Stein's Method on Stratified Spaces?

Huiling Le
University of Nottingham

- (Classical) Stein's method
  - The Stein operator and equation
  - The approach via generators
- Stein's method on manifolds
  - The Stein operator and equation
  - A fundamental condition
- Stein's method on stratified spaces
  - A version of the Stein method?
  - Related diffusions?
  - The Stein operator and equation?

(Classical) Stein's method: the Stein operator and equation

Fix probability measure P and r.v. $X \sim P$.

- Identify an operator $\mathcal{A}$ (*the Stein operator*) on a family of functions $\mathcal{F}(\mathcal{A})$ such that

$$\mathrm{E}\left[\mathcal{A}f(X)\right] = 0, \qquad \forall f \in \mathcal{F}(\mathcal{A}).$$

- Let $\mathcal{H}$ be a family of functions such that, for $h \in \mathcal{H}$, there exists $f = f_h \in \mathcal{F}(\mathcal{A})$ satisfying

$$h(x) - \mathrm{E}\left[h(X)\right] = \mathcal{A}f_h(x)$$

(*the Stein equation*).

- Then, for any other probability measure Q and r.v. $Z \sim Q$,

$$\mathsf{E}\left[h(Z)\right] - \mathsf{E}\left[h(X)\right] = \mathsf{E}\left[\mathcal{A}f_h(Z)\right].$$

In particular,

$$d_{\mathcal{H}}(Z, X) = \sup_{h \in \mathcal{H}} \left| \mathsf{E}\left[h(Z)\right] - \mathsf{E}\left[h(X)\right] \right| \leqslant \sup_{f \in \mathcal{F}(\mathcal{A})} \left| \mathsf{E}\left[\mathcal{A}f(Z)\right] \right|.$$

If $\mathcal{H}$ is the family of all Lipschitz-1-functions, the resulting $d_{\mathcal{H}}$ is the Wasserstein distance.

Example

If $X \sim N(0, \sigma^2)$, the corresponding Stein operator is

$$\mathcal{A}f(x) = \sigma^2 f'(x) - x\, f(x);$$

and the solutions to the corresponding Stein equation

$$h(x) - \mathrm{E}\left[h(X)\right] = \sigma^2 f_h'(x) - x\, f_h(x)$$

are given by

$$f_h(x) = \frac{1}{\sigma^2} e^{x^2/(2\sigma^2)} \left\{ a + \int_{-\infty}^{x} \left\{ h(u) - \mathrm{E}\left[h(X)\right]\right\} e^{-u^2/(2\sigma^2)} du \right\}.$$

Stein's method for probability measures on $\mathbb{R}^m$:

- E. Meckes (2009). On Stein's method for multivariate normal approximation, in *High Dimensional Probability V: The Luminy Volume*, C. Houdré, V. Koltchinskii, D.M. Mason and M. Peligrad eds., IMS, 153–178.
- L. Mackey & J. Gorham (2016). Multivariate Stein factors for a class of strongly log-concave distributions, *Electron. Commun. Probab.* **21**, no. 56.
- G. Mijoule, G. Reinert and Y. Swan (2019). Stein operators, kernels and discrepancies for multivariate continuous distributions. arXiv:1806.03478.

(Classical) Stein's method: the approach via generators

This is based on the classical theory of Markov processes.

For the normal distribution $N(0, \sigma^2)$, we take

$$\mathcal{L}f \equiv \mathcal{A}f' = \sigma^2 f'' - x\, f'.$$

Then,

$$\mathcal{L} = \sigma^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} - x \frac{\mathrm{d}}{\mathrm{d}x},$$

and it is the infinitesimal generator of the Ornstein-Uhlenbeck process

$$\mathrm{d}X_t = \sqrt{2}\sigma \, \mathrm{d}B_t - X_t \, \mathrm{d}t.$$

Also, the equilibrium distribution of the OU process is the target distribution $N(0, \sigma^2)$.

Stein's method on manifolds

- J. Thompson (2020). Approximation of Riemannian measures by Stein's method. arXiv:2001.009.
- A. Lewis (2021). Stein's method for probability distributions on $\mathbb{S}^1$. arXiv: 2105.13199.
- H. Le, A. Lewis, K. Bharath and C. Fallaize (2024). A diffusion approach to Stein's method on Riemannian manifolds, *Bernoulli* **30**, 1079-1104.

Stein's method on manifolds

$(\boldsymbol{M}, g)$: a complete and connected Riemannian manifold (without boundary) of dimension m.

$\rho(x, y)$: the Riemannian distance between x and y in $\boldsymbol{M}$.

