Bootstrap-based confidence sets based on directional Hadamard differentiability

February 24, 2017

Reference

Max Sommerfeld, Axel Munk, Inference for Empirical Wasserstein Distances on Finite Spaces [arXiv]

Assumptions

[1, Section 2.1]

Let $\mathcal{X} = \{x_1, \dots, x_N\}$ be finitely many points. Every probability measure on $\mathcal{X}$ is given by a vector r in

$$\mathcal{P}_{\mathcal{X}} = \left\{ r = (r_x)_{x \in \mathcal{X}} \in \mathbb{R}^{\mathcal{X}}_{>0} : \sum_{x \in \mathcal{X}} r_x = 1 \right\}$$

by $P_r(\{x\}) = r_x$. Let $r, s \in \mathcal{P}_{\mathcal{X}}$ and $\hat{r}_n$, $\hat{s}_m$ generated by i.i.d. samples $X_1, \dots, X_n \sim r$ and $Y_1, \dots, Y_m \sim s$, respectively, as $\hat{r}_n = (\hat{r}_{n,x})_{x \in \mathcal{X}}$ where $\hat{r}_{n,x} = \frac{1}{n} \#\{k : X_k = x\}$. We define the multinomial covariance matrix

$$\Sigma(r) = \begin{bmatrix} r_{x_1}(1 - r_{x_1}) & -r_{x_1} r_{x_2} & \cdots & -r_{x_1} r_{x_N} \\ -r_{x_2} r_{x_1} & r_{x_2}(1 - r_{x_2}) & \cdots & -r_{x_2} r_{x_N} \\ \vdots & \vdots & \ddots & \vdots \\ -r_{x_N} r_{x_1} & -r_{x_N} r_{x_2} & \cdots & r_{x_N}(1 - r_{x_N}) \end{bmatrix}$$

and independent Gaussian random variables $G \sim \mathcal{N}(0, \Sigma(r))$ and $H \sim \mathcal{N}(0, \Sigma(s))$. Suppose n and m are approaching infinity such that $n \wedge m \to \infty$ and $m/(n+m) \to \lambda \in (0,1)$.

Hadamard Directional Derivatives

[1, Section 2.2]

Definition. A map Φ defined on a subset $D_\Phi \subset \mathbb{R}^d$ with values in $\mathbb{R}$ is called Hadamard directionally differentiable at $u \in \mathbb{R}^d$ if there exists a map $\phi'_u : \mathbb{R}^d \to \mathbb{R}$ such that

$$\lim_{n \to \infty} \frac{\phi(u + t_n h_n) - \Phi(u)}{t_n} = \phi'_u(h)$$

for any $h \in \mathbb{R}^d$ and for arbitrary sequences t_n converging to zero from above and $h_n \to h$ such that $u + t_n h_n \in D_f$ for all $n \in \mathbb{N}$.

In contrast to usual notion of Hadamard differentiability, the derivative $h \mapsto \phi'_u(h)$ need not be linear.

Example. The absolut value $\Phi : \mathbb{R} \to \mathbb{R}$, $t \mapsto |t|$ is not in the usual sense Hadamard differentiable at $t = 0$ but directionally differentiable with the non-linear derivative $t \mapsto |t|$.

Theorem. *([1, Theorem 2]) Let Φ be a function defined on a subset F of $\mathbb{R}^d$ with values in $\mathbb{R}$, such that*

1. *Φ is Hadamard directionally differentiable at $u \in F$ with derivative $\phi'_u : F \to \mathbb{R}$ and*
2. *there is a sequence of $\mathbb{R}^d$-valued random variables X_n and a sequence of non-negative numbers $\rho_n \to \infty$ such that $\rho_n(X_n - u) \Rightarrow X$ for some random variable X taking values in F.*

Then, $\rho_n(\Phi(X_n) - \Phi(u)) \Rightarrow \phi'_u(X)$.

