Stein's Method: The Discrete Case

Elizabeth Meckes

American Institute of Mathematics
and
Case Western Reserve University

February 18, 2008

What is Stein’s method and what is it good for?

What is Stein's method and what is it good for?

- Stein's method is a technique for approximating the distribution of a random variable you care about by some model distribution (normal, Poisson, gamma, semi-circle, etc.)

What is Stein's method and what is it good for?

- Stein's method is a technique for approximating the distribution of a random variable you care about by some model distribution (normal, Poisson, gamma, semi-circle, etc.)
- The method has no *a priori* requirements for any particular structure of the random variable (e.g., it need not be a sum), or for any independence.

What is Stein's method and what is it good for?

- Stein's method is a technique for approximating the distribution of a random variable you care about by some model distribution (normal, Poisson, gamma, semi-circle, etc.)
- The method has no *a priori* requirements for any particular structure of the random variable (e.g., it need not be a sum), or for any independence.
- The method automatically produces rates of convergence.

What is Stein's method and what is it good for?

- Stein's method is a technique for approximating the distribution of a random variable you care about by some model distribution (normal, Poisson, gamma, semi-circle, etc.)
- The method has no *a priori* requirements for any particular structure of the random variable (e.g., it need not be a sum), or for any independence.
- The method automatically produces rates of convergence.
- It's useful when you already have a guess as to a good approximating distribution for your random variable.

The Characterizing Operator

The Characterizing Operator

Let X be a random variable. A characterizing operator for X is an operator T_o on some class of functions $\mathcal{A}$, such that, for any random variable Y,

$$\mathbb{E} T_o f(Y) = 0 \quad \forall f \in \mathcal{A} \qquad \text{iff} \qquad Y \stackrel{d}{=} X.$$

The Characterizing Operator

Let X be a random variable. A characterizing operator for X is an operator T_o on some class of functions $\mathcal{A}$, such that, for any random variable Y,

$$\mathbb{E}T_o f(Y) = 0 \quad \forall f \in \mathcal{A} \qquad \text{iff} \qquad Y \stackrel{d}{=} X.$$

Examples:

The Characterizing Operator

Let X be a random variable. A characterizing operator for X is an operator T_o on some class of functions $\mathcal{A}$, such that, for any random variable Y,

$$\mathbb{E}T_o f(Y) = 0 \quad \forall f \in \mathcal{A} \qquad \text{iff} \qquad Y \stackrel{d}{=} X.$$

Examples:

- Standard Normal: $T_o f(x) = f'(x) - xf(x)$ for $f : \mathbb{R} \to \mathbb{R}$.

The Characterizing Operator

Let X be a random variable. A characterizing operator for X is an operator T_o on some class of functions $\mathcal{A}$, such that, for any random variable Y,

$$\mathbb{E}T_o f(Y) = 0 \quad \forall f \in \mathcal{A} \qquad \text{iff} \qquad Y \stackrel{d}{=} X.$$

Examples:

- Standard Normal: $T_o f(x) = f'(x) - xf(x)$ for $f : \mathbb{R} \to \mathbb{R}$.
  $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[f'(x) - xf(x)\right] e^{-\frac{x^2}{2}} dx = 0$ by integration by parts.

The Characterizing Operator

Let X be a random variable. A characterizing operator for X is an operator T_o on some class of functions $\mathcal{A}$, such that, for any random variable Y,

$$\mathbb{E}T_o f(Y) = 0 \quad \forall f \in \mathcal{A} \qquad \text{iff} \qquad Y \stackrel{d}{=} X.$$

Examples:

- Standard Normal: $T_o f(x) = f'(x) - xf(x)$ for $f : \mathbb{R} \to \mathbb{R}$.
  $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[f'(x) - xf(x)\right] e^{-\frac{x^2}{2}} dx = 0$ by integration by parts.
- Poisson(λ): $T_o f(j) = \lambda f(j+1) - jf(j)$ for $f : \mathbb{N} \to \mathbb{R}$.

The Characterizing Operator

Let X be a random variable. A characterizing operator for X is an operator T_o on some class of functions $\mathcal{A}$, such that, for any random variable Y,

$$\mathbb{E}T_o f(Y) = 0 \quad \forall f \in \mathcal{A} \qquad \text{iff} \qquad Y \stackrel{d}{=} X.$$

Examples:

- Standard Normal: $T_o f(x) = f'(x) - xf(x)$ for $f : \mathbb{R} \to \mathbb{R}$.
  $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[f'(x) - xf(x)\right] e^{-\frac{x^2}{2}} dx = 0$ by integration by parts.
- Poisson(λ): $T_o f(j) = \lambda f(j+1) - jf(j)$ for $f : \mathbb{N} \to \mathbb{R}$.
- Exponential(λ): $T_o f(x) = f'(x) - \lambda f(x)$ for $f : \mathbb{R}^+ \to \mathbb{R}$ with $f(0) = 0$.

Approximation

Approximation

Let Y be a random variable of interest (e.g., a normalized sum of i.i.d random variables), and let X be a random variable with characterizing operator T_o which is conjectured to approximate Y well.

Approximation

Let Y be a random variable of interest (e.g., a normalized sum of i.i.d random variables), and let X be a random variable with characterizing operator T_o which is conjectured to approximate Y well.

Idea: instead of trying to show that $\mathbb{E}T_o f(Y) = 0$ for all $f \in \mathcal{A}$, (which is probably not true), try to show that $\mathbb{E}T_o f(Y)$ is *small* for all $f \in \mathcal{A}$. This will imply that Y is close to X in some sense.

Approximation

Let Y be a random variable of interest (e.g., a normalized sum of i.i.d random variables), and let X be a random variable with characterizing operator T_o which is conjectured to approximate Y well.

Idea: instead of trying to show that $\mathbb{E}T_o f(Y) = 0$ for all $f \in \mathcal{A}$, (which is probably not true), try to show that $\mathbb{E}T_o f(Y)$ is *small* for all $f \in \mathcal{A}$. This will imply that Y is close to X in some sense.

Specifically, the size of $\mathbb{E}T_o f(Y)$ determines the distance (in some metric) between X and Y; if Y depends on a parameter n tending to infinity, then the size of $\mathbb{E}T_o f(Y)$ determines the rate of convergence of Y to X.

The Stein Equation

The Stein Equation

Implementing this idea involves solving the Stein equation: given a function g, find f such that

$$T_o f(x) = g(x) - \mathbb{E}g(X).$$

Such an f is usually denoted $f(x) = U_o g(x)$.

The Stein Equation

Implementing this idea involves solving the Stein equation: given a function g, find f such that

$$T_o f(x) = g(x) - \mathbb{E}g(X).$$

Such an f is usually denoted $f(x) = U_o g(x)$.

Given such a function f, observe that

$$\mathbb{E}T_o f(Y) = \mathbb{E}g(Y) - \mathbb{E}g(X).$$

The Stein Equation

Implementing this idea involves solving the Stein equation: given a function g, find f such that

$$T_o f(x) = g(x) - \mathbb{E}g(X).$$

Such an f is usually denoted $f(x) = U_o g(x)$.

Given such a function f, observe that

$$\mathbb{E}T_o f(Y) = \mathbb{E}g(Y) - \mathbb{E}g(X).$$

Thus if $\mathbb{E}T_o f(Y)$ is small, then $\mathbb{E}g(Y) - \mathbb{E}g(X)$ is small.

This leads naturally to notions of distance between the random variables X and Y which can be expressed in the form

$$d(X, Y) = \sup_{\mathcal{F}} \left| \mathbb{E}g(X) - \mathbb{E}g(Y) \right|,$$

where the supremum is over some class $\mathcal{F}$ of test functions g.

