Stein's Method: Distributional Approximation and Concentration of Measure

Larry Goldstein
University of Southern California

36^{th} Midwest Probability Colloquium, 2014


Stein Mart
GRAND OPENING

Stein's method for Distributional Approximation

Goal is to approximate a given (perhaps complicated) $\mathcal{L}(W)$ by (a simpler) $\mathcal{L}(Z)$

Works in a variety of dependent situations.

Provides non-asymptotic bounds in many metrics.

Distributional Approximation

Stein's Lemma: A random variable Z has law $\mathcal{N}(0,1)$ if and only if

$$E[f'(Z)] = E[Zf(Z)] \quad \text{for all } f \in \mathcal{F},$$

the collection of bounded, absolutely continuous functions. Say.

Stein 1972, 1986.

Distributional Approximation

Stein's Lemma: A random variable Z has law $\mathcal{N}(0,1)$ if and only if

$$E[f'(Z)] = E[Zf(Z)] \quad \text{for all } f \in \mathcal{F},$$

the collection of bounded, absolutely continuous functions. Say.

Stein 1972, 1986.

Distributional Approximation

Stein's Lemma: A random variable Z has law $\mathcal{N}(0,1)$ if and only if

$$E[f'(Z)] = E[Zf(Z)] \quad \text{for all } f \in \mathcal{F},$$

the collection of bounded, absolutely continuous functions. Say.

Stein 1972, 1986.

The Stein Equation

If W is such that

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

is 'small' for many functions f, then W must be close to $\mathcal{N}(0,1)$.

Given a test function $h : \mathbb{R} \to \mathbb{R}$, evaluate $Eh(W) - Eh(Z)$ for $Z \sim \mathcal{N}(0,1)$ by solving for f in

$$f'(w) - wf(w) = h(w) - Eh(Z)$$

and then evaluating the expectation of the left hand side at W.

At first glance it looks like proceeding this way makes the problem more difficult than before.

The Stein Equation

If W is such that

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

is 'small' for many functions f, then W must be close to $\mathcal{N}(0, 1)$.

Given a test function $h : \mathbb{R} \to \mathbb{R}$, evaluate $Eh(W) - Eh(Z)$ for $Z \sim \mathcal{N}(0, 1)$ by solving for f in

$$f'(w) - wf(w) = h(w) - Eh(Z)$$

and then evaluating the expectation of the left hand side at W.

At first glance it looks like proceeding this way makes the problem more difficult than before.

Bounding Solutions of the Stein Equation

Given a test function $h : \mathbb{R} \to \mathbb{R}$, and $Z \sim \mathcal{N}(0,1)$ the magnitude of the solution f to

$$f'(w) - wf(w) = h(w) - Eh(Z)$$

and its derivatives can be given in terms of h.

For instance, if h is absolutely continuous then

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

Bounding Solutions of the Stein Equation

Given a test function $h : \mathbb{R} \to \mathbb{R}$, and $Z \sim \mathcal{N}(0,1)$ the magnitude of the solution f to

$$f'(w) - wf(w) = h(w) - Eh(Z)$$

and its derivatives can be given in terms of h.

For instance, if h is absolutely continuous then

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

Manipulating the Stein Equation

Evaluate

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

using coupling, and non-coupling methods.

With Coupling: Zero Bias, Exchangeable Pair, Size Bias, Stein Couplings.

Without Coupling: 2^{nd} order Poincaré inequalities, Malliavin Calculus.

Manipulating the Stein Equation

Evaluate

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

using coupling, and non-coupling methods.

With Coupling: Zero Bias, Exchangeable Pair, Size Bias, Stein Couplings.

Without Coupling: 2^{nd} order Poincaré inequalities, Malliavin Calculus.

Manipulating the Stein Equation

Evaluate

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

using coupling, and non-coupling methods.

With Coupling: Zero Bias, Exchangeable Pair, Size Bias, Stein Couplings.

Without Coupling: 2^{nd} order Poincaré inequalities, Malliavin Calculus.

Zero Bias Distribution

Stein Identity:

$$E[Zf(Z)] = E[f'(Z)] \quad \text{for all } f \in \mathcal{F}$$

if and only if $Z \sim \mathcal{N}(0,1)$.

