Introduction to Empirical Processes and Semiparametric Inference Lecture 14: Entropy Calculations and the Bootstrap

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Glivenko-Cantelli Preservation

In this section, we discuss methods which are useful for building up Glivenko-Cantelli (G-C) classes from other G-C classes.

Such results can be useful for establishing consistency for Z- and M-estimators and their bootstrapped versions.

It is clear from the definition of P-G-C classes, that if $\mathcal{F}$ and $\mathcal{G}$ are P-G-C, then $\mathcal{F} \cup \mathcal{G}$ and any subset thereof is also P-G-C.

The purpose of the remainder of this section is to discuss more substantive preservation results such as the following, which is a minor modification of Theorem 3 of van der Vaart and Wellner (2000):

THEOREM 1. *Suppose that $\mathcal{F}_1, \ldots, \mathcal{F}_k$ are strong P-G-C classes of functions with*

$$\max_{1 \le j \le k} \|P\|_{\mathcal{F}_j} < \infty,$$

and that $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ is continuous.

Then the class

$$\mathcal{H} \equiv \phi(\mathcal{F}_1, \ldots, \mathcal{F}_k)$$

is strong P-G-C provided it has an integrable envelope.

The following are obvious consequences of this theorem:

COROLLARY 1. *Let $\mathcal{F}$ and $\mathcal{G}$ be P-G-C classes with respective integrable envelopes F and G.*

Then the following are true:

(i) $\mathcal{F} + \mathcal{G}$ *is P-G-C.*

(ii) $\mathcal{F} \cdot \mathcal{G}$ *is P-G-C provided $P[FG] < \infty$.*

(iii) Let R be the union of the ranges of 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.

Proof. The statement (i) is obvious.

Since $(x, y) \mapsto xy$ is continuous in $\mathbb{R}^2$, statement (ii) follows from Theorem 1.

Statement (iii) also follows from the theorem since ψ has a continuous extension to $\mathbb{R}$ (Hahn-Banach Theorem), $\tilde{\psi}$, such that $\|P\tilde{\psi}(f)\|_{\mathcal{F}} = \|P\psi(f)\|_{\mathcal{F}}$.$\Box$

It is interesting to note that the “preservation of products” result in the above corollary does not hold in general for Donsker classes (although it does hold for BUEI classes).

This preservation result for G-C classes can be useful in formulating master theorems for bootstrapped Z- and M- estimators.

Consider, for example, verifying the validity of the bootstrap for a parametric Z-estimator $\hat{\theta}_n$ which is a zero of $\theta \mapsto \mathbb{P}_n \psi_\theta$, for $\theta \in \Theta$, where ψ_θ is a suitable random function.

Let $\Psi(\theta) = P\psi_\theta$, where we assume that for any sequence $\{\theta_n\} \in \Theta$, $\Psi(\theta_n) \to 0$ implies $\theta_n \to \theta_0 \in \Theta$ (i.e., the parameter is identifiable).

Usually, to obtain consistency, it is reasonable to assume that the class $\{\psi_\theta, \theta \in \Theta)$ is P-G-C. Clearly, this condition is sufficient to ensure that $\hat{\theta}_n \overset{\text{as}*}{\to} \theta_0$.

Now, under a few additional assumptions, the Z-estimator master theorem, Theorem **??** can be applied, to obtain asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

In Section 2.2.5, we made the claim that if Ψ is appropriately differentiable and the parameter is identifiable (as defined in the previous paragraph), sufficient additional conditions for this asymptotic normality to hold and for the bootstrap to be valid are

- that the $\{\psi_\theta : \theta \in \Theta\}$ is strong P-G-C with $\sup_{\theta\in\Theta} P|\psi_\theta| < \infty$,
- that $\{\psi_\theta : \theta \in \Theta, \|\theta - \theta_0\| \leq \delta\}$ is P-Donsker for some $\delta > 0$,
- and that $P\|\psi_\theta - \psi_{\theta_0}\|^2 \to 0$ as $\theta \to \theta_0$.

As we will see in Chapter 13, where we present the arguments for this result in detail, an important step in the proof of bootstrap validity is to show that the bootstrap estimate $\hat{\theta}_n^\circ$ is unconditionally consistent for θ_0.

If we use a weighted bootstrap with i.i.d. non-negative weights $\xi_1, \ldots, \xi_n$, which are independent of the data and which satisfy E$\xi_1 = 1$, then result (ii) from the above corollary tells us that

$$\mathcal{F} \equiv \{\xi\psi_\theta : \theta \in \Theta\}$$

is P-G-C.

This follows since both classes of functions $\{\xi\}$ (a trivial class with one member) and $\{\psi_\theta : \theta \in \Theta\}$ are P-G-C and since the product class $\mathcal{F}$ has an integral envelope by Lemma 8.13.

Note here that we are tacitly augmenting P to be the product probability measure of both the data and the independent bootstrap weights.

We will expand on these ideas in Section 10.3 of the next chapter for the special case where $\Theta \subset \mathbb{R}^p$ and in Chapter 13 for the more general case.

In the context of conducting uniform inference for Pf, as f ranges over a class of functions $\mathcal{F}$, the following lemma answers the question of when the limiting covariance of $\mathbb{G}_n$, indexed by $\mathcal{F}$, can be consistently estimated.

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

Although knowledge of this covariance matrix is usually not sufficient in itself to obtain inference on $\{Pf : f \in \mathcal{F}\}$, it still provides useful information:

LEMMA 1. *Let $\mathcal{F}$ be Donsker. Then $\|\hat{\sigma}(f,g) - \sigma(f,g)\|_{\mathcal{F}\cdot\mathcal{F}} \overset{\text{as}*}{\to} 0$ if and only if $P^*\|f - Pf\|_{\mathcal{F}}^2 < \infty$.*

Proof. Note that since $\mathcal{F}$ is Donsker, $\mathcal{F}$ is also G-C.

