

Preservation Results

Kosorok: 9.3 - 9.4

August 11, 2021

In chapter 9, we've focused on computing entropy for empirical process with the purpose of determining whether classes of functions are G-C, Donsker, or neither.

In sections 9.3 and 9.4, we will discuss the modifications of G-C and Donsker classes of functions under which G-C and Donsker properties are preserved.

This will allow us to describe methods for building new G-C and Donsker classes.

9.3 Glivenko-Cantelli Preservation

9.4 Donsker Preservation

Subsection 9.3: Glivenko-Cantelli Preservation

9.3: Introduction

As discussed, we aim to describe the methods in which Glivenko-Cantelli classes can be generated from other G-C classes.

These results are useful in establishing the consistency properties of Z and M estimators and their bootstrapped versions.

Our first example is trivial: for P -G-C classes \mathcal{F} and \mathcal{G} , then $\mathcal{F} \cup \mathcal{G}$ is also P -G-C, as is any subset of $\mathcal{F} \cup \mathcal{G}$.

More substantial preservation results stem from theorem 9.25, a modification of theorem 3 of van der Vaart and Wellner:

Theorem 9.25

Suppose that $\mathcal{F}_1, \dots, \mathcal{F}_k$ are strong P -G-C classes of functions with $\max_{1 \leq j \leq k} \|P\|_{\mathcal{F}_j} < \infty$ and that $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ is continuous. Then the class $\mathcal{H} \equiv \phi(\mathcal{F}_1, \dots, \mathcal{F}_k)$ is strong P -G-C provided it has an integrable envelope.

The proof is omitted in Kosorok, but presented in van der Vaart and Wellner (2000), pg. 115 - 117. We present the proof for the case when $\mathcal{F}_1, \dots, \mathcal{F}_k$ are suitably measurable.

Proof of Theorem 9.25

Assume that the classes \mathcal{F}_i are appropriately measurable. Let F_1, \dots, F_k, H be the envelopes for $\mathcal{F}_1, \dots, \mathcal{F}_k, \mathcal{H}$, respectively and define $F = F_1 \vee \dots \vee F_k$. Let $\mathcal{F} \equiv \mathcal{F}_1 \times \dots \times \mathcal{F}_k$. For $M \in (0, \infty)$, define

$$\mathcal{H}_M \equiv \{\phi(f)1\{F \leq M\} : f = (f_1, \dots, f_k) \in \mathcal{F}\}$$

Now, consider:

$$\|(\mathbb{P}_n - P)\phi(f)\|_{\mathcal{F}} \leq (\mathbb{P}_n + P)H1\{F > M\} + \|(\mathbb{P}_n - P)h\|_{\mathcal{H}_M}$$

The first term on the right converges to 0 as $M \rightarrow \infty$. Thus, our problem reduces to showing that \mathcal{H}_m is P-G-C for every fixed M . Our next steps tackle this aim. We first need VW-Lemma 2, which we state without proof.

VW-Lemma 2

Let $K \subset \mathbb{R}^k$ be compact and $\phi : K \mapsto \mathbb{R}$ be continuous. Then for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all n and for all $a_1, \dots, a_n, b_1, \dots, b_n \in K$,

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

implies that:

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

Here, the norm $\|\cdot\|$ is very general, and can be any norm on \mathbb{R}^k .

Proof of Theorem 9.25

Back in the proof of Theorem 9.25, choose a $\epsilon > 0$, and let $\delta > 0$ exist such that Lemma 2 is satisfied for $\phi : [-M, M]^k \mapsto \mathbb{R}$ for the general norm $\|\cdot\|$ being the L_1 -norm. Then for any $(f_j, g_j) \in \mathcal{F}_j, j = 1, \dots, k$:

$$\mathbb{P}_n |f_j - g_j| 1\{F_j \leq M\} < \frac{\delta}{k}, j = 1, \dots, k$$

implies that

$$\mathbb{P}_n |\phi(f_1, \dots, f_k) - \phi(g_1, \dots, g_k)| 1\{F \leq M\} < \epsilon$$

it follows then that:

$$N(\epsilon, \mathcal{H}_M, L_1(\mathbb{P}_n)) \leq \prod_{j=1}^k N\left(\frac{\delta}{k}, F_j 1\{F_j \leq M\}, L_1(\mathbb{P}_n)\right)$$

Proof of Theorem 9.25

It follows that $E^* \log N(\epsilon, \mathcal{H}_M, L_1(\mathbb{P}_n)) = o(n)$ for every $\epsilon > 0, M < \infty$.