For every mean zero variance 1 random variable W, there exists $\mathcal{L}(W^*)$ such that

$$E[Wf(W)] = E[f'(W^*)] \quad \text{for all } f \in \mathcal{F}.$$

The variable W^* is said to have the W-zero biased distribution (G. and Reinert, 1997).

Restatement of Stein's Characterization: $W \sim \mathcal{N}(0,1)$ if and only if $W^* =_d W$.

Zero Bias Distribution

Stein Identity:

$$E[Zf(Z)] = E[f'(Z)] \quad \text{for all } f \in \mathcal{F}$$

if and only if $Z \sim \mathcal{N}(0, 1)$.

For every mean zero variance 1 random variable W, there exists $\mathcal{L}(W^*)$ such that

$$E[Wf(W)] = E[f'(W^*)] \quad \text{for all } f \in \mathcal{F}.$$

The variable W^* is said to have the W-zero biased distribution (G. and Reinert, 1997).

Restatement of Stein's Characterization: $W \sim \mathcal{N}(0, 1)$ if and only if $W^* =_d W$.

Zero Bias Distribution

Stein Identity:

$$E[Zf(Z)] = E[f'(Z)] \quad \text{for all } f \in \mathcal{F}$$

if and only if $Z \sim \mathcal{N}(0,1)$.

For every mean zero variance 1 random variable W, there exists $\mathcal{L}(W^*)$ such that

$$E[Wf(W)] = E[f'(W^*)] \quad \text{for all } f \in \mathcal{F}.$$

The variable W^* is said to have the W-zero biased distribution (G. and Reinert, 1997).

Restatement of Stein's Characterization: $W \sim \mathcal{N}(0,1)$ if and only if $W^* =_d W$.

Coupling in the Stein equation

Let (W, W^*) be given on the same space, and let f be the solution to the Stein equation

$$h(w) - Eh(Z) = f'(w) - wf(w).$$

Then if $||h'|| \leq 1$ we have

$$|Eh(W) - Eh(Z)| = |E[f'(W) - Wf(W)]|$$
$$= |E[f'(W) - f'(W^*)]| \leq ||f''|| E|W - W^*| \leq 2E|W - W^*|.$$

Upper bound in terms of distance between W and its image under a transformation that leaves Z fixed. Supremum over all such functions h is so bounded, and with d_1 the Wasserstein distance, we obtain, without any assumptions on 'the dependence structure' of W,

$$d_1(W, Z) \leq 2d_1(W^*, W).$$

Coupling in the Stein equation

Let (W, W^*) be given on the same space, and let f be the solution to the Stein equation

$$h(w) - Eh(Z) = f'(w) - wf(w).$$

Then if $||h'|| \leq 1$ we have

$$|Eh(W) - Eh(Z)| = |E[f'(W) - Wf(W)]|$$
$$= |E[f'(W) - f'(W^*)]| \leq ||f''||E|W - W^*| \leq 2E|W - W^*|.$$

Upper bound in terms of distance between W and its image under a transformation that leaves Z fixed. Supremum over all such functions h is so bounded, and with d_1 the Wasserstein distance, we obtain, without any assumptions on 'the dependence structure' of W,

$$d_1(W, Z) \leq 2d_1(W^*, W).$$

Coupling in the Stein equation

Let (W, W^*) be given on the same space, and let f be the solution to the Stein equation

$$h(w) - Eh(Z) = f'(w) - wf(w).$$

Then if $||h'|| \leq 1$ we have

$$|Eh(W) - Eh(Z)| = |E[f'(W) - Wf(W)]|$$
$$= |E[f'(W) - f'(W^*)]| \leq ||f''||E|W - W^*| \leq 2E|W - W^*|.$$

