Preservation Theorems for Glivenko-Cantelli and Uniform Glivenko-Cantelli Classes

Aad van der Vaart and Jon A. Wellner

Free University and University of Washington

ABSTRACT We show that the $P-$Glivenko property of classes of functions $\mathcal{F}_1, \dots, \mathcal{F}_k$ is preserved by a continuous function φ from R^k to R in the sense that the new class of functions

$$x \to \varphi(f_1(x), \dots, f_k(x)), \quad f_i \in \mathcal{F}_i, \quad i = 1, \dots, k$$

is again a Glivenko-Cantelli class of functions if it has an integrable envelope. We also prove an analogous result for preservation of the uniform Glivenko-Cantelli property. Corollaries of the main theorem include two preservation theorems of Dudley (1998). We apply the main result to reprove a theorem of Schick and Yu (1999) concerning consistency of the NPMLE in a model for "mixed case" interval censoring. Finally a version of the consistency result of Schick and Yu (1999) is established for a general model for "mixed case interval censoring" in which a general sample space $\mathcal{Y}$ is partitioned into sets which are members of some VC-class $\mathcal{C}$ of subsets of $\mathcal{Y}$.

i

1. Glivenko - Cantelli Theorems.

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space, and suppose that $\mathcal{F} \subset \mathcal{L}_1(P)$. For such a class of functions, let $\mathcal{F}_{0,P} \equiv \{f - Pf : f \in \mathcal{F}\}$. We also let $F_{\mathcal{F}}(x) \equiv \sup_{f \in \mathcal{F}} |f(x)|$, the *envelope function* of $\mathcal{F}$. If $|f| \leq F$ for all $f \in \mathcal{F}$ with F measurable, then F is an envelope for $\mathcal{F}$.

Suppose that $X_1, X_2, \ldots$ are i.i.d. P. Glivenko-Cantelli theorems give conditions under which the empirical measure $\mathbb{P}_n$ converges uniformly to P over a class $\mathcal{F}$, either in probability (in which case we say that $\mathcal{F}$ is a *weak Glivenko-Cantelli class for P*) or almost surely:

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} \equiv \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf| \to_{a.s.*} 0 \, ;$$

in this case we say that we say that $\mathcal{F}$ is a *strong Glivenko-Cantelli class for P*. Useful sufficient conditions for a class $\mathcal{F}$ to be a strong Glivenko-Cantelli class for P are that it has an integrable envelope and either

$$\log N_{[\,]}(\epsilon, \mathcal{F}, L_1(P)) < \infty \qquad \text{for all } \ \epsilon > 0$$

or

$$\log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) = o_P^*(n) \qquad \text{for every } \ \epsilon > 0 \, .$$

Here $\mathcal{F}_M = \{f 1_{[F \leq M]} : \ f \in \mathcal{F}\}$. In the second case some additional measurability conditions are necessary; see Van der Vaart and Wellner (1996), Chapter 2.4, pages 122 - 126. In particular, the condition in Theorem 2.4.3, page 123, is shown by Giné and Zinn (1984) and Talagrand (1996) to be both necessary and sufficient, under measurability assumptions, for the class $\mathcal{F}$ to be a strong Glivenko-Cantelli class. Talagrand (1987b) gives necessary and sufficient conditions for the Glivenko-Cantelli theorem without any measurability hypotheses.

Theorem 1. (Giné and Zinn, 1984). Suppose that $\mathcal{F}$ is $L_1(P)$ bounded and nearly linearly supremum measurable for P; in particular this holds if $\mathcal{F}$ is image admissible Suslin. Then the following are equivalent:
(a) $\mathcal{F}$ is a strong Glivenko-Cantelli class for P.
(b) $\mathcal{F}$ has an envelope function $F \in L_1(P)$ and the truncated classes $\mathcal{F}_M \equiv \{f 1\{F \leq M\} : \ f \in \mathcal{F}\}$ satisfy

$$\frac{1}{n} E^* \log N(\epsilon, \mathcal{F}_M, L_r(\mathbb{P}_n)) \to 0 \ \text{ for all } \epsilon > 0, \text{ and for all } M \in (0, \infty)$$

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

2. Preservation of the Glivenko - Cantelli Property.

Our goal in this section is to present several results concerning the stability of the Glivenko-Cantelli property of one or more classes of functions under composition with functions φ. A theorem which motivated our interest is the following result of Dudley (1998a).

Theorem 2. (Dudley, 1998a). 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, $|\varphi(y)| \leq c|y| + d$ for all $y \in J$. Then $\varphi(\mathcal{F})$ is also a strong Glivenko-Cantelli class for P.

Dudley (1998a) proves this via the characterization of Glivenko-Cantelli classes due to Talagrand (1987b). Dudley (1998b) also uses Talagrand's characterization to prove the following interesting proposition.

Proposition 1. (Dudley, 1998b). 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.

Yet another proposition in this same vein is:

Proposition 2. (Giné and Zinn, 1984). Suppose that $\mathcal{F}$ is a uniformly bounded strong Glivenko-Cantelli class for P, and $g \in \mathcal{L}_1(P)$ 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.

Given 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 $x \to \varphi(f_1(x), \ldots, f_k(x))$, where $f_i \in \mathcal{F}_i$, $i = 1, \ldots, k$. Theorem 2 and Propositions 1 and 2 are all corollaries of the following theorem.

Theorem 3. 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 function.

Proof. We first assume that the classes of functions $\mathcal{F}_i$ are appropriately measurable. Let $F_1, \ldots, F_k$ and H be integrable envelopes for $\mathcal{F}_1, \ldots, \mathcal{F}_k$ and $\mathcal{H}$ respectively, and set $F = F_1 \vee \ldots \vee F_k$. For $M \in (0, \infty)$, define

$$\mathcal{H}_M \equiv \left\{ \varphi(f) 1_{[F \leq M]} : \ f = (f_1, \ldots, f_k) \in \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_k \equiv \mathcal{F} \right\}.$$

Now

$$\|(\mathbb{P}_n - P)\varphi(f)\|_{\mathcal{F}} \leq (\mathbb{P}_n + P)H1_{[F>M]} + \|(\mathbb{P}_n - P)h\|_{\mathcal{H}_M}.$$

The expectation of the first term on the right converges to 0 as $M \to \infty$. Hence it suffices to show that $\mathcal{H}_M$ is $P-$Glivenko-Cantelli for every fixed M. Let $\delta = \delta(\epsilon)$ be the δ of Lemma 2 below for $\varphi : [-M, M]^k \to R$, $\epsilon > 0$, and $\|\cdot\|$ the L_1-norm $\|\cdot\|_1$. Then for any $(f_j, g_j) \in \mathcal{F}_j$, $j = 1, \ldots, k$,

$$\mathbb{P}_n|f_j - g_j|1_{[F_j \leq M]} \leq \frac{\delta}{k}, \quad j = 1, \ldots, k$$

implies that

$$\mathbb{P}_n|\varphi(f_1, \ldots, f_k) - \varphi(g_1, \ldots, g_k)|1_{[F \leq M]} \leq \epsilon.$$

