Introduction to Empirical Processes and Semiparametric Inference Lecture 15: The Bootstrap and the Delta Method

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Bootstrap Central Limit Theorems

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 1. *The following are equivalent:*

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

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

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

THEOREM 2. *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) $\hat{\mathbb{G}}_n \underset{W}{\overset{\text{as}*}{\rightsquigarrow}} \mathbb{G}$ *in* $\ell^\infty(\mathcal{F})$.

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

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 major continuous mapping results.

The first major 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 $\hat{X}_n \underset{M}{\overset{\mathsf{P}}{\rightsquigarrow}} X$ and that the distribution of $\|X\|$ is continuous.

The next lemma reveals that pr$(\|\hat{X}_n\| \leq t | \mathcal{X}_n)$ converges uniformly to $P(\|X\| \leq t)$, in probability.

LEMMA 1. *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.

A parallel outer almost sure result holds when $\hat{X}_n \overset{\text{as}*}{\underset{M}{\rightsquigarrow}} X$.

The second major result, Theorem 3, 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, 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 $\hat{X}_n \overset{\text{P}}{\underset{M}{\leadsto}} X$*, then* $g(\hat{X}_n) \overset{\text{P}}{\underset{M}{\leadsto}} g(X)$*.*

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

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 $\underset{M}{\overset{\mathrm{as}*}{\leadsto}}$. $\square$

THEOREM 3. *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 $\mathrm{P}^*(X \in \mathbb{D}_0) = 1$,

$$g(\hat{X}_n) \underset{M}{\overset{\mathsf{P}}{\rightsquigarrow}} g(X).$$

The Bootstrap for Glivenko-Cantelli Classes

We now present several results for the bootstrap applied to Glivenko-Cantelli classes.

The primary use of these results is to assist verification of consistency of bootstrapped estimators.

The following corollary applies to a class of weighted bootstraps that includes the Bayesian bootstrap mentioned earlier:

COROLLARY 1. *Let $\mathcal{F}$ be a class of measurable functions, and let $\xi_1, \ldots,$ ξ_n be i.i.d. nonconstant, nonnegative random variables with $0 < \mathsf{E}\xi < \infty$ and independent of $X_1, \ldots, X_n$.*

Let

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

where we set $\tilde{\mathbb{P}}_n = 0$ when $\bar{\xi} = 0$.

Then the following are equivalent:

(i) $\mathcal{F}$ is strong Glivenko-Cantelli.

(ii) $\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} \stackrel{\mathrm{as}*}{\to} 0$ and $P^*\|f - Pf\|_{\mathcal{F}} < \infty$.

(iii) $\mathsf{E}_\xi \|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} \overset{\text{as}*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$.

(iv) *For every* $\eta > 0$, pr $\left(\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} > \eta \middle| \mathcal{X}_n \right) \overset{\text{as}*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$;

(v) *For every* $\eta > 0$, pr $\left(\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}} > \eta \middle| \mathcal{X}_n \right) \overset{\text{as}*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$, *for some version of* $\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}}$.

If in addition $P(\xi = 0) = 0$, *then the requirement that*

$$P^*\|f - Pf\|_{\mathcal{F}} < \infty$$

in (ii) may be dropped.

The following theorem verifies consistency of the multinomial bootstrapped empirical measure defined in Section 10.1.3, which we denote $\hat{\mathbb{P}}_n$, when $\mathcal{F}$ is strong G-C:

THEOREM 4. *Let $\mathcal{F}$ be a class of measurable functions, and let the multinomial vectors W_n in $\hat{\mathbb{P}}_n$ be independent of the data.*

Then the following are equivalent:

(i) *$\mathcal{F}$ is strong Glivenko-Cantelli;*

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

(iii) $\mathsf{E}_W\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} \overset{\text{as}*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$;

(iv) *For every* $\eta > 0$, $\text{pr}\left(\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} > \eta \,\middle|\, \mathcal{X}_n \right) \overset{\text{as}*}{\to} 0$ *and*

$P^*\|f - Pf\|_{\mathcal{F}} < \infty$;

(v) For every $\eta > 0$, pr $\left(\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}} > \eta \middle| \mathcal{X}_n \right) \overset{\text{as}*}{\to} 0$ *and*
$P^*\|f - Pf\|_{\mathcal{F}} < \infty$, *for some version of* $\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}}$.