This leads naturally to notions of distance between the random variables X and Y which can be expressed in the form

$$d(X, Y) = \sup_{\mathcal{F}} \left| \mathbb{E}g(X) - \mathbb{E}g(Y) \right|,$$

where the supremum is over some class $\mathcal{F}$ of test functions g.

Examples:

This leads naturally to notions of distance between the random variables X and Y which can be expressed in the form

$$d(X, Y) = \sup_{\mathcal{F}} \left| \mathbb{E}g(X) - \mathbb{E}g(Y) \right|,$$

where the supremum is over some class $\mathcal{F}$ of test functions g.

Examples:

- $\mathcal{F} = \{f : \|f\|_\infty \leq 1, \textit{continuous}\}$ $\longleftrightarrow$ total variation distance.

This leads naturally to notions of distance between the random variables X and Y which can be expressed in the form

$$d(X, Y) = \sup_{\mathcal{F}} \left|\mathbb{E}g(X) - \mathbb{E}g(Y)\right|,$$

where the supremum is over some class $\mathcal{F}$ of test functions g.

Examples:

- $\mathcal{F} = \{f : \|f\|_\infty \leq 1, \textit{continuous}\}$ $\longleftrightarrow$ total variation distance.
- $\mathcal{F} = \{f : \|f'\|_\infty \leq 1\}$ $\longleftrightarrow$ Wasserstein distance.

This leads naturally to notions of distance between the random variables X and Y which can be expressed in the form

$$d(X, Y) = \sup_{\mathcal{F}} \left| \mathbb{E}g(X) - \mathbb{E}g(Y) \right|,$$

where the supremum is over some class $\mathcal{F}$ of test functions g.

Examples:

- $\mathcal{F} = \{f : \|f\|_\infty \leq 1, \textit{continuous}\}$ $\longleftrightarrow$ total variation distance.
- $\mathcal{F} = \{f : \|f'\|_\infty \leq 1\}$ $\longleftrightarrow$ Wasserstein distance.
- $\mathcal{F} = \{f : \|f\|_\infty + \|f'\|_\infty \leq 1\}$ $\longleftrightarrow$ bounded Lipschitz distance.

So What?

So What?

Instead of trying to estimate the distance between X and Y directly, the problem has been reduced to trying to estimate $\mathbb{E}T_o f(Y)$ for some large class of functions f. Why is this any better?

So What?

Instead of trying to estimate the distance between X and Y directly, the problem has been reduced to trying to estimate $\mathbb{E}T_o f(Y)$ for some large class of functions f. Why is this any better?

Various techniques are in use for trying to estimate $\mathbb{E}T_o f(Y)$. Among them:

So What?

Instead of trying to estimate the distance between X and Y directly, the problem has been reduced to trying to estimate $\mathbb{E}T_o f(Y)$ for some large class of functions f. Why is this any better?

Various techniques are in use for trying to estimate $\mathbb{E}T_o f(Y)$. Among them:

- The method of exchangeable pairs (e.g. Stein's book)

So What?

Instead of trying to estimate the distance between X and Y directly, the problem has been reduced to trying to estimate $\mathbb{E}T_o f(Y)$ for some large class of functions f. Why is this any better?

Various techniques are in use for trying to estimate $\mathbb{E}T_o f(Y)$. Among them:

- The method of exchangeable pairs (e.g. Stein's book)
- The dependency graph method (e.g. Arratia, Goldstein, and Gordon or Barbour, Karoński, and Ruciński)

So What?

Instead of trying to estimate the distance between X and Y directly, the problem has been reduced to trying to estimate $\mathbb{E}T_o f(Y)$ for some large class of functions f. Why is this any better?

Various techniques are in use for trying to estimate $\mathbb{E}T_o f(Y)$. Among them:

- The method of exchangeable pairs (e.g. Stein's book)
- The dependency graph method (e.g. Arratia, Goldstein, and Gordon or Barbour, Karoński, and Ruciński)
- Size-bias coupling (e.g. Goldstein and Rinott)

So What?

Instead of trying to estimate the distance between X and Y directly, the problem has been reduced to trying to estimate $\mathbb{E}T_o f(Y)$ for some large class of functions f. Why is this any better?

Various techniques are in use for trying to estimate $\mathbb{E}T_o f(Y)$. Among them:

- The method of exchangeable pairs (e.g. Stein's book)
- The dependency graph method (e.g. Arratia, Goldstein, and Gordon or Barbour, Karoński, and Ruciński)
- Size-bias coupling (e.g. Goldstein and Rinott)
- Zero-bias coupling (e.g. Goldstein and Reinert)

So What?

Instead of trying to estimate the distance between X and Y directly, the problem has been reduced to trying to estimate $\mathbb{E}T_o f(Y)$ for some large class of functions f. Why is this any better?

Various techniques are in use for trying to estimate $\mathbb{E}T_o f(Y)$. Among them:

- The method of exchangeable pairs (e.g. Stein's book)
- The dependency graph method (e.g. Arratia, Goldstein, and Gordon or Barbour, Karoński, and Ruciński)
- Size-bias coupling (e.g. Goldstein and Rinott)
- Zero-bias coupling (e.g. Goldstein and Reinert)
- The generator method (Barbour)

The idea of the method of exchangeable pairs

The idea of the method of exchangeable pairs

- Suppose you have a random variable W which you conjecture is well-approximated by X. Make a “small random change” to W to get a new random variable W', such that $(W, W') \stackrel{d}{=} (W', W)$.

The idea of the method of exchangeable pairs

- Suppose you have a random variable W which you conjecture is well-approximated by X. Make a "small random change" to W to get a new random variable W', such that $(W, W') \stackrel{d}{=} (W', W)$.

- The goal is to bound $\left|\mathbb{E} T_o f(W)\right|$. Many characterizing operators T_o are defined using derivatives or differences. Use the fact that W and W' are close to express or approximate those derivatives or differences in terms of (W, W').

The idea of the method of exchangeable pairs

- Suppose you have a random variable W which you conjecture is well-approximated by X. Make a "small random change" to W to get a new random variable W', such that $(W, W') \stackrel{d}{=} (W', W)$.

- The goal is to bound $\left|\mathbb{E}T_o f(W)\right|$. Many characterizing operators T_o are defined using derivatives or differences. Use the fact that W and W' are close to express or approximate those derivatives or differences in terms of (W, W').

- Use the fact that W' was constructed explicitly from W together with the nesting property of conditional expectation to help evaluate/estimate the resulting espression.

Poisson approximation with exchangeable pairs

Poisson approximation with exchangeable pairs

Recall that the characterizing operator for the Poisson distribution with mean λ is defined on functions $f : \mathbb{N} \cup \{0\} \to \mathbb{R}$ and given by

$$T_o f(j) = \lambda f(j+1) - j f(j).$$

Poisson approximation with exchangeable pairs

Recall that the characterizing operator for the Poisson distribution with mean λ is defined on functions $f : \mathbb{N} \cup \{0\} \to \mathbb{R}$ and given by

$$T_o f(j) = \lambda f(j+1) - jf(j).$$

Suppose (W, W') is an exchangeable pair of random variables, and that W is conjectured to be approximately distributed as Poisson with mean λ.