It follows that

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

Thus $E^* \log N(\epsilon, \mathcal{H}_M, L_1(\mathbb{P}_n)) = o(n)$ for every $\epsilon > 0$, $M < \infty$. This implies that $E^* \log N(\epsilon, (\overline{\mathcal{H}}_M)_N, L_1(\mathbb{P}_n)) = o(n)$ for $(\overline{\mathcal{H}}_M)_N$ the functions $h1\{H \leq N\}$ for $h \in \overline{\mathcal{H}}_M$. Thus $\overline{\mathcal{H}}_M$ is strong Glivenko-Cantelli for P by Theorem 1. This concludes the proof that $\mathcal{H} = \varphi(\mathcal{F})$ is weak Glivenko-Cantelli. Because it has an integrable envelope, it is strong Glivenko-Cantelli by, e.g., Lemma 2.4.5 of Van der Vaart and Wellner (1996). This concludes the proof for appropriately measurable classes $\mathcal{F}_j$, $j = 1, \ldots, k$.

We extend the theorem to general Glivenko-Cantelli classes using separable versions as in Talagrand (1987a). (Also see van der Vaart and Wellner (1996), pages 115 - 120 for a discussion.) As shown in the preceding argument, it is not a loss of generality to assume that the classes $\mathcal{F}_i$ are uniformly bounded. Furthermore, it suffices to show that $\varphi(\mathcal{F}_1, \ldots, \mathcal{F}_k)$ is weak Glivenko-Cantelli. We first need a lemma.

Lemma 1. Any strong P-Glivenko Cantelli class $\mathcal{F}$ is totally bounded in $L_1(P)$ if and only if $\|P\|_{\mathcal{F}} < \infty$. Furthermore for any $r \in (1, \infty)$, if $\mathcal{F}$ has an envelope that is contained in $L_r(P)$, then $\mathcal{F}$ is also totally bounded in $L_r(P)$.

Proof. A class that is totally bounded is also bounded. Thus for the first statement we only need to prove that a strong Glivenko-Cantelli class $\mathcal{F}$ with $\|P\|_{\mathcal{F}} < \infty$ is totally bounded in $L_1(P)$.

It is well-known that such a class has an integrable envelope. E.g. see Giné and Zinn (1983) or Problem 2.4.1 of van der Vaart and Wellner (1996) to conclude first that $P^*\|f - Pf\|_{\mathcal{F}} < \infty$. Next the claim follows from the triangle inequality $\|f\|_{\mathcal{F}} \le \|f - Pf\|_{\mathcal{F}} + \|P\|_{\mathcal{F}}$. Thus it is no loss of generality to assume that the class $\mathcal{F}$ possesses an envelope that is finite everywhere.

Now suppose that there exists a sequence of finitely discrete probability measures P_n such that

$$L_n := \sup\{|(P_n - P)|f - g|| : f, g \in \mathcal{F}\} \to 0.$$

Then for every $\epsilon > 0$, there exists n_0 such that $L_{n_0} < \epsilon$. For this n_0 there exists a finite $\epsilon-$net $f_1, ..., f_N$ over $\mathcal{F}$ relative to the $L_1(P_{n_0})$-norm, because restricted to the support of P_{n_0} the functions f are uniformly bounded by the finite envelope and hence covering $\mathcal{F}$ in $L_1(P_{n_0})$ is like covering a compact in R^{n_0}. Now for any $f \in \mathcal{F}$ there is an f_i such that $P|f - f_i| \le L_{n_0} + P_{n_0}|f - f_i| < 2\epsilon$. It follows that $\mathcal{F}$ is totally bounded in $L_1(P)$.

To conclude the proof it suffices to select a sequence P_n. This can be constructed as a sequence of realizations of the empirical measure if we know that the class $|\mathcal{F} - \mathcal{F}|$ is P-GC. It is immediate from the definition of a Glivenko-Cantelli class that $\mathcal{F} - \mathcal{F}$ is P-GC. Next by Dudley's theorem, Theorem 2, (and also by our Theorem 3, but we have used the present lemma in the proof of this theorem to take care of measurability), the classes $(\mathcal{F} - \mathcal{F})^+$ and $(\mathcal{F} - \mathcal{F})^-$ are P-Glivenko Cantelli. Then the sum of these two classes is P-GC and hence the proof is complete.

If $\mathcal{F}$ has an envelope in $L_r(P)$, then $\mathcal{F}$ is totally bounded in $L_r(P)$ if the class $\mathcal{F}_M$ of functions $f1\{F \le M\}$ is totally bounded in $L_r(P)$ for every fixed M. The class $\mathcal{F}_M$ is P-GC by Theorem 3 and hence this class is totally bounded in $L_1(P)$. But then it is also totally bounded in $L_r(P)$, because $P|f|^r \le P|f|M^{r-1}$ for any f that is bounded by M and we can construct the ϵ-net over $\mathcal{F}_M$ in $L_1(P)$ to consist of functions that are bounded by M.   □

Because a Glivenko-Cantelli class $\mathcal{F}$ with $\|P\|_{\mathcal{F}} < \infty$ is totally bounded in $L_1(P)$ by Lemma 1, it is separable as a subset of $L_1(P)$. A minor generalization of Theorem 2.3.17 in van der Vaart and Wellner (1996) shows that there exists a bijection $f \leftrightarrow \tilde{f}$ of $\mathcal{F}$ onto a class $\tilde{\mathcal{F}} \subset L_1(P)$ such that

- $f = \tilde{f}$ $P-$almost surely for every $f \in \mathcal{F}$.
- there exists a countable subset $\mathcal{G} \subset \tilde{\mathcal{F}}$ such that for every n there exists a measurable set $N_n \subset \mathcal{X}^n$ with $P^n(N_n) = 0$

such that for all $(x_1, \ldots, x_n) \notin N_n$ and $f \in \tilde{\mathcal{F}}$ there exists $\{g_m\} \subset \mathcal{G}$ such that $P|g_m - \tilde{f}| \to 0$ and $(g_m(x_1), \ldots, g_m(x_n)) \to (\tilde{f}(x_1), \ldots, \tilde{f}(x_n))$.