A Simple Z-Estimator Master Theorem

Consider Z-estimation based on the estimating equation

$$\theta \mapsto \Psi_n(\theta) \equiv \mathbb{P}_n \psi_\theta,$$

where $\theta \in \Theta \subset \mathbb{R}^p$ and $x \mapsto \psi_\theta(x)$ is a measurable p-vector valued function for each θ.

This is a special case of the more general Z-estimation approach discussed in Section 2.2.5.

Define the map

$$\theta \mapsto \Psi(\theta) \equiv P\psi_\theta,$$

and assume $\theta_0 \in \Theta$ satisfies $\Psi(\theta_0) = 0$.

Let $\hat{\theta}_n$ be an approximate zero of Ψ_n, and let $\hat{\theta}_n^\circ$ be an approximate zero of the bootstrapped estimating equation

$$\theta \mapsto \Psi_n^\circ(\theta) \equiv \mathbb{P}_n^\circ \psi_\theta,$$

where $\mathbb{P}_n^\circ$ is either

- $\tilde{\mathbb{P}}_n$ of Corollary 1, with $\xi_1, \ldots, \xi_n$ satisfying the conditions specified in the first paragraph of Section 10.1.3 (the multiplier bootstrap),
- or $\hat{\mathbb{P}}_n$ of Theorem 4 (the multinomial bootstrap).

The goal of this section is to determine reasonably general conditions under which

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow Z,$$

where Z is mean zero normally distributed, and

$$\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) \underset{\circ}{\overset{\mathsf{P}}{\rightsquigarrow}} k_0 Z.$$

Here, we use $\underset{\circ}{\overset{\mathsf{P}}{\rightsquigarrow}}$ to denote either $\underset{\xi}{\overset{\mathsf{P}}{\rightsquigarrow}}$ or $\underset{W}{\overset{\mathsf{P}}{\rightsquigarrow}}$ depending on which bootstrap is being used, and

- $k_0 = \tau/\mu$ for the multiplier bootstrap
- while $k_0 = 1$ for the multinomial bootstrap.

One could also estimate the limiting variance rather than use the bootstrap, but there are many settings, such as least absolute deviation regression, where variance estimation may be more awkward than the bootstrap.

For theoretical validation of the bootstrap approach, we have the following theorem, which is related to Theorem 2.11 and which utilizes some of the bootstrap results of this chapter:

THEOREM 5. *Let* $\Theta \subset \mathbb{R}^p$ *be open, and assume* $\theta_0 \in \Theta$ *satisfies* $\Psi(\theta_0) = 0$.

Also assume the following:

(A) For any sequence $\{\theta_n\} \in \Theta$, $\Psi(\theta_n) \to 0$ *implies* $\|\theta_n - \theta_0\| \to 0$;

(B) The class $\{\psi_\theta : \theta \in \Theta\}$ *is strong Glivenko-Cantelli;*

(C) For some $\eta > 0$*, the class* $\mathcal{F} \equiv \{\psi_\theta : \theta \in \Theta, \|\theta - \theta_0\| \leq \eta\}$ *is Donsker and* $P\|\psi_\theta - \psi_{\theta_0}\|^2 \to 0$ *as* $\|\theta - \theta_0\| \to 0$;

(D) $P\|\psi_{\theta_0}\|^2 < \infty$ *and* $\Psi(\theta)$ *is differentiable at* θ_0 *with nonsingular derivative matrix* V_{θ_0};

(E) $\Psi_n(\hat{\theta}_n) = o_P(n^{-1/2})$ *and* $\Psi_n^\circ(\hat{\theta}_n^\circ) = o_P(n^{-1/2})$.

Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow Z \sim N\left(0, V_{\theta_0}^{-1} P[\psi_{\theta_0}\psi_{\theta_0}^T](V_{\theta_0}^{-1})^T\right)$$

and

$$\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) \underset{\circ}{\overset{\mathsf{P}}{\rightsquigarrow}} k_0 Z.$$

Condition (A) is one of several possible identifiability conditions.

Condition (B) is a sufficient condition, when combined with (A), to yield consistency of a zero of Ψ_n.

This condition is generally reasonable to verify in practice.

Condition (C) is needed for asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and is also not hard to verify in practice.

Condition (D) enables application of the delta method at the appropriate juncture in the proof.