Poisson approximation with exchangeable pairs

Recall that the characterizing operator for the Poisson distribution with mean λ is defined on functions $f : \mathbb{N} \cup \{0\} \to \mathbb{R}$ and given by

$$T_o f(j) = \lambda f(j+1) - jf(j).$$

Suppose (W, W') is an exchangeable pair of random variables, and that W is conjectured to be approximately distributed as Poisson with mean λ.

$$0 = \mathbb{E}\left[f(W')\mathbb{I}(W' = W+1) - f(W)\mathbb{I}(W = W'+1)\right]$$

Poisson approximation with exchangeable pairs

Recall that the characterizing operator for the Poisson distribution with mean λ is defined on functions $f : \mathbb{N} \cup \{0\} \to \mathbb{R}$ and given by

$$T_o f(j) = \lambda f(j+1) - j f(j).$$

Suppose (W, W') is an exchangeable pair of random variables, and that W is conjectured to be approximately distributed as Poisson with mean λ.

$$\begin{aligned} 0 &= \mathbb{E}\left[f(W')\mathbb{I}(W' = W+1) - f(W)\mathbb{I}(W = W'+1)\right] \\ &= \mathbb{E}\left[f(W+1)\mathbb{I}(W' = W+1) - f(W)\mathbb{I}(W' = W-1)\right] \end{aligned}$$

Poisson approximation with exchangeable pairs

Recall that the characterizing operator for the Poisson distribution with mean λ is defined on functions $f : \mathbb{N} \cup \{0\} \to \mathbb{R}$ and given by

$$T_o f(j) = \lambda f(j+1) - jf(j).$$

Suppose (W, W') is an exchangeable pair of random variables, and that W is conjectured to be approximately distributed as Poisson with mean λ.

$$\begin{aligned}
0 &= \mathbb{E}\left[f(W')\mathbb{I}(W' = W+1) - f(W)\mathbb{I}(W = W'+1)\right] \\
&= \mathbb{E}\left[f(W+1)\mathbb{I}(W' = W+1) - f(W)\mathbb{I}(W' = W-1)\right] \\
&= \mathbb{E}\left[f(W+1)\mathbb{P}\left[W' = W+1 \middle| W\right] - f(W)\mathbb{P}\left[W' = W-1 \middle| W\right]\right].
\end{aligned}$$

$$0 = \mathbb{E}\left[f(W+1)\mathbb{P}\left[W' = W+1 \middle| W\right] - f(W)\mathbb{P}\left[W' = W-1 \middle| W\right]\right].$$

$$0 = \mathbb{E}\left[f(W+1)\mathbb{P}\left[W' = W+1 \middle| W\right] - f(W)\mathbb{P}\left[W' = W-1 \middle| W\right]\right].$$

Suppose there is a constant c such that

$$0 = \mathbb{E}\left[f(W+1)\mathbb{P}\left[W' = W+1 \middle| W\right] - f(W)\mathbb{P}\left[W' = W-1 \middle| W\right]\right].$$

Suppose there is a constant c such that

- $\mathbb{P}\left[W' = W+1 \middle| W\right] = c\,(\lambda + E_1)$.

$$0 = \mathbb{E}\left[f(W+1)\mathbb{P}\left[W' = W+1 \middle| W \right] - f(W)\mathbb{P}\left[W' = W-1 \middle| W \right] \right].$$

Suppose there is a constant c such that

- $\mathbb{P}\left[W' = W+1 \middle| W \right] = c\,(\lambda + E_1)$.
- $\mathbb{P}\left[W' = W-1 \middle| W \right] = c\,(W + E_2)$.

$$0 = \mathbb{E}\left[f(W+1)\mathbb{P}\left[W' = W+1 \middle| W\right] - f(W)\mathbb{P}\left[W' = W-1 \middle| W\right]\right].$$

Suppose there is a constant c such that

- $\mathbb{P}\left[W' = W+1 \middle| W\right] = c\,(\lambda + E_1)$.
- $\mathbb{P}\left[W' = W-1 \middle| W\right] = c\,(W + E_2)$.

Then

$$0 = c\mathbb{E}\Big[\lambda f(W+1) - Wf(W) + E_1 f(W+1) - E_2 f(W)\Big].$$

$$0 = \mathbb{E}\left[f(W+1)\mathbb{P}\left[W' = W+1 \middle| W\right] - f(W)\mathbb{P}\left[W' = W-1 \middle| W\right]\right].$$

Suppose there is a constant c such that

- $\mathbb{P}\left[W' = W+1 \middle| W\right] = c\,(\lambda + E_1)$.
- $\mathbb{P}\left[W' = W-1 \middle| W\right] = c\,(W + E_2)$.

Then

$$0 = c\mathbb{E}\Big[\underbrace{\lambda f(W+1) - Wf(W)}_{= T_o f(W)} + E_1 f(W+1) - E_2 f(W)\Big].$$

$$0 = \mathbb{E}\left[f(W+1)\mathbb{P}\left[W' = W+1 \middle| W\right] - f(W)\mathbb{P}\left[W' = W-1 \middle| W\right]\right].$$

Suppose there is a constant c such that

- $\mathbb{P}\left[W' = W+1 \middle| W\right] = c\,(\lambda + E_1)$.
- $\mathbb{P}\left[W' = W-1 \middle| W\right] = c\,(W + E_2)$.

Then

$$0 = c\mathbb{E}\Bigg[\underbrace{\lambda f(W+1) - Wf(W)}_{= T_o f(W)} + E_1 f(W+1) - E_2 f(W)\Bigg].$$

That is,

$$\mathbb{E}T_o f(W) = \mathbb{E}\left[E_2 f(W) - E_1 f(W+1)\right].$$

Example – Poisson-binomial trials

Example – Poisson-binomial trials

Let $\{X_i\}_{i=1}^{n}$ be i.i.d. indicator random variables, with $\mathbb{P}(X_i = 1) = \frac{1}{n}$ for each i.

Example – Poisson-binomial trials

Let $\{X_i\}_{i=1}^n$ be i.i.d. indicator random variables, with $\mathbb{P}(X_i = 1) = \frac{1}{n}$ for each i.

Let $W = \sum_{i=1}^n X_i$; W is approximately distributed as a Poisson random variable with mean 1.

Example – Poisson-binomial trials

Let $\{X_i\}_{i=1}^n$ be i.i.d. indicator random variables, with $\mathbb{P}(X_i = 1) = \frac{1}{n}$ for each i.

Let $W = \sum_{i=1}^n X_i$; W is approximately distributed as a Poisson random variable with mean 1.

Make an exchangeable pair as follows:

Example – Poisson-binomial trials

Let $\{X_i\}_{i=1}^n$ be i.i.d. indicator random variables, with $\mathbb{P}(X_i = 1) = \frac{1}{n}$ for each i.

Let $W = \sum_{i=1}^n X_i$; W is approximately distributed as a Poisson random variable with mean 1.

Make an exchangeable pair as follows: Let $\{X_i'\}_{i=1}^n$ be an independent copy of the sequence $\{X_i\}$, and let $K \in \{1, \ldots, n\}$ be uniformly distributed, independent of the X_i and X_i'.

Example – Poisson-binomial trials

Let $\{X_i\}_{i=1}^n$ be i.i.d. indicator random variables, with $\mathbb{P}(X_i = 1) = \frac{1}{n}$ for each i.

Let $W = \sum_{i=1}^n X_i$; W is approximately distributed as a Poisson random variable with mean 1.

Make an exchangeable pair as follows: Let $\{X_i'\}_{i=1}^n$ be an independent copy of the sequence $\{X_i\}$, and let $K \in \{1, \ldots, n\}$ be uniformly distributed, independent of the X_i and X_i'. Let $W' = W - X_K + X_K'$.

