A mini-course

An Introduction to Stein's Method for Machine Learning

Foundations, Applications, and Geometry

Amirabbas Asadi

amir.asadi78@sharif.edu

Oct 2026

Abstract

This mini-course introduces Stein's method, connecting its foundations in probability with modern methods for sampling and inference. We will begin with Charles Stein's original paper on normal approximation, then explore Barbour's generator approach to deriving Stein operators for other distributions. Along the way, we will discuss theoretical applications, including derivation or improvement of concentration inequalities and mention other interesting results like the Fourth Moment Theorem.

In the second part, we will use the Langevin Stein operator to derive Stein variational gradient descent (SVGD), a nonparametric approach to variational inference that can also be viewed as a deterministic sampling method. Finally, we will interpret SVGD through the framework of gradient flows, revealing the geometric structure behind its particle dynamics.

Part I. Foundations of Stein's Method

We begin with the central ideas of Stein's method and how they turn distributional identities into quantitative bounds.

Normal approximation. Exploring Stein's original paper [1] and exploring how differential equations can be used to bound the error in the central limit theorem.

Barbour's generator approach. Using Markov process generators to construct Stein operators for other target distributions [2].

Theoretical applications. Exploring concentration inequalities for dependent random variables [3] and mentioning other interesting applications like the Fourth Moment Theorem [4].

Part II. From Stein Operators to Inference

We then connect these probabilistic ideas with computational statistics and machine learning, focusing on deterministic sampling and its geometry.

Stein variational gradient descent. Introducing kernelized Stein discrepancies and using the Langevin Stein operator to derive SVGD: a nonparametric variational inference method with deterministic particle updates [5].

The geometry of SVGD. Interpreting the evolution of the particle distribution as a gradient flow and understanding the geometric structure behind the algorithm [6].

Selected References

[1] Charles 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 2: Probability theory*, volume 6, pages 583–603. University of California Press, 1972.

[2] Andrew D Barbour. Stein's method for diffusion approximations. *Probability theory and related fields*, 84(3):297–322, 1990.

[3] Sourav Chatterjee. Stein's method for concentration inequalities, 2006.

[4] Ivan Nourdin and Giovanni Peccati. Stein's method on wiener chaos. *Probability Theory and Related Fields*, 145(1):75–118, 2009.

[5] Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

[6] Qiang Liu. Stein variational gradient descent as gradient flow. In *Advances in Neural Information Processing Systems*, volume 30, 2017.