ϕ: a fixed $\mathcal{C}^2$-function on $\boldsymbol{M}$ such that $\nabla\phi$ satisfies a Lipschitz condition, where ∇ is the gradient operator.

μ_ϕ: probability measure on $\boldsymbol{M}$ given by

$$\mathrm{d}\mu_\phi = \frac{1}{C_\phi} e^{-\phi} \, \mathrm{dvol},$$

assuming $C_\phi = \int_{\boldsymbol{M}} e^{-\phi} \, \mathrm{dvol} < \infty$.

r.v. $X \sim \mathrm{d}\mu_\phi$.

$\boldsymbol{M} = S^1$:

Using integration by parts,

$$f_h(x) = C_\phi e^{\phi(x)} \left\{ a + \int_{-\pi}^{x} (h(y) - \mathrm{E}\left[h(X)\right]) \, \mathrm{d}\mu_\phi(y) \right\}$$

solves the Stein equation for $\mathrm{d}\mu_\phi$

$$h(x) - \mathrm{E}\left[h(X)\right] = f_h'(x) - \phi'(x) \, f_h(x),$$

with the Stein operator $\mathcal{A}f = f' - \phi' \, f$.

For example, $\phi(x) = -c \, \cos(x - x_0)$ corresponds to the von Mises distribution $M(x_0, c)$, so that the Stein operator for $M(x_0, c)$ is

$$\mathcal{A}f(x) = f'(x) + c \, \sin(x - x_0) \, f(x).$$

General $\boldsymbol{M}$:

Consider the (uniformly elliptic) diffusion on $\boldsymbol{M}$, given by the solution of the Itô stochastic differential equation

$$\mathrm{d}X_t = \mathrm{d}B_t^{\boldsymbol{M}} - \frac{1}{2}\nabla\phi(X_t)\,\mathrm{d}t,$$

$B_t^{\boldsymbol{M}}$: BM on $\boldsymbol{M}$.

- The infinitesimal generator for this diffusion is the self-adjoint operator

$$\mathcal{L}_\phi = \frac{1}{2}\left\{\Delta - \langle\nabla\phi, \nabla\rangle\right\},$$

  Δ: the Laplace-Beltrami operator of $(\boldsymbol{M}, g)$.

- If there is a constant $\kappa > 0$ such that

$$\mathrm{Ric}(x) + \mathrm{Hess}^{\phi}(x) \geqslant -\kappa\, g(x), \quad \forall x \in \boldsymbol{M},$$

  then X_t is conservative, where

  Ric: the Ricci curvature tensor;
  Hess^{ϕ}: the Hessian of ϕ.

- $\mathrm{d}\mu_{\phi}$ is the unique equilibrium measure for X_t.

Stein's method on manifolds: the Stein operator and equation for $\mathrm{d}\mu_\phi$

Assume

- $X \sim \mathrm{d}\mu_\phi = e^{-\phi}\,\mathrm{dvol}\,/\,C_\phi$;
- $\mathrm{E}\,[\rho(X, x)] < \infty$ for some $x \in \boldsymbol{M}$;
- $X_{x,t}$: a diffusion determined by

$$\mathrm{d}X_t = \mathrm{d}B_t^{\boldsymbol{M}} - \frac{1}{2}\nabla\phi(X_t)\,\mathrm{d}t,$$

  starting from x.

For a given h, define

$$f_h(x) = \int_0^\infty \{\mathrm{E}\,[h(X)] - \mathrm{E}\,[h(X_{x,t})]\}\,\mathrm{d}t.$$

Assume that

$$\mathrm{Ric} + \mathrm{Hess}^{\phi} \geqslant 2\kappa\, g \tag{$*$}$$

for a constant $\kappa > 0$, and that $h \in \mathcal{C}_0(\boldsymbol{M})$ is a Lipschitz function. Then, the Stein equation for $\mathrm{d}\mu_\phi$ is

$$h(x) - \mathsf{E}\left[h(X)\right] = \mathcal{L}_\phi f_h(x).$$

In particular, $\mathcal{L}_\phi$ is the Stein operator for $\mathrm{d}\mu_\phi$.

By carefully analysing the bound for $\mathcal{L}_\phi(f_h)$, the results relating to Stein's method for Euclidean r.v.'s can be generalised to Riemannian manifolds.