Hence $\dot{\mathcal{F}} \equiv \{\dot{f} : f \in \mathcal{F}\}$ is G-C, where for any $f \in \mathcal{F}$, $\dot{f} = f - Pf$.

Now we first assume that $P^*\|f - Pf\|_{\mathcal{F}}^2 < \infty$.

By Theorem 1, $\dot{\mathcal{F}} \cdot \dot{\mathcal{F}}$ is also G-C.

Uniform consistency of $\hat{\sigma}$ now follows since, for any $f, g \in \mathcal{F}$,

$$\hat{\sigma}(f, g) - \sigma(f, g) = (\mathbb{P}_n - P)\dot{f}\dot{g} - \mathbb{P}_n\dot{f}\mathbb{P}_n\dot{g}.$$

Assume next that

$$\|\hat{\sigma}(f, g) - \sigma(f, g)\|_{\mathcal{F}\cdot\mathcal{F}} \stackrel{\text{as}*}{\rightarrow} 0.$$

This implies that $\dot{\mathcal{F}} \cdot \dot{\mathcal{F}}$ is G-C.

Now Lemma 8.13 implies that

$$P^*\|f - Pf\|_{\mathcal{F}}^2 = P^*\|fg\|_{\dot{\mathcal{F}}\cdot\dot{\mathcal{F}}} < \infty.\Box$$

We close this section with the following theorem that provides several interesting necessary and sufficient conditions for $\mathcal{F}$ to be strong G-C:

THEOREM 2. *Let $\mathcal{F}$ be a class of measurable functions.*

Then the following are equivalent:

(i) $\mathcal{F}$ *is strong* P*-G-C;*

(ii) $\mathsf{E}^*\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0$ *and* $\mathsf{E}^*\|f - Pf\|_{\mathcal{F}} < \infty$*;*

(iii) $\|\mathbb{P}_n - P\|_{\mathcal{F}} \xrightarrow{\mathsf{P}} 0$ *and* $\mathsf{E}^*\|f - Pf\|_{\mathcal{F}} < \infty$*.*

Donsker Preservation

In this section, we describe several techniques for building Donsker classes from other Donsker classes.

The first theorem, Theorem 3, gives results for subsets, pointwise closures and symmetric convex hulls of Donsker classes.

The second theorem, Theorem 4, presents a very powerful result for Lipschitz functions of Donsker classes.

The corollary that follows presents consequences of this theorem that are quite useful in statistical applications.

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 all $f : \mathcal{X} \mapsto \mathbb{R}$ for which there exists a sequence $\{f_m\} \in \mathcal{F}$ such that $f_m \to f$ both pointwise (i.e., for every argument $x \in \mathcal{X}$) and in $L_2(P)$.

Similarly, let $\overline{\text{sconv}}^{(P,2)}\,\mathcal{F}$ be the pointwise and $L_2(P)$ closure of sconv $\mathcal{F}$ defined in Section 9.1.1.

THEOREM 3. *Let $\mathcal{F}$ be a P-Donsker class.*

Then

(i) For any $\mathcal{G} \subset \mathcal{F}$, $\mathcal{G}$ is P-Donsker.

(ii) $\overline{\mathcal{F}}^{(P,2)}$ is P-Donsker.

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

The following theorem, Theorem 2.10.6 of VW, is one of the most useful Donsker preservation results for statistical applications:

THEOREM 4. *Let $\mathcal{F}_1, \ldots, \mathcal{F}_k$ be Donsker classes with $\max_{1\leq i\leq k} \|P\|_{\mathcal{F}_i} < \infty$.*

Let $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ satisfy

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

for every $f, g \in \mathcal{F}_1 \times \cdots \times \mathcal{F}_k$ and $x \in \mathcal{X}$ and for some constant $c < \infty$.

Then $\phi \circ (\mathcal{F}_1, \ldots, \mathcal{F}_k)$ is Donsker provided $\phi \circ f$ is square integrable for at least one $f \in \mathcal{F}_1 \times \cdots \times \mathcal{F}_k$.

COROLLARY 2. *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,* $\mathcal{F} \cdot \mathcal{G}$ *is Donsker.*

(iv) *If* $\psi : \overline{R} \mapsto \mathbb{R}$ *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 a uniformly bounded, measurable function, then* $\mathcal{F} \cdot g$ *is Donsker.*

The Bootstrap

The purpose of this chapter is to obtain consistency results for bootstrapped empirical processes.

Much of the bootstrap results for such estimators will be deferred to later chapters where we discuss the functional delta method, Z-estimation and M-estimation.

We do, however, present one specialized result for parametric Z-estimators in Section 3 of this chapter as a practical illustration of bootstrap techniques.

The best choice of bootstrap weights for a given statistical application is also an important question, and the answer depends on the application.

While the multinomial bootstrap is conceptually simple, its use in survival analysis applications may result in too much tied data.

In the presence of censoring, it is even possible that a bootstrap sample could be drawn that consists of only censored observations.

To avoid complications of this kind, it may be better to use the Bayesian bootstrap (Rubin, 1981).

The weights for the Bayesian bootstrap are

$$\xi_1/\bar{\xi}, \ldots, \xi_n/\bar{\xi},$$

where $\xi_1, \ldots, \xi_n$ are i.i.d. standard exponential (mean and variance 1), independent of the data $X_1, \ldots, X_n$, and where $\bar{\xi} \equiv n^{-1} \sum_{i=1}^n \xi_i$.

Since these weights are strictly positive, all observations are represented in each bootstrap realization, and the aforementioned problem with tied data won't happen unless the original data has ties.

Both the multinomial and Bayesian bootstraps are included in the bootstrap weights we discuss in this chapter.

The multinomial weighted bootstrap is sometimes called the *nonparametric bootstrap* since it amounts to sampling from the empirical distribution, which is a nonparametric estimate of the true distribution.

