A Gentle Introduction to Stein's Method for Normal Approximation I

Larry Goldstein

University of Southern California

Introduction to Stein's Method for Normal Approximation

1. Much activity since Stein's 1972 paper
2. Any introduction, and this one in particular, is necessarily somewhat 'biased', a selective tour of a larger area than the one presented


Stein Mart
GRAND OPENING

Introduction to Stein's Method for Normal Approximation

I. Background, Stein Identity, Equation, Bounds

II. Size Bias Couplings

III. Exchangeable Pair, Zero Bias Couplings

IV. Local dependence, Nonsmooth functions

De Moive/Laplace Theorem, 1733/1820

Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with distribution

$$P(X_1 = 1) = P(X_1 = -1) = 1/2,$$

and let

$$S_n = X_1 + \cdots + X_n \quad \text{and} \quad W_n = S_n/\sqrt{n}.$$

Then $W_n \to_d Z$ with $Z \sim \mathcal{N}(0,1)$, that is, for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} P(W_n \leq x) = \Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2. Then

$$W_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \to_d Z \quad \text{as } n \to \infty.$$

May relax the identical distribution assumption by imposing the Lindeberg Condition.

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be independent and identically distributed with mean μ and variance σ^2. Then

$$W_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \to_d Z \quad \text{as } n \to \infty.$$

If in addition $E|X_1|^3 < \infty$, a bound on the distance between the distribution function F_n of W_n and Φ is provided by the Berry-Esseen Theorem,

$$\sup_{-\infty<x<\infty} |F_n(x) - \Phi(x)| \leq \frac{cE|X_1|^3}{\sigma^3\sqrt{n}}$$

where $c \leq 0.7056$ [Shevtsova (2006)].

Simple Random Sampling

Let $\mathcal{A} = \{a_1, \ldots, a_N\}$ be real numbers, and $X_1, \ldots, X_n$ a simple random sample of $\mathcal{A}$ of size $0 < n < N$.

Prove $W_n = (S_n - ES_n)/\sqrt{\text{Var}(S_n)}$ satisfies

$$W_n \to_d Z \quad \text{as } n \to \infty.$$

As the variables are dependent, the classical CLT does not apply.

Simple Random Sampling

Proved by Hájek in 1960, over 200 years after Laplace.

Demonstrating that the sum of a simple random sample is asymptotically normal, due to the dependence, seems to be somewhat harder than handling the independent case.

Approach by Stein's Method

Without loss of generality, suppose that

$$\sum_{a \in \mathcal{A}} a = 0.$$

Let $X', X'', X_2, \ldots, X_{n+1}$ be a simple random sample of size $n+1$ of $\mathcal{A}$, and set

$$W' = X' + \sum_{i=2}^{n} X_i \quad \text{and} \quad W'' = X'' + \sum_{i=2}^{n} X_i.$$

The pair W', W'' is exchangeable, that is,

$$(W', W'') =_d (W'', W').$$

Approach by Stein's Method

In addition to (W', W'') being an exchangeable pair,

$$E[X'|W'] = \frac{1}{n}W', \quad \text{and} \quad E[X''|W'] = -\frac{1}{N-n}W'.$$

As $W'' - W' = X'' - X'$, with $2 < n < N-1$,

$$E[W''|W'] = (1-\lambda)W' \quad \text{for } \lambda = \tfrac{N}{n(N-n)} \in (0,1).$$

We call such a (W', W'') a *Stein pair*. We are handling the random variables directly.

Approach by Stein's Method

The construction of this Stein pair indicates that Stein's method might be used to prove Hájek's theorem, and additionally provide a Berry-Esseen bound.

Bounds to the normal using these types of approaches typically reduce to the computation of, or bounds on, low order moments, perhaps even only on variances of certain quantities.

Under the principle 'there is no such thing as a free lunch', such variance computations may be difficult.

Stein's Lemma

Characterization of the normal distribution:

$$Z \sim \mathcal{N}(0,1)$$

if and only if

$$E[Zf(Z)] = E[f'(Z)]$$

for all absolutely continuous functions f such that $E|f'(Z)| < \infty$.

Stein's Lemma

Characterization of the $\mathcal{N}(0,1)$ distribution:

$$E[Zf(Z)] = E[f'(Z)] \quad \text{if and only if } Z \sim \mathcal{N}(0,1).$$

Note that the normal $\mathcal{N}(0,1)$ density function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

satisfies the 'dual' equation

$$z\phi(z) = -\phi'(z).$$

Proof of Stein's Lemma

If direction: Assume Z is standard normal.