By an adaptation of a theorem due to Talagrand (1987a) (see Theorem 2.3.15 in van der Vaart and Wellner (1996)) a class $\mathcal{F}$ is weak Glivenko-Cantelli if and only if the class $\tilde{\mathcal{F}}$ is weak Glivenko-Cantelli and $\sup_{f \in \mathcal{F}} \mathbb{P}_n|f - \tilde{f}| \to 0$ in outer probability. Construct a "pointwise separable version" $\tilde{\mathcal{F}}_i$ for each of the classes $\mathcal{F}_i$. The classes $\tilde{\mathcal{F}}_i$ possess enough measurability to make the preceding argument work; in particular "pointwise separable version" in the above sense is sufficient for the nearly linearly supremum measurable hypothesis of Giné and Zinn (1984) for both $\mathcal{F}_1, \ldots, \mathcal{F}_k$ and $\varphi(\mathcal{F}_1, \ldots, \mathcal{F}_k)$. Thus the class $\varphi(\tilde{\mathcal{F}}_1, \ldots, \tilde{\mathcal{F}}_k)$ is Glivenko-Cantelli for P.

Now by Lemma 2 there exists for every $\epsilon > 0$ a $\delta > 0$ such that

$$\mathbb{P}_n|f_j - \tilde{f}_j| < \frac{\delta}{k}, \qquad j = 1, \ldots, k,$$

implies

$$\mathbb{P}_n|\varphi(f_1, \ldots, f_k) - \varphi(\tilde{f}_1, \ldots, \tilde{f}_k)| < \epsilon.$$

The theorem follows. □

Lemma 2. Suppose that $\varphi : K \to R$ is continuous and $K \subset R^k$ is compact. Then for every $\epsilon > 0$ there exists $\delta > 0$ 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 it can be $\|x\|_r = \left(\sum_{i=1}^{k} |x_i|^r \right)^{1/r}$, $r \in [1, \infty)$ or $\|x\|_\infty \equiv \max_{1 \leq i \leq k} |x_i|$ for $x = (x_1, \ldots, x_k) \in R^k$.

Proof. Let U_n be uniform on $\{1, \ldots, n\}$, and set $X_n = a_{U_n}$, $Y_n = b_{U_n}$. Then we can write

$$\frac{1}{n} \sum_{i=1}^{n} \|a_i - b_i\| = E\|X_n - Y_n\|$$

and

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

Hence it suffices to show that for every $\epsilon > 0$ there exists $\delta > 0$ such that for all (X, Y) random vectors in $K \subset R^k$,

$$E\|X - Y\| < \delta \qquad \text{implies} \qquad E|\varphi(X) - \varphi(Y)| < \epsilon\,.$$

Suppose not. Then for some $\epsilon > 0$ and for all $m = 1, 2, \dots$ there exists (X_m, Y_m) such that

$$E\|X_m - Y_m\| < \frac{1}{m}, \qquad E|\varphi(X_m) - \varphi(Y_m)| \geq \epsilon\,.$$

But since $\{(X_m, Y_m)\}$ is tight, there exists $(X_{m'}, Y_{m'}) \to_d (X, Y)$. Then it follows that

$$E\|X - Y\| = \lim_{m'\to\infty} E\|X_{m'} - Y_{m'}\| = 0$$

so that $X = Y$ a.s., while on the other hand

$$0 = E|\varphi(X) - \varphi(Y)| = \lim_{m'\to\infty} E|\varphi(X_{m'}) - \varphi(Y_{m'})| \geq \epsilon > 0\,.$$

This contradiction means that the desired implication holds. □

Another potentially useful preservation theorem is one based on building up Glivenko-Cantelli classes from the restrictions of a class of functions to elements of a partition of the sample space. The following theorem is related to the results of Van der Vaart (1996) for Donsker classes.

Theorem 4. 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 $i \neq j$. Suppose that $\mathcal{F}_j \equiv \{f1_{\mathcal{X}_j} : f \in \mathcal{F}\}$ is $P-$Glivenko-Cantelli for each j, and $\mathcal{F}$ has an integrable envelope function F. Then $\mathcal{F}$ is itself $P-$Glivenko-Cantelli.

Proof. Since

$$f = f\sum_{j=1}^{\infty}1_{\mathcal{X}_j} = \sum_{j=1}^{\infty}f1_{\mathcal{X}_j}\,,$$

it follows that

$$E^*\|\mathbb{P}_n - P\|_{\mathcal{F}} \leq \sum_{j=1}^{\infty}E^*\|\mathbb{P}_n - P\|_{\mathcal{F}_j} \to 0$$

by the dominated convergence theorem since each term in the sum converges to zero by the hypothesis that each $\mathcal{F}_j$ is $P-$ Glivenko-Cantelli, and we have

$$E^* \|\mathbb{P}_n - P\|_{\mathcal{F}_j} \leq E^* \mathbb{P}_n(F1_{\mathcal{X}_j}) + P(F1_{\mathcal{X}_j}) \leq 2P(F1_{\mathcal{X}_j})$$

where $\sum_{j=1}^{\infty} P(F1_{\mathcal{X}_j}) = P(F) < \infty$. $\square$

3. Preservation of the Uniform Glivenko - Cantelli Property.

We say that $\mathcal{F}$ is a strong uniform Glivenko-Cantelli class if for all $\epsilon > 0$

$$\sup_{P \in \mathcal{P}(\mathcal{X},\mathcal{A})} Pr_P^* \left(\sup_{m \geq n} \|\mathbb{P}_m - P\|_{\mathcal{F}} > \epsilon \right) \to 0 \qquad \text{as } n \to \infty$$

where $\mathcal{P}(\mathcal{X},\mathcal{A})$ is the set of all probability measures on $(\mathcal{X},\mathcal{A})$. For $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$, $n = 1, 2, \ldots$, and $r \in (0, \infty)$, we define on $\mathcal{F}$ the pseudo-distances

$$
\begin{aligned}
e_{x,r}(f,g) &= \left\{ n^{-1} \sum_{i=1}^{n} |f(x_i) - g(x_i)|^r \right\}^{r^{-1} \wedge 1}, \\
e_{x,\infty}(f,g) &= \max_{1 \leq i \leq n} |f(x_i) - g(x_i)|, \qquad f, g \in \mathcal{F}.
\end{aligned}
$$

Let $N(\epsilon, \mathcal{F}, e_{x,r})$ denote the $\epsilon-$covering number of $(\mathcal{F}, e_{x,r})$, $\epsilon > 0$. Then define, for $n = 1, 2, \ldots$, $\epsilon > 0$, and $r \in (0, \infty]$, the quantities

$$N_{n,r}(\epsilon, \mathcal{F}) = \sup_{x \in \mathcal{X}^n} N(\epsilon, \mathcal{F}, e_{x,r}).$$

Theorem 5. (Dudley, Giné, and Zinn (1991)). Suppose that $\mathcal{F}$ is a class of uniformly bounded functions such that $\mathcal{F}$ is image admissible Suslin. Then the following are equivalent:

(a) $\mathcal{F}$ is a strong uniform Glivenko-Cantelli class.

(b)
$$\frac{\log N_{n,r}(\epsilon, \mathcal{F})}{n} \to 0 \qquad \text{for all } \epsilon > 0$$

for some (all) $r \in (0, \infty]$.

