THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 384 11: Theoretical Statistics

Stein's method: introduction

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

https://psarkar.github.io/teaching

Stein's method

- Stein's method identifies an operator, such that the distribution of interest (we will only do Gaussian) is the **unique fixed point**.
- Define $\mathcal{A}f(x) = f'(x) - xf(x)$.
- This is also known as a *Characterizing operator*.

Notation

- Define $\|g\| := \sup_x |g(x)|$
- An absolutely continuous function f is one that has a derivative f' (which is Lebesgue integrable) almost everywhere. So f can be represented as $f(x) = f(a) + \int_a^x f'(t)dt$
- $\Phi(x) = P(Z \leq x)$ where $Z \sim N(0, 1)$.

Stein's Lemma - properties of the Characterizing Operator

Lemma 1
Define $\mathcal{A}f(x) = f'(x) - xf(x)$*. We have:*

- *If* $Z \sim N(0,1)$*, then* $Ef(Z) = 0$ *for all absolutely continuous functions* f *with* $E[|f'(Z)|] < \infty$*.*
- *If for some random variable* W*,* $E\mathcal{A}f(W) = 0$*, and for absolutely continuous functions* f *with* $\|f'\| < \infty$*, then* $W \sim N(0,1)$*.*

Properties of solutions

Lemma 2
Let $\Phi(x)$ denote the CDF of a Gaussian random variable evaluated at x. The unique bounded solution to the differential equation

$$f_x'(w) - wf_x(w) = 1(w \leq x) - \Phi(x)$$

is given by:

$$\begin{aligned} f_x(w) &= e^{w^2/2} \int_w^\infty e^{-t^2/2}(\Phi(x) - 1(t \leq x))dt \\ &= -e^{w^2/2} \int_{-\infty}^w e^{-t^2/2}(\Phi(x) - 1(t \leq x))dt \end{aligned}$$

Proof of lemma 2

- Let's drop the subscript for a second.
- First note that

$$\frac{d}{dt} e^{-t^2/2} f(t) = e^{-t^2/2}(f'(t) - tf(t))$$

- So

$$\begin{aligned}
e^{t^2/2} \frac{d}{dt} \left(e^{-t^2/2} f(t) \right) &= 1(t \leq x) - \Phi(x) \\
\frac{d}{dt} e^{-t^2/2} f(t) &= e^{-t^2/2} \left(1(t \leq x) - \Phi(x)\right) \\
e^{-w^2/2} f(w) &= \int_w^\infty e^{-t^2/2} \left(\Phi(x) - 1(t \leq x)\right) dt + C \\
&= -\int_{-\infty}^w e^{-t^2/2} \left(\Phi(x) - 1(t \leq x)\right) dt + C
\end{aligned}$$

Continuing

- Note that we are looking at the only bounded solution. So $C = 0$.
- But if $C = 0$, is f bounded?

$$\begin{aligned}&\left| e^{w^2/2} \int_w^\infty e^{-t^2/2} \left(\Phi(x) - 1(t \leq x) \right) dt \right| \\ &\leq \sqrt{2\pi} e^{w^2/2} (1 - \Phi(w))\end{aligned}$$

- Now use the fact that

$$1 - \Phi(w) \leq \min\left\{ \frac{1}{2}, \frac{1}{w\sqrt{2\pi}} \right\} e^{-w^2/2}$$

- So $\|f\| \leq \sqrt{\pi/2}$

Proof of Stein's lemma

- First note that, we have

$$e^{-z^2/2} = \int_z^\infty ue^{-u^2/2} du = -\int_{-\infty}^z ue^{-u^2/2} du$$

- So now, using Fubini's theorem, since $E|f'(Z)|$ is finite,

$$\begin{aligned}
&E[f'(Z)] = \frac{1}{\sqrt{2\pi}} \int f'(z)e^{-z^2/2} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty f'(z) \int_z^\infty ue^{-u^2/2} dudz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 f'(z) \int_{-\infty}^z ue^{-u^2/2} dudz \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty ue^{-u^2/2} \int_0^u f'(z) dzdu - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 ue^{-u^2/2} \int_u^0 f'(z) dzdu \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty ue^{-u^2/2} (f(u) - f(0)) du + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 ue^{-u^2/2} (f(u) - f(0)) du \\
&= E[Zf(Z)]
\end{aligned}$$

Stein's lemma - other direction

- Note that $f_x(w)$ for $x \in \mathbb{R}$ are all absolutely continuous functions wtih $\|f'\| < \infty$. Hence using the condition of this part of the lemma, $E\mathcal{A}f_x(W) = 0$, $E[1(w \leq x) - \Phi(x)] = P(W \leq x) - \Phi(x) = 0$
- So, $\forall x$, $P(W \leq x) = \Phi(x)$, i.e. $W \sim N(0,1)$.

General setup

- Define $\mathcal{H}$ as a family of *test* functions, X, Y are two random variables.
- Define the distance between the distributions of X, Y as

$$d_{\mathcal{H}}(X, Y) = \sup_{h \in \mathcal{H}} |Eh(X) - Eh(Y)|$$

- When $\mathcal{H} = \{1(. \leq x) | x \in \mathbb{R}\}$, we have the Kolmogorov metric. Denote by d_K.
- When $\mathcal{H} = \{h : |h(x) - h(y)| \leq |x - y|\}$, we have the Wasserstein metric, denoted by d_W
- We have $d_K(X, Z) \leq \sqrt{2Cd_W(X, Z)}$, where C is the bound on the density of Z

General solution

- Let

$$f_h'(w) - wf_h(w) = h(x) - \Phi(h)$$

where $\Phi(h) = Eh(Z)$.

