we find that

$$N(\epsilon, \mathcal{H}_M, L_1(\mathbb{P}_n)) \le \prod_{j=1}^k N(\frac{\delta}{k}, \mathcal{F}_j 1_{[F_j \le M]}, L_1(\mathbb{P}_n))\,.$$

Theorem 3. (van der Vaart and Wellner). Suppose that $\mathcal{F}$ is a class of functions on $(\mathcal{X}, \mathcal{A}, P)$, and $\{\mathcal{X}_i\}$ is a partition of $\mathcal{X}$: $\cup_{i=1}^\infty \mathcal{X}_i = \mathcal{X}$, $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$ for $i \ne j$. Suppose that $\mathcal{F}_j \equiv \{f 1_{\mathcal{X}_j} : \ f \in \mathcal{F}\}$ is $P-$Glivenko-Cantelli for each j, and $\mathcal{F}$ has an integrable envelope function F. Then $\mathcal{F}$ is itself $P-$Glivenko-Cantelli.

Theorem 4. (Dudley). Suppose that $\mathcal{F}$ is a strong Glivenko- Cantelli class of functions for P on $(\mathcal{X}, \mathcal{A})$. Then the symmetric convex hull class

$$\mathcal{G} \equiv \{g = \sum_{i=1}^k c_i f_i : \quad f_i \in \mathcal{F}, c_i \in R, \sum_1^k |c_i| \le 1\}$$

is a strong Glivenko-Cantelli class for P, and so is the class $\overline{\mathcal{G}}$ of functions which are both the pointwise limit and the $L_1(P)$ limit of sequences in $\mathcal{G}$.

Remark 1.
Similar theorems for preservation of **uniform Glivenko-Cantelli classes** related to the theorem of Dudley, Giné, and Zinn (1991) characterizing such classes.

Remark 2.
What operations preserve **Donsker classes**?
Answer: Lipschitz functions $\varphi : R^k \rightarrow R$; see e.g. Van der Vaart and Wellner (1996), section 2.10, pages 190 - 203.

4. Interval Censoring.

Case 1: $Y \sim F$, $T \sim G$, Y independent of T.
We observe $X \equiv (T, 1_{[Y \le T]}) \equiv (T, \Delta)$.

$$(\Delta | T) \sim \text{Bernoulli}(F(T)) .$$

Problem: Estimate the distribution function F of Y.
Solution: Groeneboom (1989), Groeneboom and Wellner (1992).

Picture 1.

Case 2: $Y \sim F$, $\underline{T} \equiv (T_1, T_2) \sim G$, Y independent of $\underline{T}$.
We observe

$$X \equiv (T_1, T_2, 1_{[Y \le T_1]}, 1_{[T_1 < Y \le T_2]}, 1_{[T_2 < Y]}) \equiv (\underline{T}, \underline{\Delta}) .$$

Note that

$$(\underline{\Delta} | \underline{T}) \sim \text{Mult}_3(1, (F(T_1), F(T_2) - F(T_1), 1 - F(T_2))) .$$

Problem: Estimate the distribution function F of Y.
Solution: Groeneboom and Wellner (1992), Groeneboom (1996).

Picture 2.

"Mixed Case" Interval Censoring: $Y \sim F$ independent of $(\underline{T}_K, K)$, with $\underline{T}_K = (T_{K1}, \ldots, T_{KK})$, $T_{K,0} = -\infty$, $T_{K,K+1} = \infty$.
We observe $X \equiv (\underline{T}_K, \underline{\Delta}_K, K)$ where

$$\Delta_{Kj} \equiv 1_{(T_{K,j-1}, T_{K,j}]}(Y) , \qquad j = 1, \ldots , K+1$$

so that

$$(\underline{\Delta}_K | \underline{T}_K, K) \sim Mult_{K+1}(1, \underline{\Delta F}_K)$$

where $(\Delta F)_{K,j} \equiv F(T_{K,j}) - F(T_{K,j-1})$.
Problem: Estimate the distribution function F of Y.
Solution: Schick and Yu (2000).

Picture 3.

Theorem 5. (Schick and Yu). If $E(K) < \infty$, then the nonparametric MLE $\widehat{F}_n$ of F satisfies

$$\int |\widehat{F}_n - F| d\mu \rightarrow_{a.s.} 0$$

where, for $B \in \mathcal{B}_1$,

$$\mu(B) \equiv \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(T_{k,j} \in B | K = k) .$$

Note that μ is a finite measure if $E(K) < \infty$.

Theorem 6. (van der Vaart and Wellner).

$$H(p_{\widehat{F}_n}, p_{F_0}) \rightarrow_{a.s.} 0$$

and

$$\frac{1}{2}\left\{\int |\widehat{F}_n - F_0| d\tilde{\mu}\right\}^2 \le H^2(p_{\widehat{F}_n}, p_{F_0}) \rightarrow_{a.s.} 0$$

where

$$\tilde{\mu}(B) \equiv \sum_{k=1}^{\infty} P(K = k) \frac{1}{k} \sum_{j=1}^{k} P(T_{k,j} \in B | K = k) .$$

Note that $\tilde{\mu}$ is always a finite measure.

Proof of Theorem 6: By van de Geer (1993), (1996),

$$\begin{aligned} H^2(p_{\widehat{F}_n}, p_{F_0}) &\le (\mathbb{P}_n - P_0)(J(p_{\widehat{F}_n}/p_{F_0})) \\ &\le \|\mathbb{P}_n - P_0\|_{\mathcal{J}} \end{aligned}$$

where

$$J(t) \equiv \begin{cases} (t-1)/(t+1), & t \ge 0, \\ -1, & t < 0 \end{cases}$$

and

$$\mathcal{J} \equiv \{J(p_F/p_{F_0}) : \ F \in \mathcal{F}\} .$$

Thus if we can show that $\mathcal{J}$ is a Glivenko-Cantelli class of functions, consistency in the Hellinger metric follows.

