Stein's method and Continuous symmetries

Elizabeth Meckes
AIM / Cornell

March 29, 2007

Stein's method of exchangeable pairs for normal approximation ~ Classical version

Lemma (Stein): Let $Z \sim N(0,1)$. Then

1. For all $f \in C_c^1(\mathbb{R})$,

$$\mathbb{E}[f'(Z) - Zf(Z)] = 0.$$

2. If Y is a random variable s.t.

$$\mathbb{E}[f'(Y) - Yf(Y)] = 0$$

for all $f \in C_b^1(\mathbb{R})$, then $Y \sim N(0,1)$.

3. For $g \in C_c(\mathbb{R})$, the function

$$U_0 g(t) := e^{t^2/2} \int_{-\infty}^{t} [g(x) - \mathbb{E}g(Z)] e^{-x^2/2} \, dx$$

is a solution to the differential equation

$$f'(x) - xf(x) = g(x) - \mathbb{E}g(Z).$$

<u>Lemma 2</u> (Stein): Let $g:\mathbb{R}\to\mathbb{R}$ be a given bounded continuous function, and let
$$U_0 g(t) = e^{t^2/2}\int_{-\infty}^{t}[g(x)-\mathbb{E}g(Z)]e^{-x^2/2}\,dx.$$

Then

1. $\|U_0 g\|_\infty \le \sqrt{2\pi}\,\|g\|_\infty$
2. $\|(U_0 g)'\|_\infty \le 4\|g\|_\infty$
3. $\|(U_0 g)''\|_\infty \le 2\|g'\|_\infty.$

These two lemmas are the main ingredients for:

<u>Abstract normal approximation theorem</u> (Stein): Let (W,W') be an exchangeable pair, and suppose $\mathbb{E}W=0$, $\mathbb{E}W^2=1$, and there is a $\lambda>0$ s.t.
$$\mathbb{E}[W'-W\mid W] = -\lambda W.$$
Then
$$|\mathbb{E}g(W)-\mathbb{E}g(Z)| \le \frac{2}{\lambda}\|g\|_\infty\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2\mid W])} + \frac{\|g'\|_\infty}{2\lambda}\mathbb{E}|W'-W|^3.$$

This theorem is useful in many situations posessing some type of symmetry.

Recall: $W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$. For each i, let X_i' be an independent copy of X_i. Pick $I \in \{1, \ldots, n\}$ at random, independent of everything. Define

$$W' = W - \frac{1}{\sqrt{n}} X_I + \frac{1}{\sqrt{n}} X_I'.$$

Or: say W is a permutation statistic depending on a random permutation π. Make an exchangeable pair by picking a random transposition τ independent of π: define

$$\pi' = \tau \circ \pi$$

$$W' = W(\pi')$$

In each of these examples, the symmetries used are discrete.

Consider the following example:

Let X be a random point on the $S^{n-1} \subseteq \mathbb{R}^n$ (i.e., distributed according to surface area measure.)

Define $W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$. Then $W \approx N(0,1)$.

Note that by the given normalization, $\mathbb{E} X_i = 0$, $\mathbb{E} X_i^2 = 1$. Here the summands are not independent, since we have $\sum X_i^2 = n$. But the sphere has a big continuous group of symmetries; this should help.

The idea for how to do this is based on a result of Stein (1995).

<u>Theorem</u> (M): Suppose that for each $\varepsilon \in (0, \varepsilon_0)$, (W, W_ε) is an exchangeable pair s.t.

1. $\frac{1}{\varepsilon^2} \mathbb{E}[W_\varepsilon - W \mid W] = -\lambda W + o(\varepsilon)\alpha(W)$

2. $\frac{1}{\varepsilon^2} \mathbb{E}[(W_\varepsilon - W)^2 \mid W] = 2\lambda\sigma^2 + E\sigma^2 + o(\varepsilon)\beta(W)$

3. $\frac{1}{\varepsilon^2} \mathbb{E}|W_\varepsilon - W|^3 = o(1)$,

where $\lambda > 0$, $\sigma > 0$, and α, β are measurable functions s.t. $\mathbb{E}|\alpha(\sigma^{-1}W)| < \infty$, $\mathbb{E}|\beta(\sigma^{-1}W)| < \infty$, and E is a random variable. Then

$$d_{T.V.}(W, Z) \le \frac{1}{\lambda} \mathbb{E}|E|$$

where $Z \sim N(0, \sigma^2)$

Example:

Let $X \sim \text{unif}(\sqrt{n}\, S^{n-1})$.

