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.

A collection of exercises from Greg Terlov and Partha Dey:

- `https://sites.google.com/a/illinois.edu/gterlov`

An introduction to Stein's method

3. Multivariate and process approximation

`https://tinyurl.com/bzxe3tew`

Nathan Ross (University of Melbourne)

Classical multivariate CLT

Assume

- $X_1, X_2, \ldots$ iid random vectors of dimension d,
- $\mathbb{E}X_1 = 0$, $\mathrm{Cov}(X_{1i}, X_{1j}) = \sigma_{ij}$,
- $W = W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$.

Write $\sigma = (\sigma_{ij})_{1 \le i,j \le d}$.

Then, for $Z_\sigma \sim \mathrm{MVNormal}_d\big(0, \sigma\big)$,

$$W_n \xrightarrow[n \to \infty]{dist} Z_\sigma.$$

Rate of convergence

Assume

- $X_1, X_2, \ldots$ iid random vectors of dimension d,
- $\mathbb{E}X_1 = 0$, $\mathrm{Cov}(X_{1i}, X_{1j}) = \sigma_{ij}$,
- $W = W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$.

For simplicity, assume $\sigma = (\sigma_{ij})_{1 \le i,j \le d}$ is invertible.

For $Z_\sigma \sim \mathrm{MVNormal}_d\big(0, \sigma\big)$, and K convex,

$$|\mathbb{P}(W_n \in K) - \mathbb{P}(Z_\sigma \in K)| \le C_d \frac{\|\sigma^{-3/2}\|_2 \mathbb{E}\big[|X_1|^3\big]}{\sqrt{n}}.$$

Sleight of Hand

For random vectors W and Z of dimension d, define the smooth test function distance between their distributions by

$$d_{\mathcal{H}}(W, Z) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)|,$$

where $\mathcal{H}$ is the set of functions $\mathbb{R}^d \to \mathbb{R}$ bounded in absolute value by one, with three partial derivatives which are bounded in absolute value by one and continuous.

Proposition: If $Z_\sigma \sim \mathrm{MVNormal}_d(0, \sigma)$, then for convex K,

$$|\mathbb{P}(W \in K) - \mathbb{P}(Z_\sigma \in K)| \leq C_d \big(\|\sigma^{-3/2}\|_2 \, d_{\mathcal{H}}(W, Z_\sigma)\big)^{1/4}.$$

Stein's Method Framework

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

1. Characterising 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 $d_{\mathcal{H}}$, take $h \in \mathcal{H}$.

Stein's Method for MVNormal Approximation

For d-dimensional random vector W and $Z_\sigma \sim \mathrm{MVNormal}_d(0, \sigma)$,

$$d_{\mathcal{H}}(W, Z_\sigma) \leq \sup_{f \in \mathcal{F}} \left| \mathbb{E}\left[\sum_{i,j} \sigma_{ij} \, \partial_{ij} f(W) - \sum_i W_i \, \partial_i f(W) \right] \right|,$$

where ∂ denotes partial derivative(s), and

$$\mathcal{F} := \left\{ f : \|\partial_i f\|_\infty \leq 1; \|\partial_{ij} f\|_\infty \leq 1/2; \|\partial_{ijk} f\|_\infty \leq 1/3 \right\}.$$

Exchangeable pairs bound

Assume

- W is a d-dimensional random vector with invertible covariance matrix σ.
- (W, W') is an exchangeable pair, satisfying

$$\mathbb{E}[W' - W|W] = -\Lambda W,$$

  for some invertible matrix Λ.

Then, writing $D = W' - W$, for $Z_\sigma \sim \mathrm{MVNormal}_d(0, \sigma)$,

$$d_{\mathcal{H}}(W, Z_\sigma) \leq \left|\left|\Lambda^{-1}\right|\right|_1 \left(\sum_{i,j} \sqrt{\mathrm{Var}\left(\mathbb{E}[D_i D_j | W]\right)} + \sum_{i,j,k} \mathbb{E}\left|D_i D_j D_k\right| \right).$$

Application

$X_1, X_2, \ldots, X_n$ are i.i.d. Bernoulli(p).

$$W_1 = \frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}}, \quad W_2 = \frac{\sum_{i=1}^n (X_i X_{i+1} - p^2)}{\sqrt{np^2(1-p)}},$$

$$\sigma = \begin{pmatrix} 1 & 2\sqrt{p} \\ 2\sqrt{p} & 1+3p \end{pmatrix}.$$

Construct W' by resampling a uniformly chosen coordinate I.

$$D_1 = W_1' - W_1 = (X_I' - X_I)/\sqrt{np(1-p)},$$
$$D_2 = W_2' - W_2 = \frac{X_I' X_{I+1} + X_{I-1} X_I' - X_I X_{I+1} - X_{I-1} X_I}{\sqrt{np^2(1-p)}}.$$

$$\mathbb{E}[W' - W | W] = -\frac{1}{n} \begin{pmatrix} 1 & 0 \\ -2\sqrt{p} & 2 \end{pmatrix} W =: -\Lambda W.$$

Stein's method for processes

Z is a process, e.g., Brownian motion, or Poisson point process.

- The characterizing operator $\mathcal{A}$ is the generator of a Markov process $(X_t)_{t\geq 0}$ with semigroup $(P_t)_{t\geq 0}$:

$$P_t f(x) = \mathbb{E}[f(X_t)|X_0 = x],$$

  and stationary distribution Z.
- The distribution of Z is characterized by

$$\mathbb{E}\mathcal{A}f(Z) = 0.$$

- Dynkin-type formula gives the Stein solution

$$f_h(x) = -\int_0^\infty P_t \tilde{h}(x) dt.$$

- Bounds on derivatives of f_h follow by couplings or direct computations.

Some notes

- Refinements of this method have obtained some of the best known dependence on the dimension of the constant in the CLT setting of sums of iid random variables.
- Can also develop local dependence and size-biasing for multivariate normal approximation (Chen, Goldstein, Shao 2011, Chapter 12).

Exercises

- Apply the exchangeable pairs theorem to obtain a CLT with rate of convergence for the joint number of k runs for $k = 1, 2, \ldots, d$. See (Reinert and Röllin 2009).
- Apply the exchangeable pairs theorem to obtain a CLT with rate of convergence for the joint number of edges, two-stars, and triangles in an Erdős-Rényi random graph. See (Reinert and Röllin 2010).

References and further reading

Multivariate normal approximation:

- Chen, Goldstein, Shao (2011). Normal approximation by Stein's method. **Chapter 12**