Then

$$\mathbb{P}\left[W' = W + 1 \middle| W\right] = \mathbb{P}\left[X_K = 0, X_K' = 1 \middle| W\right]$$

Then

$$\begin{aligned}\mathbb{P}\left[W' = W + 1 \middle| W\right] &= \mathbb{P}\left[X_K = 0, X'_K = 1 \middle| W\right] \\ &= \left(\frac{n - W}{n}\right)\left(\frac{1}{n}\right)\end{aligned}$$

Then

$$\begin{aligned}\mathbb{P}\left[W' = W + 1 \middle| W\right] &= \mathbb{P}\left[X_K = 0, X'_K = 1 \middle| W\right] \\ &= \left(\frac{n - W}{n}\right)\left(\frac{1}{n}\right) \\ &= \frac{1}{n} - \frac{1}{n}\left(\frac{W}{n}\right),\end{aligned}$$

Then

$$\begin{aligned}\mathbb{P}\left[W' = W + 1 \middle| W\right] &= \mathbb{P}\left[X_K = 0, X_K' = 1 \middle| W\right] \\ &= \left(\frac{n-W}{n}\right)\left(\frac{1}{n}\right) \\ &= \frac{1}{n} - \frac{1}{n}\underbrace{\left(\frac{W}{n}\right)}_{=:\, E_1},\end{aligned}$$

Then

$$\mathbb{P}\left[W' = W + 1 \middle| W\right] = \mathbb{P}\left[X_K = 0, X'_K = 1 \middle| W\right]$$
$$= \left(\frac{n - W}{n}\right)\left(\frac{1}{n}\right)$$
$$= \frac{1}{n} - \frac{1}{n}\underbrace{\left(\frac{W}{n}\right)}_{=: E_1},$$

and

$$\mathbb{P}\left[W' = W - 1 \middle| W\right] = \mathbb{P}\left[X_K = 1, X'_K = 0 \middle| W\right]$$

Then

$$\begin{aligned}\mathbb{P}\left[W' = W+1 \middle| W\right] &= \mathbb{P}\left[X_K = 0, X'_K = 1 \middle| W\right] \\ &= \left(\frac{n-W}{n}\right)\left(\frac{1}{n}\right) \\ &= \frac{1}{n} - \frac{1}{n}\underbrace{\left(\frac{W}{n}\right)}_{=:\,E_1},\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\left[W' = W-1 \middle| W\right] &= \mathbb{P}\left[X_K = 1, X'_K = 0 \middle| W\right] \\ &= \left(\frac{W}{n}\right)\left(1 - \frac{1}{n}\right)\end{aligned}$$

Then

$$\mathbb{P}\left[W' = W + 1 \middle| W\right] = \mathbb{P}\left[X_K = 0, X_K' = 1 \middle| W\right]$$
$$= \left(\frac{n - W}{n}\right)\left(\frac{1}{n}\right)$$
$$= \frac{1}{n} - \frac{1}{n}\underbrace{\left(\frac{W}{n}\right)}_{=:\, E_1},$$

and

$$\mathbb{P}\left[W' = W - 1 \middle| W\right] = \mathbb{P}\left[X_K = 1, X_K' = 0 \middle| W\right]$$
$$= \left(\frac{W}{n}\right)\left(1 - \frac{1}{n}\right)$$
$$= \frac{W}{n} - \frac{1}{n}\left(\frac{W}{n}\right).$$

Then

$$\mathbb{P}\left[W' = W + 1 \middle| W\right] = \mathbb{P}\left[X_K = 0, X_K' = 1 \middle| W\right]$$
$$= \left(\frac{n - W}{n}\right)\left(\frac{1}{n}\right)$$
$$= \frac{1}{n} - \frac{1}{n}\underbrace{\left(\frac{W}{n}\right)}_{=: E_1},$$

and

$$\mathbb{P}\left[W' = W - 1 \middle| W\right] = \mathbb{P}\left[X_K = 1, X_K' = 0 \middle| W\right]$$
$$= \left(\frac{W}{n}\right)\left(1 - \frac{1}{n}\right)$$
$$= \frac{W}{n} - \frac{1}{n}\underbrace{\left(\frac{W}{n}\right)}_{=: E_2}.$$

Then

$$\begin{aligned}
\mathbb{P}\left[W' = W+1 \middle| W\right] &= \mathbb{P}\left[X_K = 0, X'_K = 1 \middle| W\right] \\
&= \left(\frac{n-W}{n}\right)\left(\frac{1}{n}\right) \\
&= \frac{1}{n} - \underbrace{\frac{1}{n}\left(\frac{W}{n}\right)}_{=:\, E_1},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}\left[W' = W-1 \middle| W\right] &= \mathbb{P}\left[X_K = 1, X'_K = 0 \middle| W\right] \\
&= \left(\frac{W}{n}\right)\left(1-\frac{1}{n}\right) \\
&= \frac{W}{n} - \underbrace{\frac{1}{n}\left(\frac{W}{n}\right)}_{=:\, E_2}.
\end{aligned}$$

$$\Longrightarrow \quad c = \frac{1}{n}.$$

We had

$$\mathbb{E}T_o f(W) = \mathbb{E}\left[f(W)E_2 - f(W+1)E_1\right],$$

We had

$$\mathbb{E}T_o f(W) = \mathbb{E}\left[f(W)E_2 - f(W+1)E_1\right],$$

and thus for this W,

$$\mathbb{E}T_o f(W) = \mathbb{E}\left[\frac{W}{n}\Big(f(W) - f(W+1)\Big)\right].$$

We had

$$\mathbb{E} T_o f(W) = \mathbb{E}\left[f(W)E_2 - f(W+1)E_1\right],$$

and thus for this W,

$$\mathbb{E} T_o f(W) = \mathbb{E}\left[\frac{W}{n}\Big(f(W) - f(W+1)\Big)\right].$$

Now, if X is a Poisson random variable with mean 1 and $g : \mathbb{N} \cup \{0\} \to [0,1]$ is given, one can solve the Stein equation

$$T_o f(j) = g(j) - \mathbb{E} g(X),$$

We had

$$\mathbb{E} T_o f(W) = \mathbb{E}\left[f(W)E_2 - f(W+1)E_1\right],$$

and thus for this W,

$$\mathbb{E} T_o f(W) = \mathbb{E}\left[\frac{W}{n}\Big(f(W) - f(W+1)\Big)\right].$$

Now, if X is a Poisson random variable with mean 1 and $g : \mathbb{N} \cup \{0\} \to [0, 1]$ is given, one can solve the Stein equation

$$T_o f(j) = g(j) - \mathbb{E}g(X),$$

and the solution $f =: U_o g$ has the property

$$\sup_j \big|f(j+1) - f(j)\big| \leq 1.$$

Taking $f = U_o g$ above thus gives

$$\left|\mathbb{E}g(W) - \mathbb{E}g(X)\right| \leq \mathbb{E}\left(\frac{W}{n}\right) = \frac{1}{n}.$$

Taking $f = U_o g$ above thus gives

$$\left|\mathbb{E}g(W) - \mathbb{E}g(X)\right| \leq \mathbb{E}\left(\frac{W}{n}\right) = \frac{1}{n}.$$

That is,

$$d_{TV}(W, X) \leq \frac{1}{n}.$$

Exchangeable pairs for normal approximation

Exchangeable pairs for normal approximation

$$0 = \mathbb{E}\left[(W' - W)\big(f(W') + f(W)\big)\right]$$

Exchangeable pairs for normal approximation

$$\begin{aligned}
0 &= \mathbb{E}\left[(W' - W)\big(f(W') + f(W)\big)\right] \\
&= \mathbb{E}\left[(W' - W)\big(f(W') - f(W)\big) + 2(W' - W)f(W)\right]
\end{aligned}$$