Upper bound in terms of distance between W and its image under a transformation that leaves Z fixed. Supremum over all such functions h is so bounded, and with d_1 the Wasserstein distance, we obtain, without any assumptions on 'the dependence structure' of W,

$$d_1(W, Z) \leq 2d_1(W^*, W).$$

Zero Bias Coupling

If $X_1, \ldots, X_n$ are independent, mean zero random variables with variances $\sigma_1^2, \ldots, \sigma_n^2$ and let W be their sum, then for I an independent random index with distribution $P(I = i) = \sigma_i^2$,

$$W^* = W - X_I + X_I^*.$$

Then

$$d_1(W, Z) \leq 2d_1(W^*, W) \leq \frac{1}{\sigma^3} \sum_{i=1}^{n} E|X_i|^3,$$

an L^1 Berry-Esseen theorem with correct rate, and, by judiciously coupling X_I and X_I^*, with best known constant. (G. 2010, Tyurin 2010)

Zero Bias Coupling

If $X_1, \ldots, X_n$ are independent, mean zero random variables with variances $\sigma_1^2, \ldots, \sigma_n^2$ and let W be their sum, then for I an independent random index with distribution $P(I = i) = \sigma_i^2$,

$$W^* = W - X_I + X_I^*.$$

Then

$$d_1(W, Z) \leq 2d_1(W^*, W) \leq \frac{1}{\sigma^3} \sum_{i=1}^{n} E|X_i|^3,$$

an L^1 Berry-Esseen theorem with correct rate, and, by judiciously coupling X_I and X_I^*, with best known constant. (G. 2010, Tyurin 2010)

Exchangeable Pair Coupling

For $\lambda \in (0,1)$ suppose (W, W') is an exchangeable pair of variance 1 variables that satisfies

$$E[W'|W] = (1-\lambda)W.$$

Then (without any independence conditions) the Kolmogorov (L^∞) distance between W and $\mathcal{N}(0,1)$ is bounded by

$$\frac{2}{\lambda}\sqrt{\mathrm{Var}\left(E((W'-W)^2|W)\right)} + \frac{1}{(2\pi)^{1/4}}\sqrt{\frac{1}{\lambda}E|W'-W|^3}$$

We call (W, W') a Stein pair. Stein, (1986).

Hoeffding's Statistic: Exchangeable Pair

Let $A \in \mathbb{R}^{n\times n}$ satisfy $\sum_{i,j} a_{ij} = 0$, and let $\pi \in \mathcal{S}_n$ be a random permutation. Set

$$W_\pi = \sum_{i=1}^{n} a_{i\pi(i)}.$$

Simple random sampling, permutation tests.

Let τ_{ij} transpose i and j, (I, J) uniform over all unequal pairs, independent of π, and $\pi' = \pi\tau_{IJ}$. Then $(W_\pi, W_{\pi'})$ is a Stein pair. (Stein, Ho and Chen 1978, Chen 2013)

May consider other permutation distributions, e.g. π is chosen randomly from all fixed point free involutions. (G. and Rinott 2003)

We may write $W_\pi = \mathrm{tr}(AP)$ for P a permutation matrix. Can apply same technique for traces of random matrices on classical groups, group characters. Meckes 2008, Fulman 2006.

Hoeffding's Statistic: Exchangeable Pair

Let $A \in \mathbb{R}^{n \times n}$ satisfy $\sum_{i,j} a_{ij} = 0$, and let $\pi \in \mathcal{S}_n$ be a random permutation. Set

$$W_\pi = \sum_{i=1}^{n} a_{i\pi(i)}.$$

Simple random sampling, permutation tests.
Let τ_{ij} transpose i and j, (I, J) uniform over all unequal pairs, independent of π, and $\pi' = \pi\tau_{IJ}$. Then $(W_\pi, W_{\pi'})$ is a Stein pair. (Stein, Ho and Chen 1978, Chen 2013)

May consider other permutation distributions, e.g. π is chosen randomly from all fixed point free involutions. (G. and Rinott 2003)

We may write $W_\pi = \mathrm{tr}(AP)$ for P a permutation matrix. Can apply same technique for traces of random matrices on classical groups, group characters. Meckes 2008, Fulman 2006.