For the definition of the image admissible Suslin property see Dudley (1984), sections 10.3 and 11.1. The following theorem gives natural sufficient conditions for preservation of the uniform Glivenko-Cantelli theorem.

Theorem 6. Suppose that $\mathcal{F}_1, \dots, \mathcal{F}_k$ are classes of uniformly bounded functions on $(\mathcal{X}, \mathcal{A})$ such that $\mathcal{F}_1, \dots, \mathcal{F}_k$ are image-admissible Suslin and strong uniform Glivenko-Cantelli classes. Suppose that $\varphi : R^k \to R$ is continuous. Then $\mathcal{H} \equiv \varphi(\mathcal{F}_1, \dots, \mathcal{F}_k)$ is a strong uniform Glivenko-Cantelli class.

Proof. It follows from Lemma 2 that for any $\epsilon > 0$ there exists $\delta > 0$ such that for any $f_j, g_j \in \mathcal{F}_j$, $j = 1, \dots, k$, $x \in \mathcal{X}^n$,

$$e_{x,1}(f_j, g_j) \leq \frac{\delta}{k}, \quad j = 1, \dots, k$$

implies that

$$e_{x,1}(\varphi(f_1, \dots, f_k), \varphi(g_1, \dots, g_k)) \leq \epsilon.$$

It follows that

$$N_{n,1}(\epsilon, \mathcal{H}) \leq \prod_{j=1}^{k} N_{n,1}(\frac{\delta}{k}, \mathcal{F}_j),$$

and hence that

$$\frac{1}{n} \log N_{n,1}(\epsilon, \mathcal{H}) \leq \sum_{j=1}^{k} \frac{1}{n} \log N_{n,1}(\frac{\delta}{k}, \mathcal{F}_j) \to 0$$

where the convergence follows from part (b) of Theorem 5 with $r = 1$. If we show that $\mathcal{H} = \varphi(\mathcal{F})$ is image admissible Suslin, then the conclusion follows from (b) implies (a) in Theorem 5. We give a proof of a slightly stronger statement in the following lemma.   □

Lemma 3. If $\mathcal{F}_1, \dots, \mathcal{F}_k$ are image admissible Suslin via $(\mathcal{Y}_i, \mathcal{S}_i, T_i)$, $i = 1, \dots, k$, and $\phi : R^k \to R$ is measurable, then $\varphi(\mathcal{F}_1, \dots, \mathcal{F}_k)$ is image admissible Suslin via $(\prod_{i=1}^{k} \mathcal{Y}_i, \prod_{i=1}^{k} \mathcal{S}_i, T)$ for $T(y_1, \dots, y_k) = \varphi(T_1 y_1, \dots, T_k y_k)$.

Proof. There exist Polish spaces $\mathcal{Z}_i$ and measurable maps $g_i : \mathcal{Z}_i \to \mathcal{Y}_i$ which are onto. Then $g : \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_k \to \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_k$ defined by $g(z_1, \dots, z_k) = (g_1(z_1), \dots, g_k(z_k))$ is onto and measurable for $\mathcal{S}_1 \times \cdots \times \mathcal{S}_k$. So $(\prod_i \mathcal{Y}_i, \prod_i \mathcal{S}_i)$ is Suslin. It suffices to check that T is onto and ψ defined by the map

$$\psi(x, y_1, \dots, y_k) = \varphi(T_1 y_1(x), \dots, T_k y_k(x))$$

is measurable. Obviously T is onto, and ψ is measurable because each map $(x, y_i) \mapsto T_i y_i(x)$ is measurable on $\mathcal{X} \times \mathcal{Y}_i$ and hence on $\mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_k$, and hence

$$(x, y_1, \dots, y_k) \mapsto (T_1 y_1(x), \dots, T_k y_k(x))$$

is measurable from $\mathcal{X} \times \mathcal{Y}$ to R^k.   □

4. Nonparametric Maximum Likelihood Estimation: a general result.

Now we prove a general result for nonparametric maximum likelihood estimation in a class of densities. The main proposition in this section is related to results of Pfanzagl (1988) and Van de Geer (1993), (1996). Suppose that $\mathcal{P}$ is a class of densities with respect to a fixed $\sigma-$finite measure μ on a measurable space $(\mathcal{X}, \mathcal{A})$. Suppose that $X_1, \ldots, X_n$ are i.i.d. P_0 with density $p_0 \in \mathcal{P}$. Let

$$\widehat{p}_n \equiv \operatorname{argmax} \mathbb{P}_n \log p \,.$$

For $0 < \alpha \le 1$, let $\varphi_\alpha(t) = (t^\alpha - 1)/(t^\alpha + 1)$ for $t \ge 0$, $\varphi(t) = -1$ for $t < 0$. Then φ_α is bounded and continuous for each $\alpha \in (0, 1]$. For $0 < \beta < 1$ define

$$h_\beta^2(p, q) \equiv 1 - \int p^\beta q^{1-\beta} d\mu \,.$$

Note that $h_{1/2}(p, q) \equiv h(p, q)$ is the Hellinger distance between p and q, and by Hölder's inequality, $h_\beta(p, q) \ge 0$ with equality if and only if $p = q$ a.e. μ.

Proposition 3. Suppose that $\mathcal{P}$ is convex. Then

$$h_{1-\alpha/2}^2(\widehat{p}_n, p_0) \le (\mathbb{P}_n - P_0) \left(\varphi_\alpha(\frac{\widehat{p}_n}{p_0}) \right) \,.$$

In particular, when $\alpha = 1$ we have, with $\varphi \equiv \varphi_1$,

$$h^2(\widehat{p}_n, p_0) \equiv h_{1/2}^2(\widehat{p}_n, p_0) \le (\mathbb{P}_n - P_0) \left(\varphi(\frac{\widehat{p}_n}{p_0}) \right) \,.$$

Corollary 1. Suppose that $\{\varphi(p/p_0) : p \in \mathcal{P}\}$ is a P_0 Glivenko-Cantelli class. Then for each $0 < \alpha \le 1$, $h_{1-\alpha/2}(\widehat{p}_n, p_0) \rightarrow_{a.s.} 0$.