Exchangeable pairs for normal approximation

$$\begin{aligned}
0 &= \mathbb{E}\left[(W' - W)\big(f(W') + f(W)\big)\right] \\
&= \mathbb{E}\left[(W' - W)\big(f(W') - f(W)\big) + 2(W' - W)f(W)\right] \\
&= \mathbb{E}\Big[(W' - W)^2 f'(W) + 2(W' - W)f(W) + R\Big]
\end{aligned}$$

Exchangeable pairs for normal approximation

$$\begin{aligned}
0 &= \mathbb{E}\left[(W'-W)\big(f(W')+f(W)\big)\right] \\
&= \mathbb{E}\left[(W'-W)\big(f(W')-f(W)\big)+2(W'-W)f(W)\right] \\
&= \mathbb{E}\left[(W'-W)^2 f'(W)+2(W'-W)f(W)+R\right] \\
&= \mathbb{E}\left[f'(W)\mathbb{E}\left[(W'-W)^2\middle|W\right]+2f(W)\mathbb{E}\left[W'-W\middle|W\right]+R\right].
\end{aligned}$$

Exchangeable pairs for normal approximation

$$\begin{aligned}
0 &= \mathbb{E}\left[(W' - W)\big(f(W') + f(W)\big)\right] \\
&= \mathbb{E}\left[(W' - W)\big(f(W') - f(W)\big) + 2(W' - W)f(W)\right] \\
&= \mathbb{E}\left[(W' - W)^2 f'(W) + 2(W' - W)f(W) + R\right] \\
&= \mathbb{E}\left[f'(W)\mathbb{E}\left[(W' - W)^2 \middle| W\right] + 2f(W)\mathbb{E}\left[W' - W \middle| W\right] + R\right].
\end{aligned}$$

Now, suppose that there is a $\lambda \in (0, 1)$ such that

Exchangeable pairs for normal approximation

$$
\begin{aligned}
0 &= \mathbb{E}\left[(W' - W)\big(f(W') + f(W)\big)\right] \\
&= \mathbb{E}\left[(W' - W)\big(f(W') - f(W)\big) + 2(W' - W)f(W)\right] \\
&= \mathbb{E}\left[(W' - W)^2 f'(W) + 2(W' - W)f(W) + R\right] \\
&= \mathbb{E}\left[f'(W)\mathbb{E}\left[(W' - W)^2 \middle| W\right] + 2f(W)\mathbb{E}\left[W' - W \middle| W\right] + R\right].
\end{aligned}
$$

Now, suppose that there is a $\lambda \in (0, 1)$ such that

- $\mathbb{E}\left[W' - W \middle| W\right] = -\lambda W$

Exchangeable pairs for normal approximation

$$
\begin{aligned}
0 &= \mathbb{E}\left[(W' - W)\big(f(W') + f(W)\big)\right] \\
&= \mathbb{E}\left[(W' - W)\big(f(W') - f(W)\big) + 2(W' - W)f(W)\right] \\
&= \mathbb{E}\left[(W' - W)^2 f'(W) + 2(W' - W)f(W) + R\right] \\
&= \mathbb{E}\left[f'(W)\mathbb{E}\left[(W' - W)^2 \middle| W\right] + 2f(W)\mathbb{E}\left[W' - W \middle| W\right] + R\right].
\end{aligned}
$$

Now, suppose that there is a $\lambda \in (0, 1)$ such that

- $\mathbb{E}\left[W' - W \middle| W\right] = -\lambda W$
- $\mathbb{E}\left[(W' - W)^2 \middle| W\right] = 2\lambda + E.$ (E is a random variable.)

Exchangeable pairs for normal approximation

$$
\begin{aligned}
0 &= \mathbb{E}\left[(W'-W)\big(f(W')+f(W)\big)\right] \\
&= \mathbb{E}\left[(W'-W)\big(f(W')-f(W)\big)+2(W'-W)f(W)\right] \\
&= \mathbb{E}\left[(W'-W)^2 f'(W)+2(W'-W)f(W)+R\right] \\
&= \mathbb{E}\left[f'(W)\mathbb{E}\left[(W'-W)^2\middle|W\right]+2f(W)\mathbb{E}\left[W'-W\middle|W\right]+R\right].
\end{aligned}
$$

Now, suppose that there is a $\lambda \in (0,1)$ such that

- $\mathbb{E}\left[W'-W\middle|W\right] = -\lambda W$
- $\mathbb{E}\left[(W'-W)^2\middle|W\right] = 2\lambda + E.$ (E is a random variable.)

Then

$$
0 = 2\lambda\mathbb{E}\left[f'(W) - Wf(W) + \frac{f'(W)E+R}{2\lambda}\right].
$$

Exchangeable pairs for normal approximation

$$
\begin{aligned}
0 &= \mathbb{E}\left[(W'-W)\big(f(W')+f(W)\big)\right] \\
&= \mathbb{E}\left[(W'-W)\big(f(W')-f(W)\big)+2(W'-W)f(W)\right] \\
&= \mathbb{E}\left[(W'-W)^2 f'(W)+2(W'-W)f(W)+R\right] \\
&= \mathbb{E}\left[f'(W)\mathbb{E}\left[(W'-W)^2\middle|W\right]+2f(W)\mathbb{E}\left[W'-W\middle|W\right]+R\right].
\end{aligned}
$$

Now, suppose that there is a $\lambda \in (0,1)$ such that

- $\mathbb{E}\left[W'-W\middle|W\right] = -\lambda W$
- $\mathbb{E}\left[(W'-W)^2\middle|W\right] = 2\lambda + E.$ (E is a random variable.)

Then

$$
0 = 2\lambda\mathbb{E}\Bigg[\underbrace{f'(W)-Wf(W)}_{T_o f(W)} + \frac{f'(W)E+R}{2\lambda}\Bigg].
$$

Exchangeable pairs for normal approximation

$$\begin{aligned} 0 &= \mathbb{E}\left[(W' - W)\big(f(W') + f(W)\big)\right] \\ &= \mathbb{E}\left[(W' - W)\big(f(W') - f(W)\big) + 2(W' - W)f(W)\right] \\ &= \mathbb{E}\left[(W' - W)^2 f'(W) + 2(W' - W)f(W) + R\right] \\ &= \mathbb{E}\left[f'(W)\mathbb{E}\left[(W' - W)^2 \middle| W\right] + 2f(W)\mathbb{E}\left[W' - W \middle| W\right] + R\right]. \end{aligned}$$

Now, suppose that there is a $\lambda \in (0, 1)$ such that

- $\mathbb{E}\left[W' - W \middle| W\right] = -\lambda W$
- $\mathbb{E}\left[(W' - W)^2 \middle| W\right] = 2\lambda + E.$ (E is a random variable.)

Then

$$0 = 2\lambda\mathbb{E}\Big[\underbrace{f'(W) - Wf(W)}_{T_o f(W)} + \frac{f'(W)E + R}{2\lambda}\Big].$$

That is, $\mathbb{E}T_o f(W) = -\frac{1}{2\lambda}\mathbb{E}\left[f'(W)E + R\right]$.