Hoeffding's Statistic: Exchangeable Pair

Let $A \in \mathbb{R}^{n\times n}$ satisfy $\sum_{i,j} a_{ij} = 0$, and let $\pi \in \mathcal{S}_n$ be a random permutation. Set

$$W_\pi = \sum_{i=1}^{n} a_{i\pi(i)}.$$

Simple random sampling, permutation tests.
Let τ_{ij} transpose i and j, (I, J) uniform over all unequal pairs, independent of π, and $\pi' = \pi\tau_{IJ}$. Then $(W_\pi, W_{\pi'})$ is a Stein pair. (Stein, Ho and Chen 1978, Chen 2013)

May consider other permutation distributions, e.g. π is chosen randomly from all fixed point free involutions. (G. and Rinott 2003)

We may write $W_\pi = \mathrm{tr}(AP)$ for P a permutation matrix. Can apply same technique for traces of random matrices on classical groups, group characters. Meckes 2008, Fulman 2006.

Hoeffding's Statistic: Exchangeable Pair

Let $A \in \mathbb{R}^{n \times n}$ satisfy $\sum_{i,j} a_{ij} = 0$, and let $\pi \in \mathcal{S}_n$ be a random permutation. Set

$$W_\pi = \sum_{i=1}^{n} a_{i\pi(i)}.$$

Simple random sampling, permutation tests.
Let τ_{ij} transpose i and j, (I, J) uniform over all unequal pairs, independent of π, and $\pi' = \pi\tau_{IJ}$. Then $(W_\pi, W_{\pi'})$ is a Stein pair. (Stein, Ho and Chen 1978, Chen 2013)

May consider other permutation distributions, e.g. π is chosen randomly from all fixed point free involutions. (G. and Rinott 2003)

We may write $W_\pi = \mathrm{tr}(AP)$ for P a permutation matrix. Can apply same technique for traces of random matrices on classical groups, group characters. Meckes 2008, Fulman 2006.

Size Bias Coupling

For $W \geq 0$ with $EW = \mu$ finite and positive, we say W^s has the W-size bias distribution when

$$E[Wf(W)] = \mu E[f(W^s)] \quad \text{for all } f \in \mathcal{F}.$$

If W is the sum of indicators $X_1, \ldots, X_n$, choose one proportional to its mean $P(I = i) = EX_i/EW$, independently of W. Set X_I to be equal to one, sample the remaining indicators from their joint conditional distribution on this event.

Size Bias Coupling

For $W \geq 0$ with $EW = \mu$ finite and positive, we say W^s has the W-size bias distribution when

$$E[Wf(W)] = \mu E[f(W^s)] \quad \text{for all } f \in \mathcal{F}.$$

If W is the sum of indicators $X_1, \ldots, X_n$, choose one proportional to its mean $P(I = i) = EX_i/EW$, independently of W. Set X_I to be equal to one, sample the remaining indicators from their joint conditional distribution on this event.

Size Bias Wasserstein Bound

Let Y be a nonnegative random variable with non zero mean μ and positive, finite variance σ^2. If Y^s on the same space as Y has the Y-size bias distribution, then for the standardized version W of Y,

$$d_1(W, Z) \leq \frac{\mu}{\sigma^2}\sqrt{\mathrm{Var}\left(E(Y^s - Y|Y)\right)} + \frac{\mu}{\sigma^3}E|Y^s - Y|^2.$$

Chen and Röllin (2010).

E.g. vertex degree in random graphs, occupancy models. With more work, can also achieve optimal rate L^∞ bounds.

Size Bias Wasserstein Bound

Let Y be a nonnegative random variable with non zero mean μ and positive, finite variance σ^2. If Y^s on the same space as Y has the Y-size bias distribution, then for the standardized version W of Y,

$$d_1(W, Z) \leq \frac{\mu}{\sigma^2}\sqrt{\mathrm{Var}\left(E(Y^s - Y|Y)\right)} + \frac{\mu}{\sigma^3}E|Y^s - Y|^2.$$

Chen and Röllin (2010).

E.g. vertex degree in random graphs, occupancy models. With more work, can also achieve optimal rate L^∞ bounds.

Second Order Poincaré Inequality

Stein identity for $\mathcal{N}(0,1)$,

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

Zero bias distribution for mean zero, variance one variable W,

$$E[Wf(W)] = E[f'(W^*)].$$

Might also hope to modify right hand side by finding $T(W)$ such that

$$E[Wf(W)] = E[T(W)f'(W)].$$

Cacoullos and Papathanasiou, 1992.