We also note that the asymptotic results of this chapter are all first order, and in this situation the limiting results do not vary among those schemes that satisfy the stated conditions.

A more refined analysis of differences between weighting schemes is beyond the scope of this chapter, but such differences may be important in small samples.

A good reference for higher order properties of the bootstrap is Hall (1992).

Throughout this chapter, we will sometimes for simplicity omit the subscript when referring to a representative of an i.i.d. sample.

For example, we may use E$|\xi|$ to refer to E$|\xi_1|$, where ξ_1 is the first member of the sample $\xi_1, \ldots, \xi_n$.

The context will make the meaning clear.

Multiplier Central Limit Theorems

In this section, we present a multiplier central limit theorem that forms the basis for the bootstrap results of this chapter, and we also present an interesting corollary.

For a real random variable ξ, recall from Section 2.2.3 the quantity $\|\xi\|_{2,1} \equiv \int_0^\infty \sqrt{\text{pr}(|\xi| > x)}dx$.

Exercise 10.5.1 below verifies this is a norm slightly larger than $\|\cdot\|_2$.

Also recall that δ_{X_i} is the probability measure that assigns a mass of 1 to X_i so that $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ and $\mathbb{G}_n = n^{-1/2}\sum_{i=1}^n(\delta_{X_i} - P)$.

THEOREM 5. *(Multiplier central limit theorem) Let $\mathcal{F}$ be a class of measurable functions, and let $\xi_1, \ldots, \xi_n$ be i.i.d. random variables with mean zero, variance 1, and with $\|\xi\|_{2,1} < \infty$, independent of the sample data $X_1, \ldots, X_n$.*

Let

$$\mathbb{G}_n' \equiv n^{-1/2} \sum_{i=1}^{n} \xi_i(\delta_{X_i} - P)$$

and

$$\mathbb{G}_n'' \equiv n^{-1/2} \sum_{i=1}^{n} (\xi_i - \bar{\xi})\delta_{X_i},$$

where $\bar{\xi} \equiv n^{-1} \sum_{i=1}^{n} \xi_i$.

Then the following are equivalent:

(i) $\mathcal{F}$ *is* P*-Donsker;*

(ii) $\mathbb{G}'_n$ *converges weakly to a tight process in* $\ell^\infty(\mathcal{F})$*;*

(iii) $\mathbb{G}'_n \rightsquigarrow \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$*;*

(iv) $\mathbb{G}''_n \rightsquigarrow \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$*.*

We now present the following interesting corollary which shows the possibly unexpected result that the multiplier empirical process is asymptotically independent of the usual empirical process, even though the same data $X_1, \ldots, X_n$ are used in both processes:

COROLLARY 3. *Assume the conditions of Theorem 5 hold and that $\mathcal{F}$ is Donsker.*

Then $(\mathbb{G}_n, \mathbb{G}'_n, \mathbb{G}''_n) \rightsquigarrow (\mathbb{G}, \mathbb{G}', \mathbb{G}',)$ in $[\ell^\infty(\mathcal{F})]^3$, where $\mathbb{G}$ and $\mathbb{G}'$ are independent P-Brownian bridges.

Proof. By the preceding theorem, the three processes are asymptotically tight marginally and hence asymptotically tight jointly.

Since the first process is uncorrelated with the second process, the limiting distribution of the first process is independent of the limiting distribution of the second process.

Note that by definition,

$$\mathbb{G}_n' - \mathbb{G}_n'' = \bar{\xi}\mathbb{G}_n,$$

and we now have that $\|\mathbb{G}_n' - \mathbb{G}_n''\|_{\mathcal{F}} \xrightarrow{\mathsf{P}} 0$, and thus the remainder of the corollary follows.$\Box$

The above multiplier processes will now be studied conditional on the data.

This yields in-probability and outer-almost-sure conditional multiplier central limit theorems.

These results are one step closer to the bootstrap validity results of the next section.

For a metric space $(\mathbb{D}, d)$, define $BL_1(\mathbb{D})$ to be the space of all functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1, i.e., $\|f\|_\infty \leq 1$ and $|f(x) - f(y)| \leq d(x, y)$ for all $x, y \in \mathbb{D}$.

In the current set-up, $\mathbb{D} = \ell^\infty(\mathcal{F})$, for some class of measurable functions $\mathcal{F}$, and d is the corresponding uniform metric.

As we did in Section 2.2.3, we will use BL_1 as shorthand for $BL_1(\ell^\infty(\mathcal{F}))$.

We now present the in-probability conditional multiplier central limit theorem:

THEOREM 6. *Let $\mathcal{F}$ be a class of measurable functions, and let $\xi_1, \ldots,$ ξ_n be i.i.d. random variables with mean zero, variance 1, and $\|\xi\|_{2,1} < \infty$, independent of the sample data $X_1, \ldots, X_n$.*

Let $\mathbb{G}'_n$, $\mathbb{G}''_n$ and $\bar{\xi}$ be as defined in Theorem 5.

Then the following are equivalent:

(i) $\mathcal{F}$ *is Donsker;*

(ii) $\mathbb{G}'_n \underset{\xi}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$ *and* $\mathbb{G}'_n$ *is asymptotically measurable.*

(iii) $\mathbb{G}''_n \underset{\xi}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$ *and* $\mathbb{G}''_n$ *is asymptotically measurable.*

In the above theorem, E_ξ denotes taking the expectation conditional on $X_1, \ldots, X_n$.

Note that for a continuous function $h : \ell^\infty(\mathcal{F}) \mapsto \mathbb{R}$, if we fix $X_1, \ldots, X_n$, then

$$(a_1, \ldots, a_n) \mapsto h(n^{-1/2} \sum_{i=1}^{n} a_i(\delta_{X_i} - P))$$

is a measurable map from $\mathbb{R}^n$ to $\mathbb{R}$, provided $\|f(X) - Pf\|_{\mathcal{F}}^{*} < \infty$ almost surely.

