Stein couplings for normal approximation

Adrian Röllin

National University of Singapore

June, 2010

Outline

A Stein identity

We have the following fact for a r.v. Z:

$$Z \sim \mathrm{N}(0,1)$$

$$\Updownarrow$$

$$\mathbb{E} Z f(Z) = \mathbb{E} f'(Z) \qquad \forall f$$

What can we do with this identity?

$$\mathbb{E}Zf(Z) = \mathbb{E}f'(Z)$$

f	Stein identity
$f(x) = 1$	$\mathbb{E}Z = 0$
$f(x) = x$	$\mathbb{E}Z^2 = 1$
$\Longrightarrow$ *all moments recursively*	
$f(x) = e^{\lambda x}$	$m'(\lambda) = \lambda m(\lambda)$
$\Longrightarrow \mathbb{P}[Z > x] \leqslant e^{-x^2/2}$	
$f(x) = \int_0^x g'(t)^2 dt$	$\mathbb{E}Z \int_0^Z g'(t)^2 dt = \mathbb{E}g'(Z)^2$
$\Longrightarrow \operatorname{Var} g(Z) \leqslant \mathbb{E}g'(Z)^2$	

The basic idea

- Assume $\mathbb{E}W = 0$ and $\operatorname{Var} W = 1$ for some r.v. W.
- If
$$\mathbb{E}Wf(W) \approx \mathbb{E}f'(W) \qquad \forall f$$
then we expect that W has similar properties as $\mathrm{N}(0,1)$
- In addition deduce closeness of W to Z (*Stein's method*).

An formal identity

- Assume existence of two r.v.s. T_1 and T_2 such that

$$\mathbb{E}Wf(W) = \mathbb{E}T_1 f'(W + T_2). \qquad (*)$$

- Ideally we should have

$$T_1 \approx 1, \qquad T_2 \approx 0$$

- If $(*)$ is true with $T_1 = 1$ we call this *zero-biasing*.
- If W is a functional of a Gaussian field, use Malliavin calculus to obtain $(*)$ with $T_2 = 0$.
- Our approach makes use of both, T_1 and T_2.

Outline

Auxiliary randomisation

- Let W' be a perturbation of W, let $D := W' - W$.
- Then

$$Gf(W') - Gf(W) = GD\int_0^1 f'(W + sD)ds$$

where G is any random variable.

- Hence, with $U \sim U(0,1)$,

$$\mathbb{E}\big\{Gf(W') - Gf(W)\big\} = \mathbb{E}\big\{GDf'(W + UD)\big\}$$

$$= \mathbb{E}GDf'(W) + \mathrm{O}\big(\|f''\|\mathbb{E}|GD^2|\big)$$

$$= \mathbb{E}f'(W) + \mathrm{O}\Big(\|f'\|\mathbb{E}\big|\mathbb{E}^W GD - 1\big| + \|f''\|\mathbb{E}|GD^2|\Big)$$

Stein coupling

Definition

We call a triple (W, W', G) of square integrable r.v.s a **Stein coupling** if

$$\mathbb{E}\{Gf(W') - Gf(W)\} = \mathbb{E}Wf(W) \qquad \forall f$$

- Note that the above identity implies

$$\mathbb{E}W = 0, \qquad \operatorname{Var} W = \mathbb{E}GD$$

- We hence have

$$\mathbb{E}Wf(W) = \mathbb{E}f'(W) + \mathrm{O}\Big(\|f'\|\mathbb{E}\big|\mathbb{E}^W GD - 1\big| + \|f''\|\mathbb{E}|GD^2|\Big)$$

An abstract theorem

Theorem

Let (W, W', G) *be a Stein coupling with* $\operatorname{Var} W = 1$. *Then*

$$\int_{-\infty}^{\infty} \left|\mathbb{P}[W \leqslant x] - \Phi(x)\right| dx \leqslant 0.8\sqrt{\operatorname{Var} \mathbb{E}^{W}(GD)} + \mathbb{E}|GD^2|$$

$$\sup_x \left|\mathbb{P}[W \leqslant x] - \Phi(x)\right| \leqslant 2\sqrt{\operatorname{Var} \mathbb{E}^{W}(GD)} + 8\|G\|\|D\|^2$$

Outline

Independent summands

- Let $W = \sum_{i=1}^{n} X_i$, where $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 < \infty$.
- Let I be independent and uniform on $\{1, \ldots, n\}$.
- Take

$$W' = W - X_I, \qquad G = -nX_I.$$

- Then,

$$\begin{aligned}\mathbb{E}\big\{Gf(W') - Gf(W)\big\} &= -\sum_{i=1}^{n} \mathbb{E}\big\{X_i f(W - X_i) - X_i f(W)\big\} \\ &= \sum_{i=1}^{n} \mathbb{E}\big\{X_i f(W)\big\} = \mathbb{E}Wf(W).\end{aligned}$$

- The following applications are of similar flavour.