Second Order Poincaré Inequality

Stein identity for $\mathcal{N}(0,1)$,

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

Zero bias distribution for mean zero, variance one variable W,

$$E[Wf(W)] = E[f'(W^*)].$$

Might also hope to modify right hand side by finding $T(W)$ such that

$$E[Wf(W)] = E[T(W)f'(W)].$$

Cacoullos and Papathanasiou, 1992.

Second Order Poincaré Inequality

Stein identity for $\mathcal{N}(0,1)$,

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

Zero bias distribution for mean zero, variance one variable W,

$$E[Wf(W)] = E[f'(W^*)].$$

Might also hope to modify right hand side by finding $T(W)$ such that

$$E[Wf(W)] = E[T(W)f'(W)].$$

Cacoullos and Papathanasiou, 1992.

Stein Coefficient

For mean zero, variance 1 W, given $T(W)$ such that

$E[Wf(W)] = E[T(W)f'(W)]$, which hence satisfies $E[T] = 1$,

use in the Stein equation as

$$\begin{aligned} Eh(W) - Eh(Z) &= E[f'(W) - Wf(W)] \\ &= E[(1 - T(W))f'(W)] \le ||f'||E|1 - T(W)|. \end{aligned}$$

For instance, taking $0 \le h \le 1$, we obtain

$$d_{\mathrm{TV}}(W, Z) \le 2E|1 - T(W)| \le 2\sqrt{\mathrm{Var}(T)}.$$

Gaussian Inequalities

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, and $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I_n})$.

Poincaré Inequality:

$$\mathrm{Var}(f(\mathbf{g})) \leq E\left[||\nabla f(\mathbf{g})||^2\right].$$

Logarithmic Sobolev Inequality

$$\mathrm{Ent}(f(\mathbf{g})^2) \leq 2E\left[||\nabla f(\mathbf{g})||^2\right],$$

where

$$\mathrm{Ent}(f(\mathbf{g})) = E[f(\mathbf{g}) \log f(\mathbf{g})] - E[f(\mathbf{g})] \log[Ef(\mathbf{g})].$$

Gaussian Inequalities

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, and $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I_n})$.

Poincaré Inequality:

$$\mathrm{Var}(f(\mathbf{g})) \leq E\left[||\nabla f(\mathbf{g})||^2\right].$$

Logarithmic Sobolev Inequality

$$\mathrm{Ent}(f(\mathbf{g})^2) \leq 2E\left[||\nabla f(\mathbf{g})||^2\right],$$

where

$$\mathrm{Ent}(f(\mathbf{g})) = E[f(\mathbf{g})\log f(\mathbf{g})] - E[f(\mathbf{g})]\log[Ef(\mathbf{g})].$$

Gaussian Inequalities

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, and $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I_n})$.

Poincaré Inequality:

$$\mathrm{Var}(f(\mathbf{g})) \leq E\left[||\nabla f(\mathbf{g})||^2\right].$$

Logarithmic Sobolev Inequality

$$\mathrm{Ent}(f(\mathbf{g})^2) \leq 2E\left[||\nabla f(\mathbf{g})||^2\right],$$

where

$$\mathrm{Ent}(f(\mathbf{g})) = E[f(\mathbf{g})\log f(\mathbf{g})] - E[f(\mathbf{g})]\log[Ef(\mathbf{g})].$$

Gaussian Inequalities

Can prove the Poincaré Inequality by showing that

$$\mathrm{Var}(f(\mathbf{g})) = E[T] \quad \text{where} \quad T = \int_0^\infty e^{-t} \langle \nabla f(\mathbf{g}), \widehat{E} \nabla f(\widehat{\mathbf{g}}_t) \rangle dt$$

with

$$\widehat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1 - e^{-2t}}\widehat{\mathbf{g}},$$

and $\widehat{\mathbf{g}}$ an independent copy of $\mathbf{g}$, and $\widehat{E}$ expectation with respect to $\widehat{\mathbf{g}}$. Applying the Cauchy-Schwarz inequality one obtains $E[T] \leq E\|\nabla f(\mathbf{g})\|^2$.