This implies that $E^* \log N(\epsilon, (\overline{\mathcal{H}}_M)_N, L_1(\mathbb{P}_n)) = o(n)$ where $(\overline{\mathcal{H}}_M)_N$ contains the functions $h1\{H \leq N\}$ for $h \in \overline{\mathcal{H}}_M$. Thus, $\overline{\mathcal{H}}_M$ is strong G-C for P , and \mathcal{H} is weak G-C.

Because \mathcal{H} has an integrable envelope and is weak G-C, it is thus strong G-C by VW-Lemma 2.4.5 (1996).

This concludes the proof for the case when the classes $\mathcal{F}_1, \dots, \mathcal{F}_k$ are suitably measurable. A generalization is available in van der Vaart and Wellner (2000).

Corollary 9.26 gives obvious consequences of theorem 9.25:

Corollary 9.26

Let \mathcal{F} and \mathcal{G} be P -G-C classes with respective integrable envelopes F and G . Then the following hold:

- (i) $\mathcal{F} + \mathcal{G}$ is P -G-C.
- (ii) $\mathcal{F} \cdot \mathcal{G}$ is P -G-C provided that $P[FG] < \infty$
- (iii) Let R be the union of the ranges of the functions in \mathcal{F} , and let $\psi : \overline{R} \mapsto \mathbb{R}$ be continuous. Then $\psi(\mathcal{F})$ is P -G-C provided it has an integrable envelope.

A short proof is presented on the following slide.

Proof of Corollary 9.26

- (i) $F + G < \infty$ provides a trivial integrable envelope for class $\mathcal{F} + \mathcal{G}$
- (ii) Since $(x, y) \mapsto xy$ is continuous in \mathbb{R}^2 , theorem 9.25 applies directly.
- (iii) Follows from 9.25 since there exists a continuous extension of ψ , $\tilde{\psi}$, such that $\|P\tilde{\psi}(f)\|_{\mathcal{F}} = \|P\psi(f)\|_{\mathcal{F}}$.

Example of Corollary 9.26

Note that the "preservation of product" G-C result, part (ii) of 9.26 does not hold for Donsker classes in general.

This result is particularly useful for formulating master theorems for bootstrapped Z and M estimators.

Consider the following example of this, continued on the next slide:

Example of Corollary 9.26

Consider verifying the validity of the bootstrap for the parametric Z -estimator, $\hat{\theta}_n$, obtained as the zero of $\theta \mapsto \mathbb{P}_n \psi_\theta$ for $\theta \in \Theta$ for an arbitrary random function ψ_θ . Let $\Psi(\theta) = P\psi_\theta$. Assume that the parameter is identifiable, i.e. for $\{\theta_n\} \in \Theta$, $\Psi(\theta_n) \rightarrow 0$ implies $\theta_n \rightarrow \theta_0$. To obtain consistency, we can assume that the class $\{\psi_\theta, \theta \in \Theta\}$ is P -G-C, yielding $\hat{\theta}_n \xrightarrow{as*} \theta_0$

Using arguments laid out in section 2.2.5, if Ψ is appropriately differentiable and the parameter is identifiable, sufficient additional conditions for asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ are that $\{\psi_\theta : \theta \in \Theta\}$ is strong P -G-C, that $\{\psi_\theta : \theta \in \Theta, \|\theta - \theta_0\| \leq \delta\}$ is P -Donsker for some $\delta > 0$, and that $P\|\psi_\theta - \psi_{\theta_0}\|^2 \rightarrow 0$ as $\theta \rightarrow \theta_0$.

Example of Corollary 9.26

As we will see in later chapters, an important step in showing the validity of the bootstrap estimate $\hat{\theta}_n^\circ$ is to show that it is unconditionally consistent for θ_0 .