This last inequality is tacitly assumed so that the empirical processes under investigation reside in $\ell^\infty(\mathcal{F})$.

Thus the expectation E_ξ in conclusions (ii) and (iii) is proper.

The following lemma is a conditional multiplier central limit theorem for i.i.d. Euclidean data:

LEMMA 2. *Let $Z_1, \ldots, Z_n$ be i.i.d. Euclidean random vectors, with $\mathsf{E}Z = 0$ and $\mathsf{E}\|Z\|^2 < \infty$, independent of the i.i.d. sequence of real random variables $\xi_1, \ldots, \xi_n$ with $\mathsf{E}\xi = 0$ and $\mathsf{E}\xi^2 = 1$.*

Then, conditionally on $Z_1, Z_2, \ldots$, $n^{-1/2} \sum_{i=1}^n \xi_i Z_i \rightsquigarrow N(0, \text{cov}Z)$, for almost all sequences $Z_1, Z_2, \ldots$.

Proof sketch of Theorem 6. Basically, the above lemma is used for finite-dimensional convergence of the conditional multiplier process.

Then the multiplier CLT is used for asymptotical equicontinuity.

We now present the outer-almost-sure conditional multiplier central limit theorem:

THEOREM 7. *Assume the conditions of Theorem 6.*

Then the following are equivalent:

(i) $\mathcal{F}$ *is Donsker and* $P^*\|f - Pf\|_{\mathcal{F}}^2 < \infty$*;*

(ii) $\mathbb{G}_n' \underset{\xi}{\overset{\text{as}*}{\rightsquigarrow}} \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$.

(iii) $\mathbb{G}_n'' \underset{\xi}{\overset{\text{as}*}{\rightsquigarrow}} \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$.

Bootstrap Central Limit Theorems

Theorems 6 and 7 will now be used to prove Theorems 8 and 9 from Page 38 of Section 2.2.3. Recall that the multinomial bootstrap is obtained by resampling from the data $X_1, \ldots, X_n$, with replacement, n times to obtain a bootstrapped sample $X_1^*, \ldots, X_n^*$. The empirical measure $\hat{\mathbb{P}}_n^*$ of the bootstrapped sample has the same distribution—given the data—as the measure $\hat{\mathbb{P}}_n \equiv n^{-1}\sum_{i=1}^n W_{ni}\delta_{X_i}$, where $W_n \equiv (W_{n1}, \ldots, W_{nn})$ is a multinomial$(n, n^{-1}, \ldots, n^{-1})$ deviate independent of the data. As in Section 2.2.3, let $\hat{\mathbb{P}}_n \equiv n^{-1}\sum_{i=1}^n W_{ni}\delta_{X_i}$ and $\hat{\mathbb{G}}_n \equiv \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n)$. Also recall the definitions $\tilde{\mathbb{P}}_n \equiv n^{-1}\sum_{i=1}^n (\xi/\bar{\xi})\delta_{X_i}$ and $\tilde{\mathbb{G}}_n \equiv \sqrt{n}(\mu/\tau)(\tilde{\mathbb{P}}_n - \mathbb{P}_n)$, where the weights $\xi_1, \ldots, \xi_n$ are i.i.d. nonnegative, independent of $X_1, \ldots, X_n$, with mean $0 < \mu < \infty$ and variance $0 < \tau^2 < \infty$, and with $\|\xi\|_{2,1} < \infty$. When $\bar{\xi} = 0$, we define $\tilde{\mathbb{P}}_n$ to be zero. Note that the weights $\xi_1, \ldots, \xi_n$ in this section must have μ subtracted from them and then divided by τ before they satisfy the criteria of the multiplier weights in the previous section.

THEOREM 8. *The following are equivalent:*

(i) $\mathcal{F}$ *is* P*-Donsker.*

(ii) $\hat{\mathbb{G}}_n \underset{W}{\overset{\mathsf{P}}{\rightsquigarrow}} \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$ *and the sequence* $\hat{\mathbb{G}}_n$ *is asymptotically measurable.*

(iii) $\tilde{\mathbb{G}}_n \underset{\xi}{\overset{\mathsf{P}}{\rightsquigarrow}} \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$ *and the sequence* $\tilde{\mathbb{G}}_n$ *is asymptotically measurable.*

Proof of Theorem 8 (Page 38). The equivalence of (i) and (ii) follows from Theorem 3.6.1 of VW, which proof we omit. We note, however, that a key component of this proof is a clever approximation of the multinomial weights with i.i.d. Poisson mean 1 weights. We will use this approximation in our proof of Theorem **??** below.

We now prove the equivalence of (i) and (iii). Let $\xi_i^\circ \equiv \tau^{-1}(\xi_i - \mu)$, $i = 1, \ldots, n$, and define $\mathbb{G}_n^\circ \equiv n^{-1/2}\sum_{i=1}^n(\xi_i^\circ - \bar{\xi}^\circ)\delta_{X_i}$, where $\bar{\xi}^\circ \equiv n^{-1}\sum_{i=1}^n \xi_i^\circ$. The basic idea is to show the asymptotic equivalence of $\tilde{\mathbb{G}}_n$ and $\mathbb{G}_n^\circ$. Then Theorem 6 can be used to establish the desired result. Accordingly,

$$\mathbb{G}_n^\circ - \tilde{\mathbb{G}}_n = \left(1 - \frac{\mu}{\bar{\xi}}\right)\mathbb{G}_n^\circ = \left(\frac{\bar{\xi}}{\mu} - 1\right)\tilde{\mathbb{G}}_n. \qquad (1)$$