Proposition 3
Let W be a random variable and $Z \sim N(0,1)$, then

$$d_{\mathcal{H}}(W,Z) = \sup_{h \in \mathcal{H}} \left| E\left[f_h'(W) - Wf_h'(W) \right] \right|$$

Lemma 2 for general function families

Lemma 4
Let $\Phi(x)$ denote the CDF of a Gaussian random variable evaluated at x. The unique bounded solution to the differential equation

$$f_h'(w) - wf_h(w) = h(x) - \Phi(h)$$

is given by:

$$\begin{aligned} f_h(w) &= e^{w^2/2} \int_w^\infty e^{-t^2/2}(\Phi(h) - h(t))dt \\ &= -e^{w^2/2} \int_{-\infty}^w e^{-t^2/2}(\Phi(h) - h(t))dt \end{aligned}$$

- *For bounded h, $\|f_h\| \leq \sqrt{\pi/2}\|h(.) - \Phi(h)\|$, $\|f_h'\| \leq 2\|h(.) - \Phi(h)\|$*
- *For absolutely continuous h $\|f_h\|, \|f_h''\| \leq 2\|h'\|$, $\|f_h'\| \leq \sqrt{2/\pi}\|h'\|$*

General theorem for bounding d_W

Theorem 5
Let W be a random variable and $Z \sim N(0,1)$. Define $\mathcal{F} = \{f : \|f\|, \|f''\| \leq 2, \|f'\| \leq \sqrt{2/\pi}\}$. Then we have,

$$d_W(W,Z) \leq \sup_{f \in \mathcal{F}} \left| E[f'(W) - Wf(W)] \right|$$

Proof.

- First use Lemma 4 to construct f_h.
- Using Prop 3, $d_{\mathcal{H}}(W,Z) = \sup_{h \in \mathcal{H}} \left| E\left[f_h'(W) - Wf_h'(W) \right] \right|$.
- But $f_h \in \mathcal{F}$ for $h \in \mathcal{H}$, hence the upper bound.

$\square$

Lets apply this

Theorem 6
Let $X_1, \ldots, X_n$ be independent mean zero random variables with $EX_i^2 = 1$ and $EX_i^4 < \infty$. Let $W = \sum_i X_i/\sqrt{n}$.

$$d_W(W, Z) \leq \frac{1}{n^{3/2}} \sum_i E[|X_i|^3] + \frac{\sqrt{2/\pi}}{n} \sqrt{\sum_i EX_i^4}$$

- Define $W_i = \sum_{j \neq i} X_j$.
- Claim 1: $E[X_i f(W)] = E[X_i(f(W) - f(W_i))]$

Proof cont.

- So, we have:

$$E[Wf(W)] = \sum_i \frac{X_i(f(W) - f(W_i))}{\sqrt{n}} \tag{1}$$

- So, lets do a Taylor expansion. Let ξ_i be in between W_i and W.

$$f(W_i) = f(W) + (W_i - W)f'(W) + \frac{(W_i - W)^2}{2} f''(\xi_i)$$

- So we have:

$$X_i(f(W) - f(W_i)) = -X_i(W_i - W)f'(W) - X_i \frac{(W_i - W)^2}{2} f''(\xi_i)$$

- So the RHS of (1) is:

$$\frac{1}{n} \sum_i X_i^2 f'(W) - \frac{1}{n^{3/2}} \sum_i X_i^3 f''(\xi_i)$$

Proof cont.

- So, now we will evaluate $E[Wf(W) - f'(W)]$

$$
\begin{aligned}
\left|E[Wf(W) - f'(W)]\right| &\leq \sqrt{\frac{2}{\pi}} E \left| \frac{\sum_i X_i^2}{n} - 1 \right| + \frac{2}{n^{3/2}} E \left| \sum_i X_i^3 \right| \\
&\leq \sqrt{\frac{2}{\pi}} \sqrt{\mathrm{var}(\sum_i X_i^2/n)} + \frac{2}{n^{3/2}} \sum_i E|X_i^3| \\
&\leq \sqrt{\frac{2}{n\pi}} \sqrt{\sum_i EX_i^4} + \frac{2}{n^{3/2}} \sum_i E|X_i^3|
\end{aligned}
$$

- Last line follows from $\mathrm{var}(X_i^2) \leq EX_i^4$.

Proof cont.

- Note that while the d_W converges at the $n^{-1/2}$ rate, the upper bound on d_K gives a suboptimal $1/n^{1/4}$ rate.
- It is possible to get Berry Esseen, but you will need a bit more careful analysis, since f_h is not *as smooth*.
- The main reason of using the Wasserstein metric is that the solutions f_h when h belongs to $1-$Lipschitz functions are twice differentiable, while in contrast, when h are indicators over a half line, they are only differentiable once.
- These notes were based on material from 1) L.H.Y. Chen et. al., Normal Approximation by Stein's Method, Probability and Its Applications and 2) Nathan Ross's "Fundamentals of Stein's method" in Probability Surveys.