Measuring Sample Quality with Stein's Method

An attempt at quantifying Monte-Carlo efficiency by Gorham and Mackey (2015)

Luis Sierra Muntané
luis.sierra@mail.utoronto.ca

Bayesian Reading Group
University of Toronto DoSS

9th December 2025

Outline

Our goal is to calculate quantities of the form $\mathbb{E}_P[h(X)]$ where P is an unknown target distribution form the samples of Q as

$$\mathbb{E}_Q[h(X)] = \sum_{i=1}^{n} q(x_i)h(x_i).$$

We will be following (Gorham and Mackey, 2015).

- Revisiting Stein's Lemma
- Finding distances between distributions
- How Stein's method provides an answer
- Constructing Stein operators
- Calculating the discrepancies
- Experiments

Stein's Lemma

For X mean-zero random variable and g a weakly differentiable and L_1 integrable function, we have that

$$X \text{ standard Gaussian} \quad \Longrightarrow \quad \mathrm{Cov}(X, g(X)) = \mathbb{E}[g'(X)],$$

Stein's Lemma

For X mean-zero random variable and g a weakly differentiable and L_1 integrable function, we have that

$$X \text{ standard Gaussian} \quad \Longleftrightarrow \quad \mathrm{Cov}(X, g(X)) = \mathbb{E}[g'(X)],$$

Stein's Lemma

For X mean-zero random variable and g a weakly differentiable and L_1 integrable function, we have that

$$X \text{ standard Gaussian} \quad \Longleftrightarrow \quad \mathrm{Cov}(X, g(X)) = \mathbb{E}[g'(X)],$$

$$\mathbb{E}[g'(X) - Xg(X)] = 0,$$

Stein's Lemma

For X mean-zero random variable and g a weakly differentiable and L_1 integrable function, we have that

$$X \text{ standard Gaussian} \qquad \Longleftrightarrow \qquad \mathrm{Cov}(X, g(X)) = \mathbb{E}[g'(X)],$$

$$\mathbb{E}[g'(X) - Xg(X)] = 0, \qquad \longrightarrow \qquad \mathbb{E}[\mathcal{A}g(X)] = 0$$

Stein's Lemma

For X mean-zero random variable and g a weakly differentiable and L_1 integrable function, we have that

$$X \text{ standard Gaussian} \quad \Longleftrightarrow \quad \mathrm{Cov}(X, g(X)) = \mathbb{E}[g'(X)],$$

$$\mathbb{E}[g'(X) - Xg(X)] = 0, \qquad \longrightarrow \qquad \mathbb{E}[\mathcal{A}g(X)] = 0$$

$$\begin{aligned} X \sim \mathrm{Gamma}(\alpha, \beta) \quad &\Longleftrightarrow \quad \mathbb{E}[Xf'(X) + (\alpha - \beta X)f(X)] = 0 \\ X \sim \mathrm{Poisson}(\lambda) \quad &\Longleftrightarrow \quad \mathbb{E}[\lambda f(X+1) - Xf(X)] = 0 \\ X \sim \mathrm{Binomial}(n, p) \quad &\Longleftrightarrow \quad \mathbb{E}[(1-p)Xf(X) - p(n-X)f(X+1)] \end{aligned}$$

Stein's Lemma

For X mean-zero random variable and g a weakly differentiable and L_1 integrable function, we have that

$$X \text{ standard Gaussian} \quad \Longleftrightarrow \quad \mathrm{Cov}(X, g(X)) = \mathbb{E}[g'(X)],$$

$$\mathbb{E}[g'(X) - Xg(X)] = 0, \qquad \longrightarrow \qquad \mathbb{E}[\mathcal{A}g(X)] = 0$$

$$\begin{aligned} X \sim \mathrm{Gamma}(\alpha, \beta) \quad &\Longleftrightarrow \quad \mathbb{E}[Xf'(X) + (\alpha - \beta X)f(X)] = 0 \\ X \sim \mathrm{Poisson}(\lambda) \quad &\Longleftrightarrow \quad \mathbb{E}[\lambda f(X+1) - Xf(X)] = 0 \\ X \sim \mathrm{Binomial}(n, p) \quad &\Longleftrightarrow \quad \mathbb{E}[(1-p)Xf(X) - p(n-X)f(X+1)] \end{aligned}$$

$$X \sim P \quad \Longleftrightarrow \quad \mathbb{E}_P[(\mathcal{A}f)(X)] = 0.$$

Quality Measures for samples

Target distribution P with open convex support $\mathcal{X} \subseteq \mathbb{R}^d$. We approximate P with Q.

