Stein's Method: Past and Future

Yevgeniy Kovchegov
Oregon State University

Joint work with
Wasamon Jantai
Chulalongkorn University

Introduction.

Disclaimer: for clarity purposes, we may sometimes omit specifying what function spaces or metrics are being considered.

Observe that for $Z \overset{d}{\sim} \mathcal{N}(0,1)$, and $f \in \mathcal{C}^1(\mathbb{R})$ satisfying $E\Big[|f'(Z)|\Big] < \infty$,

$$E\Big[f'(Z) - Z\,f(Z)\Big] = 0$$

as integration by parts

$$\int_{-\infty}^{\infty} f'(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = -f(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} xf(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx$$

yields $E\Big[f'(Z)\Big] = E\Big[Z\,f(Z)\Big]$.

Observe that $\forall\epsilon > 0\ \exists a > 0$ s.t. $\int_a^\infty |f'(t)|e^{-t^2/2}dt < \epsilon/2$.

Thus, for $x > 0$ sufficiently large so that $|f(a)|e^{-x^2/2} < \epsilon/2$,

$$\left|f(x)e^{-x^2/2}\right| = e^{-x^2/2}\left|f(a) + \int_a^x f'(t)dt\right| \le |f(a)|e^{-x^2/2} + \int_a^\infty |f'(t)|e^{-t^2/2}dt < \epsilon.$$

Introduction.

Observe that for $Z \overset{d}{\sim} \mathcal{N}(0,1)$, and $f \in \mathcal{C}^1(\mathbb{R})$ satisfying $E\Big[|f'(Z)|\Big] < \infty$,

$$E\Big[f'(Z) - Z\,f(Z)\Big] = 0.$$

Stein's Lemma. Random variable $X \overset{d}{\sim} \mathcal{N}(0,1)$ if and only if

$$E\Big[f'(X) - X\,f(X)\Big] = 0$$

for all $f \in \mathcal{C}^1(\mathbb{R})$ satisfying $E\Big[|f'(Z)|\Big] < \infty$ for $Z \sim \mathcal{N}(0,1)$.

Sketch Proof: $X \overset{d}{\sim} \mathcal{N}(0,1)$ iff $E[g(X)] = E[g(Z)]$ for all g from a class of functions such as $\Big\{\cos(\alpha x)\Big\}_{\alpha\in\mathbb{R}} \bigcup \Big\{\sin(\alpha x)\Big\}_{\alpha\in\mathbb{R}}$.

For each such g, one finds $f \in \mathcal{C}^1(\mathbb{R})$ by solving the first-order linear ODE

$$f'(x) - xf(x) = g(x) - E[g(Z)].$$

Hence, $E[g(X)] = E[g(Z)]$ for all functions g in the class iff $E\Big[f'(X) - X\,f(X)\Big] = 0$ for all corresponding f.

Introduction.

Let $\mathcal{L} =$ Lipschitz functions and $\mathcal{D}_M = \Big\{ f : \|f\|_\infty, \|f'\|_\infty, \|f''\|_\infty < M \Big\}$ for a given $M > 0$. Then, $\exists C_M > 0$ such that for any random variable X such that the following expectations are finite,

$$\sup_{g\in\mathcal{L}} \Big| E[g(X)] - E[g(Z)] \Big| \le C_M \sup_{f\in\mathcal{D}_M} \Big| E\Big[f'(X) - X\, f(X) \Big] \Big|,$$

where $Z \overset{d}{\sim} \mathcal{N}(0,1)$.

Thus, a sequence of random variables $X_n \xrightarrow{d} \mathcal{N}(0,1)$ if and only if

$$E\Big[f'(X_n) - X_n\, f(X_n) \Big] \to 0 \qquad \forall f \in \mathcal{D}_M.$$

Stein-Markov operator: $Af(x) = f'(x) - x\, f(x)$

Stein operator: $\mathcal{A}f(x) = f''(x) - x\, f'(x)$

Summary: $X_n \xrightarrow{d} \mathcal{N}(0,1) \;\Leftrightarrow\; E[Af(X_n)] \to 0 \;\; \forall f \in \mathcal{D}_M$

$\Leftrightarrow \;\; E[\mathcal{A}f(X_n)] \to 0 \;\; \forall f$ s.t. $\|f'\|_\infty, \|f''\|_\infty, \|f'''\|_\infty < M$.

Multivariate normal.