Equality to variance hints that T may be a Stein coefficient.

Gaussian Inequalities

Can prove the Poincaré Inequality by showing that

$$\mathrm{Var}(f(\mathbf{g})) = E[T] \quad \text{where} \quad T = \int_0^\infty e^{-t} \langle \nabla f(\mathbf{g}), \widehat{E} \nabla f(\widehat{\mathbf{g}}_t) \rangle dt$$

with

$$\widehat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1 - e^{-2t}}\,\widehat{\mathbf{g}},$$

and $\widehat{\mathbf{g}}$ an independent copy of $\mathbf{g}$, and $\widehat{E}$ expectation with respect to $\widehat{\mathbf{g}}$. Applying the Cauchy-Schwarz inequality one obtains $E[T] \le E\|\nabla f(\mathbf{g})\|^2$.

Equality to variance hints that T may be a Stein coefficient.

Second Order Poincaré Inequality

Let $f(\mathbf{g})$ have mean μ and variance σ^2, and $N \sim \mathcal{N}(\mu, \sigma^2)$. Then

$$d_{\mathrm{TV}}(f(\mathbf{g}), N) \leq \frac{2}{\sigma^2} \sqrt{\mathrm{Var}\left(\int_0^\infty e^{-t} \langle \nabla f(\mathbf{g}), \widehat{E} \nabla f(\widehat{\mathbf{g}}_t) \rangle dt \right)}$$

Chatterjee (2009), Nourdin, Peccati and Reinert (2009).

E.g. Let $f(\mathbf{x}) = d^2(\mathbf{x}, C)$, the shortest squared distance between $\mathbf{x}$ and a closed convex set C. Then $\nabla f(\mathbf{x}) = 2(\mathbf{x} - \Pi_C(\mathbf{x}))$, and

$$\mathrm{Var}(f(\mathbf{g})) \leq 4Ed^2(\mathbf{g}, C) \quad \text{and} \quad d_{\mathrm{TV}}(f(\mathbf{g}), N) \leq \frac{16\sqrt{Ed^2(\mathbf{g}, C)}}{\sigma^2}$$

Second Order Poincaré Inequality

Let $f(\mathbf{g})$ have mean μ and variance σ^2, and $N \sim \mathcal{N}(\mu, \sigma^2)$. Then

$$d_{\mathrm{TV}}(f(\mathbf{g}), N) \leq \frac{2}{\sigma^2}\sqrt{\mathrm{Var}\left(\int_0^\infty e^{-t}\langle \nabla f(\mathbf{g}), \widehat{E}\nabla f(\widehat{\mathbf{g}}_t)\rangle dt\right)}$$

Chatterjee (2009), Nourdin, Peccati and Reinert (2009).

E.g. Let $f(\mathbf{x}) = d^2(\mathbf{x}, C)$, the shortest squared distance between $\mathbf{x}$ and a closed convex set C. Then $\nabla f(\mathbf{x}) = 2(\mathbf{x} - \Pi_C(\mathbf{x}))$, and

$\mathrm{Var}(f(\mathbf{g})) \leq 4Ed^2(\mathbf{g}, C)$ and $d_{\mathrm{TV}}(f(\mathbf{g}), N) \leq \dfrac{16\sqrt{Ed^2(\mathbf{g}, C)}}{\sigma^2}$

Second Order Poincaré Inequality

Let $f(\mathbf{g})$ have mean μ and variance σ^2, and $N \sim \mathcal{N}(\mu, \sigma^2)$. Then

$$d_{\mathrm{TV}}(f(\mathbf{g}), N) \leq \frac{2}{\sigma^2}\sqrt{\mathrm{Var}\left(\int_0^\infty e^{-t}\langle \nabla f(\mathbf{g}), \widehat{E}\nabla f(\widehat{\mathbf{g}}_t)\rangle dt\right)}$$

Chatterjee (2009), Nourdin, Peccati and Reinert (2009).

