An Introduction to Stein's Method

Liam Hodgkinson

ACEMS

Objective

To show that two random elements X and Z are close in distribution:

$$\mathbb{E}f(X) \approx \mathbb{E}f(Z) \text{ for } f \in \mathcal{F}.$$

What you can do

By the Central Limit Theorem,

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

But how good is this approximation for *finite n*?

What you can do

Theorem (Berry-Esseen)

Let $X_1, X_2, \ldots$ *be iid w/* $\mathbb{E}X_i = 0$ *and* $\mathbb{E}X_i^2 = 1$.

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

$$\left|\mathbb{E}f(\sqrt{n}\,\bar{X}_n) - \mathbb{E}f(Z)\right| \leq \frac{4\|f'\|_\infty \mathbb{E}|X_i^3|}{\sqrt{n}}$$

ACEMS

What you can do

Theorem (Schuhmacher-Stucki)

Let Ξ, H be Gibbs processes with conditional intensities ν and λ respectively, with respect to μ. Then

$$\begin{aligned}&|\mathbb{E}f(\Xi) - \mathbb{E}f(H)| \\ &\quad \leq C(\lambda)\|f\|_\infty \int \mathbb{E}|\nu(x|\Xi) - \lambda(x|\Xi)|\mu(dx).\end{aligned}$$

An analytic approach

Estimating $f(b) - f(a)$:

An analytic approach

Estimating $f(b) - f(a)$: apply Taylor's Theorem!

An analytic approach

Estimating $f(b) - f(a)$: apply Taylor's Theorem!

$$f(b) - f(a) \underset{\text{FTC}}{=} \int_a^b f'(x)\mathrm{d}x \underset{\text{MVT}}{=} (b-a)f'(\xi).$$

bound the derivative $\rightarrow$ *bound the difference.*

An analytic approach

More generally,

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) = \int_{\gamma[\boldsymbol{x},\boldsymbol{y}]} \nabla f(\boldsymbol{r}) \cdot \mathrm{d}\boldsymbol{r},$$

for any curve $\gamma[x, y]$ **from** x **to** y.

$\longrightarrow$ Taylor's Theorem in multiple variables

The stochastic analogue

A ‘curve’ between two random variables

The stochastic analogue

A ‘curve’ between two random variables

— a stochastic process!

The stochastic analogue

Theorem (Ito's Lemma)

Let X_t satisfy the stochastic differential equation $\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t$,

$$\mathbb{E}f(X_t) - \mathbb{E}f(X_0) = \int_0^t \mathbb{E}\mathcal{A}f(X_s)\mathrm{d}s,$$

where for $\Sigma(x) = \sigma(x)\sigma(x)^\top$,

$$\mathcal{A}f(x) = \frac{1}{2}\sum_{i,j=1}^{n} \Sigma_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_{i=1}^{n} \mu_i(x)\frac{\partial}{\partial x_i} f(x).$$

The stochastic analogue

More generally,

Theorem (Dynkin's Formula)

Let X_t *be a Markov process with* **generator** $\mathcal{A}$. *Then*

$$\mathbb{E}f(X_t) - \mathbb{E}f(X_0) = \int_0^t \mathbb{E}\mathcal{A}f(X_s)\mathrm{d}s.$$

ACEMS

Barbour's generator approach

Suppose X_t has equilibrium distribution Z.

If $X_0 \sim X$, then

$$\mathbb{E}f(X_t) - \mathbb{E}f(X_0) = \int_0^t \mathbb{E}\mathcal{A}f(X_s)\mathrm{d}s.$$

Barbour's generator approach

Suppose X_t has equilibrium distribution Z.

If $X_0 \sim X$, then

$$\mathbb{E}f(X_t) - \mathbb{E}f(X) = \int_0^t \mathbb{E}\mathcal{A}f(X_s)\mathrm{d}s.$$

Barbour's generator approach

Suppose X_t has equilibrium distribution Z.

If $X_0 \sim X$, then

$$\mathbb{E}f(Z) - \mathbb{E}f(X) = \int_0^{\infty} \mathbb{E}\mathcal{A}f(X_s)\mathrm{d}s.$$

Barbour's generator approach

Proposition (Ethier & Kurtz, Proposition 1.5)

$$\int_0^\infty \mathbb{E}\mathcal{A}f(X_s)\mathrm{d}s = \mathbb{E}\mathcal{A}g_f(X_0)$$

where g_f is the function

$$g_f(x) = \int_0^\infty \mathbb{E}_x f(X_s) - \mathbb{E}f(Z)\mathrm{d}s$$

Barbour's generator approach

Suppose X_t has equilibrium distribution Z.

If $X_0 \sim X$, then

$$\mathbb{E}f(X) - \mathbb{E}f(Z) = \mathbb{E}\mathcal{A}g_f(X).$$

where $g_f(x) = -\int_0^\infty \mathbb{E}_x f(X_s) - \mathbb{E}f(Z)\mathrm{d}s.$

Barbour's generator approach

Suppose X_t has equilibrium distribution Z.