Break $Ef'(Z)$ into two parts to handle separately,

$$
\begin{aligned}
Ef'(Z) &= \int_{-\infty}^{\infty} f'(z)\phi(z)dz \\
&= \int_{0}^{\infty} f'(z)\phi(z)dz + \int_{-\infty}^{0} f'(z)\phi(z)dz.
\end{aligned}
$$

Proof of Stein's Lemma

Consider the first integral. Use Fubini's theorem to change the order of integration to obtain

$$\begin{aligned}
\int_0^\infty f'(z)\phi(z)dz &= \int_0^\infty f'(z)\int_z^\infty y\phi(y)dydz \\
&= \int_0^\infty \int_z^\infty f'(z)y\phi(y)dydz \\
&= \int_0^\infty \int_0^y f'(z)y\phi(y)dzdy \\
&= \int_0^\infty [f(y)-f(0)]y\phi(y)dy.
\end{aligned}$$

Proof of Stein's Lemma

Similarly, for the second integral,

$$\int_{-\infty}^{0} f'(z)\phi(z)dz = \int_{-\infty}^{0} [f(y) - f(0)]y\phi(y)dy.$$

Combining gives

$$Ef'(Z) = EZ[f(Z) - f(0)] = EZf(Z).$$

Stein's Lemma, $\mathcal{N}(\mu, \sigma^2)$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $(X-\mu)/\sigma \sim \mathcal{N}(0,1)$, so

$$E\left(\frac{X-\mu}{\sigma}\right) g\left(\frac{X-\mu}{\sigma}\right) = Eg'\left(\frac{X-\mu}{\sigma}\right).$$

Letting $f(x) = g((x-\mu)/\sigma)$, we have

$$f'(x) = \sigma^{-1} g'((x-\mu)/\sigma) \quad \text{or} \quad \sigma f'(x) = g'((x-\mu)/\sigma),$$

so, in general $X \sim \mathcal{N}(\mu, \sigma^2)$ if and only if

$$E(X-\mu)f(X) = \sigma^2 E f'(X).$$

Stein's Method: Basic Idea

If $\mathcal{F}$ is a sufficiently large class of functions, and

$$E(X-\mu)f(X) = \sigma^2 E f'(X) \quad \text{for all } f \in \mathcal{F}$$

then $X \sim \mathcal{N}(\mu, \sigma^2)$.

Hence, if

$$E(W-\mu)f(W) \approx \sigma^2 E f'(W) \quad \text{for all } f \in \mathcal{F}$$

then $W \approx \mathcal{N}(\mu, \sigma^2)$.

Hence, we would like to show that

$$E[(W-\mu)f(W) - \sigma^2 f'(W)] \approx 0.$$

Distance to Normality

Let X and Y be random variables. Many distances between the distributions of X and Y can be given by

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

for some class of functions $\mathcal{H}$.

Distance may also be given in terms of functions of the distribution functions F and G, of X and Y, respectively.

Kolmogorov, or L^∞ distance

Letting X and Y have distribution functions F and G, respectively,

$$\begin{aligned} ||F - G||_\infty &= \sup_{-\infty < z < \infty} |F(z) - G(z)| \\ &= \sup_{h \in \mathcal{H}} |Eh(X) - Eh(Y)| \end{aligned}$$

for

$$\mathcal{H} = \{\mathbf{1}_{(-\infty, z]}(x) : z \in \mathbb{R}\}.$$

Wasserstein, or L^1 distance

$$\begin{aligned} ||F-G||_1 &= \int_{-\infty}^{\infty} |F(z)-G(z)|dz \\ &= \sup_{h\in\mathcal{H}} |Eh(X)-Eh(Y)| \end{aligned}$$

for

$$\mathcal{H} = \{h : |h(x)-h(y)| \leq |x-y|\}.$$

Have also that

$$||F-G||_1 = \inf E|X-Y|$$

over all couplings of X and Y on a joint space with the given marginals.

Total Variation Distance

Total variation distance

$$||F - G||_{\mathsf{TV}} = \sup_{h \in \mathcal{H}} |Eh(X) - Eh(Y)|$$

where

$$\mathcal{H} = \{h : 0 \leq h \leq 1\}.$$

From the Stein Identity to the Stein Equation

Can measure distance from W to a standard normal Z by

$$Eh(W) - Nh$$

for $Nh = Eh(Z)$ over $\mathcal{H}$.

Stein identity says discrepancy between W and Z is also reflected in

$$E[f'(W) - Wf(W)].$$

From the Stein Identity to the Stein Equation

Can measure distance from W to normal Z by

$$Eh(W) - Nh$$

for $Nh = Eh(Z)$ over $\mathcal{H}$.

Stein identity says discrepancy between W and Z is also reflected in

$$E[f'(W) - Wf(W)].$$

Equating these quantities at w yields the Stein Equation:

$$f'(w) - wf(w) = h(w) - Nh.$$

Stein Equation

$$f'(w) - wf(w) = h(w) - Nh. \qquad (1)$$

Goal: Given $h \in \mathcal{H}$, compute (a bound) on

$$Eh(W) - Nh.$$

Approach: Given h, solve (1) for f and instead compute

$$E\left[f'(W) - Wf(W)\right].$$

Apriori, this appears harder. It also requires bounds on f in terms of h.