First, assume that $\mathcal{F}$ is Donsker. Since the weights $\xi_1^\circ, \ldots, \xi_n^\circ$ satisfy the conditions of the unconditional multiplier central limit theorem, we have that $\mathbb{G}_n^\circ \rightsquigarrow \mathbb{G}$. Theorem 6 also implies that $\mathbb{G}_n^\circ \underset{\xi}{\overset{\mathsf{P}}{\rightsquigarrow}} \mathbb{G}$. Now (1) can be applied to verify that $\|\tilde{\mathbb{G}}_n - \mathbb{G}_n^\circ\|_{\mathcal{F}} \xrightarrow{\mathsf{P}} 0$, and thus $\tilde{\mathbb{G}}_n$ is asymptotically measurable and

$$\sup_{h \in BL_1} \left| \mathsf{E}_\xi h(\mathbb{G}_n^\circ) - \mathsf{E}_\xi h(\tilde{\mathbb{G}}_n) \right| \xrightarrow{\mathsf{P}} 0.$$

Thus (i)$\Rightarrow$(iii).

Second, assume that $\tilde{\mathbb{G}}_n \underset{\xi}{\overset{\mathsf{P}}{\rightsquigarrow}} \mathbb{G}$. It is not hard to show, arguing as we did in the proof of Theorem 6 for the implication (ii)$\Rightarrow$(i), that $\tilde{\mathbb{G}}_n \rightsquigarrow \mathbb{G}$ in $\ell^\infty(\mathcal{F})$ unconditionally. By applying (1) again, we now have that $\|\mathbb{G}_n^\circ - \tilde{\mathbb{G}}_n\|_{\mathcal{F}} \xrightarrow{\mathsf{P}} 0$, and thus $\mathbb{G}_n^\circ \rightsquigarrow \mathbb{G}$ in $\ell^\infty(\mathcal{F})$ unconditionally. The unconditional multiplier central limit theorem now verifies that $\mathcal{F}$ is Donsker, and thus (iii)$\Rightarrow$(i).$\Box$

THEOREM 9. *The following are equivalent:*

(i) $\mathcal{F}$ *is* P*-Donsker and* $P^*\left[\sup_{f\in\mathcal{F}}(f(X) - Pf)^2\right] < \infty$*.*

(ii) $\hat{\mathbb{G}}_n \underset{W}{\overset{\text{as}*}{\leadsto}} \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$.

(iii) $\tilde{\mathbb{G}}_n \underset{\xi}{\overset{\text{as}*}{\leadsto}} \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$.

Proof of Theorem 9 (Page 40). The equivalence of (i) and (ii) follows from Theorem 3.6.2 of VW, which proof we again omit. We now prove the equivalence of (i) and (iii).

First, assume (i). Then $\mathbb{G}_n^\circ \underset{\xi}{\overset{\text{as}*}{\leadsto}} \mathbb{G}$ by Theorem 7. Fix $\rho > 0$, and note that by using the first equality in (1), we have for any $h \in BL_1$ that

$$\left| h(\tilde{\mathbb{G}}_n) - h(\mathbb{G}_n^\circ) \right| \quad \leq \quad 2 \times \mathbf{1}\left\{ \left| 1 - \frac{\mu}{\bar{\xi}} \right| > \rho \right\} + (\rho \|\mathbb{G}_n^\circ\|_{\mathcal{F}}) \wedge 1 \tag{2}$$

The first term on the right $\overset{\text{as}*}{\rightarrow} 0$. Since the map $\|\cdot\|_{\mathcal{F}} \wedge 1 : \ell^\infty(\mathcal{F}) \mapsto \mathbb{R}$ is in BL_1, we have by Theorem 7 that

$\mathsf{E}_\xi \left[(\rho \|\mathbb{G}_n^\circ\|_{\mathcal{F}}) \wedge 1 \right] \overset{\text{as}*}{\rightarrow} \mathsf{E}\left[\|\rho \mathbb{G}\|_{\mathcal{F}} \wedge 1 \right]$. Let the sequence $0 < \rho_n \downarrow 0$

converge slowly enough so that that the first term on the right in (2) $\overset{\text{as}*}{\rightarrow} 0$ after replacing ρ with ρ_n. Since E $[\|\rho_n \mathbb{G}\|_{\mathcal{F}} \wedge 1] \rightarrow 0$, we can apply E$_\xi$ to both sides of (2)—after replacing ρ with ρ_n—to obtain

$$\sup_{h \in BL_1} \left| h(\tilde{\mathbb{G}}_n) - h(\mathbb{G}_n^\circ) \right| \overset{\text{as}*}{\rightarrow} 0.$$

Combining the fact that $h(\mathbb{G}_n^\circ)^* - h(\mathbb{G}_n^\circ)_* \overset{\text{as}*}{\rightarrow} 0$ with additional applications of (2) yields $h(\tilde{\mathbb{G}}_n)^* - h(\tilde{\mathbb{G}}_n)_* \overset{\text{as}*}{\rightarrow} 0$. Since h was arbitrary, we have established that $\tilde{\mathbb{G}}_n \underset{\xi}{\overset{\text{as}*}{\rightsquigarrow}} \mathbb{G}$, and thus (iii) follows.