Bootstrap

[1, Appendix A]

We denote by $\hat{r}_n^*$ and $\hat{s}_m^*$ some bootstrap versions of $\hat{r}_n$ and $\hat{s}_m$. More precisely, let $\hat{r}_n^*$ a measurable function of $X_1, \ldots, X_n$ and random weights $W_1, \ldots, W_n$, independent of the data and analogously for $\hat{s}_m^*$. The bootstrap is consistent if the limiting distribution of

$$\rho_{n,m} \{(\hat{r}_n, \hat{s}_m) - (r, s)\} \Rightarrow (\sqrt{\lambda}G, \sqrt{1-\lambda}H)$$

is consistently estimated by the law of

$$\rho_{n,m}\{(\hat{r}_n^*, \hat{s}_m^*) - (\hat{r}_n, \hat{s}_m)\}.$$

To make this precise, we define for $A \subset \mathbb{R}^d$, the set of bounded Lipschitz-1 functions

$$BL_1(A) = \left\{ f : A \to \mathbb{R} : \sup_{x \in A} |f(x)| \leq 1, \, |f(x_1) - f(x_2)| \leq \|x_1 - x_2\| \right\},$$

where $\|\cdot\|$ is the Euclidean norm. We say that the bootstrap versions $(\hat{r}_n^*, \hat{s}_m^*)$ are onsistent if

$$\sup_{f \in BL_1(\mathbb{R}^{\mathcal{X}} \times \mathbb{R}^{\mathcal{X}})} \left| \mathbb{E}\left[f(\rho_{n,m}\{(\hat{r}_n^*, \hat{s}_m^*) - (\hat{r}_n, \hat{s}_m)\} | \, X_1, \ldots, X_n, Y_1, \ldots, Y_m \right] - \mathbb{E}\left[f((\sqrt{\lambda}G, \sqrt{1-\lambda}H)) \right] \right|$$

converges to zero in probability.

Bootstrap for Directionally Differentiable Functions

[1, Appendix A]

Let $\Phi : F \subset \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be the function that we want to bootstrap, with its directional derivative $\phi_p : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$. In this paper, $\Phi(r, s) = W_p^p(r, s)$ and $\phi_p(h_1, h_2) = \max_{u \in \Phi_p} \langle u, h_2 - h_1 \rangle$.

The most straightforward way to bootstrap $\Phi(\hat{r}_n, \hat{s}_m)$ is to simply plug-in $\hat{r}_n^*$ and $\hat{s}_m^*$, that is, approximate $\rho_{n,m}\{\Phi(\hat{r}_n, \hat{s}_m) - \Phi(r, s)\}$ by the law of

$$\rho_{n,m}\{\Phi(\hat{r}_n^*, \hat{s}_n^*) - \Phi(\hat{r}_n, \hat{s}_m)\}$$

conditional on the data. If Φ were Hadamard differentiable, this approach yields a consistent bootstrap, but this is not in general true for if Φ were only directionally Hadamard differentiable.

There are two approaches: First is to re-sample fewer than n or m, as in part 2 in the following Theorem. Second is to plug in $\rho_{n,m}\{(\hat{r}_n^*, \hat{s}_n^*) - (\hat{r}_n, \hat{s}_m)\}$ into the derivative of the function ϕ_p.

Theorem. *([1, Theorem 5]) Let $\hat{r}_n^*$ and $\hat{s}_m^*$ be consistent bootstrap versions of $\hat{r}_n$ and $\hat{s}_n$. Then,*

1. *The plug-in bootstrap $\rho_{n,m}\{\Phi(\hat{r}_n^*, \hat{s}_n^*) - \Phi(\hat{r}_n, \hat{s}_m)\}$ is not consistent when $\Phi = W_p^p$, that is,*

$$\sup_{f \in BL_1(\mathbb{R})} \mathbb{E}\left[f(\rho_{n,m}\{\Phi(\hat{r}_n^*, \hat{s}_m^*) - \Phi(\hat{r}_n, \hat{s}_m)\}) | \, X_1, \ldots, X_n, Y_1, \ldots, Y_m \right] - \mathbb{E}\left[f(\rho_{n,m}\{\Phi(\hat{r}_n, \hat{s}_m) - \Phi(r, s)\}) \right]$$

   does not converges to zero in probability.

2. *Under the null hypothesis $r = s$, the derivative plug-in*

$$\phi_p(\rho_{n,m}\{(\hat{r}_n^*, \hat{s}_m^*) - (\hat{r}_n, \hat{s}_m)\})$$

   is consistent, that is

$$\sup_{f \in BL_1(\mathbb{R})} \mathbb{E}\left[f(\phi_p(\rho_{n,m}\{(\hat{r}_n^*, \hat{s}_m^*) - (\hat{r}_n, \hat{s}_m)\})) | \, X_1, \ldots, X_n, Y_1, \ldots, Y_m \right] - \mathbb{E}\left[f(\rho_{n,m}\{\Phi(\hat{r}_n, \hat{s}_m) - \Phi(r, s)\}) \right]$$

   does not converges to zero in probability.

