An introduction to Stein's method

Nathan Ross (University of Melbourne)

Three lectures

1. Basics and normal approximation
2. Poisson approximation
3. Multivariate and process approximation

References:

- Ross (2011). Fundamentals of Stein's method. *Probability Surveys*.
- Barbour, Holst, Janson (1992). Poisson approximation.
- Chen, Goldstein, Shao (2011). Normal approximation by Stein's method.

An introduction to Stein's method

1. Basics and normal approximation

Nathan Ross (University of Melbourne)

Classical CLT

Assume

- $X_1, X_2, \dots$ iid,
- $\mathbb{E}X_i = 0$, $\mathrm{Var}(X_i) = 1$,
- $W = W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$.

Then, for $Z \sim \mathrm{Normal}(0, 1)$ and $x \in \mathbb{R}$,

$$\mathbb{P}(W_n \leq x) \to \mathbb{P}(Z \leq x).$$

Berry-Esseen Theorem

Assume

- $X_1, X_2, \ldots$ iid,
- $\mathbb{E}X_i = 0$, $\mathrm{Var}(X_i) = 1$, and $\mathbb{E}\left[|X_i|^3\right] < \infty$,
- $W = W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$.

Then, for $Z \sim \mathrm{Normal}(0,1)$ and $x \in \mathbb{R}$,

$$\left|\mathbb{P}(W_n \leq x) - \mathbb{P}(Z \leq x)\right| \leq \frac{\mathbb{E}\left[|X_1|^3\right]}{\sqrt{n}}.$$

Dependency Graph

We say $X_1, X_2, \ldots, X_n$ has dependency graph G with vertices $\{1, \ldots, n\}$ if

- for any two disjoint subsets $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_m\}$ of vertices, with no edges between the two sets,

$$(X_{i_1}, \ldots, X_{i_k}) \text{ is independent of } (X_{j_1}, \ldots, X_{j_m}).$$

Q: Under what conditions on G is $\sum_{i=1}^{n} X_i$ close in distribution to a normal distribution?

Sleight of Hand

For random variables W and Z, define the Wasserstein distance between their distributions by

$$d_{\mathrm{Lip}}(W, Z) = \sup_{h \in \mathrm{Lip}_1} |\mathbb{E}h(W) - \mathbb{E}h(Z)|.$$

Proposition: If Z has density bounded by C, then

$$|\mathbb{P}(W \leq x) - \mathbb{P}(Z \leq x)| \leq \sqrt{(2C)d_{\mathrm{Lip}}(W, Z)}.$$

Stein's Method Framework

Three steps to Stein's method for a given target distribution of Z.

1. Characterizing operator $\mathcal{A} = \mathcal{A}_Z$ on real valued functions:

$$\mathbb{E}\mathcal{A}f(X) = 0 \text{ wide class of functions } f \iff X \stackrel{d}{=} Z.$$

2. For given h, find Stein solution f_h:

$$\mathcal{A}f_h(x) = h(x) - \mathbb{E}h(Z) =: \tilde{h}(x).$$

3. Use structure of W and properties of f_h to bound

$$|\mathbb{E}\mathcal{A}f_h(W)| = |\mathbb{E}h(W) - \mathbb{E}h(Z)|.$$

For bound on d_{Lip}, take $h \in \mathrm{Lip}_1$.

Stein's Method for Normal Approximation

For random variable W and $Z \sim \mathrm{Normal}(0,1)$,

$$d_{\mathrm{Lip}}(W,Z) \leq \sup_{f \in \mathcal{F}} |\mathbb{E}[f'(W) - Wf(W)]|,$$

where

$$\mathcal{F} := \left\{ f : \|f\|_\infty \leq 2;\ \|f''\|_\infty \leq 2;\ \|f'\|_\infty \leq \sqrt{2/\pi} \right\}.$$

Dependency Graph Bound

Assume

- $X_1, X_2, \ldots, X_n$ has dependency graph with maximum degree bounded above by D.
- $\mathbb{E}X_i = 0$ and $\mathbb{E}\left[X_i^4\right] < \infty$.
- $W = \frac{1}{\sigma}\sum_{i=1}^n X_i$, where $\sigma^2 = \mathrm{Var}\left(\sum_{i=1}^n X_i\right)$.

Then, for $Z \sim \mathrm{Normal}(0,1)$,

$$d_{\mathrm{Lip}}(W, Z) \leq 3(D+1)^2\left(\frac{1}{\sigma^3}\sum_{i=1}^n \mathbb{E}\left[|X_i|^3\right] + \frac{1}{\sigma^2}\sqrt{\sum_{i=1}^n \mathbb{E}\left[X_i^4\right]}\right).$$

Exercises

- (Easy) Apply the previous result directly to W equal to the (centered and scaled) number of triangles in an Erdős-Rényi random graph with m vertices and edge probability p, to derive a CLT for $p = p_m \asymp m^{-\alpha}$ for $0 \leq \alpha < 2/9$.

- (Hard) In the setting of the previous exercise, derive a CLT in a greater range of α by bounding $|\mathbb{E}[f'(W) - Wf(W)]|$ directly using additional structure of W and Taylor expansion.

References and Further Reading

Basic introduction:

- Ross (2011). Fundamentals of Stein's method. *Probability Surveys*. **Sections 1, 2, 3.1 and 3.2**.

Research monograph:

- Chen, Goldstein, Shao (2011). Normal approximation by Stein's method. **Sections 1.3, 2.1, 2.2, 4.7**.