If we use a weighted bootstrap with i.i.d non-negative weights ξ_1, \dots, ξ_n which are independent of the data and satisfy that $E\xi_1 = 1$, then (ii) of corollary 9.26 gives us that $\mathcal{F} \equiv \{\xi\psi_\theta : \theta \in \Theta\}$ is P -G-C.

This follows since both the class $\{\xi\}$ and $\{\psi_\theta : \theta \in \Theta\}$ are P -G-C and since the product class \mathcal{F} has an integrable envelope by lemma 8.13.

Lemma 9.27 Introduction

Lemma 9.27 gives a useful result on covariance estimation. The goal is to describe the conditions under which the limiting covariance of \mathbb{G}_n , indexed by \mathcal{F} , can be consistently estimated.

Recall that this covariance value is $\sigma(f, g) \equiv Pfg - PfPg$ and has an estimator $\hat{\sigma}(f, g) \equiv \mathbb{P}_n fg - \mathbb{P}_n f \mathbb{P}_n g$.

Although knowledge of the covariance matrix is not sufficient to obtain inference on $\{Pf : f \in \mathcal{F}\}$, the information it does contain is still useful.

Lemma 9.27

Let \mathcal{F} be Donsker. Then $\|\hat{\sigma}(f, g) - \sigma(f, g)\|_{\mathcal{F} \cdot \mathcal{F}} \xrightarrow{\text{as}^*} 0$ if and only if $P^* \|f - Pf\|_{\mathcal{F}}^2 < \infty$

Proof

\Leftarrow Note since \mathcal{F} is Donsker, it is also G-C. Hence $\dot{\mathcal{F}} \equiv \{\dot{f} : f \in \mathcal{F}\}$ is also G-C, where for any f , $\dot{f} \equiv f - Pf$. Assume that $P^* \|f - Pf\|_{\mathcal{F}}^2 < \infty$. Theorem 9.25 provides that $\dot{\mathcal{F}} \cdot \dot{\mathcal{F}}$ is also G-C. Uniform consistency of $\hat{\sigma}$ follows since for any $f, g \in \mathcal{F}$, $\hat{\sigma}(f, g) - \sigma(f, g) = (\mathbb{P}_n - P)\dot{f}\dot{g} - \mathbb{P}_n\dot{f}\mathbb{P}_n\dot{g}$.

\Rightarrow Now, assume that $\|\hat{\sigma}(f, g) - \sigma(f, g)\|_{\mathcal{F} \cdot \mathcal{F}} \xrightarrow{\text{as}^*} 0$. Thus, $\dot{\mathcal{F}} \cdot \dot{\mathcal{F}}$ is G-C. Lemma 8.13 thus implies that $P^* \|f - Pf\|_{\mathcal{F}}^2 = P^* \|fg\|_{\dot{\mathcal{F}} \cdot \dot{\mathcal{F}}} < \infty$.

Theorem 9.28

Theorem 9.28 gives equivalent conditions for a functional class \mathcal{F} to be strong Glivenko-Cantelli.

Theorem 9.28

Let \mathcal{F} be a class of measurable functions. Then the following are equivalent:

- (i) \mathcal{F} is strong P -G-C
- (ii) $E^* \|\mathbb{P}_n - P\|_{\mathcal{F}} \rightarrow 0$ and $E^* \|f - Pf\|_{\mathcal{F}} < \infty$
- (iii) $\|\mathbb{P}_n - P\|_{\mathcal{F}} \xrightarrow{P} 0$ and $E^* \|f - Pf\|_{\mathcal{F}} < \infty$

Proof of Theorem 9.28

Proof Since $\mathbb{P}_n - P$ does not change when the class \mathcal{F} is replaced by $\{f - Pf : f \in \mathcal{F}\}$, we can assume that $\|P\|_{\mathcal{F}} = 0$, WLOG.