For example, let $k = 1, 2$,

$$\mathcal{H}_k = \{h \in \mathcal{C}^k(\boldsymbol{M}) \mid h \text{ is Lipschitz with } C_i(h) = 1, i = 0, \cdots, k\}.$$

(*i*) For $Z \sim \mathrm{d}\mu_\psi$ satisfying the corresponding $(*)$,

$$d_{\mathcal{H}_1}(Z, X) \leqslant \frac{1}{\kappa} \mathsf{E}\left[|\nabla(\psi - \phi)(Z)|\right].$$

(*ii*) For a general Z on $\boldsymbol{M}$,

$$d_{\mathcal{H}_2}(Z, X) \leqslant \frac{2}{\kappa} \eta \, \mathsf{E}\left[\rho(Z, X)\right],$$

where η is a positive constant depending only on ϕ and the geometry of $\boldsymbol{M}$.

Examples.

(i) Take $\boldsymbol{M} = S^m$. Then, the distance $d_{\mathcal{H}_1}$ between two von Mises-Fisher distributions $X_1 \sim M(x_1, c_1)$ and $X_2 \sim M(x_2, c_2)$ is bounded by

$$d_{\mathcal{H}_1}(X_1, X_2) \leqslant \frac{|c_2 x_2 - c_1 x_1|}{2\kappa} \sum_{i=1}^{2} \{\rho(x^*, x_i) + \mathrm{E}\left[\rho(x_i, X_i)\right]\},$$

where

$$x^* = \frac{c_2 x_2 - c_1 x_1}{|c_2 x_2 - c_1 x_1|}.$$

(ii) Take $\boldsymbol{M} = SO(m)$ with the bi-invariant metric determined by $\mathrm{tr}(AB)$ for skew-symmetric A, B.

Take $\phi(S) = -c\,\mathrm{tr}(S_0 S)$ with $S_0 \in SO(m)$ and the constant $c > 0$. Then, $\mathrm{d}\mu_\phi$ is a von Mises-Fisher distribution on $SO(m)$.

If Z is a uniform random variable on $SO(m)$ and

$$\mathrm{Hess}^\phi \geqslant \left(2\kappa - \frac{(m-2)}{4}\right) g$$

for some $\kappa > 0$, then

$$d_{\mathcal{H}_1}(Z, X) \leqslant \frac{c}{\kappa}\,\mathrm{E}\left[\sqrt{m - \mathrm{tr}(Z^2)}\right].$$

Stein's method on manifolds: a fundamental condition

$$\mathrm{Ric} + \mathrm{Hess}^{\phi} \geqslant 2\kappa\, g \qquad (*)$$

for a constant $\kappa > 0$.

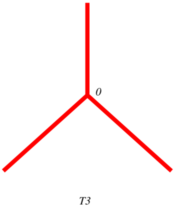
Assume that the condition $(*)$ holds. Then, there is a pair of coupled diffusions $(X_{x,t}, Y_{y,t})$, both with generator $\mathcal{L}_\phi$, s.t.

$$\mathrm{E}\left[\rho(X_{x,t}, Y_{y,t})^p\right] \leqslant \rho(x,y)^p\, e^{-p\kappa t}, \quad p \geqslant 1.$$

Stein's method on stratified spaces

Assume: top dimensional strata are joined by co-dimensional one strata and each co-dimension one stratum lies on the boundary of more than 2 top dimensional strata, e.g. spiders, open books and tree spaces etc.

It is sufficient to concentrate on spiders.



For a fixed integer $N > 2$, let Γ be the space defined by

$$\Gamma = \{\boldsymbol{x} = (x, i) : \quad x \geqslant 0; \quad i = 1, \cdots, N\},$$

where we identify $(0, i)$, $i = 1, \cdots, N$, and call it O.

Let ϕ_i, $i = 1, \cdots, N$, be N $\mathcal{C}^2([0, \infty))$ functions with $\phi_i(0) = \phi_1(0)$ and with $c_i = \int_0^\infty e^{-\phi_i(x)} \, \mathrm{d}x < \infty$. Then,

$$\mathrm{d}\mu_\phi(\boldsymbol{x}) = \frac{\alpha_i}{\sum_{k=1}^{N} \alpha_k c_k} e^{-\phi_i(x)} \, \mathrm{d}x. \qquad \text{for } \boldsymbol{x} = (x, i)$$

is a probability measure on Γ, where $\alpha_i > 0$ such that

$$\sum_{i=1}^{N} \alpha_i = 1.$$

Stein's method on stratified spaces: a version of the Stein method?