Second, assume (iii). Fix $\rho > 0$, and note that by using the second equality in (1), we have for any $h \in BL_1$ that

$$\left| h(\mathbb{G}_n^\circ) - h(\tilde{\mathbb{G}}_n) \right| \quad \leq \quad 2 \times \mathbf{1}\left\{ \left| \frac{\bar{\xi}}{\mu} - 1 \right| > \rho \right\} + \left(\rho \|\tilde{\mathbb{G}}_n\|_{\mathcal{F}} \right) \wedge 1.$$

Since the first term on the right $\overset{\text{as}*}{\rightarrow} 0$, we can use virtually identical arguments to those used in the previous paragraph—but with the roles of $\mathbb{G}_n^\circ$ and $\tilde{\mathbb{G}}_n$ reversed—to obtain that $\mathbb{G}_n^\circ \underset{\xi}{\overset{\text{as}*}{\rightsquigarrow}} \mathbb{G}$. Now Theorem 7 yields that $\mathcal{F}$ is Donsker, and thus (i) follows.$\square$

Continuous Mapping Results for the Bootstrap

We now assume a more general set-up, where $\hat{X}_n$ is a bootstrapped process in a Banach space $(\mathbb{D}, \|\cdot\|)$ and is composed of the sample data $\mathcal{X}_n \equiv (X_1, \ldots, X_n)$ and a random weight vector $M_n \in \mathbb{R}^n$ independent of $\mathcal{X}_n$. We do not require that $X_1, \ldots, X_n$ be i.i.d. In this section, we obtain two continuous mapping results. The first result, Proposition 1, is a simple continuous mapping results for the very special case of Lipschitz continuous maps. It is applicable to both the in-probability or outer-almost-sure versions of bootstrap consistency. An interesting special case is the map $g(x) = \|x\|$. In this case, the proposition validates the use of the bootstrap to construct asymptotically uniformly valid confidence bands for $\{Pf : f \in \mathcal{F}\}$ whenever Pf is estimated by $\mathbb{P}_n f$ and $\mathcal{F}$ is P-Donsker. Now assume that $\hat{X}_n \underset{M}{\overset{\mathrm{P}}{\rightsquigarrow}} X$ and that the distribution of $\|X\|$ is continuous. Lemma 3 towards the end of this section reveals that pr$(\|\hat{X}_n\| \leq t|\mathcal{X}_n)$ converges uniformly to $P(\|X\| \leq t)$, in probability. A parallel outer almost sure result holds when $\hat{X}_n \underset{M}{\overset{\text{as}*}{\rightsquigarrow}} X$.

The second result, Theorem 10, is a considerably deeper result for general continuous maps applied to bootstraps which are consistent in probability. Because of this generality, we must require certain measurability conditions on the map $M_n \mapsto \hat{X}_n$. Fortunately, based on the discussion in the paragraph following Theorem 6 above, these measurability conditions are easily satisfied when either $\hat{X}_n = \hat{\mathbb{G}}_n$ or $\hat{X}_n = \tilde{\mathbb{G}}_n$. It appears that other continuous mapping results for bootstrapped empirical processes hold, such as for bootstraps which are outer almost surely consistent, but such results seem to be very challenging to verify.

PROPOSITION 1. *Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces, X a tight random variable on $\mathbb{D}$, and $g : \mathbb{D} \mapsto \mathbb{E}$ Lipschitz continuous. We have the* *following:*

(i) If $\hat{X}_n \underset{M}{\overset{\mathsf{P}}{\leadsto}} X$*, then* $g(\hat{X}_n) \underset{M}{\overset{\mathsf{P}}{\leadsto}} g(X)$*.*

(ii) If $\hat{X}_n \underset{M}{\overset{\mathsf{as}*}{\leadsto}} X$*, then* $g(\hat{X}_n) \underset{M}{\overset{\mathsf{as}*}{\leadsto}} g(X)$*.*

Proof. Let $c_0 < \infty$ be the Lipschitz constant for g, and, without loss of generality, assume $c_0 \geq 1$. Note that for any $h \in BL_1(\mathbb{E})$, the map $x \mapsto h(g(x))$ is an element of $c_0 BL_1(\mathbb{D})$. Thus

$$\begin{aligned} \sup_{h \in BL_1(\mathbb{E})} \left| \mathsf{E}_M h(g(\hat{X}_n)) - \mathsf{E} h(g(X)) \right| &\leq \sup_{h \in c_0 BL_1(\mathbb{D})} \left| \mathsf{E}_M h(\hat{X}_n) - \mathsf{E} h(X) \right| \\ &= c_0 \sup_{h \in BL_1(\mathbb{D})} \left| \mathsf{E}_M h(\hat{X}_n) - \mathsf{E} h(X) \right|, \end{aligned}$$

and the desired result follows by the respective definitions of $\underset{M}{\overset{\mathsf{P}}{\leadsto}}$ and $\underset{M}{\overset{\mathsf{as}*}{\leadsto}}$. $\square$

THEOREM 10. *Let* $g : \mathbb{D} \mapsto \mathbb{E}$ *be continuous at all points in* $\mathbb{D}_0 \subset \mathbb{D}$*,* *where* $\mathbb{D}$ *and* $\mathbb{E}$ *are Banach spaces and* $\mathbb{D}_0$ *is closed. Assume that* $M_n \mapsto h(\hat{X}_n)$ *is measurable for every* $h \in C_b(\mathbb{D})$ *outer almost surely. Then if* $\hat{X}_n \underset{M}{\overset{\mathsf{P}}{\rightsquigarrow}} X$ *in* $\mathbb{D}$*, where* X *is tight and* $\mathrm{P}^*(X \in \mathbb{D}_0) = 1$*,* $g(\hat{X}_n) \underset{M}{\overset{\mathsf{P}}{\rightsquigarrow}} g(X)$*.*

Proof. As in the proof of the implication (ii)⇒(i) of Theorem 6, we can argue that $\hat{X}_n \rightsquigarrow X$ unconditionally, and thus $g(\hat{X}_n) \rightsquigarrow g(X)$ unconditionally by the standard continuous mapping theorem. Moreover, we can replace $\mathbb{E}$ with its closed linear span so that the restriction of g to $\mathbb{D}_0$ has an extension $\tilde{g} : \mathbb{D} \mapsto \mathbb{E}$ which is continuous on all of $\mathbb{D}$ by Dugundji's extension theorem (Theorem 11 below). Thus $(g(\hat{X}_n), \tilde{g}(\hat{X}_n)) \rightsquigarrow (g(X), \tilde{g}(X))$, and hence $g(\hat{X}_n) - \tilde{g}(\hat{X}_n) \xrightarrow{\mathsf{P}} 0$. Therefore we can assume without loss of generality that g is continuous on all of $\mathbb{D}$. We can also assume without loss of generality that $\mathbb{D}_0$ is a separable Banach space since X is tight. Hence $\mathbb{E}_0 \equiv g(\mathbb{D}_0)$ is also a separable Banach space.