Solution to the Stein Equation

For h such that Nh exists, write

$$f'(w) - wf(w) = h(w) - Nh \tag{2}$$

as

$$e^{w^2/2}\frac{d}{dw}\left(e^{-w^2/2}f(w)\right) = h(w) - Nh$$

and

$$f(w) = e^{w^2/2}\int_{-\infty}^{w}[h(z) - Nh]e^{-z^2/2}dz$$

is the unique bounded solution to (2). Note also that f has one more derivative than h.

Solution to the Stein Equation

Since $Eh(Z) - Nh = 0$,

$$
\begin{aligned}
f(w) &= e^{w^2/2} \int_{-\infty}^{w} [h(u) - Nh] e^{-u^2/2} du \\
&= -e^{w^2/2} \int_{w}^{\infty} [h(u) - Nh] e^{-u^2/2} du.
\end{aligned}
$$

Bounds on the Solution to the Stein Equation

Example: If f is the unique bounded solution to the Stein equation for a bounded function h, then

$$||f||_\infty \leq \sqrt{\frac{\pi}{2}}||h - Nh||_\infty.$$

Bounds on Solution to the Stein Equation

For $w \leq 0$, we have

$$\begin{aligned} |f(w)| &= e^{w^2/2} \left| \int_{-\infty}^{w} [h(u) - Nh] e^{-u^2/2} du \right| \\ &\leq \sup_{x \leq 0} |h(x) - Nh| e^{w^2/2} \int_{-\infty}^{w} e^{-u^2/2} du, \end{aligned}$$

while

$$\frac{d}{dw} e^{w^2/2} \int_{-\infty}^{w} e^{-u^2/2} du = 1 + w e^{w^2/2} \int_{-\infty}^{w} e^{-u^2/2} du > 0,$$

increasing over $(-\infty, 0]$, so maximum is attained at $w = 0$.

Bounds on Solution to the Stein Equation

For $w \leq 0$,

$$\begin{aligned} |f(w)| &\leq \sup_{x \leq 0} |h(x) - Nh| e^{w^2/2} \int_{-\infty}^{w} e^{-u^2/2} du \\ &\leq \sup_{x \leq 0} |h(x) - Nh| \int_{-\infty}^{0} e^{-u^2/2} du \\ &= \sup_{x \leq 0} |h(x) - Nh| \sqrt{2\pi} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-u^2/2} du \right) \\ &= \sup_{x \leq 0} |h(x) - Nh| \sqrt{\frac{\pi}{2}}. \end{aligned}$$

Similar bound for $w \geq 0$.

Bounds on Solution of the Stein Equation

If h is bounded

$$||f||_\infty \leq \sqrt{\frac{\pi}{2}}||h - Nh||_\infty,$$

and absolutely continuous

$$||f'||_\infty \leq 2||h - Nh||_\infty,$$

with $||h'||_\infty < \infty$,

$$||f''||_\infty \leq 2||h'||_\infty.$$

Indicator Bounds

For $h_z(w) = \mathbf{1}_{(-\infty,z]}(w)$ [Chen and Shao (2005)] (Singapore lecture notes)

$$0 < f(w) \leq \frac{\sqrt{2\pi}}{4} \quad \text{and} \quad |f'(w)| \leq 1.$$

Also will use one additional bound for 'smoothed indicators' which decrease from 1 to zero over interval of some small length $\lambda > 0$.

Proof of Stein's Lemma

Only if direction. Suppose $E[f'(Z)] = E[Zf(Z)]$ for all absolutely continuous functions f with $E|f'(Z)| < \infty$.

Let $f(w)$ be the solution to the Stein Equation for $h(w) = \mathbf{1}_{(-\infty,z]}(w)$. Then f is absolutely continuous and $||f'||_\infty < \infty$. Hence

$$E[\mathbf{1}_{(-\infty,z]}(Z) - \Phi(z)] = E[f'(Z) - Zf(Z)] = 0,$$

that is

$$P(Z \leq z) = E\mathbf{1}_{(-\infty,z]}(Z) = \Phi(z) \quad \text{for all } z \in \mathbb{R}.$$

Generator Approach to the Stein Equation

[Barbour (1990), Götze (1991)]

Note that with one more derivative

$$(\mathcal{A}f)(w) = \sigma^2 f''(w) - wf'(w)$$

is the generator of the Ornstein-Uhlenbeck process W_t,

$$dW_t = \sqrt{2}\sigma dB_t - W_t dt$$

which has the normal $\mathcal{N}(0, \sigma^2)$ as its unique stationary distribution.

Generator Approach

Letting T_t be the transition operator of the Ornstein-Uhlenbeck process $W_t, t \geq 0$,

$$(T_t h)(x) = E\left[h(W_t)|W_0 = x\right],$$

a solution to $\mathcal{A}f = h - Nh$ is given by

$$f = -\int_0^\infty T_t h dt.$$

Technique works also in higher dimensional, and more abstract spaces.

Coming Attractions

II. Size Bias Couplings

III. Exchangeable Pair, Zero Bias Couplings

IV. Local dependence, Nonsmooth functions