It can be checked, using integration by parts, that, for given H on Γ, the F_h on Γ defined by

$$F_{h,i}(x) = e^{\phi_i(x)} \int_0^x e^{-\phi_i(t)} \left(H_i(t) - \frac{\alpha_i}{\sum_{j=1}^N \alpha_j c_j} \int_0^\infty H_i(u)\, e^{-\phi_i(u)}\, \mathrm{d}u \right) \mathrm{d}t,$$

where $F_{h,i}(x) = F_h(\boldsymbol{x})$ for $\boldsymbol{x} = (x, i)$, solves the Stein equation

$$F_h'(\boldsymbol{x}) - \phi'(\boldsymbol{x})\, F_h(\boldsymbol{x}) = H(\boldsymbol{x}) - \mathrm{E}\left[H(X)\right],$$

where $X \sim \mathrm{d}\mu_\phi$, i.e.

$$\frac{\mathrm{d}F_{h,i}(x)}{\mathrm{d}x} - \frac{\mathrm{d}\phi_i(x)}{\mathrm{d}x} F_{h,i}(x) = H_i(x) - \frac{\alpha_i}{\sum_{j=1}^N \alpha_j c_j} \int_0^\infty H_i(u)\, e^{-\phi_i(u)}\, \mathrm{d}u.$$

Stein's method on stratified spaces: related diffusions?

Let L_i be the operator on $(0, \infty)$ given by

$$L_i F_i(x) = \frac{1}{2}\left\{\frac{\mathrm{d}^2 F_i(x)}{\mathrm{d}x^2} - \frac{\mathrm{d}\phi_i(x)}{\mathrm{d}x}\frac{\mathrm{d}F_i(x)}{\mathrm{d}x}\right\}, \qquad \text{for } x > 0,$$

and define the operator $\mathcal{L}_\phi$ on $\mathcal{C}^\infty(\Gamma)$ by

$$\mathcal{L}_\phi F(\boldsymbol{x}) = L_i F_i(x) \qquad \text{for } \boldsymbol{x} = (x, i),$$

where the domain $\mathcal{D}(\mathcal{L}_\phi)$ consists of functions $F \in \mathcal{C}^\infty(\Gamma)$ satisfying the condition

$$\rho(F) = \sum_{i=1}^{N} \alpha_i \frac{\mathrm{d}F_i}{\mathrm{d}x}(0) = 0.$$

The $\mathcal{L}_\phi$ generates a Markov process $X(t) = (x(t), i(t))$ on Γ, which, inside each leg I_i, is a diffusion process governed by L_i. Then, there is a BM $B(t)$ and a continuous increasing process $\ell(t)$ such that

$$\mathrm{d}x(t) = \mathrm{d}B(t) - \frac{1}{2}\frac{\mathrm{d}\phi_{i(t)}}{\mathrm{d}x}(x(t))\,\mathrm{d}t + \mathrm{d}\ell(t),$$

where $\ell(t)$ increases only when $x(t) = 0$.

- The process almost surely spends zero time at O.
- If $X(0) = O$, α_k is the probability that the process moves into I_k next.
- The Itô formula becomes

$$\mathrm{d}F(X(t)) = \frac{\mathrm{d}F_{i(t)}}{\mathrm{d}x}(x(t))\,\mathrm{d}B(t) + \mathcal{L}_\phi F(X(t))\,\mathrm{d}t + \rho(F)\,\mathrm{d}\ell(t).$$

Stein's method on stratified spaces: the Stein operator and equation?

- $\mathrm{d}\mu_\phi$ is the invariant distribution of $X(t)$.

IF there is a pair of coupled Markov processes $(X_{\mathbf{x}}(t), Y_{\mathbf{y}}(t))$, both with generator $\mathcal{L}_\phi$, s.t.

$$\mathsf{E}\left[d(X_{\mathbf{x}}(t), Y_{\mathbf{y}}(t))^p\right] \leqslant d(\mathbf{x}, \mathbf{y})^p\, e^{-p\kappa t}, \quad p = 1, 2,$$

for some $\kappa > 0$, then

- $\mathcal{L}_\phi$ is the Stein operator for $\mathrm{d}\mu_\phi$.
- The Stein equation for $\mathrm{d}\mu_\phi$:

$$H(\mathbf{x}) - \mathsf{E}\left[H(X)\right] = \mathcal{L}_\phi F_h(\mathbf{x}),$$

  with

$$F_h(\mathbf{x}) = \int_0^\infty \left(\mathsf{E}\left[H(X)\right] - \mathsf{E}\left[H(X_{\mathbf{x}}(t))\right]\right) \mathrm{d}t.$$

- The Stein equation can be used to study discrepancies between random variables.

BUT,

- unlike the case for manifolds, it is unclear under what conditions we can construct a pair of Markov processes with the required exponential decay (or perhaps weaker) property;
- the difficulty arises due to the 'local time' term in the Itô formula for the distance function.