Proof of Proposition 3. It follows from convexity of $\mathcal{P}$ and convexity of $t \rightarrow \varphi_\alpha(1/t)$ that

$$(1) \qquad 0 \le \mathbb{P}_n \varphi_\alpha(\widehat{p}_n/p_0) = (\mathbb{P}_n - P_0) \varphi_\alpha(\widehat{p}_n/p_0) + P_0 \varphi_\alpha(\widehat{p}_n/p_0) \,;$$

see van der Vaart and Wellner (1996) page 330, and Pfanzagl (1988), pages 141 - 143. Now we show that

$$(2) \qquad P_0 \varphi_\alpha(p/p_0) = \int \frac{p^\alpha - p_0^\alpha}{p^\alpha + p_0^\alpha} dP_0 \leq - \left(1 - \int p_0^\beta p^{1-\beta} d\mu \right)$$

for $\beta = 1 - \alpha/2$. Note that this holds if and only if

$$-1 + 2 \int \frac{p^\alpha}{p_0^\alpha + p^\alpha} p_0 d\mu \leq -1 + \int p_0^\beta p^{1-\beta} d\mu \, ,$$

or

$$\int p_0^\beta p^{1-\beta} d\mu \geq 2 \int \frac{p^\alpha}{p_0^\alpha + p^\alpha} p_0 d\mu \, .$$

But his holds if

$$p_0^\beta p^{1-\beta} \geq 2 \frac{p^\alpha p_0}{p_0^\alpha + p^\alpha} \, .$$

With $\beta = 1 - \alpha/2$, this becomes

$$\frac{1}{2}(p_0^\alpha + p^\alpha) \geq p_0^{\alpha/2} p^{\alpha/2} = \sqrt{p_0^\alpha p^\alpha} \, ,$$

and this holds by the arithmetic mean - geometric mean inequality. Thus (2) holds. Combining (2) with (1) yields the claim of the proposition. $\qquad\qquad\square$

5. Example: a result of Schick and Yu.

Our goal in this section is to give another proof of the consistency result of Schick and Yu (1999) for the Non-Parametric Maximum Likelihood Estimator (NPMLE) $\widehat{F}_n$ for "mixed case" interval censored data. Our proof is based on the inequality of the preceding section, and is similar in spirit to results of Van de Geer (1993), (1996).

Suppose that Y is a random variable taking values in $R^+ = [0, \infty)$ with distribution function $F \in \mathcal{F} = \{$all df's F on $R^+\}$. Unfortunately we are not able to observe Y itself. What we do observe is a vector of times $T_K = (T_{K,1}, \ldots, T_{K,K})$ where K, the number of times, is itself random, and the interval $(T_{K,j-1}, T_{K,j}]$ into which Y falls (with $T_{K,0} \equiv 0$, $T_{K,K+1} \equiv \infty$). More formally, we assume that K is an integer-valued random variable, and $\underline{T} = \{T_{k,j}, j = 1, \ldots, k, k = 1, 2, \ldots\}$, is a triangular array of "potential observation times", and that Y and $(K, \underline{T})$ are independent. Let $X = (\Delta_K, T_K, K)$, with a possible value $x = (\delta_k, t_k, k)$, where $\Delta_k =$

$(\Delta_{k,1}, \dots, \Delta_{k,k})$ with $\Delta_{k,j} = 1_{(T_{k,j-1}, T_{k,j}]}(Y)$, $j = 1, 2, \dots, k+1$, and T_k is the kth row of the triangular array $\underline{T}$. Suppose we observe n i.i.d. copies of X; $X_1, X_2, \dots, X_n$, where $X_i = (\Delta^{(i)}_{K^{(i)}}, T^{(i)}_{K^{(i)}}, K^{(i)})$, $i = 1, 2, \dots, n$. Here $(Y^{(i)}, \underline{T}^{(i)}, K^{(i)})$, $i = 1, 2, \dots$ are the underlying i.i.d. copies of $(Y, \underline{T}, K)$.

We first note that conditionally on K and T_K, the vector Δ_K has a multinomial distribution:

$$(\Delta_K | K, T_K) \sim \text{Multinomial}_{K+1}(1, \Delta F_K)$$

where

$$\Delta F_K \equiv (F(T_{K,1}), F(T_{K,2}) - F(T_{K,1}), \dots, 1 - F(T_{K,K})).$$

Suppose for the moment that the distribution G_k of $(T_K | K = k)$ has density g_k and $p_k \equiv P(K = k)$. Then a density of X is given by

$$(3) \quad p_F(x) \equiv p_F(\delta_k, t_k, k) = \prod_{j=1}^{k+1} (F(t_{k,j}) - F(t_{k,j-1}))^{\delta_{k,j}} g_k(t_k) p_k$$

where $t_{k,0} \equiv 0$, $t_{k,k+1} \equiv \infty$. In general,

$$
\begin{aligned}
p_F(x) \equiv p_F(\delta_k, t_k, k) &= \prod_{j=1}^{k+1} (F(t_{k,j}) - F(t_{k,j-1}))^{\delta_{k,j}} \\
(4) \qquad &= \sum_{j=1}^{k+1} \delta_{k,j}(F(t_{k,j}) - F(t_{k,j-1}))
\end{aligned}
$$

is a density of X with respect to the dominating measure ν where ν is determined by the joint distribution of $(K, \underline{T})$, and it is this version of the density of X with which we will work throughout the rest of the paper. Thus the log-likelihood function for F of $X_1, \dots, X_n$ is given by

$$
\begin{aligned}
\frac{1}{n} l_n(F | \underline{X}) &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K^{(i)}+1} \Delta^{(i)}_{K,j} \log \left(F(T^{(i)}_{K^{(i)},j}) - F(T^{(i)}_{K^{(i)},j-1}) \right) \\
&= \mathbb{P}_n m_F
\end{aligned}
$$

where

$$
\begin{aligned}
m_F(X) &= \sum_{j=1}^{K+1} \Delta_{K,j} \log (F(T_{K,j}) - F(T_{K,j-1})) \\
&\equiv \sum_{j=1}^{K+1} \Delta_{K,j} \log (\Delta F_{K,j})
\end{aligned}
$$

and where we have ignored the terms not involving F. We also note that, with $P_0 \equiv P_{F_0}$,

$$P_0 m_F(X) = P_0 \left(\sum_{j=1}^{K+1} \Delta F_{0,K,j} \log \left(\Delta F_{K,j} \right) \right).$$

The Nonparametric Maximum Likelihood Estimator (NPMLE) $\widehat{F}_n$ is the distribution function $\widehat{F}_n(t)$ which puts all its mass at the observed time points and maximizes the log-likelihood $l_n(F|\underline{X})$. It can be calculated via the iterative convex minorant algorithm proposed in Groeneboom and Wellner (1992) for case 2 interval censored data.