Consider an m-dimensional normal $Z \stackrel{d}{\sim} \mathcal{N}(0, \Sigma)$ with $m \times m$ covariance matrix Σ.

Stein-Markov operator: for $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^m$,

$$Af(\mathbf{x}) = \nabla^T \Sigma \mathbf{f}(\mathbf{x}) - \mathbf{x}^T \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m.$$

Stein operator: for $f : \mathbb{R}^m \to \mathbb{R}$,

$$\mathcal{A}f(\mathbf{x}) = \nabla^T \Sigma \nabla f(\mathbf{x}) - \mathbf{x}^T \nabla f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m.$$

For $Z \stackrel{d}{\sim} \mathcal{N}(0, \Sigma)$ and $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^m$, integration by part yields

$$E[\nabla^T \Sigma \mathbf{f}(Z)] = (2\pi)^{-d/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\mathbf{x}^T \Sigma^{-1} \mathbf{x}/2} \, \nabla^T \Sigma \mathbf{f}(\mathbf{x}) \, dx_1 \dots dx_m$$

$$= (2\pi)^{-d/2} \Sigma \mathbf{f}(\mathbf{x}) e^{-\mathbf{x}^T \Sigma^{-1} \mathbf{x}/2} \Big|_{-\infty}^{\infty} \dots \Big|_{-\infty}^{\infty} + (2\pi)^{-d/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\mathbf{x}^T \Sigma^{-1} \mathbf{x}/2} \, \mathbf{x}^T \mathbf{f}(\mathbf{x}) dx_1 \dots dx_m$$

Multivariate normal.

Consider an m-dimensional normal $Z \stackrel{d}{\sim} \mathcal{N}(0, \Sigma)$ with $m \times m$ positive definite covariance matrix Σ.

Stein-Markov operator: for $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^m$,

$$Af(\mathbf{x}) = \nabla^T \Sigma \mathbf{f}(\mathbf{x}) - \mathbf{x}^T \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m.$$

Stein operator: for $f : \mathbb{R}^m \to \mathbb{R}$,

$$\mathcal{A}f(\mathbf{x}) = \nabla^T \Sigma \nabla f(\mathbf{x}) - \mathbf{x}^T \nabla f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m.$$

For $Z \stackrel{d}{\sim} \mathcal{N}(0, \Sigma)$ and $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^m$, integration by part yields

$$E[Af(Z)] = E\left[\nabla^T \Sigma \mathbf{f}(Z) - Z^T \mathbf{f}(Z)\right] = 0$$

and for $f : \mathbb{R}^m \to \mathbb{R}$,

$$E[\mathcal{A}f(Z)] = E\left[\nabla^T \Sigma \nabla f(Z) - Z^T \nabla f(Z)\right] = 0.$$

Multivariate normal.

Consider an m-dimensional normal $Z \overset{d}{\sim} \mathcal{N}(0, \Sigma)$ with $m \times m$ co-variance matrix Σ.

Stein operator: for $f : \mathbb{R}^m \to \mathbb{R}$,

$$\mathcal{A}f(\mathbf{x}) = \nabla^T \Sigma \nabla f(\mathbf{x}) - \mathbf{x}^T \nabla f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m.$$

Multidimensional Stein's Lemma. Random variable $X \overset{d}{\sim} \mathcal{N}(0, \Sigma)$ if and only if $E\Big[\mathcal{A}f(X)\Big] = 0$ for all $f : \mathbb{R}^m \to \mathbb{R}$ in $\mathcal{C}^2(\mathbb{R})$ satisfying $E\Big[|\nabla^T \Sigma \nabla f(Z)|\Big] < \infty$ for $Z \sim \mathcal{N}(0, \Sigma)$.

Convergence criterion. $X_n \xrightarrow{d} \mathcal{N}(0, \Sigma) \quad \Leftrightarrow \quad E[\mathcal{A}f(X_n)] \to 0$ $\forall f$ with bounded first, second, and third partial derivatives.

Brownian bridge.

Let $B(t)$ be a standard Brownian motion. The process

$$W(t) = B(t) - tB(1), \quad t \in [0,1],$$

is called a Brownian bridge.

Observe that for $0 \le s \le t \le 1$, $Cov\Big(W(s), W(t)\Big) = s(1-t)$.