3. *Let $\hat{r}_n^{**}$ and $\hat{s}_m^{**}$ be consistent bootstrap versions of $\hat{r}_n$ and $\hat{s}_m$ that are obtained via re-sampling k observations with $k/n \to 0$ and $k/m \to 0$. That is, if*

$$\sup_{f \in BL_1(\mathbb{R})} \mathbb{E}\left[f(\sqrt{k}\{(\hat{r}_n^{**}, \hat{s}_m^{**}) - (\hat{r}_n, \hat{s}_m)\}) | \, X_1, \ldots, X_n, Y_1, \ldots, Y_m \right] - \mathbb{E}\left[f((\sqrt{\lambda}G, \sqrt{1-\lambda}H)) \right]$$

   converges to zero in probability, then the plug-in bootstrap with $\hat{r}_n^{**}$ and $\hat{s}_m^{**}$ is consistent, that is

$$\sup_{f \in BL_1(\mathbb{R})} \mathbb{E}\left[f(\phi_p(\sqrt{k}\{(\hat{r}_n^{**}, \hat{s}_m^{**}) - (\hat{r}_n, \hat{s}_m)\})) | \, X_1, \ldots, X_n, Y_1, \ldots, Y_m \right] - \mathbb{E}\left[f(\rho_{n,m}\{\Phi(\hat{r}_n, \hat{s}_m) - \Phi(r, s)\}) \right]$$

   converges to zero in probability.

Wasserstein Distance on Finite Spaces

[1, Section 1]

The Wasserstein distance of order p between two probability measures $r, s \in \mathcal{P}_\mathcal{X}$ then becomes

$$W_p(r,s) = \left\{ \min_{w \in \Pi(r,s)} \sum_{x,x' \in \mathcal{X}} d^p(x,x') w_{x,x'} \right\}^{1/p},$$

where $\Pi(r,s)$ is the set of all probability measures on $\mathcal{X} \times \mathcal{X}$ with marginal distributions r and s.

Main Result

[1, Section 2.1]

We define the convex sets

$$\Phi_p^* = \left\{ u \in \mathbb{R}^\mathcal{X} : \, u_x - u_{x'} \leq d^p(x,x'), \, x, x' \in \mathcal{X} \right\}.$$

Φ_p^* is the convex set of dual solutions to the Wasserstein problem depending on the metric d only.

Theorem. *([1, Theorem 1], 3) Let* $\rho_{n,m} = \sqrt{\frac{nm}{n+m}}$*. If* $r = s$ *and* n *and* m *are approaching infinity such that* $n \wedge m \to \infty$ *and* $m/(n+m) \to \lambda \in (0,1)$ *we have*

$$\rho_{n,m}^{1/p} W_p(\hat{r}_n, \hat{s}_m) \Rightarrow \left\{ \max_{u \in \Phi_p^*} \langle G, u \rangle \right\}^{\frac{1}{p}}.$$

Directional Derivative of the Wasserstein Distance

[1, Section 2.2]

We define

$$\Phi_p^*(r,s) = \left\{ (u,v) \in \mathbb{R}^\mathcal{X} \times \mathbb{R}^\mathcal{X} : \, \langle u, r \rangle + \langle v, s \rangle = W_p^p(r,s), \, u_x + v_{x'} \leq d^p(x,x'), x, x' \in \mathcal{X} \right\}$$

Theorem. *([1, Theorem 3]) THe functional* $(r,s) \mapsto W_p^p(r,s)$ *is directionally Hadamard differentiable at all* $(r,s) \in \mathcal{P}_\mathcal{X} \times \mathcal{P}_\mathcal{X}$ *with derivative*

$$(h_1, h_2) \mapsto \max_{(u,v) \in \Phi_p^*(r,s)} - \left(\langle u, h_1 \rangle + \langle v, h_2 \rangle \right).$$

References

[1] M. Sommerfeld and A. Munk. Inference for Empirical Wasserstein Distances on Finite Spaces. *ArXiv e-prints*, October 2016.