Step 1. The class of functions $\mathcal{F}_1 \equiv \{p_F : \ F \in \mathcal{F}\}$ is a Glivenko-Cantelli class.
Proof: This follows from Theorem 3 (with $\mathcal{X}_j \equiv \{x = (\underline{\delta}, \underline{t}_k, k) : k = j\}$), and Theorem 4 (with $\mathcal{F}$ the VC class of indicators of sets $(-\infty, t]$ and $\overline{\mathcal{G}}$ the class of functions of bounded variation).

Step 2. The class of functions $\mathcal{F}_2 \equiv \{p_F/p_{F_0} : \ F \in \mathcal{F}\}$ is a Glivenko-Cantelli class.
Proof: $1/p_{F_0} \in L_1(P_0)$ and the functions p_F are uniformly bounded, so this follows from Corollary 3 and Step 1.

Step 3. The class of functions $\mathcal{J} = \{J(p_F/p_{F_0}) : \ F \in \mathcal{F}\}$ is a Glivenko-Cantelli class.
Proof: This follows from Theorem 3 with $\varphi = J$ and step 2.

Step 4.

$$H^2(p_{\widehat{F}_n}, p_{F_0}) \le d_{TV}(p_{\widehat{F}_n}, p_{F_0}) \le \sqrt{2} H(p_{\widehat{F}_n}, p_{F_0})$$

where

$$\begin{aligned} &H^2(p_{\widehat{F}_n}, p_{F_0}) \\ &= \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k+1} \int \{[\widehat{F}_n(y_{k,j}) - \widehat{F}_n(y_{k,j-1})]^{1/2} \\ &\qquad - [F_0(y_{k,j}) - F_0(y_{k,j-1})]^{1/2}\}^2 dG_k(y) \end{aligned}$$

and

$$\begin{aligned} &d_{TV}(p_{\widehat{F}_n}, p_{F_0}) \\ &= \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k+1} \int |[\widehat{F}_n(y_{k,j}) - \widehat{F}_n(y_{k,j-1})] \\ &\qquad - [F_0(y_{k,j}) - F_0(y_{k,j-1})]| dG_k(y) \\ &\ge \int |\widehat{F}_n - F_0| d\tilde{\mu} . \end{aligned}$$

A General Model:

- Let Y take values in $(\mathcal{Y}, \mathcal{B})$, $Y \sim Q$.

- Suppose that $C_K \equiv (C_{K1}, \ldots, C_{K,K})$ where $\{C_{K,j}\}_{j=1}^K$ form a partition of $\mathcal{Y}$, and $(\underline{C}_K, K)$ is independent of Y.

- Suppose we observe
$$X \equiv (\underline{\Delta}_K, \underline{C}_K, K)$$
where $\Delta_{K,j} \equiv 1_{C_{K,j}}(Y)$, so that
$$(\underline{\Delta}_K | \underline{C}_K, K) \sim \text{Mult}_k(1, \underline{Q}_K)$$
where $\underline{Q}_K \equiv (Q(C_{K,1}, \ldots, Q(C_{K,K}))$.

Picture 4.

Theorem 7. (van der Vaart and Wellner). If all $C_{K,j} \in \mathcal{C}$, a VC collection of subsets of $\mathcal{X}$, then the nonparametric maximum likelihood estimator $\widehat{Q}_n$ of Q_0 satisfies
$$H(p_{\widehat{Q}_n}, p_{Q_0}) \rightarrow_{a.s.} 0$$
and
$$\int |\widehat{Q}_n(c) - Q_0(c)| d\mu(c) \rightarrow_{a.s.} 0$$
where, for $B \in \Sigma$, a $\sigma-$field of subsets of the space of sets where the $C_{k,j}$ takes values,
$$\mu(B) \equiv \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(C_{k,j} \in B | K = k) \,.$$

5. Problems

1. How to characterize $\widehat{Q}_n$?
2. How to compute $\widehat{Q}_n$? Fast algorithms?
3. Global rates of convergence? That is, how fast does $H(p_{\widehat{Q}_n}, p_{Q_0})$ converge to zero?
4. Local rates of convergence? For fixed sets C how fast does $\widehat{Q}_n(C) - Q_0(C)$ converge to zero?

Shuguang Song, Ph.D. dissertation in progress at U.W.

6. Selected References.

- Dudley, R. M. (1998). Consistency of M-estimators and one-sided bracketing. In *Proceedings of the First International Conference on High Dimensional Probability*, 33- 58. E. Eberlein, M. Hahn, and M. Talagrand, editors. Birkhäuser, Basel.

- Dudley, R. M., Giné, E., and Zinn, J. (1991). Uniform and universal Glivenko-Cantelli classes. *J. Theoretical Probab.* **4**, 485 - 510.

- Giné, E., and Zinn, J. (1984). Some limit theorems for empirical processes. *Ann. Probabilty* **12**, 929-989.

- Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Statist.* **27**, 45 - 55.

- van de Geer, S. (1993). Hellinger consistency of certain nonparametric maximum likelihood estimators. estimator in mixture models. *Ann. Statist.* **21**, 14 - 44.

- van de Geer, S. (1996). Rates of convergence for the maximum likelihood estimator in mixture models. *J. Nonparametric Statist.* **6**, 293 - 310.