Hence, for $0 \le t_1, \ldots, t_m \le 1$, random vector

$$\mathbf{w} = \Big(W(t_1), \ldots, W(t_m)\Big)^T \overset{d}{\sim} \mathcal{N}(0, \Sigma)$$

with the covariance matrix

$$\Sigma = \Big((t_i \wedge t_j)(1 - t_i \vee t_j)\Big)_{i,j}$$

and the Stein operator for $f : \mathbb{R}^m \to \mathbb{R}$ given by

$$\mathcal{A}f(\mathbf{w}) = \nabla^T \Sigma \nabla f(\mathbf{w}) - \mathbf{w}^T \nabla f(\mathbf{w})$$

$$= -\langle \nabla f(\mathbf{w}), \mathbf{w} \rangle + \sum_{i,j} (t_i \wedge t_j)(1 - t_i \vee t_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{w}).$$

Fréchet derivatives.

Consider Banach spaces V and W. For $x \in U \subseteq V$ and $f : U \to W$, a bounded linear operator $D = Df(x)[\cdot] : V \to W$ is said to be the Fréchet derivative if

$$\lim_{\|h\| \to 0} \frac{\left\| f(x+h) - f(x) - Dh \right\|}{\|h\|} = 0.$$

A bounded bilinear operator $D^2 = D^2 f(x)[\cdot, \cdot] : V \times V \to W$ is said to be the second-order Fréchet derivative if

$$\lim_{\|k\| \to 0} \frac{\left\| Df(x+k)[y] - Df(x)[y] - D^2[y,k] \right\|}{\|k\|} = 0$$

uniformly for bounded $y \in V$.

Then,

$$f(x+a) = f(x) + Df(x)[a] + D^2 f(x)[a,a] + \epsilon[f;a]\|a\|^2,$$

where $\|\epsilon[f;a]\| \to 0$ as $\|a\| \to 0$.

Fréchet derivatives.

Consider Banach spaces V and W. For $x \in U \subseteq V$ and $f : U \to W$, a bounded linear operator $D = Df(x)[\cdot] : V \to W$ is said to be the Fréchet derivative if

$$\lim_{\|h\| \to 0} \frac{\left\| f(x+h) - f(x) - Dh \right\|}{\|h\|} = 0.$$

A bounded bilinear operator $D^2 = D^2 f(x)[\cdot, \cdot] : V \times V \to W$ is said to be the second-order Fréchet derivative if

$$\lim_{\|k\| \to 0} \frac{\left\| Df(x+k)[y] - Df(x)[y] - D^2[y,k] \right\|}{\|k\|} = 0$$

uniformly for bounded $y \in V$.

For $V = D[0,1]$ and $W = \mathbb{R}$, iterating the Riesz-Markov-Kakutani Theorem yields

$$D^2 f(y)[h_1, h_2] = \int_0^1 \int_0^1 h_1(t_1) h_2(t_2) \, d\varphi_y(t_1, t_2),$$

where the Borel measure $|\varphi_y(A_1 \times A_2)| \leq \varphi(A_1 \times A_2)$ is uniformly bounded for all $y \in D[0,1]$.

Functional Stein operators.

Let $B(t)$ be a standard Brownian motion and $W(t)$ be a Brownian bridge.

In 1990, A. D. Barbour proposed the following functional Stein operators. For a given $s \in [0,1]$, let $J_s(t) = \mathbf{1}_{t \geq s} \in D[0,1]$.

Brownian motion: For $f : D[0,1] \to \mathbb{R}$ and $\mathbf{u} \in D[0,1]$, let

$$\mathcal{A}f(\mathbf{u}) = -Df(\mathbf{u})[\mathbf{u}] + \int_0^1 D^2 f(\mathbf{u})[J_s, J_s]\, ds.$$

$E[\mathcal{A}f(B_n)] \to 0$ for all f in a certain exotic metric space $\Rightarrow$ $B_n \xrightarrow{d}$ BM

Brownian bridge: For $f : D[0,1] \to \mathbb{R}$ and $\mathbf{u} \in D[0,1]$, let

$$\mathcal{A}f(\mathbf{u}) = -Df(\mathbf{u})[\mathbf{u}] + \int_0^1 D^2 f(\mathbf{u})[J_s - I, J_s - I]\, ds, \quad \text{where } I(t) = t.$$

$E[\mathcal{A}f(W_n)] \to 0$ for all f in a certain exotic metric space $\Rightarrow$ $W_n \xrightarrow{d}$ BB

Functional Stein operators.

Consider Brownian bridge $W(t) = B(t) - tB(1)$ ($t \in [0,1]$). For a given $s \in [0,1]$, let $J_s(t) = \mathbf{1}_{t \geq s} \in D[0,1]$.