(i) \implies (ii): By lemma 8.13, \mathcal{F} being P -G-C implies $E^* \|f\|_{\mathcal{F}} < \infty$. Pick $0 < M < \infty$, and note that:

$$E^* \|\mathbb{P}_n - P\|_{\mathcal{F}} \leq E^* \|(\mathbb{P}_n - P)f \times 1\{F \leq M\}\|_{\mathcal{F}} + 2E^* [F \times 1\{F > M\}]$$

applying corollary 9.26's product preservation to the class $\mathcal{F} \cdot 1\{F > M\}$ yields that it is strong P -G-C, and thus $E^* \|(\mathbb{P}_n - P)f \times 1\{F \leq M\}\|_{\mathcal{F}} \rightarrow 0$ for any M , while the second term can be made arbitrarily small by increasing M . Thus, $E^* \|\mathbb{P}_n - P\|_{\mathcal{F}} \rightarrow 0$

(ii) \implies (iii): Follows directly from the first condition of (ii).

(iii) \implies (i): By assuming that the envelope F is integrable, lemma 8.16 yields that there is a version of $\|\mathbb{P}_n - P\|_{\mathcal{F}}^*$ that converges outer almost surely to a constant. The first condition of (iii) guarantees that this constant must be zero.

Subsection 9.4: Donsker Preservation

Theorem 9.29

We consider techniques for building Donsker classes from other Donsker classes. Theorem 9.29, to follow, gives such results for subsets, pointwise closure, and symmetric convex hulls.

For a class \mathcal{F} of real-valued, measurable functions on the sample space \mathcal{X} , let $\overline{\mathcal{F}}^{(P,2)}$ be the set of $f : \mathcal{X} \mapsto \mathbb{R}$ for which there exists a sequence $\{f_m\} \in \mathcal{F}$ such that $f_m \rightarrow f$ pointwise and in $L_2(P)$.

Let $\text{sconv}\mathcal{F}$ be defined as all functions f which can be represented as $f = \sum_{i=1}^m \alpha_i f_i$ for constants α_i satisfying $\sum_{i=1}^m |\alpha_i| = 1$ and $f_i \in \mathcal{F}$.

Finally, let $\overline{\text{sconv}}^{(P,2)}\mathcal{F}$ be the pointwise and $L_2(P)$ closure of the class $\text{sconv}\mathcal{F}$.

Theorem 9.29

Let \mathcal{F} be a P -Donsker class. Then:

- (i) Any $\mathcal{G} \subset \mathcal{F}$ is P -Donsker.
- (ii) $\overline{\mathcal{F}}^{(P,2)}$ is P -Donsker.
- (iii) $\overline{\text{sconv}}^{(P,2)} \mathcal{F}$ is P -Donsker.

Proof to follow.

Proof

(i) Recall that weak convergence relies on marginal convergence plus asymptotic equicontinuity. Since equicontinuity is dependent on the maximum modulus of continuity, which cannot increase on a smaller set, the result holds.

(ii) Similar to our proof of theorem 9.28, we can assume that both \mathcal{F} and $\overline{\mathcal{F}}^{(P,2)}$ are mean zero classes. For a class \mathcal{G} , denote the modulus of continuity as:

$$M_{\mathcal{G}}(\delta) \equiv \sup_{f, g \in \mathcal{G} : \|f - g\|_{P,2} < \delta} |\mathbb{G}_n(f - g)|$$

Proof (cont).

(ii) fix $\delta > 0$. We can choose $f, g \in \overline{\mathcal{F}}^{(P,2)}$ such that $|\mathbb{G}_n(f - g)|$ is arbitrarily close to $M_{\overline{\mathcal{F}}^{(P,2)}}(\delta)$ and $\|f - g\|_{P,2} < \delta$.

Now choose $f_*, g_* \in \mathcal{F}$ such that $\|f - f_*\|_{P,2}$ and $\|g - g_*\|_{P,2}$ are arbitrarily small given the data. Since δ is arbitrary, we obtain that asymptotic equicontinuity in probability of $\overline{\mathcal{F}}^{(P,2)}$ follows from the asymptotic equicontinuity of $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$.

(iii) is found in VW, theorem 2.10.3. The proof involves proving that the symmetric convex hull is pre-Gaussian, followed by an application the almost sure representation theorem using perfect maps. Following the Kosorok text, we omit the proof here.

Theorem 9.30

Theorem 9.30, stated without proof, is theorem 2.10.6 of VW, and one of the most useful Donsker preservation results for statistical applications.