E.g. Let $f(\mathbf{x}) = d^2(\mathbf{x}, C)$, the shortest squared distance between $\mathbf{x}$ and a closed convex set C. Then $\nabla f(\mathbf{x}) = 2(\mathbf{x} - \Pi_C(\mathbf{x}))$, and

$$\mathrm{Var}(f(\mathbf{g})) \leq 4Ed^2(\mathbf{g}, C) \quad \text{and} \quad d_{\mathrm{TV}}(f(\mathbf{g}), N) \leq \frac{16\sqrt{Ed^2(\mathbf{g}, C)}}{\sigma^2}$$

Stein's method and Malliavin Calculus

For a mean zero function F of a Gaussian isonormal process $\{X(h) : h \in \mathfrak{H}\}$ in $\mathbb{D}^{1,2}$ (derivative, moment), using $F = LL^{-1}F, L = -\delta D$, we have the expression

$$T = \langle DF, -DL^{-1}F \rangle_{\mathfrak{H}}$$

where D is the Malliavin derivative, δ its adjoint, and L is the Ornstein-Uhlenbeck generator.

Stein's method and Malliavin Calculus

Mean zero L^2 functions F of X have an orthogonal Wiener-Ito Chaos expansion

$$F = \sum_{q=1}^{\infty} J_q.$$

The Stein coefficient

$$T = \langle DF, -DL^{-1}F \rangle_{\mathfrak{H}}$$

is particularly tractable when F is an element J_q of a fixed Wiener Chaos for $q \geq 1$.

Hence, can obtain bounds to the normal in total variation for, say, multiple stochastic integrals of Brownian Motion.

Stein's method and Malliavin Calculus

Mean zero L^2 functions F of X have an orthogonal Wiener-Ito Chaos expansion

$$F = \sum_{q=1}^{\infty} J_q.$$

The Stein coefficient

$$T = \langle DF, -DL^{-1}F \rangle_{\mathfrak{H}}$$

is particularly tractable when F is an element J_q of a fixed Wiener Chaos for $q \geq 1$.
Hence, can obtain bounds to the normal in total variation for, say, multiple stochastic integrals of Brownian Motion.

Stein's method and Malliavin Calculus

Nourdin and Peccati 2009:

Theorem

Let F belong to the q^{th} Wiener chaos of a Brownian motion for some $q \geq 2$. Then

$$d_{\mathrm{TV}}(F, Z) \leq 2|1 - EF^2| + 2\sqrt{\frac{q-1}{3q}}\sqrt{EF^4 - 3(EF^2)^2}.$$

Stein's method for non Normal distributions

Poisson $Z \sim \mathcal{P}(\lambda)$, characterizing equation

$$E[Zf(Z)] = \lambda E[f(Z+1)].$$

Recalling the size bias transformation, $W \sim \mathcal{P}(\lambda)$ for some $\lambda > 0$ if and only if W satisfies the distributional fixed point equation

$$W =_d W^s - 1.$$

If W^s has the W-size biased distribution, and is defined on the same space as W, then (Chen 1975)

$$d_{\mathrm{TV}}(W,Z) \leq (1-e^{-\lambda})E|W-(W^s-1)|$$

Applications to head runs, DNA sequence matching, etc. (Arratia, G., Gordon 1989, Barbour et at. 1992)

Stein's method for non Normal distributions

Poisson $Z \sim \mathcal{P}(\lambda)$, characterizing equation

$$E[Zf(Z)] = \lambda E[f(Z+1)].$$

Recalling the size bias transformation, $W \sim \mathcal{P}(\lambda)$ for some $\lambda > 0$ if and only if W satisfies the distributional fixed point equation

$$W =_d W^s - 1.$$

If W^s has the W-size biased distribution, and is defined on the same space as W, then (Chen 1975)

$$d_{\mathrm{TV}}(W, Z) \leq (1 - e^{-\lambda}) E|W - (W^s - 1)|$$

Applications to head runs, DNA sequence matching, etc. (Arratia, G., Gordon 1989, Barbour et at. 1992)

Poisson Subset Numbers

Let n be an integer and $A_0, \ldots, A_k$ be uniformly and independently chosen random subsets of fixed sizes $a_0, \ldots, a_k$ of $\{1, \ldots, n\}$. Compute a bound to the Poisson for

$$W = |\bigcap_{j=0}^{k} A_j| = \sum_{\alpha=1}^{n} \mathbf{1}(\alpha \in \bigcap_{j=0}^{k} A_j)$$

Size bias by choosing α uniformly and independently, and for all j such that $A_j \not\ni \alpha$, swap α into A_j by kicking out a uniformly and independently chosen element of A_j, forming W^α. Now

$$W - (W^\alpha - 1) = X_\alpha + \sum_{\beta \neq \alpha} (X_\beta - X_\beta^\alpha) \geq 0,$$

so absolute value in bound can be removed. Simple moment calculation results.