We show the intuition behind A. D. Barbour's functional Stein operator for $W(t)$,

$$\mathcal{A}f(\mathbf{u}) = -Df(\mathbf{u})[\mathbf{u}] + \int_0^1 D^2 f(\mathbf{u})[J_s - I, J_s - I]\, ds, \quad \text{where } I(t) = t.$$

Here,

$$\int_0^1 D^2 f(\mathbf{u})[J_s - I, J_s - I]\, ds = \int_0^1\int_0^1\int_0^1 (J_s(t_1) - t_1)(J_s(t_2) - t_2)\, ds\, d\varphi_{\mathbf{u}}(t_1, t_2)$$

with

$$\int_0^1 (J_s(t_1) - t_1)(J_s(t_2) - t_2)\, ds = t_1 \wedge t_2 - t_1 t_2 = (t_1 \wedge t_2)(1 - t_1 \vee t_2),$$

and therefore,

$$\mathcal{A}f(\mathbf{u}) = -Df(\mathbf{u})[\mathbf{u}] + \int_0^1\int_0^1 (t_1 \wedge t_2)(1 - t_1 \vee t_2)\, d\varphi_{\mathbf{u}}(t_1, t_2).$$

Functional Stein operators.

Consider Brownian bridge $W(t) = B(t) - tB(1)$ ($t \in [0,1]$). For a given $s \in [0,1]$, let $J_s(t) = \mathbf{1}_{t \geq s} \in D[0,1]$.

We show the intuition behind A. D. Barbour's functional Stein operator for $W(t)$. Here,

$$\mathcal{A}f(\mathbf{u}) = -Df(\mathbf{u})[\mathbf{u}] + \int_0^1 \int_0^1 (x_1 \wedge x_2)(1 - x_1 \vee x_2)\, d\varphi_{\mathbf{u}}(x_1, x_2).$$

Now, recall that for $0 \leq t_1, \ldots, t_m \leq 1$, random vector

$$\mathbf{w} = \Big(W(t_1), \ldots, W(t_m)\Big)^T \overset{d}{\sim} \mathcal{N}(0, \Sigma)$$

with the covariance matrix

$$\Sigma = \Big((t_i \wedge t_j)(1 - t_i \vee t_j)\Big)_{i,j}$$

and the Stein operator for $f : \mathbb{R}^m \to \mathbb{R}$ given by

$$\mathcal{A}f(\mathbf{w}) = \nabla^T \Sigma \nabla f(\mathbf{w}) - \mathbf{w}^T \nabla f(\mathbf{w})$$

$$= -\langle \nabla f(\mathbf{w}), \mathbf{w} \rangle + \sum_{i,j} (t_i \wedge t_j)(1 - t_i \vee t_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{w}).$$

Liggett's limit theorem.

The following is a special case of the limit theorem from T. M. Liggett's Ph.D. thesis, which by coincidence, was reproved in my own Ph.D. thesis (in the context of a percolation problem).

Theorem. Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables such that, for some $p \in (0,1)$,

$$P(X_i = 1) = p \quad \text{and} \quad P(X_i = -1) = 1 - p.$$

Then, for $S(t) = \sum_{k:k\le t} X_k$,

$$Y_n(t) = \left(\frac{1}{\sqrt{2n}} S(2nt) \;\middle|\; S(2n) = 0 \right)_{t\in[0,1]} \xrightarrow{d} \text{BB}.$$

In my student's (W. Jantai's) Ph.D. thesis, we showed that $E[\mathcal{A}f(Y_n)] \to 0$ for twice Fréchet differentiable $f : D[0,1] \to \mathbb{R}$ with bounded $D^2 f$, i.e., $\exists K_f > 0$ s.t. $\|D^2 f(\mathbf{u})\| < K_f(1 + \|\mathbf{u}\|)$.

Our contribution: we extended the Stein's method for the sums of exchangeable random variables.

Fixing the functional Stein operator approach.