Stein’s abstract normal approximation theorem

Stein's abstract normal approximation theorem

Theorem (Stein)

Let (W, W') be an exchangeable pair of random variables with $\mathbb{E}W^2 = 1$ and $\mathbb{E}\left[W' - W \middle| W\right] = -\lambda W$ for some $\lambda \in (0,1)$. Let $\Delta = W' - W$, and let $h : \mathbb{R} \to \mathbb{R}$ be bounded with piecewise continuous derivative h'. Then for Z a standard normal random variable,

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \leq \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

A few quick words about the proof

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \le \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

A few quick words about the proof

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \leq \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

- We had $\mathbb{E}T_o f(W) = -\frac{1}{2\lambda}\mathbb{E}\big[f'(W)E + R\big]$. Taking $f = U_o h$ (the solution to the Stein equation) gives

$$\mathbb{E}h(Z) - \mathbb{E}h(W) = \frac{1}{2\lambda}\mathbb{E}\big[f'(W)E + R\big].$$

A few quick words about the proof

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \le \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

- We had $\mathbb{E}T_o f(W) = -\frac{1}{2\lambda}\mathbb{E}\left[f'(W)E + R\right]$. Taking $f = U_o h$ (the solution to the Stein equation) gives

$$\mathbb{E}h(Z) - \mathbb{E}h(W) = \frac{1}{2\lambda}\mathbb{E}\left[f'(W)E + R\right].$$

- $E = \mathbb{E}\left[\Delta^2 \middle| W\right] - \mathbb{E}\Delta^2$, and thus

$$\frac{1}{2\lambda}\mathbb{E}\left|f'(W)E\right| \le \frac{\|f'\|_\infty}{2\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)}.$$

A few quick words about the proof

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \leq \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\operatorname{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

- We had $\mathbb{E}T_o f(W) = -\frac{1}{2\lambda}\mathbb{E}\big[f'(W)E + R\big]$. Taking $f = U_o h$ (the solution to the Stein equation) gives

$$\mathbb{E}h(Z) - \mathbb{E}h(W) = \frac{1}{2\lambda}\mathbb{E}\big[f'(W)E + R\big].$$

- $E = \mathbb{E}\left[\Delta^2 \middle| W\right] - \mathbb{E}\Delta^2$, and thus

$$\frac{1}{2\lambda}\mathbb{E}|f'(W)E| \leq \frac{\|f'\|_\infty}{2\lambda}\sqrt{\operatorname{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)}.$$

- $R = \Delta\big[f(W') - f(W)\big] - \Delta^2 f'(W)$, thus

$$\frac{1}{2\lambda}\mathbb{E}|R| \leq \frac{\|f''\|_\infty}{4\lambda}\mathbb{E}|\Delta|^3.$$

A few quick words about the proof (cont.)

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \leq \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

A few quick words about the proof (cont.)

$$\left|\mathbb{E}h(W)-\mathbb{E}h(Z)\right| \leq \frac{\|h-\mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2\middle|W\right]\right)}+\frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

Lemma (Stein)

If $f = U_o h$, then

A few quick words about the proof (cont.)

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \leq \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

Lemma (Stein)

If $f = U_o h$, then

- $\|f\|_\infty \leq \sqrt{\frac{\pi}{2}}\|h - \mathbb{E}h(Z)\|_\infty$

A few quick words about the proof (cont.)

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \leq \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

Lemma (Stein)

If $f = U_o h$, then

- $\|f\|_\infty \leq \sqrt{\frac{\pi}{2}}\|h - \mathbb{E}h(Z)\|_\infty$
- $\|f'\|_\infty \leq 2\|h - \mathbb{E}h(Z)\|_\infty$

A few quick words about the proof (cont.)

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \leq \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

Lemma (Stein)

If $f = U_o h$, then

- $\|f\|_\infty \leq \sqrt{\frac{\pi}{2}}\|h - \mathbb{E}h(Z)\|_\infty$
- $\|f'\|_\infty \leq 2\|h - \mathbb{E}h(Z)\|_\infty$
- $\|f''\|_\infty \leq 2\|h'\|_\infty$.

A few quick words about the proof (cont.)

$$\left|\mathbb{E}h(W)-\mathbb{E}h(Z)\right| \le \frac{\|h-\mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2\middle|W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

Lemma (Stein)

If $f = U_o h$, then

- $\|f\|_\infty \le \sqrt{\frac{\pi}{2}}\|h-\mathbb{E}h(Z)\|_\infty$
- $\|f'\|_\infty \le 2\|h-\mathbb{E}h(Z)\|_\infty$
- $\|f''\|_\infty \le 2\|h'\|_\infty.$

Thus

$$\begin{aligned}\left|\mathbb{E}h(W)-\mathbb{E}h(Z)\right| \\ &\le \frac{\|f'\|_\infty}{2\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2\middle|W\right]\right)} + \frac{\|f''\|_\infty}{4\lambda}\mathbb{E}|\Delta|^3\end{aligned}$$

A few quick words about the proof (cont.)

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \le \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.$$

Lemma (Stein)

If $f = U_o h$, then

- $\|f\|_\infty \le \sqrt{\frac{\pi}{2}}\|h - \mathbb{E}h(Z)\|_\infty$
- $\|f'\|_\infty \le 2\|h - \mathbb{E}h(Z)\|_\infty$
- $\|f''\|_\infty \le 2\|h'\|_\infty$.

Thus

$$\begin{aligned}\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \\ &\le \frac{\|f'\|_\infty}{2\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|f''\|_\infty}{4\lambda}\mathbb{E}|\Delta|^3 \\ &\le \frac{\|h - \mathbb{E}h(Z)\|_\infty}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|h'\|}{2\lambda}\mathbb{E}|\Delta|^3.\end{aligned}$$

Stein's method for multivariate normal approximation
(joint work with Sourav Chatterjee)

Stein's method for multivariate normal approximation

(joint work with Sourav Chatterjee)

Some notation:

Stein's method for multivariate normal approximation
(joint work with Sourav Chatterjee)

Some notation:

- Let $Z = (Z_1, \ldots, Z_k)$ be a standard normal random vector (i.e., the Z_i are i.i.d. standard Gaussians).

Stein's method for multivariate normal approximation
(joint work with Sourav Chatterjee)

Some notation:

- Let $Z = (Z_1, \ldots, Z_k)$ be a standard normal random vector (i.e., the Z_i are i.i.d. standard Gaussians).
- For a smooth function $f : \mathbb{R}^k \to \mathbb{R}$, define

Stein's method for multivariate normal approximation
(joint work with Sourav Chatterjee)

Some notation:

- Let $Z = (Z_1, \ldots, Z_k)$ be a standard normal random vector (i.e., the Z_i are i.i.d. standard Gaussians).
- For a smooth function $f : \mathbb{R}^k \to \mathbb{R}$, define
  - $M_1(f) := \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} = \sup_{x \in \mathbb{R}^k} \left| \nabla f(x) \right|$

Stein's method for multivariate normal approximation
(joint work with Sourav Chatterjee)

Some notation:

- Let $Z = (Z_1, \ldots, Z_k)$ be a standard normal random vector (i.e., the Z_i are i.i.d. standard Gaussians).
- For a smooth function $f : \mathbb{R}^k \to \mathbb{R}$, define
  - $M_1(f) := \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} = \sup_{x \in \mathbb{R}^k} \left|\nabla f(x)\right|$
  - $M_2(f) := \sup_{x \neq y} \frac{|\nabla f(x) - \nabla f(y)|}{|x - y|} = \sup_{x \in \mathbb{R}^k} \|\text{Hess}\, f(x)\|_{op}$

Stein's method for multivariate normal approximation
(joint work with Sourav Chatterjee)

Some notation:

- Let $Z = (Z_1, \ldots, Z_k)$ be a standard normal random vector (i.e., the Z_i are i.i.d. standard Gaussians).
- For a smooth function $f : \mathbb{R}^k \to \mathbb{R}$, define
  - $M_1(f) := \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} = \sup_{x \in \mathbb{R}^k} \left|\nabla f(x)\right|$
  - $M_2(f) := \sup_{x \neq y} \frac{|\nabla f(x) - \nabla f(y)|}{|x - y|} = \sup_{x \in \mathbb{R}^k} \|\mathrm{Hess}\, f(x)\|_{op}$
  - $M_3(f) := \sup_{x \neq y} \frac{\|\mathrm{Hess}\, f(x) - \mathrm{Hess}\, f(y)\|_{op}}{|x - y|}$

Multivariate characterizing operator

Multivariate characterizing operator

A characterizing operator for Z is:

$$T_o f(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle .$$

Multivariate characterizing operator

A characterizing operator for Z is:

$$T_o f(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle .$$

(Compare with $T_o f(x) = f'(x) - x f(x)$ in the univeriate case.)