Poisson Subset Numbers

Let n be an integer and $A_0, \ldots, A_k$ be uniformly and independently chosen random subsets of fixed sizes $a_0, \ldots, a_k$ of $\{1, \ldots, n\}$. Compute a bound to the Poisson for

$$W = |\bigcap_{j=0}^{k} A_j| = \sum_{\alpha=1}^{n} \mathbf{1}(\alpha \in \bigcap_{j=0}^{k} A_j)$$

Size bias by choosing α uniformly and independently, and for all j such that $A_j \not\ni \alpha$, swap α into A_j by kicking out a uniformly and independently chosen element of A_j, forming W^α. Now

$$W - (W^\alpha - 1) = X_\alpha + \sum_{\beta \neq \alpha} (X_\beta - X_\beta^\alpha) \geq 0,$$

so absolute value in bound can be removed. Simple moment calculation results.

Poisson Subset Numbers

Let n be an integer and $A_0, \ldots, A_k$ be uniformly and independently chosen random subsets of fixed sizes $a_0, \ldots, a_k$ of $\{1, \ldots, n\}$. Compute a bound to the Poisson for

$$W = |\bigcap_{j=0}^{k} A_j| = \sum_{\alpha=1}^{n} \mathbf{1}(\alpha \in \bigcap_{j=0}^{k} A_j)$$

Size bias by choosing α uniformly and independently, and for all j such that $A_j \not\ni \alpha$, swap α into A_j by kicking out a uniformly and independently chosen element of A_j, forming W^α. Now

$$W - (W^\alpha - 1) = X_\alpha + \sum_{\beta \neq \alpha} (X_\beta - X_\beta^\alpha) \geq 0,$$

so absolute value in bound can be removed. Simple moment calculation results.

Exponential

For a non-negative random variable W with finite non zero mean we say W^e has the W-equilibrium transformation when $W^e =_d UW^s$, where U and W^s are independent, $U \sim \mathcal{U}[0,1]$ and W^s has the W-size biased distribution.

The variable W is exponential if and only if $W =_d W^e$.

Bounds (one example) from Peköz and Röellin (2011), if $EW^2 < \infty$ then with $Z \sim \mathrm{Exp}(1)$,

$$d_1(W, Z) \leq 2E|W^e - W|.$$

Exponential Approximation

Bounds in theorem of Yaglom on exponential limit for population size Z_n of generation n of critical Galton-Watson branching process GW conditioned on non-extinction, offspring distribution ν. Peköz and Röellin (2011), $Z \sim \mathrm{Exp}(1)$, offspring distribution variance σ^2,

$$d_1(\mathcal{L}(2Z_n/(\sigma^2 n))|Z_n > 0), Z) = O\left(\frac{\log n}{n}\right).$$

Exponential Approximation

Galton Watson Tree GW population Z_n generating n, size biased GW^s, from Lyons, Pemantle and Peres (1995). Start with single individual v_0 at time 0. In generation $n \geq 0$, pick an individual v_n uniformly to have offspring according ν^s, all others with ν, gives 'spine' $v_0, v_1, \ldots$ of tree that never dies out.

Number S_n of individuals in generation n of GW^s is Z_n^s, where Z_n is corresponding number in GW. Individual v_n is uniform over all individuals in generation n, and

$$\mathcal{L}(\mathrm{GW}^s | v_n \text{ is left most individual}) = \mathcal{L}(\mathrm{GW} | Z_n > 0)$$

Distribution of R_n, number of Z_n to the right of v_n is both approximately $\mathcal{L}(Z_n | Z_n > 0)$ and UZ_n^s, hence exponential.

Next Stop

Second floor of the Stein Mart: Concentration inequalities.