By Proposition 3 with $\alpha = 1$ and $\varphi \equiv \varphi_1$ as before, it follows that

$$h^2(p_{\widehat{F}_n}, p_{F_0}) \leq (\mathbb{P}_n - P_0) \left(\varphi(p_{\widehat{F}_n}/p_{F_0}) \right)$$

where φ is bounded and continuous from R to R. Now the collection of functions

$$\mathcal{G} \equiv \{p_F : F \in \mathcal{F}\}$$

is easily seen to be a Glivenko-Cantelli class of functions: this can be seen by first applying Theorem 4 to the collections $\mathcal{G}_k$, $k = 1, 2, \ldots$ obtained from $\mathcal{G}$ by restricting to the sets $K = k$. Then for fixed k, the collections $\mathcal{G}_k = \{p_F(\delta, t_k, k) : F \in \mathcal{F}\}$ are P_0 Glivenko-Cantelli classes since $\mathcal{F}$ is a uniform Glivenko-Cantelli class, and since the functions p_F are continuous transformations of the classes of functions $x \to \delta_{k,j}$ and $x \to F(t_{k,j})$ for $j = 1, \ldots, k+1$, and hence $\mathcal{G}$ is $P-$Glivenko-Cantelli by Theorem 3. Note that the single function p_{F_0} is trivially P_0- Glivenko-Cantelli since it is uniformly bounded, and the single function $(1/p_{F_0})$ is also P_0- GC since $P_0(1/p_{F_0}) < \infty$. Thus by Proposition 2 with $g = (1/p_{F_0})$ and $\mathcal{F} = \mathcal{G} = \{p_F : F \in \mathcal{F}\}$, it follows that $\mathcal{G}' \equiv \{p_F/p_{F_0} : F \in \mathcal{F}\}$ is P_0-Glivenko-Cantelli. Finally another application of Theorem 3 shows that the collection

$$\mathcal{H} \equiv \{L_1(p_F/p_{F_0}) : F \in \mathcal{F}\} = \{\varphi(p_F/p_{F_0}) : F \in \mathcal{F}\}$$

is also P_0-Glivenko-Cantelli. When combined with Proposition 3, this yields the following theorem:

Theorem 7. The NPMLE $\widehat{F}_n$ satisfies

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

To relate this result to a recent theorem of Schick and Yu (1999), it remains only to understand the relationship between their $L_1(\mu)$ and the Hellinger metric h between p_F and p_{F_0}.

Let $\mathcal{B}$ denote the collection of Borel sets in R. On $\mathcal{B}$ we define measures μ and $\widetilde{\mu}$ as follows: For $B \in \mathcal{B}$,

$$(5) \qquad \mu(B) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(T_{k,j} \in B | K = k) ,$$

and

$$(6) \qquad \widetilde{\mu}(B) = \sum_{k=1}^{\infty} P(K = k) \frac{1}{k} \sum_{j=1}^{k} P(T_{k,j} \in B | K = k) .$$

Let d be the $L_1(\mu)$ metric on the class $\mathcal{F}$; thus for $F_1, F_2 \in \mathcal{F}$,

$$d(F_1, F_2) = \int |F_1(t) - F_2(t)| d\mu(t) .$$

The measure μ was introduced by Schick and Yu (1999); note that μ is a finite measure if $E(K) < \infty$. Note that $d(F_1, F_2)$ can also be written in terms of an expectation as:

$$(7) \qquad d(F_1, F_2) = E_{(K,\underline{T})} \left[\sum_{j=1}^{K} |(F_1(T_{K,j}) - F_2(T_{K,j})| \right] .$$

As Schick and Yu (1999) observed, consistency of the NPMLE $\widehat{F}_n$ in $L_1(\mu)$ holds under virtually no further hypotheses.

Theorem 8. (Schick and Yu). Suppose that $E(K) < \infty$. Then $d(\widehat{F}_n, F_0) \rightarrow_{a.s.} 0$.

Proof. We will show that Theorem 8 follows from Theorem 7 and the following Lemma.

Lemma 4.

$$\frac{1}{2} \left\{ \int |\widehat{F}_n - F_0| d\widetilde{\mu} \right\}^2 \le h^2(p_{\widehat{F}_n}, p_{F_0}) .$$

Proof. We know that

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

where, with $y_{k,0} = -\infty$, $y_{k,k+1} = \infty$,

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

while

$$d_{TV}(p_{\widehat{F}_n}, p_{F_0}) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k+1} \int |[\widehat{F}_n(y_{k,j}) - \widehat{F}_n(y_{k,j-1})]$$
$$- [F_0(y_{k,j}) - F_0(y_{k,j-1})]| dG_k(y).$$

Note that

$$\sum_{j=1}^{k+1} |[\widehat{F}_n(y_{k,j}) - \widehat{F}_n(y_{k,j-1})] - [F_0(y_{k,j}) - F_0(y_{k,j-1})]|$$

$$= \sum_{j=1}^{k+1} |(\widehat{F}_n - F_0)(y_{k,j-1}, y_{k,j}]|$$

$$\geq \max_{1 \leq j \leq k+1} |\widehat{F}_n(y_{k,j}) - F_0(y_{k,j})|,$$

so integrating across this inequality with respect to $G_k(y)$ yields

$$\sum_{j=1}^{k+1} \int |[\widehat{F}_n(y_{k,j}) - \widehat{F}_n(y_{k,j-1})] - [F_0(y_{k,j}) - F_0(y_{k,j-1})]| \, dG_k(y)$$

$$\geq \max_{1 \leq j \leq k} \int |\widehat{F}_n(y_{k,j}) - F_0(y_{k,j})| \, dG_{k,j}(y_{k,j})$$

$$\geq \frac{1}{k} \sum_{j=1}^{k} \int |\widehat{F}_n(y_{k,j}) - F_0(y_{k,j})| \, dG_{k,j}(y_{k,j}).$$

By multiplying across by $P(K = k)$ and summing over k, this yields

$$d_{TV}(p_{\widehat{F}_n}, p_{F_0}) \geq \int |\widehat{F}_n - F_0| d\tilde{\mu},$$

and hence

$$(8) \qquad h^2(p_{\widehat{F}_n}, p_{F_0}) \;\geq\; \frac{1}{2} \left\{ \int |\widehat{F}_n - F_0| d\tilde{\mu} \right\}^2.$$

$\square$

The measure $\tilde{\mu}$ figuring in Lemma 4 is not the same as the measure μ of Schick and Yu (1999) because of the factor $1/k$. Note that this factor means that the measure $\tilde{\mu}$ is always a finite measure, even if $E(K) = \infty$. It is clear that

$$\tilde{\mu}(B) \leq \mu(B)$$

for every Borel set B, and that $\mu \prec\prec \tilde{\mu}$. The following lemma (Lemma 2.2 of Schick and Yu (1999)) together with Lemma 4 shows that Theorem 7 implies the result of Schick and Yu once again:

Lemma 5. Suppose that μ and $\tilde{\mu}$ are two finite measures, and that $g, g_1, g_2, \ldots$ are measurable functions with range in $[0, 1]$. Suppose that μ is absolutely continuous with respect to $\tilde{\mu}$. Then $\int |g_n - g| d\tilde{\mu} \to 0$ implies that $\int |g_n - g| d\mu \to 0$.

Proof. Write

$$\int |g_n - g| d\mu = \int |g_n - g| \frac{d\mu}{d\tilde{\mu}} d\tilde{\mu}$$

and use the dominated convergence theorem applied to a.e. convergent subsequences.   □

6. Example: generalizing the result of Schick and Yu.

Our goal in this section is to give a generalization of the consistency result of Schick and Yu (1999).