Theorem 9.30

Let $\mathcal{F}_1, \dots, \mathcal{F}_k$ be Donsker classes with $\max_{1 \leq i \leq k} \|P\|_{\mathcal{F}_i} < \infty$. Let $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ satisfy:

$$|\phi \circ f(x) - \phi \circ g(x)|^2 \leq c^2 \sum_{i=1}^k (f_i(x) - g_i(x))^2 \quad (1)$$

for every $f, g \in \mathcal{F}_1 \times \dots \times \mathcal{F}_k$ and $x \in \mathcal{X}$ for some constant $c < \infty$. Then the class $\phi \circ (\mathcal{F}_1, \dots, \mathcal{F}_k)$ is Donsker provided $\phi \circ f$ is square integrable for at least one $f \in \mathcal{F}_1 \times \dots \times \mathcal{F}_k$.

Note that equation (1) is satisfied for Lipschitz functions ϕ .

Theorem 9.30 Example

Suppose \mathcal{F} is Donsker with $\|P\|_{\mathcal{F}} < \infty$ and $f \geq \delta$ for some constant $\delta > 0$ for every $f \in \mathcal{F}$, then $1/\mathcal{F} = \{1/f : f \in \mathcal{F}\}$ is Donsker.

Proof

The result follows directly from the fact that $\phi : x \mapsto \frac{1}{x}$ is Lipschitz on (δ, ∞) . Since:

$$|\phi(x_1) - \phi(x_2)| = \left| \frac{1}{x_1} - \frac{1}{x_2} \right| = \frac{|x_1 - x_2|}{x_1 x_2} \leq \frac{1}{\delta^2} |x_1 - x_2|$$

The main tool in the proof of Theorem 9.30 is "Gaussianization".

Given random variables ξ_1, \dots, ξ_n with standard normal distribution independent of the data X_1, \dots, X_n , Chapter 2.9 covering multiplier CLT in VW gives us that the conditional and unconditional asymptotic behavior of the empirical process is related to the behavior of the process:

$$\mathbb{Z}_n \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \delta_{X_i}$$

where δ_{X_i} is the dirac measure.

Theorem 9.30 Discussion

Given fixed values X_1, \dots, X_n , the process $\{\mathbb{Z}_n(f) : f \in L_2(P)\}$ is Gaussian with zero mean and standard deviation metric:

$$\sigma_\xi(\mathbb{Z}_n(f) - \mathbb{Z}_n(g)) = \left(\frac{1}{n} \sum_{i=1}^n (f(X_i) - g(X_i))^2 \right)^{1/2}$$

which is equal to $L_2(\mathbb{P}_n)$ semimetric. This allows for use of comparison principles for Gaussian processes to derive the result. Three lemmas, with proofs, are given in preparation to prove 9.30. One of the these, VW-Lemma 2.10.14, is of independent interest. We state VW-Lemma 2.10.14 on the following slide.

VW-Lemma 2.10.14

Let \mathcal{F} be a P -Donsker class with $\|P\|_{\mathcal{F}} < \infty$. Then the class $\mathcal{F}^2 \equiv \{f^2 : f \in \mathcal{F}\}$ is G-C in probability: $\|\mathbb{P}_n - P\|_{\mathcal{F}^2}^* \xrightarrow{P} 0$. If, in addition, $P^*F^2 < \infty$ for some envelope function F , then \mathcal{F}^2 is also G-C almost surely and in mean.

The proof of VW-Lemma 2.10.14 relies on the Gaussianization technique discussed previously, but is omitted here due to length.