Condition (E) is a specification of the level of approximation permitted in obtaining the zeros of the estimating equations.

Example (Problem 10.5.5): Assume that, given the covariate $Z \in \mathbb{R}^p$, Y is Bernoulli with probability of success $e^{\theta^T Z}/(1+e^{\theta^T Z})$, where $\theta \in \Theta = \mathbb{R}^p$ and $\mathrm{E}[ZZ^T]$ is positive definite.

Assume that we observe an i.i.d. sample $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ generated from this model with true parameter $\theta_0 \in \mathbb{R}^p$.

We can show that the conditions of Theorem 5 are satisfied for Z-estimators based on

$$\psi_\theta(y,z) = Z\left(Y - \frac{e^{\theta^T Z}}{1+e^{\theta^T Z}}\right).$$

The Functional Delta Method

We now build on the presentation of the functional delta method given in Section 2.2.4.

Recall the concept of Hadamard differentiability.

The key result of Section 2.2.4 is that the delta method and its bootstrap counterpart work provided the map ϕ is Hadamard differentiable tangentially to a suitable set $\mathbb{D}_0$.

We first discuss the two main theorems of the functional delta method,

- the functional delta method for Hadamard differentiable maps (Theorem 2.8) and
- the conditional analog for the bootstrap (Theorem 2.9).

We then give several important examples of Hadamard differentiable maps of use in statistics, along with specific illustrations of how those maps are utilized.

THEOREM 6. *For normed spaces $\mathbb{D}$ and $\mathbb{E}$, let $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$ be Hadamard-differentiable at θ tangentially to $\mathbb{D}_0 \subset \mathbb{D}$.*

Assume that $r_n(X_n - \theta) \rightsquigarrow X$ for some sequence of constants $r_n \to \infty$, where X_n takes its values in $\mathbb{D}_\phi$, and X is a tight process taking its values in $\mathbb{D}_0$.

Then

$$r_n(\phi(X_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(X).$$

Proof. Consider the map

$$h \mapsto r_n(\phi(\theta + r_n^{-1}h) - \phi(\theta)) \equiv g_n(h),$$

and note that it is defined on the domain

$$\mathbb{D}_n \equiv \{h : \theta + r_n^{-1}h \in \mathbb{D}_\phi\}$$

and satisfies $g_n(h_n) \to \phi'_\theta(h)$ for every $h_n \to h \in \mathbb{D}_0$ with $h_n \in \mathbb{D}_n$.

Thus the conditions of the extended continuous mapping theorem (Theorem 7.24) are satisfied by $g(\cdot) = \phi'_\theta(\cdot)$.

Hence conclusion (i) of that theorem implies

$$g_n(r_n(X_n - \theta)) \rightsquigarrow \phi'_\theta(X).\square$$

Before presenting the next theorem,

- define $\mathbb{X}_n(X_n)$ to be a sequence of random elements in a normed space $\mathbb{D}$ based on the data sequence $\{X_n, n \geq 1\}$,
- and let $\hat{\mathbb{X}}_n(X_n, W_n)$ be a bootstrapped version of $\mathbb{X}_n$ based on both the data sequence and a sequence of weights $W = \{W_n, n \geq 1\}$.

Note that the proof of this theorem utilizes the bootstrap continuous mapping theorem above (Theorem 3).

THEOREM 7. *For normed spaces $\mathbb{D}$ and $\mathbb{E}$, let $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$ be Hadamard-differentiable at μ tangentially to $\mathbb{D}_0 \subset \mathbb{D}$, with derivative ϕ'_μ.*

Suppose

- *$\mathbb{X}_n$ and $\hat{\mathbb{X}}_n$ have values in $\mathbb{D}_\phi$, with $r_n(\mathbb{X}_n - \mu) \rightsquigarrow \mathbb{X}$,*
- *where $\mathbb{X}$ is tight and takes its values in $\mathbb{D}_0$ for some sequence of constants $0 < r_n \to \infty$,*
- *the maps $W_n \mapsto h(\hat{\mathbb{X}}_n)$ are measurable for every $h \in C_b(\mathbb{D})$ outer almost surely,*
- *and where $r_n c(\hat{\mathbb{X}}_n - \mathbb{X}_n) \underset{W}{\overset{\mathsf{P}}{\rightsquigarrow}} \mathbb{X}$, for a constant $0 < c < \infty$.*

Then

$$r_n c(\phi(\hat{\mathbb{X}}_n) - \phi(\mathbb{X}_n)) \underset{W}{\overset{\mathsf{P}}{\rightsquigarrow}} \phi'_\mu(\mathbb{X}).$$

Delta Method Examples: Composition

Recall from Section 2.2.4 the map $\phi : \mathbb{D}_\phi \mapsto D[0,1]$, where $\phi(f) = 1/f$ and $\mathbb{D}_\phi = \{f \in D[0,1] : |f| > 0\}$.