Suppose that Y is a random variable taking values in $\mathcal{Y}$. Suppose that Y has distribution Q on the measurable space $(\mathcal{Y}, \mathcal{B})$. Unfortunately we are not able to observe Y itself. What we do observe is a vector of random sets $C_K = (C_{K,1}, \ldots, C_{K,K})$ where K, the number of sets is itself random, and the set $\{C_{K,j}\}_{j=1}^K$ form a partition of $\mathcal{Y}$: $C_{K,j} \cap C_{K,j'} = \emptyset$ if $j \neq j'$ and $\cup_{j=1}^K C_{K,j} = \mathcal{Y}$. More formally, we assume that K is an integer-valued random variable, and $\underline{C} = \{C_{k,j}, j = 1, \ldots, k, k = 1, 2, \ldots\}$, is a triangular array of "random sets", and that Y and $(K, \underline{C})$ are independent. Let $X = (\Delta_K, C_K, K)$, with a possible value $x = (\delta_k, c_k, k)$, where $\Delta_k = (\Delta_{k,1}, \ldots, \Delta_{k,k})$ with $\Delta_{k,j} = 1_{C_{k,j}}(Y)$, $j = 1, 2, \ldots, k$, and C_k is the k-th row of the triangular array $\underline{C}$. Suppose we observe n i.i.d. copies of X; $X_1, X_2, \ldots, X_n$, where $X_i = (\Delta_{K^{(i)}}^{(i)}, C_{K^{(i)}}^{(i)}, K^{(i)})$, $i = 1, 2, \ldots, n$. Here $(Y^{(i)}, \underline{C}^{(i)}, K^{(i)})$, $i = 1, 2, \ldots$ are the underlying i.i.d. copies of $(Y, \underline{C}, K)$.

We first note that conditionally on K and C_K, the vector Δ_K has a multinomial distribution:

$$(\Delta_K | K, C_K) \sim \text{Multinomial}_K(1, Q_K)$$

where

$$Q_K \equiv (Q(C_{K,1}), Q(C_{K,2}), \ldots, Q(C_{K,K})) \, .$$

Suppose for the moment that the distribution G_k of $(C_K|K = k)$ has density g_k and $p_k \equiv P(K = k)$. Then a density of X is given by

$$(9) \qquad p_Q(x) \equiv p_Q(\delta_k, c_k, k) = \prod_{j=1}^{k} Q(c_{k,j})^{\delta_{k,j}} g_k(c_k) p_k .$$

In general,

$$(10) \qquad p_Q(x) \equiv p_Q(\delta_k, c_k, k) = \prod_{j=1}^{k} Q(c_{k,j})^{\delta_{k,j}} = \sum_{j=1}^{k} \delta_{k,j} Q(c_{k,j})$$

is a density of X with respect to the dominating measure ν where ν is determined by the joint distribution of $(K, \underline{C})$, and it is this version of the density of X with which we will work throughout the rest of the paper. Thus the log-likelihood function for Q of $X_1, \ldots, X_n$ is given by

$$\frac{1}{n} l_n(Q|\underline{X}) \quad = \quad \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K^{(i)}} \Delta_{K,j}^{(i)} \log Q(C_{K^{(i)},j}^{(i)}) = \mathbb{P}_n m_Q$$

where

$$m_Q(X) \quad = \quad \sum_{j=1}^{K} \Delta_{K,j} \log Q(C_{K,j})$$

and where we have ignored the terms not involving Q. We also note that, with $P_0 \equiv P_{Q_0}$,

$$P_0 m_Q(X) = P_0 \left(\sum_{j=1}^{K} Q_0(C_{K,j}) \log Q(C_{K,j}) \right) .$$

The Nonparametric Maximum Likelihood Estimator (NPMLE) $\widehat{Q}_n$ is a probability measure $\widehat{Q}_n$ which maximizes the log-likelihood $l_n(Q|\underline{X})$. Here we will bypass the many interesting existence, characterization, and computational issues connected with the NPMLE $\widehat{Q}_n$, and focus instead on the issue of consistency once we have the NPMLE in hand.

By Proposition 3 with $\alpha = 1$ it follows that

$$h^2(p_{\widehat{Q}_n}, p_{Q_0}) \leq (\mathbb{P}_n - P_0) \left(\varphi(p_{\widehat{Q}_n}/p_{Q_0}) \right) .$$

where φ is bounded and continuous. Now the collection of functions

$$\mathcal{G} \equiv \{ p_Q : \ Q \in \mathcal{Q} \}$$

where $\mathcal{Q}$ is the collection of all probability distributions on $(\mathcal{Y}, \mathcal{B})$, is easily seen to be a Glivenko-Cantelli class of functions: this can be seen by first applying Theorem 4 to the collections $\mathcal{G}_k$, $k = 1, 2, \ldots$ obtained from $\mathcal{G}$ by restricting to the sets $K = k$. The next step is to show that the collections

$$\mathcal{G}_k \equiv \{p_Q(\cdot, \cdot, k) : Q \in \mathcal{Q}\}$$

are P_0-Glivenko-Cantelli for each k. To this end, suppose that $(\mathcal{Y}, \mathcal{B})$ be a measurable space and let $\mathcal{C}$ be a universal GC - class of sets in $\mathcal{Y}$ (i.e. $Q-$GC for every probability measure Q on $(\mathcal{Y}, \mathcal{B})$). Let $(T, \mathcal{T}, G)$ be a probability space, and let $t \mapsto C_t$ be a map with values in $\mathcal{C}$.

Lemma 6. Suppose that the collection $\{\{t : y \in C_t\} : y \in \mathcal{Y}\}$ is $G-$GC. Then the collection of functions

$$t \mapsto Q(C_t)$$

where Q ranges over all probability measures Q on $\mathcal{Y}$ is $G-$GC.

Remark: By Assouad (1983), Proposition 2.12, page 246, together with Corollary 1.10, page 241, it follows that $\{\{t \in T : y \in C_t\} : y \in \mathcal{Y}\}$ is a VC-class of subsets of T if and only if $\{C_t : t \in T\}$ is a VC-class of subsets of $\mathcal{Y}$. Thus if we assume that all the subsets $C = C_t$ arising in the partitions are elements of a VC-class $\mathcal{C}$, then, subject to being suitably measurable, the hypothesis of Lemma 6 is satisfied (and the class in question is even universal Glivenko-Cantelli).