If $X_0 \sim X$, then

$$\mathbb{E}f(X) - \mathbb{E}f(Z) = \mathbb{E}\mathcal{A}g_f(X).$$

where $g_f(x) = -\int_0^\infty \mathbb{E}_x f(X_s) - \mathbb{E}f(Z)\mathrm{d}s$.

This is the foundation of Stein's method

Examples for X_t

Ornstein-Uhlenbeck: $Z \sim \mathcal{N}(0,1)$:

$$\mathrm{d}X_t = -X_t\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$$

The **generator** is

$$\mathcal{A}f(x) = f''(x) - xf'(x).$$

Examples for X_t

Langevin SDE: Z with density π:

$$\mathrm{d}X_t = \nabla \log \pi(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$$

The **generator** is

$$\mathcal{A}f(x) = \Delta f(x) + \nabla f(x) \cdot \nabla \log \pi(x).$$

This might look familiar...

Stein's trick

For the generator $\mathcal{A}$:

1. $\mathbb{E}\mathcal{A}f(X) = 0$ for all $f \in \mathcal{C}^\infty \iff X \sim Z$
2. g_f is the unique solution to the **Stein equation**

$$f(x) - \mathbb{E}f(Z) = \mathcal{A}g_f(x).$$

Stein's trick

For a Stein operator $\mathcal{A}$ for Z:

1. $\mathbb{E}\mathcal{A}f(X) = 0$ for all $f \in \mathcal{F} \iff X \sim Z$
2. g_f is the unique solution to the **Stein equation**

$$f(x) - \mathbb{E}f(Z) = \mathcal{A}g_f(x).$$

Stein's trick

For a **Stein operator** $\mathcal{A}$ for Z:

1. $\mathbb{E}\mathcal{A}f(X) = 0$ for all $f \in \mathcal{F} \iff X \sim Z$
2. g_f is the unique solution to the **Stein equation**

$$f(x) - \mathbb{E}f(Z) = \mathcal{A}g_f(x).$$

Stein's Method: bound $\sup_{f\in\mathcal{F}} |\mathbb{E}\mathcal{A}g_f(X)|$

"*The* Stein operator"

For Z with density π, the generator of the Langevin SDE is

$$\mathcal{A}f(x) = \Delta f(x) + \nabla f(x) \cdot \nabla \log \pi(x)$$

“*The* Stein operator”

For Z with density π, “the” Stein operator for Z is

$$\mathcal{A}\phi(x) = \nabla \cdot \phi(x) + \phi(x) \cdot \nabla \log \pi(x)$$

Kernelised Stein Discrepancy

If $\mathcal{H}$ is a reproducing kernel Hilbert space

$$\sup_{\phi \in \mathcal{H}} |\mathbb{E}\mathcal{A}\phi(X)| = \mathbb{E}K(X, X'),$$

where K is a certain kernel (Liu et al, 2016).

A direct way to estimate discrepancy between a sample and a distribution

ACEMS

Proof of the Berry-Esseen Theorem

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

Let $\mathcal{A}f(x) = f'(x) - xf(x)$

Let $W = n^{-1/2}\sum_{i=1}^{n} X_i = \sqrt{n}\,\bar{X}_n$

Proof of the Berry-Esseen Theorem

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

Let $\mathcal{A}f(x) = f'(x) - xf(x)$

Let $W = n^{-1/2}\sum_{i=1}^{n} X_i = \sqrt{n}\,\bar{X}_n$

Let $W_i = n^{-1/2}\sum_{j=1, j\neq i}^{n} X_j = W - n^{-1/2}X_i$

Proof of the Berry-Esseen Theorem

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

Let $\mathcal{A}f(x) = f'(x) - xf(x)$

Let $W = n^{-1/2}\sum_{i=1}^{n} X_i = \sqrt{n}\,\bar{X}_n$

Let $W_i = n^{-1/2}\sum_{j=1, j\neq i}^{n} X_j = W - n^{-1/2}X_i$

Taylor's Theorem:

$$f(W) - f(W_i) = \frac{1}{\sqrt{n}}X_i f'(W_i) + \frac{1}{n}X_i^2 f''(\xi_i)$$

Proof of the Berry-Esseen Theorem

Lemma (Stein's Lemma)

Let $\mathcal{A}f(x) = f'(x) - xf(x)$ be the Stein operator of $\mathcal{N}(0,1)$. Then

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

Proof of the Berry-Esseen Theorem

Theorem (Berry-Esseen)

Let $X_1, X_2, \ldots$ *be iid w/* $\mathbb{E}X_i = 0$ *and* $\mathbb{E}X_i^2 = 1$.

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

$$|\mathbb{E}f(W) - \mathbb{E}f(Z)| \leq \frac{4\|f'\|_\infty \mathbb{E}|X_i^3|}{\sqrt{n}}$$

For any distribution there are infinitely many such X_t.

What makes a good one?

Contraction rates

Rapid convergence of X_t to stationarity

Contraction rates

Rapid convergence of X_t to stationarity

- If X_t is exponentially ergodic, then $\|g_f\|_\infty \leq C\|f\|_\infty$
- If X_t is hypercontractive with constant λ, then $\|\nabla g_f\|_\infty \leq \lambda\|\nabla f\|_\infty$.