Fix $\epsilon > 0$. There now exists a compact $K \subset \mathbb{E}_0$ such that pr$(g(X) \notin K) < \epsilon$. By Theorem 12 below, the proof of which is given in Section 10.4, we know there exists an integer $k < \infty$, elements $z_1, \ldots, z_k \in C[0,1]$, continuous functions $f_1, \ldots, f_k : \mathbb{E} \mapsto \mathbb{R}$, and a Lipschitz continuous function $J : \overline{\text{lin}}(z_1, \ldots, z_k) \mapsto \mathbb{E}$, such that the map $x \mapsto T_\epsilon(x) \equiv J\left(\sum_{j=1}^k z_j f_j(x)\right)$ has domain $\mathbb{E}$ and range $\subset \mathbb{E}$ and satisfies $\sup_{x \in K} \|T_\epsilon(x) - x\| < \epsilon$. Let $BL_1 \equiv BL_1(\mathbb{E})$. We now have

$$\begin{aligned}
&\sup_{h\in BL_1}\left|\mathsf{E}_M h(g(\hat{X}_n)) - \mathsf{E}h(g(X))\right| \\
&\quad\leq \sup_{h\in BL_1}\left|\mathsf{E}_M h(T_\epsilon g(\hat{X}_n)) - \mathsf{E}h(T_\epsilon g(X))\right| \\
&\quad\quad + \mathsf{E}_M\left\{\left\|T_\epsilon g(\hat{X}_n) - g(\hat{X}_n)\right\| \wedge 2\right\} + \mathsf{E}\left\{\|T_\epsilon g(X) - g(X)\| \wedge 2\right\}.
\end{aligned}$$

However, the outer expectation of the second term on the right converges to the third term, as $n \to \infty$, by the usual continuous mapping theorem. Thus, provided

$$\sup_{h\in BL_1}\left|\mathsf{E}_M h(T_\epsilon g(\hat{X}_n)) - \mathsf{E}h(T_\epsilon g(X))\right| \quad \stackrel{\mathsf{P}}{\to} \quad 0, \qquad (3)$$

we have that

$$
\begin{aligned}
\limsup_{n\to\infty} \mathsf{E}^* \left\{ \sup_{h\in BL_1} \left| \mathsf{E}_M h(g(\hat{X}_n)) - \mathsf{E}h(g(X)) \right| \right\} \qquad (4) \\
\leq \quad & 2\mathsf{E}\left\{ \|T_\epsilon g(X) - g(X)\| \wedge 2 \right\} \\
\leq \quad & 2\mathsf{E}\left\| \{T_\epsilon g(X) - g(X)\} \mathbf{1}\{g(X) \in K\} \right\| + 4\mathrm{pr}(g(X) \notin K) \\
< \quad & 6\epsilon.
\end{aligned}
$$

Now note that for each $h \in BL_1$, $h\left(J\left(\sum_{j=1}^k z_j a_j\right)\right) = \tilde{h}(a_1, \ldots, a_k)$ for all $(a_1, \ldots, a_k) \in \mathbb{R}^k$ and some $\tilde{h} \in c_0 BL_1(\mathbb{R}^k)$, where $1 \leq c_0 < \infty$ (this follows since J is Lipschitz continuous and

$\left\|\sum_{j=1}^k z_j a_j\right\| \leq \max_{1\leq j\leq k} |a_j| \times \sum_{j=1}^k \|z_j\|$). Hence

$$\sup_{h\in BL_1} \left| \mathsf{E}_M h(T_\epsilon g(\hat{X}_n)) - \mathsf{E} h(T_\epsilon g(X))\right| \tag{5}$$

$$\leq \sup_{h\in c_0 BL_1(\mathbb{R}^k)} \left| \mathsf{E}_M h(u(\hat{X}_n)) - \mathsf{E} h(u(X))\right|$$

$$= c_0 \sup_{h\in BL_1(\mathbb{R}^k)} \left| \mathsf{E}_M h(u(\hat{X}_n)) - \mathsf{E} h(u(X))\right|,$$

where $x \mapsto u(x) \equiv (f_1(g(x)), \ldots, f_k(g(x)))$. Fix any $v : \mathbb{R}^k \mapsto [0, 1]$ which is Lipschitz continuous (the Lipschitz constant may be > 1). Then, since $\hat{X}_n \rightsquigarrow X$ unconditionally, $\mathsf{E}^* \left\{ \mathsf{E}_M v(u(\hat{X}_n))^* - \mathsf{E}_M v(u(\hat{X}_n))_* \right\} \leq$ $\mathsf{E}^* \left\{ v(u(\hat{X}_n))^* - v(u(\hat{X}_n))_* \right\} \to 0$, where sub- and super- script $*$ denote measurable majorants and minorants, respectively, with respect to the joint probability space of $(\mathcal{X}_n, M_n)$. Thus

$$\left| \mathsf{E}_M v(u(\hat{X}_n)) - \mathsf{E}_M v(u(\hat{X}_n))^* \right| \quad \xrightarrow{\mathsf{P}} \quad 0. \tag{6}$$

Note that we are using at this point the outer almost sure measurability of $M_n \mapsto v(u(\hat{X}_n))$ to ensure that $E_M v(u(\hat{X}_n))$ is well defined, even if the resulting random expectation is not itself measurable.