Multivariate characterizing operator

A characterizing operator for Z is:

$$T_o f(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle .$$

(Compare with $T_o f(x) = f'(x) - xf(x)$ in the univeriate case.)

One can write down an explicit integral formula for the solution $f = U_o h$ to the Stein equation

$$\Delta f(x) - \langle x, \nabla f(x) \rangle = h(x) - \mathbb{E}h(Z),$$

as in the univariate context.

Multivariate characterizing operator

A characterizing operator for Z is:

$$T_o f(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle .$$

(Compare with $T_o f(x) = f'(x) - xf(x)$ in the univeriate case.)

One can write down an explicit integral formula for the solution $f = U_o h$ to the Stein equation

$$\Delta f(x) - \langle x, \nabla f(x) \rangle = h(x) - \mathbb{E}h(Z),$$

as in the univariate context.

Lemma (C-M)

Let $f = U_o h$ be as above. Then

Multivariate characterizing operator

A characterizing operator for Z is:

$$T_o f(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle .$$

(Compare with $T_o f(x) = f'(x) - xf(x)$ in the univeriate case.)

One can write down an explicit integral formula for the solution $f = U_o h$ to the Stein equation

$$\Delta f(x) - \langle x, \nabla f(x) \rangle = h(x) - \mathbb{E}h(Z),$$

as in the univariate context.

Lemma (C-M)

Let $f = U_o h$ be as above. Then

- $\sup_{x \in \mathbb{R}^k} \| Hess\, f(x) \|_{H.S.} \leq M_1(h)$.

Multivariate characterizing operator

A characterizing operator for Z is:

$$T_o f(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle .$$

(Compare with $T_o f(x) = f'(x) - x f(x)$ in the univeriate case.)

One can write down an explicit integral formula for the solution $f = U_o h$ to the Stein equation

$$\Delta f(x) - \langle x, \nabla f(x) \rangle = h(x) - \mathbb{E} h(Z),$$

as in the univariate context.

Lemma (C-M)

Let $f = U_o h$ be as above. Then

- $\sup_{x \in \mathbb{R}^k} \| Hess\, f(x) \|_{H.S.} \leq M_1(h)$.
- $M_3(f) \leq \frac{\sqrt{2\pi}}{4} M_2(h)$.

An abstract normal approximation theorem

An abstract normal approximation theorem

Theorem (C-M)

Let X and X' be two random vectors in $\mathbb{R}^k$ with $X \stackrel{d}{=} X'$.

An abstract normal approximation theorem

Theorem (C-M)

Let X and X' be two random vectors in $\mathbb{R}^k$ with $X \stackrel{d}{=} X'$. Suppose there is a constant λ such that

$$\frac{1}{\lambda}\mathbb{E}\left[X' - X \middle| X\right] = -X.$$

An abstract normal approximation theorem

Theorem (C-M)

Let X and X' be two random vectors in $\mathbb{R}^k$ with $X \stackrel{d}{=} X'$. Suppose there is a constant λ such that

$$\frac{1}{\lambda}\mathbb{E}\left[X' - X \middle| X\right] = -X.$$

Define the random matrix E by

$$\frac{1}{2\lambda}\mathbb{E}\left[(X' - X)(X' - X)^T \middle| X\right] = \sigma^2 I_k + \mathbb{E}\left[E \middle| X\right].$$

An abstract normal approximation theorem

Theorem (C-M)

Let X and X' be two random vectors in $\mathbb{R}^k$ with $X \stackrel{d}{=} X'$. Suppose there is a constant λ such that

$$\frac{1}{\lambda}\mathbb{E}\left[X' - X \middle| X\right] = -X.$$

Define the random matrix E by

$$\frac{1}{2\lambda}\mathbb{E}\left[(X' - X)(X' - X)^T \middle| X\right] = \sigma^2 I_k + \mathbb{E}\left[E \middle| X\right].$$

Then if $h \in C^2(\mathbb{R}^k)$ with $M_1(h) < \infty$ and $M_2(h) < \infty$,

An abstract normal approximation theorem

Theorem (C-M)

Let X and X' be two random vectors in $\mathbb{R}^k$ with $X \stackrel{d}{=} X'$. Suppose there is a constant λ such that

$$\frac{1}{\lambda}\mathbb{E}\left[X' - X \middle| X\right] = -X.$$

Define the random matrix E by

$$\frac{1}{2\lambda}\mathbb{E}\left[(X' - X)(X' - X)^T \middle| X\right] = \sigma^2 I_k + \mathbb{E}\left[E \middle| X\right].$$

Then if $h \in C^2(\mathbb{R}^k)$ with $M_1(h) < \infty$ and $M_2(h) < \infty$,

$$\left|\mathbb{E}h(X) - \mathbb{E}h(\sigma Z)\right| \leq \frac{1}{\sigma}M_1(h)\mathbb{E}\|E\|_{H.S.} + \frac{\sqrt{2\pi}M_2(h)}{24\sigma\lambda}\mathbb{E}|X' - X|^3.$$

Example – A multivariate CLT

Example – A multivariate CLT

Let $\{X_i\}_{i=1}^n$ be i.i.d random vectors in $\mathbb{R}^k$, and assume that $\mathbb{E}X = 0$ and $\mathbb{E}XX^T = I_k$.

Example – A multivariate CLT

Let $\{X_i\}_{i=1}^n$ be i.i.d random vectors in $\mathbb{R}^k$, and assume that $\mathbb{E}X = 0$ and $\mathbb{E}XX^T = I_k$. Define

$$S = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

Example – A multivariate CLT

Let $\{X_i\}_{i=1}^n$ be i.i.d random vectors in $\mathbb{R}^k$, and assume that $\mathbb{E}X = 0$ and $\mathbb{E}XX^T = I_k$. Define

$$S = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

Define S' as follows. Choose $K \in \{1, \ldots, n\}$ uniformly at random, independent of $\{X_i\}_{i=1}^n$.