Consider $0 \le t_1, \dots, t_m \le 1$ and $g : \mathbb{R}^m \to \mathbb{R}$ in $C^2[0,1]^m$, and define $f : D[0,1] \to \mathbb{R}$ as follows:

$$f(\mathbf{u}) = g\big(\mathbf{u}(t_1), \dots, \mathbf{u}(t_m)\big) \text{ for } \mathbf{u} \in D[0,1].$$

Such $f(\mathbf{u})$ is twice Fréchet differentiable $f : D[0,1] \to \mathbb{R}$ with bounded $D^2 f$, and

$$\begin{aligned}\mathcal{A}f(\mathbf{u}) &= -Df(\mathbf{u})[\mathbf{u}] + \int_0^1 D^2 f(\mathbf{u})[J_s - I, J_s - I]\, ds \\ &= -\langle \nabla g(\mathbf{u}), \mathbf{u} \rangle + \sum_{i,j} (t_i \wedge t_j)(1 - t_i \vee t_j) \frac{\partial^2}{\partial x_i \partial x_j} g(\mathbf{u}).\end{aligned}$$

Hence, showing $\mathcal{A}f(Y_n) \to 0$ as $n \to 0$ implies

$$\big(Y_n(t_1), \dots, Y_n(t_m)\big)^T \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad \text{where } \Sigma = \Big((t_i \wedge t_j)(1 - t_i \vee t_j)\Big)_{i,j}.$$

Thus, showing the convergence of multidimensional distribution. Together with tightness, this yields

$$Y_n \xrightarrow{d} \mathrm{BB}.$$

What future holds for the Stein's method.

I believe that the future of the Stein's method is in extending the Stein's method in multivariate setting. The 1-D Gaussian Free Fields (GFF) is either standard Brownian motion or the Brownian Bridge. By the analogy to this work, next step is showing convergence of random fields to the GFF with Stein's method.

Another aspect is of interest to me is that the eigenfunctions of Stein operators are orthogonal polynomials. They are the solutions of Sturm-Liouville differential or difference equations $\mathcal{A}Q_j = \lambda_j Q_j$ for the corresponding eigenvalues λ_j.

• For $\mathcal{N}(0,1)$ distribution, Q_j are the Hermite polynomials.

• For multivariate $\mathcal{N}(0,\Sigma)$ distribution, it is multivariate Hermite-like polynomials: for $j=(j_1,\ldots,j_m)\in\mathbb{Z}_+^m$,

$$Q_j(\mathbf{x}) = (-1)^{j_1+\cdots+j_m} e^{\frac{\mathbf{x}^T\Sigma^{-1}\mathbf{x}}{2}} \frac{\partial^{j_1+\cdots+j_m}}{\partial^j x_1^{j_1}\ldots\partial x_m^{j_m}} e^{-\frac{\mathbf{x}^T\Sigma^{-1}\mathbf{x}}{2}}.$$

• For Gamma(α,λ) distribution, the Stein-Markov operator

$$Af(x) = xf'(x) + (\alpha-\lambda x)f(x),$$

the Stein operator $\mathcal{A}f = Af'$, and Q_j are the Laguerre polynomials.

What future holds for the Stein's method.

- For Beta(a, b) distribution, the Stein-Markov operator

$$Af(x) = x(1-x)f'(x) + (a - (a+b)x)f(x),$$

the Stein operator $\mathcal{A}f = Af'$, and Q_j are the Jacobi polynomials.

- For Poi(λ) distribution, the Stein-Markov operator

$$Af(x) = f(x) - \frac{x}{\lambda} f(x-1),$$

the Stein operator $\mathcal{A}f = A\Delta f$, where $\Delta f(x) = f(x+1) - f(x)$ denotes the forward difference, and Q_j are the Charlier polynomials.

- For Bin(n, p) distribution, the Stein-Markov operator

$$Af(x) = (1-p)x\nabla f(x) + \Big(np - x\Big) f(x),$$

where $\nabla f(x) = f(x) - f(x-1)$ is the backward difference, the Stein operator $\mathcal{A}f = A\Delta f$, and Q_j are the Krawtchouk polynomials.

Finally, it will be beneficial to consider functional Stein operators for Gamma, Poisson, Meixner, and other Lévy processes.

References.

• Thomas M. Liggett, "Weak Convergence of Conditioned Sums of Independent Random Vectors" Ph.D. thesis, Stanford University (1969)

• Y. K., "Brownian bridge, percolation and related processes" Ph.D. thesis, Stanford University (2002)

• Wasamon Jantai, "Proving limit theorems for associated random variables via Stein's method" Ph.D. thesis, Oregon State University (2021)

• Andrew D. Barbour "Stein's method for diffusion approximations" PTRF (1990)

• Wim Schoutens, "Stochastic Processes and Orthogonal Polynomials" Springer, NY (2000)

• Y. K. and Nese Yildiz, "Orthogonal polynomials for semiparametric instrumental variables model" ESAIM: Probability and Statistics (2015)