Linear functionals in the classic urn model

- Distribute m balls uniformly and independently into n urns:
$$\xi_i = \#\text{balls in urn } i.$$
- Let
$$W = \sum_{i=1}^{n} h(\xi_i)$$
for some arbitrary function h with $\mathbb{E}h(\xi_1) = 0$.
- Examples are
$$\begin{aligned} h(x) &= I[x = 0] && \#\text{empty urns} \\ h(x) &= I[x = m_0] && \#\text{urns with exactly } m_0 \text{ balls} \\ h(x) &= I[x \leqslant m_0] && \#\text{urns not exceeding } m_0 \\ h(x) &= (x - m_0)I[x > m_0] && \#\text{excess balls} \end{aligned}$$

A Stein coupling

- Pick a random urn I uniformly among the urns.
- Remove the content of that urn and distribute into the other urns.
- Let

$$\xi_j^I = \#\text{balls in urn } j \text{ after re-distributing urn } I$$

and

$$W' = \sum_{j \neq I} h(\xi_j^I)$$

- Then, with $G = -nh(\xi_I)$, we obtain a Stein coupling:

$$-\mathbb{E}Gf(W) = \sum_{i=1}^{n} \mathbb{E}h(\xi_i)f(W) = \mathbb{E}Wf(W)$$

$$\mathbb{E}Gf(W') = -\sum_{i=1}^{n} \mathbb{E}^{I=i}h(\xi_i)f(W') = 0$$

Final result

Theorem

With $\sigma^2 = \operatorname{Var} W$ and if $m/n \asymp 1$, we have

$$\sup_x \left| \mathbb{P}[W/\sigma \leqslant x] - \Phi(x) \right| \leqslant C \left(\frac{n\|\Delta h\|^3 \ln(n\|h\|)^6}{\sigma^3} + \frac{\ln(n\|h\|)}{\sigma^2} \right).$$

- If $m/n \to \lambda > 0$ and h is fixed with $\|\Delta h\| < \infty$ but not affine, then $\operatorname{Var} W \asymp n$ and $\sup |\cdot| = \mathrm{O}(\log^6(n)/\sqrt{n})$.

Remark: localisation

- The Stein coupling "localizes" the dependence:

$$\operatorname{Var} W = n\mathbb{E}\left(h(\xi_I)^2 + h(\xi_I) \sum_{j \neq I} h(\xi_j) \right)$$

$$\|$$

$$\mathbb{E} GD = n\mathbb{E}\left(h(\xi_I)^2 + h(\xi_I) \sum_{j \text{ affected}} \big(h(\xi_j) - h(\xi_j^I) \big) \right)$$

Components in subcritical Erdős-Rényi random graph

- Let H be an Erdős-Rényi random graph with n vertices and edge probability $p = \lambda/n$ with fixed $\lambda < 1$.
- We are interested in

$$W = \sum_{i,j \in V(H)} h(H, i, j)$$

where $h(H, i, j)$ is a function of the graph and any two of its vertices such that

(*i*) $h(H, i, j) = h(H, j, i)$,
(*ii*) $h(H, i, j)$ only depends on the components of i and j,
(*iii*) $h(H, i, j) = 0$ if i and j are note in the same component.

- E.g. $h(H, i, j) = \mathrm{I}[i \text{ and } j \text{ are in the same component}]$,
$h(H, i, j) = \mathrm{I}[i \text{ and } j \text{ are in the same cycle}]$
$h(H, i) = 1/\text{size of component of } i$.

A Stein coupling

- Write $C(i)$ for the component containing i and

$$W = \sum_i \sum_{j \in C(i)} h(H, i, j) =: \sum_i X_i^H$$

- Without loss of generality we may assume $\mathbb{E}X_i^H = 0$.
- Let I be independent and uniform on all the vertices.
- Define the new graph H' by resampling the edges of all the pairs containing at least one of its endpoints in $C(I)$ (i.e. 'resample' the component $C(I)$).
- Let

$$W' = \sum_i \sum_{j \in C(i)} h(H', i, j) = \sum_i X_i^{H'}$$

- Note that X_I^H is independent of W'.
- Hence, $G = -nX_I$ makes the triple a Stein coupling.

Final result

Theorem

With $\sigma^2 = \mathrm{Var}\, W$, *we have*

$$\sup_x \Big| \mathbb{P}[W/\sigma \leqslant x] - \Phi(x) \Big| \leqslant C \left(\frac{n \log^{11}(n\|h\|)\|\Delta h\|}{\sigma^3} + \frac{\ln(n\|h\|)^2}{\sigma^2} \right).$$

- If $\sigma^2 \asymp n$, then we obtain a rate of $\log(n)^{11}/\sqrt{n}$.
- Note that $\mathscr{L}(H^I) = \mathscr{L}(H)$, but they are not exchangeable.

General Stein couplings

- *There is always a Stein coupling*. Let W be a r.v. with $\mathbb{E}W = 0$. Then

$$(W, 0, -W) \text{ is a Stein coupling.}$$

- *Given a coupling* (W, W'), *there is always a (formal) G:*

$$G = -W + \mathbb{E}(W|W') - \mathbb{E}\big(\mathbb{E}(W|W')|W\big) + \mathbb{E}\Big(\mathbb{E}\big(\mathbb{E}(W|W')|W\big)|W'\Big) - \dots + \dots$$

Indeed, if the series converges almost surely,

$$\mathbb{E}\big\{Gf(W') - Gf(W)\big\} = \mathbb{E}Wf(W).$$

Conclusions

- Once a Stein coupling is found, normal approximation becomes an exercise in bounding some moments (may not be easy).
- The advantage of the couplings presented in this talk is that they work for general linear functionals.
- From a Stein coupling, other properties of W may also be obtained (variance, moderate and large deviations, concentration of measure, Poincaré-type inequalities. . .).
- *Reference: L. H. Y. Chen and A. Röllin (2010, arXiv) "Stein couplings for normal approximation"*