Goal: to have a measure of the quality of the samples Q:

Quality Measures for samples

Target distribution P with open convex support $\mathcal{X} \subseteq \mathbb{R}^d$. We approximate P with Q.

Goal: to have a measure of the quality of the samples Q:

- Detects convergence $Q_m \to P$
- Detects Q not converging to P
- Is computationally feasible

Quality Measures for samples

Target distribution P with open convex support $\mathcal{X} \subseteq \mathbb{R}^d$. We approximate P with Q.

Goal: to have a measure of the quality of the samples Q:

- Detects convergence $Q_m \to P$
- Detects Q not converging to P
- Is computationally feasible

A possible way to ensure these is by using an *integral probability metric*

Quality Measures for samples

Target distribution P with open convex support $\mathcal{X} \subseteq \mathbb{R}^d$. We approximate P with Q.

Goal: to have a measure of the quality of the samples Q:

- Detects convergence $Q_m \to P$
- Detects Q not converging to P
- Is computationally feasible

A possible way to ensure these is by using an *integral probability metric*

$$d_{\mathcal{H}}(Q, P) = \sup_{h \in \mathcal{H}} |\mathbb{E}_Q[h(X)] - \mathbb{E}_P[h(X)]|$$

Stein's Method

How to characterize convergence in distribution classically (Stein, 1972):

1. Find a real-valued operator $\mathcal{T} : \mathcal{G} \to \mathbb{R}$ characterizing P in the sense that
$$\mathbb{E}_P[(\mathcal{T}g)(Z)] = 0 \quad \forall g \in \mathcal{G}.$$
Together, $\mathcal{T}, \mathcal{G}$ define the *Stein discrepancy*
$$\mathcal{S}(Q, \mathcal{T}, \mathcal{G}) := \sup_{g \in \mathcal{G}} |\mathbb{E}_Q[(\mathcal{T}g)(X)]| = d_{\mathcal{G}}(Q, P).$$
2. Lower bound (Consistency)
3. Upper bound $\mathcal{S}(Q, \mathcal{T}, \mathcal{G})$ to demonstrate convergence to zero. (Reliability)

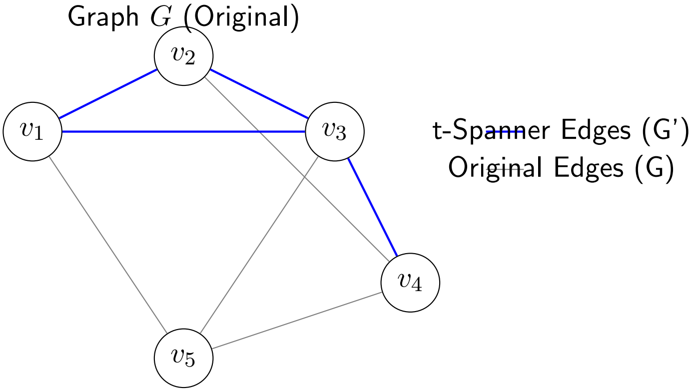
How to construct $\mathcal{A}$

Infinitesimal generator. For a Markov process $(Z_t)_{t\geq 0}$

$$\lim_{t\to 0}\frac{1}{t}\left(\mathbb{E}[u(x)\,|\,Z_0 = x] - u(x)\right)$$

Bounding the Stein Discrepancy

Frame Title



Graph G' (t-Spanner for $t \geq 1$)

Extensions

Replacing with a Kernel Approach (Gorham and Mackey, 2017) Diffusion (Gorham et al., 2019)

References

J. Gorham and L. Mackey. Measuring sample quality with stein's method. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

J. Gorham and L. Mackey. Measuring sample quality with kernels. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1292–1301. PMLR, Aug 06–11 2017. URL https://proceedings.mlr.press/v70/gorham17a.html.

J. Gorham, A. B. Duncan, S. J. Vollmer, and L. W. Mackey. Measuring sample quality with diffusions. *Annals of Applied Probability*, 29(5):2884–2928, 2019. doi: 10.1214/19-AAP1467.

C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume II: Probability Theory*, pages 583–602. University of California Press, 1972.