Proof. For Q the Dirac measure at y (unit mass at y), the function $t \mapsto Q(C_t)$ becomes

$$t \mapsto 1_{C_t}(y).$$

By assumption, the set of all such functions, with y ranging over $\mathcal{Y}$, is $G-$GC. Then so is the set of all functions

$$t \mapsto \frac{1}{m} \sum_{i=1}^{m} 1_{C_t}(y_i), \qquad y_1, \ldots, y_m \in \mathcal{Y}, m \in \mathbb{N}.$$

Let $Y_1, \ldots, Y_m$ be an i.i.d. sample from a given Q. Because $\mathcal{C}$ is $Q-$GC we have

$$\sup_{C \in \mathcal{C}} \left| \frac{1}{m} \sum_{i=1}^{m} 1\{Y_i \in C\} - P(C) \right| \to 0 \qquad \text{a.s.}^{*}.$$

Then there exists a sequence $y_1, y_2, \ldots$ in $\mathcal{Y}$ such that

$$\sup_{C \in \mathcal{C}} \left| \frac{1}{m} \sum_{i=1}^{m} 1\{y_i \in C\} - P(C) \right| \to 0,$$

and consequently

$$\sup_{t \in T} \Big| \frac{1}{m} \sum_{i=1}^{m} 1\{y_i \in C_t\} - P(C_t) \Big| \to 0 \,.$$

It follows that the maps $t \to P(C_t)$ are the (uniform) limits of sequences of functions from a $G-$GC class. Hence they are $G-$GC. $\square$

Note on measurability: It is assumed implicitly that the sets $\{t : y \in C_t\}$ are measurable in T for every $y \in \mathcal{Y}$. The proof shows that the maps $t \mapsto P(C_t)$ are then also measurable.

It follows that for fixed k, the collections $\mathcal{G}_k = \{p_Q(\delta, c_k, k) : Q \in \mathcal{Q}\}$ are P_0-Glivenko-Cantelli, and since the functions p_Q are continuous transformations of the classes of functions $x \to \delta_{k,j}$ and $x \to Q(c_{k,j})$ for $j = 1, \ldots, k$, and hence $\mathcal{G}$ is $P-$Glivenko-Cantelli by Theorem 3. Note that the single function p_{Q_0} is trivially P_0-Glivenko-Cantelli since it is uniformly bounded. Now another application of Theorem 3 shows that the collection

$$\mathcal{H} \equiv \{L_1(p_Q/p_{Q_0}) : Q \in \mathcal{Q}\} = \{\varphi(p_Q/p_{Q_0}) : Q \in \mathcal{Q}\}$$

is also P_0-Glivenko-Cantelli. When combined with Proposition 3, this yields the following theorem.

Theorem 9. If all $C_{K,j} \in \mathcal{C}$, a VC collection of subsets of $\mathcal{X}$, then the NPMLE $\widehat{Q}_n$ satisfies

$$h(p_{\widehat{Q}_n}, p_{Q_0}) \to_{a.s.} 0 \,.$$

To obtain a statement analogous to the theorem of Schick and Yu (1999), let Σ denote a sigma algebra for the space $\mathcal{C}$ of subsets: we are assuming that $P(\cap_{j=1}^{K}[C_{K,j} \in \mathcal{C}]) = 1$. Thus $(\mathcal{C}, \Sigma)$ is a measurable space. On Σ we define the measure μ as follows:

$$(11) \quad \mu(B) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(C_{k,j} \in B | K = k), \qquad B \in \Sigma.$$

Note that

$$d_{TV}(p_{\widehat{Q}_n}, p_{Q_0}) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} \int |\widehat{Q}_n(c_{k,j}) - Q_0(c_{k,j})| dG_k(c)$$

$$(12) \qquad\qquad = \int |\widehat{Q}_n(c) - Q_0(c)| \, d\mu(c) \equiv d(\widehat{Q}_n, Q_0) \,.$$

Here is a generalization of the theorem of Schick and Yu (1999).

Theorem 10. The NPMLE $\widehat{Q}_n$ satisfies $d(\widehat{Q}_n, Q_0) \to_{a.s.} 0$.

Proof. Theorem 10 follows from Theorem 9, (12) and the observation that we then have

$$d(\widehat{Q}_n, Q_0) = d_{TV}(p_{\widehat{Q}_n}, p_{Q_0}) \leq \sqrt{2} h(p_{\widehat{Q}_n}, p_{Q_0}).$$

$\square$

Example 1. (Mixed case interval censoring in R^2). Suppose that $Y \sim Q_0$ takes values in R^{+2}, and the partitions $\{C_{K,j}\}_{j=1}^{K}$ are obtained as the natural rectangles formed from two increasing collections $0 \leq S_{K_1,1} \leq \ldots \leq S_{K_1,K}$ and $0 \leq T_{K_2,1} \leq \ldots \leq T_{K_2,K}$ on the two coordinate axes. Then $K = (K_1 + 1)(K_2 + 1)$, and all the elements of the partitions are elements of the VC-class of rectangles in R^2. It follows that the NPMLE $\widehat{Q}_n$ of Q_0 satisfies

$$\int_{s,t \in R^{+2}} |\widehat{Q}_n(s,t] - Q_0(s,t]| d\mu(s,t) \to 0 \qquad \text{a.s.}^* .$$

References

: Assouad, P. (1983). Densité et dimension. *Annales de l'Institut Fourier* **33**, 233 - 282.

: Dudley, R. M. (1984). *A Course on Empirical Processes. Ecole d'été de probabilités de St.-Flour, 1982.* Lecture Notes in Mathematics **1097**, 1-142, Springer, Berlin.

: Dudley, R. M. (1998a). "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. (1998b). Personal communication.

: 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. Probability* **12**, 929 - 989.

: Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* DMV Seminar Band 19, Birkhäuser, Basel.

: Pfanzagl, J. (1988). "Consistency of maximum likelihood estimators for certain nonparametric families, in particular: mixtures." *J. Statist. Planning and Inference* **19**, 137 - 158.

: SCHICK, A. AND YU, Q. (1999), "Consistency of the GMLE with Mixed Case Interval-Censored Data," *Scand. J. Statist.*, to appear.

: TALAGRAND, M. (1987a). "Measurability problems for empirical measures." *Ann. Probability* **15**, 204 - 212.

: TALAGRAND, M. (1987b). "The Glivenko-Cantelli problem." *Ann. Probability* **15**, 837 - 870.

: TALAGRAND, M. (1996). "The Glivenko-Cantelli problem, ten years later." *J. Theor. Prob.* **9**, 371 - 384.

: VAN DE GEER, S. (1993). "Hellinger consistency of certain nonparametric maximum likelihood estimators." *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.

: VAN DER VAART, A. W. (1996). New Donsker classes. *Ann. Probability* **24**, 2128 - 2140.

: VAN DER VAART, A. W., AND WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer-Verlag, New York.

DEPARTMENT OF MATHEMATICS   DEPARTMENT OF STATISTICS
FREE UNIVERSITY   UNIVERSITY OF WASHINGTON
DE BOELELAAN 1081A   P.O. BOX 354322
1081 HV AMSTERDAM   SEATTLE, WASHINGTON 98195-4322
THE NETHERLANDS   U.S.A.
e-mail: aad@cs.vu.nl   *e-mail: jaw@stat.washington.edu*