Example – A multivariate CLT

Let $\{X_i\}_{i=1}^n$ be i.i.d random vectors in $\mathbb{R}^k$, and assume that $\mathbb{E}X = 0$ and $\mathbb{E}XX^T = I_k$. Define

$$S = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

Define S' as follows. Choose $K \in \{1, \ldots, n\}$ uniformly at random, independent of $\{X_i\}_{i=1}^n$. Let $\{X_i'\}$ be an independent copy of the collection $\{X_i\}$,

Example – A multivariate CLT

Let $\{X_i\}_{i=1}^n$ be i.i.d random vectors in $\mathbb{R}^k$, and assume that $\mathbb{E}X = 0$ and $\mathbb{E}XX^T = I_k$. Define

$$S = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

Define S' as follows. Choose $K \in \{1, \ldots, n\}$ uniformly at random, independent of $\{X_i\}_{i=1}^n$. Let $\{X_i'\}$ be an independent copy of the collection $\{X_i\}$, and let

$$S' = S - \frac{1}{\sqrt{n}} X_K + \frac{1}{\sqrt{n}} X_K'.$$

Now,

$$\mathbb{E}\left[S' - S \middle| S\right] = \frac{1}{\sqrt{n}}\mathbb{E}\left[X'_K - X_K \middle| S\right]$$

Now,

$$\begin{aligned}\mathbb{E}\left[S' - S \middle| S\right] &= \frac{1}{\sqrt{n}}\mathbb{E}\left[X'_K - X_K \middle| S\right] \\ &= \frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left[X'_i - X_i \middle| S\right]\end{aligned}$$

Now,

$$\begin{aligned}\mathbb{E}\left[S' - S \middle| S\right] &= \frac{1}{\sqrt{n}}\mathbb{E}\left[X'_K - X_K \middle| S\right] \\ &= \frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left[X'_i - X_i \middle| S\right] \\ &= -\frac{1}{n^{3/2}}\mathbb{E}\left[\sum_{i=1}^{n} X_i \middle| S\right]\end{aligned}$$

Now,

$$
\begin{aligned}
\mathbb{E}\left[S' - S \middle| S\right] &= \frac{1}{\sqrt{n}}\mathbb{E}\left[X'_K - X_K \middle| S\right] \\
&= \frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left[X'_i - X_i \middle| S\right] \\
&= -\frac{1}{n^{3/2}}\mathbb{E}\left[\sum_{i=1}^{n} X_i \middle| S\right] \\
&= -\frac{1}{n}S.
\end{aligned}
$$

Now,

$$\begin{aligned}
\mathbb{E}\left[S' - S \middle| S\right] &= \frac{1}{\sqrt{n}}\mathbb{E}\left[X'_K - X_K \middle| S\right] \\
&= \frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left[X'_i - X_i \middle| S\right] \\
&= -\frac{1}{n^{3/2}}\mathbb{E}\left[\sum_{i=1}^{n} X_i \middle| S\right] \\
&= -\frac{1}{n}S.
\end{aligned}$$

$$\Longrightarrow \quad \lambda = \frac{1}{n}$$

Write $X_i = (X_i^1, \ldots, X_i^k)$.

Write $X_i = (X_i^1, \ldots, X_i^k)$.

$$\mathbb{E}\left[(S' - S)_j (S' - S)_\ell \middle| S\right] = \frac{1}{n} \mathbb{E}\left[\left((X_K')^j - X_K^j\right)\left((X_K')^\ell - X_K^\ell\right) \middle| S\right]$$

Write $X_i = (X_i^1, \ldots, X_i^k)$.

$$\begin{aligned}\mathbb{E}\left[(S'-S)_j(S'-S)_\ell \middle| S\right] &= \frac{1}{n}\mathbb{E}\left[\left((X_K')^j - X_K^j\right)\left((X_K')^\ell - X_K^\ell\right)\middle| S\right] \\ &= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[\left((X_i')^j - X_i^j\right)\left((X_i')^\ell - X_i^\ell\right)\middle| S\right]\end{aligned}$$

Write $X_i = (X_i^1, \ldots, X_i^k)$.

$$\begin{aligned}
\mathbb{E}\left[(S' - S)_j (S' - S)_\ell \middle| S\right] &= \frac{1}{n}\mathbb{E}\left[\left((X_K')^j - X_K^j\right)\left((X_K')^\ell - X_K^\ell\right)\middle| S\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left[\left((X_i')^j - X_i^j\right)\left((X_i')^\ell - X_i^\ell\right)\middle| S\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left(\delta_{j\ell} + \mathbb{E}\left[X_i^j X_i^\ell \middle| S\right]\right)
\end{aligned}$$

Write $X_i = (X_i^1, \ldots, X_i^k)$.

$$\begin{aligned}
\mathbb{E}\left[(S' - S)_j (S' - S)_\ell \middle| S\right] &= \frac{1}{n} \mathbb{E}\left[\left((X_K')^j - X_K^j\right)\left((X_K')^\ell - X_K^\ell\right) \middle| S\right] \\
&= \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}\left[\left((X_i')^j - X_i^j\right)\left((X_i')^\ell - X_i^\ell\right) \middle| S\right] \\
&= \frac{1}{n^2} \sum_{i=1}^n \left(\delta_{j\ell} + \mathbb{E}\left[X_i^j X_i^\ell \middle| S\right]\right) \\
&= \frac{2}{n} \delta_{j\ell} + \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}\left[X_i^j X_i^\ell - \delta_{j\ell} \middle| S\right]
\end{aligned}$$

Write $X_i = (X_i^1, \dots, X_i^k)$.

$$
\begin{aligned}
\mathbb{E}\left[(S'-S)_j(S'-S)_\ell \middle| S\right] &= \frac{1}{n}\mathbb{E}\left[\left((X_K')^j - X_K^j\right)\left((X_K')^\ell - X_K^\ell\right)\middle| S\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[\left((X_i')^j - X_i^j\right)\left((X_i')^\ell - X_i^\ell\right)\middle| S\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n \left(\delta_{j\ell} + \mathbb{E}\left[X_i^j X_i^\ell \middle| S\right]\right) \\
&= \frac{2}{n}\delta_{j\ell} + \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[X_i^j X_i^\ell - \delta_{j\ell}\middle| S\right]
\end{aligned}
$$

That is,

$$
\frac{n}{2}\mathbb{E}\left[(S'-S)(S'-S)^T \middle| S\right] = I_k + \frac{1}{2n}\mathbb{E}\left[\sum_{i=1}^n \left(X_i X_i^T - I_k\right)\middle| S\right].
$$

Write $X_i = (X_i^1, \ldots, X_i^k)$.

$$\begin{aligned}
\mathbb{E}\left[(S'-S)_j(S'-S)_\ell \middle| S\right] &= \frac{1}{n}\mathbb{E}\left[\left((X_K')^j - X_K^j\right)\left((X_K')^\ell - X_K^\ell\right)\middle| S\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[\left((X_i')^j - X_i^j\right)\left((X_i')^\ell - X_i^\ell\right)\middle| S\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n \left(\delta_{j\ell} + \mathbb{E}\left[X_i^j X_i^\ell \middle| S\right]\right) \\
&= \frac{2}{n}\delta_{j\ell} + \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[X_i^j X_i^\ell - \delta_{j\ell}\middle| S\right]
\end{aligned}$$

That is,

$$\frac{n}{2}\mathbb{E}\left[(S'-S)(S'-S)^T \middle| S\right] = I_k + \underbrace{\frac{1}{2n}\mathbb{E}\left[\sum_{i=1}^n \left(X_i X_i^T - I_k\right)\middle| S\right]}_{=:\, E}.$$

Bounding $\mathbb{E}\|E\|_{H.S.}$ using standard techniques, and writing $\mathbb{E}\left|S' - S\right|^3$ in terms of the X_i fairly easily yields:

$$\left|\mathbb{E}h(S) - \mathbb{E}h(Z)\right| \leq \frac{M_1(h)}{2\sqrt{n}}\sqrt{\mathbb{E}|X_1|^4 - k} + \frac{\sqrt{2\pi}}{3\sqrt{n}}M_2(h)\mathbb{E}|X_1|^3.$$