Corollary 9.31 follows directly from theorem 9.30:

Corollary 9.31

Let \mathcal{F} and \mathcal{G} be Donsker classes. Then:

- (i) $\mathcal{F} \cup \mathcal{G}$ and $\mathcal{F} + \mathcal{G}$ are Donsker.
- (ii) If $\|P\|_{\mathcal{F} \cup \mathcal{G}} < \infty$, then the classes of pairwise infima, $\mathcal{F} \wedge \mathcal{G}$, and pairwise suprema, $\mathcal{F} \vee \mathcal{G}$, are both Donsker.
- (iii) If \mathcal{F} and \mathcal{G} are both uniformly bounded, then $\mathcal{F} \cdot \mathcal{G}$ is Donsker.
- (iv) If $\psi : \bar{R} \mapsto \mathbb{R}$ is Lipschitz continuous, where R is the range of functions in \mathcal{F} and $\|\psi(f)\|_{P,2} < \infty$ for at least one $f \in \mathcal{F}$, then $\psi(\mathcal{F})$ is Donsker.
- (v) If $\|P\|_{\mathcal{F}} < \infty$ and g is uniformly bounded and measurable, then $\mathcal{F} \cdot g$ is Donsker.

Proof

For any measurable function f , let $\dot{f} \equiv f - Pf$. Define $\dot{\mathcal{F}} \equiv \{\dot{f} : f \in \mathcal{F}\}$ and $\dot{\mathcal{G}} \equiv \{\dot{g} : g \in \mathcal{G}\}$. Note that for any $f \in \mathcal{F}$ and $g \in \mathcal{G}$, $\mathbb{G}_n f = \mathbb{G}_n \dot{f}$ and $\mathbb{G}_n(f + g) = \mathbb{G}_n(\dot{f} + \dot{g})$. Hence $\mathcal{F} \cup \mathcal{G}$ is Donsker if and only if $\dot{\mathcal{F}} \cup \dot{\mathcal{G}}$ is Donsker, and $\mathcal{F} + \mathcal{G}$ if and only if $\dot{\mathcal{F}} + \dot{\mathcal{G}}$ is Donsker.

(i) Note that $\|P\|_{\dot{\mathcal{F}} \cup \dot{\mathcal{G}}} = 0$. Since $(x, y) \mapsto x + y$ is Lipschitz continuous on \mathbb{R}^2 , theorem 9.30 provides that $\dot{\mathcal{F}} + \dot{\mathcal{G}}$, and thus $\mathcal{F} + \mathcal{G}$, are Donsker. Since $\dot{\mathcal{F}} \cup \dot{\mathcal{G}} \subset \{\dot{\mathcal{F}} \cup \{0\}\} \cup \{\dot{\mathcal{G}} \cup \{0\}\}$, $\dot{\mathcal{F}} \cup \dot{\mathcal{G}}$ as well as $\mathcal{F} \cup \mathcal{G}$ are Donsker.

Proof (cont).

(ii) Left as exercise. Consider $\mathcal{F} \vee \mathcal{G}$. Then for fixed x , arbitrary $f_1, f_2 \in \mathcal{F}$ and arbitrary $g_1, g_2 \in \mathcal{G}$:

$$\begin{aligned} & |\phi(f_1(x), g_1(x)) - \phi(f_2(x), g_2(x))|^2 \\ &= |\max\{f_1(x), g_1(x)\} - \max\{f_2(x), g_2(x)\}|^2 \\ &\leq |f_1(x) - f_2(x)|^2 + |g_1(x) - g_2(x)|^2 \end{aligned}$$

satisfying (1) of theorem 9.30, and the result follows. The proof for $\mathcal{F} \wedge \mathcal{G}$ is similar.

(iii) Since $(x, y) \mapsto xy$ is Lipschitz continuous on bounded subsets of \mathbb{R}^2 , application of theorem 9.30 yields that $\mathcal{F} \cdot \mathcal{G}$ is Donsker.

Proof (cont).

(v) Note that for any $f_1, f_2 \in \mathcal{F}$ we have that

$$|f_1(x)g(x) - f_2(x)g(x)| \leq \|g\|_\infty |f_1(x) - f_2(x)|$$

hence taking $\phi(x, y) = xy$, $\phi \circ \{\mathcal{F}, \{g\}\}$ is Lipschitz continuous and $\mathcal{F} \cdot g$ is thus Donsker.

Sections 9.3 and 9.4 present preservation results for Glivenko-Cantelli and Donsker properties, respectively.

Preservations results are useful for generating G-C and Donsker classes from their pre-existing counterparts.

The results discussed in these two sections will be useful in Chapter 10, where we will discuss the validity of bootstrap in the empirical process setting.