In that section, we established that ϕ was Hadamard differentiable, tangentially to $D[0,1]$, with derivative at $\theta \in \mathbb{D}_\phi$ equal to $h \mapsto -h/\theta^2$.

This is a simple example of the following general composition result:

LEMMA 2. *Let*

$$g : B \subset \bar{\mathbb{R}} \equiv [-\infty, \infty] \mapsto \mathbb{R}$$

be differentiable with derivative continuous on all closed subsets of B*, and let*

$$\mathbb{D}_\phi = \{A \in \ell^\infty(\mathcal{X}) : \{R(A)\}^\delta \subset B \text{ for some } \delta > 0\},$$

where $\mathcal{X}$ *is a set,* $R(C)$ *denotes the range of the function* $C \in \ell^\infty(\mathcal{X})$*, and* D^δ *is the* δ*-enlargement of the set* D*.*

Then $A \mapsto g \circ A$ *is Hadamard-differentiable as a map from* $\mathbb{D}_\phi \subset \ell^\infty(\mathcal{X})$ *to* $\ell^\infty(\mathcal{X})$*, at every* $A \in \mathbb{D}_\phi$*.*

The derivative is given by ${\phi'}_A(\alpha) = g'(A)\alpha$*, where* g' *is the derivative of* g*.*

Before giving the proof, we briefly return to our simple example of the reciprocal map $A \mapsto 1/A$.

The differentiability of this map easily generalizes from $D[0,1]$ to $\ell^\infty(\mathcal{X})$, for arbitrary $\mathcal{X}$, provided we restrict the domain of the reciprocal map to

$$\mathbb{D}_\phi = \{A \in \ell^\infty(\mathcal{X}) : \inf_{x \in \mathcal{X}} |A(x)| > 0\}.$$

This follows after applying Lemma 2 to the set $B = [-\infty, 0) \cup (0, \infty]$.

Proof of Lemma 2. Note that $\mathbb{D} = \ell^\infty(\mathcal{X})$ in this case, and that the tangent set for the derivative is all of $\mathbb{D}$.

Let t_n be any real sequence with $t_n \to 0$, let $\{h_n\} \in \ell^\infty(\mathcal{X})$ be any sequence converging to $\alpha \in \ell^\infty(\mathcal{X})$ uniformly, and define $A_n = A + t_n h_n$.

Then, by the conditions of the theorem, there exists a closed $B_1 \subset B$ such that $\{R(A) \cup R(A_n)\}^\delta \subset B_1$ for some $\delta > 0$ and all n large enough.

Hence

$$\sup_{x\in\mathcal{X}}\left|\frac{g(A(x)+t_nh_n(x))-g(A(x))}{t_n}-g'(A(x))\alpha(x)\right|\to 0,$$

as $n\to\infty$, since continuous functions on closed sets are bounded, and thus g' is uniformly continuous on B_1.$\Box$

Delta Method Example: Integration

For an $M < \infty$ and an interval $[a, b] \in \bar{\mathbb{R}}$, let $BV_M[a, b]$ be the set of all functions $A \in D[a, b]$ with total variation $|A(0)| + \int_{(a,b]} |dA(s)| \leq M$.

In this section, we consider, for given functions $A \in D[a, b]$ and $B \in BV_M[a, b]$ and domain $\mathbb{D}_M \equiv D[a, b] \times BV_M[a, b]$, the maps $\phi : \mathbb{D}_M \mapsto \mathbb{R}$ and $\psi : \mathbb{D}_M \mapsto D[a, b]$ defined by

$$\phi(A, B) = \int_{(a,b]} A(s)dB(s) \quad \text{and} \quad \psi(A, B)(t) = \int_{(a,t]} A(s)dB(s). \tag{1}$$

The following lemma verifies that these two maps are Hadamard differentiable:

LEMMA 3. *For each fixed $M < \infty$, the maps $\phi : \mathbb{D}_M \mapsto \mathbb{R}$ and $\psi : \mathbb{D}_M \mapsto D[a, b]$ defined in (1) are Hadamard differentiable at each $(A, B) \in \mathbb{D}_M$ with $\int_{(a,b]} |dA| < \infty$.*

The derivatives are given by

$$\begin{aligned} \phi'_{A,B}(\alpha, \beta) &= \int_{(a,b]} A d\beta + \int_{(a,b]} \alpha dB, \quad \textit{and} \\ \psi'_{A,B}(\alpha, \beta)(t) &= \int_{(a,t]} A d\beta + \int_{(a,t]} \alpha dB. \end{aligned}$$

Note that in the above lemma we define

$$\int_{(a,t]} A d\beta = A(t)\beta(t) - A(a)\beta(a) - \int_{(a,t]} \beta(s-) dA(s)$$

so that the integral is well defined even when β does not have bounded variation.

We now look at a statistical applications of Lemma 3 to the two-sample Wilcoxon rank sum statistic.

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from distributions F and G on the reals.

If $\mathbb{F}_m$ and $\mathbb{G}_n$ are the respective empirical distribution functions, the Wilcoxon rank sum statistic for comparing F and G has the form

$$T_1 = m \int_{\mathbb{R}} (m\mathbb{F}_m(x) + n\mathbb{G}_n(x)) d\mathbb{F}_m(x).$$

If we temporarily assume that F and G are continuous, then

$$\begin{aligned} T_1 &= mn \int_{\mathbb{R}} \mathbb{G}_n(x) d\mathbb{F}_m(x) + m^2 \int_{\mathbb{R}} \mathbb{F}_m(x) d\mathbb{F}_m(x) \\ &= mn \int_{\mathbb{R}} \mathbb{G}_n(x) d\mathbb{F}_m(x) + \frac{m^2 + m}{2} \\ &\equiv mnT_2 + \frac{m^2 + m}{2}, \end{aligned}$$

where T_2 is the Mann-Whitney statistic.

When F or G have atoms, the relationship between the Wilcoxon and Mann-Whitney statistics is more complex.

We will now study the asymptotic properties of the Mann-Whitney version of the rank sum statistic, T_2.

For arbitrary F and G, $T_2 = \phi(\mathbb{G}_n, \mathbb{F}_m)$, where ϕ is as defined in Lemma 3.

Note that F, G, $\mathbb{F}_m$ and $\mathbb{G}_n$ all have total variation ≤ 1.

Thus Lemma 3 applies, and we obtain that the Hadamard derivative of ϕ at $(A, B) = (G, F)$ is the map

$$\phi'_{G,F}(\alpha, \beta) = \int_{\mathbb{R}} G d\beta + \int_{\mathbb{R}} \alpha dF,$$

which is continuous and linear over $\alpha, \beta \in D[-\infty, \infty]$.

If we assume that $m/(m+n) \to \lambda \in [0, 1]$, as $m \wedge n \to \infty$, then

$$\sqrt{\frac{mn}{m+n}} \left(\begin{array}{c} \mathbb{G}_n - G \\ \mathbb{F}_m - F \end{array} \right) \rightsquigarrow \left(\begin{array}{c} \sqrt{\lambda}\, \mathbb{B}_1(G) \\ \sqrt{1-\lambda}\, \mathbb{B}_2(F) \end{array} \right),$$

where $\mathbb{B}_1$ and $\mathbb{B}_2$ are independent standard Brownian bridges.

Hence $\mathbb{G}_G(\cdot) \equiv \mathbb{B}_1(G(\cdot))$ and $\mathbb{G}_F(\cdot) \equiv \mathbb{B}_2(F(\cdot))$ both live in $D[-\infty, \infty]$.

Now Theorem 6 yields

$$T_2 \rightsquigarrow \sqrt{\lambda} \int_{\mathbb{R}} G d\mathbb{G}_F + \sqrt{1-\lambda} \int_{\mathbb{R}} \mathbb{G}_F dG,$$

as $m \wedge n \to \infty$.

When $F = G$ and F is continuous, this limiting distribution is mean zero normal with variance $1/12$.

The delta method bootstrap, Theorem 7, is also applicable and can be used to obtain an estimate of the limiting distribution under more general hypotheses on F and G.