Now, for every subsequence n', there exists a further subsequence n'' such that $\hat{X}_{n''} \underset{M}{\overset{\mathsf{as}*}{\rightsquigarrow}} X$. This means that for this subsequence, the set B of data subsequences $\{\mathcal{X}_{n''} : n \geq 1\}$ for which $\mathsf{E}_M v(u(\hat{X}_{n''})) - \mathsf{E} v(u(X)) \to 0$ has inner probability 1. Combining this with (6) and Proposition **??**, we obtain that $\mathsf{E}_M v(u(\hat{X}_n)) - \mathsf{E} v(u(X)) \xrightarrow{\mathsf{P}} 0$. Since v was an arbitrary real, Lipschitz continuous function on $\mathbb{R}^k$, we now have by Part (i) of Lemma 3 below followed by Lemma 4 below, that

$$\sup_{h \in BL_1(\mathbb{R}^k)} \left| \mathsf{E}_M h(u(\hat{X}_n)) - \mathsf{E} h(u(X)) \right| \xrightarrow{\mathsf{P}} 0.$$

Combining this with (5), we obtain that (3) is satisfied. The desired result now follows from (4), since $\epsilon > 0$ was arbitrary.$\square$

THEOREM 11. *(Dugundji's extension theorem) Let X be an arbitrary metric space, A a closed subset of X, L a locally convex linear space (which includes Banach vector spaces), and $f : A \mapsto L$ a continuous map. Then there exists a continuous extension of f, $F : X \mapsto L$. Moreover, $F(X)$ lies in the closed linear span of the convex hull of $f(A)$.*

Proof. This is Theorem 4.1 of Dugundji (1951), and the proof can be found therein.$\square$

THEOREM 12. *Let $\mathbb{E}_0 \subset \mathbb{E}$ be Banach spaces with $\mathbb{E}_0$ separable and $\overline{\mathit{lin}}\mathbb{E}_0 \subset \mathbb{E}$. Then for every $\epsilon > 0$ and every compact $K \subset \mathbb{E}_0$, there* *exists an integer* $k < \infty$*, elements* $z_1, \ldots, z_k \in C[0,1]$*, continuous functions* $f_1, \ldots, f_k : \mathbb{E} \mapsto \mathbb{R}$*, and a Lipschitz continuous function* $J : \overline{\text{lin}}(z_1, \ldots, z_k) \mapsto \mathbb{E}$*, such that the map* $x \mapsto T_\epsilon(x) \equiv J\left(\sum_{j=1}^k z_j f_j(x)\right)$ *has domain* $\mathbb{E}$ *and range* $\subset \mathbb{E}$*, is continuous, and satisfies* $\sup_{x \in K} \|T_\epsilon(x) - x\| < \epsilon$*.*

The proof of this theorem is given in Section 10.4. For the next two lemmas, we use the usual partial ordering on $\mathbb{R}^k$ to define relations between points, e.g., for any $s, t \in \mathbb{R}^k$, $s \leq t$ is equivalent to $s_1 \leq t_1, \ldots, s_k \leq t_k$.

LEMMA 3. *Let* X_n *and* X *be random variables in* $\mathbb{R}^k$ *for all* $n \geq 1$*. Define* $\mathcal{S} \subset [\mathbb{R} \cup \{-\infty, \infty\}]^k$ *to be the set of all continuity points of* $t \mapsto F(t) \equiv \text{pr}(X \leq t)$ *and* H *to be the set of all Lipschitz continuous functions* $h : \mathbb{R}^k \mapsto [0,1]$ *(the Lipschitz constants may be* > 1*). Then, provided the expectations are well defined, we have:*

(i) If $\mathsf{E}[h(X_n)|\mathcal{Y}_n] \overset{\mathsf{P}}{\to} \mathsf{E}h(X)$ *for all* $h \in H$*, then* $\sup_{t\in A}|\mathsf{pr}(X_n \le t|\mathcal{Y}_n) - F(t)| \overset{\mathsf{P}}{\to} 0$ *for all closed* $A \subset \mathcal{S}$*;*

(ii) If $\mathsf{E}[h(X_n)|\mathcal{Y}_n] \overset{\mathsf{as}*}{\to} \mathsf{E}h(X)$ *for all* $h \in H$*, then* $\sup_{t\in A}|\mathsf{pr}(X_n \le t|\mathcal{Y}_n) - F(t)| \overset{\mathsf{as}*}{\to} 0$ *for all closed* $A \subset \mathcal{S}$*.*

Proof. Let $t_0 \in \mathcal{S}$. For every $\delta > 0$, there exists $h_1, h_2 \in H$, such that $h_1(u) \le \mathbf{1}\{u \le t_0\} \le h_2(u)$ for all $u \in \mathbb{R}^k$ and $\mathsf{E}[h_2(X) - h_1(X)] < \delta$. Under the condition in (i), we therefore have that $\mathsf{pr}(X_n \le t_0|\mathcal{Y}_n) \overset{\mathsf{P}}{\to} F(t_0)$, since δ was arbitrary. The conclusion of (i) follows since this convergence holds for all $t_0 \in \mathcal{S}$, since both $\mathsf{pr}(X_n \le t|\mathcal{Y}_n)$ and $F(t)$ are monotone in t with range $\subset [0, 1]$, and since $[0, 1]$ is compact. The proof for Part (ii) follows similarly.$\Box$

LEMMA 4. *Let* $\{F_n\}$ *and* F *be distribution functions on* $\mathbb{R}^k$*, and let* $\mathcal{S} \subset [\mathbb{R} \cup \{-\infty, \infty\}]^k$ *be the set of all continuity points of* F*. Then the* *following are equivalent:*

(i) $\sup_{t \in A} |F_n(t) - F(t)| \to 0$ *for all closed* $A \subset \mathcal{S}$.

(ii) $\sup_{h \in BL_1(\mathbb{R}^k)} \left| \int_{\mathbb{R}^k} h(dF_n - dF) \right| \to 0$.

The relatively straightforward proof is saved as Exercise 10.5.3.