Consider $W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$.

I $W \mathrel{\dot\sim} N(0,1)$ via the classical method

Pick $I \in \{1, \dots, n\}$ at random, independent of X. Define

$$X' = (X_1, \dots X_{I-1}, -X_I, X_{I+1}, \dots, X_n);$$

$$W' = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i' = W - \frac{2}{\sqrt{n}} X_I .$$

$$\begin{aligned}
\mathbb{E}[W' - W \mid W] &= -\frac{2}{\sqrt{n}} \mathbb{E}[X_I \mid W] \\
&= -\frac{2}{n^{3/2}} \mathbb{E}\Big[\sum_{i=1}^{n} X_i \mid W\Big] \\
&= -\frac{2}{n} W \Rightarrow \lambda = \frac{2}{n}
\end{aligned}$$

Now we need

- $\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2 \mid W])}$
- $\mathbb{E}|W'-W|^3$.

$$(W'-W)^2 = \frac{4}{n} X_I^2$$

$$\Rightarrow \mathbb{E}[(W'-W)^2 \mid W] = \frac{4}{n^2}\sum_{i=1}^{n} \mathbb{E}[X_i^2 \mid W]$$

$$= \frac{4}{n^2}\mathbb{E}\Big[\underbrace{\sum_{i=1}^{n} X_i^2}_{\equiv n} \mid W\Big]$$

$$= \frac{4}{n}$$

$$\Rightarrow \mathrm{Var}(\mathbb{E}[(W'-W)^2 \mid W]) = 0.$$

Next, $\mathbb{E}|W'-W|^3 = \frac{8}{n^{3/2}}\mathbb{E}|X_I|^3$

$$= \frac{8}{n^{3/2}}\mathbb{E}|X_1|^3$$

$$\mathbb{E}|X_1|^3 \le (\mathbb{E}|X_1|^4)^{3/4} = \left(\frac{3n}{n+2}\right)^{3/4} \le 3^{3/4}.$$

Stein's theorem says:

$$|\mathbb{E}g(W) - \mathbb{E}g(Z)| \le \frac{1}{2\lambda}\left(\mathbb{E}|W'-W|^3\right)\|g'\|_\infty$$

Now, $\lambda = \frac{2}{n}$, $\mathbb{E}|W'-W|^3 \le \frac{8 \cdot 3^{3/4}}{n^{3/2}}$

$$\Rightarrow |\mathbb{E}g(W) - \mathbb{E}g(Z)| \le \frac{\|g'\|_\infty}{\sqrt{n}}\left(2 \cdot 3^{3/4}\right);$$

i.e., $d_W(W, Z) \le \frac{2 \cdot 3^{3/4}}{\sqrt{n}}$

Wasserstein distance

Good news: we proved a theorem.

Bad news: this isn't the right rate.

II $W \approx N(0,1)$ using the full symmetries of S^{n-1}.

Let
$$R_\varepsilon = \begin{bmatrix} \sqrt{1-\varepsilon^2} & \varepsilon \\ -\varepsilon & \sqrt{1-\varepsilon^2} \end{bmatrix} \oplus I_{n-2}.$$

Pick $U \in \mathcal{O}_n$ at random (distributed according to Haar measure), independent of X. Define

$$X_\varepsilon = U R_\varepsilon U^T X;$$

X_ε is a rotation of X by $\arcsin(\varepsilon)$ in a random 2-dim. subspace of $\mathbb{R}^n$.

To apply the theorem, need:

- $\mathbb{E}[W_\varepsilon - W \mid W]$
- $\mathbb{E}[(W_\varepsilon - W)^2 \mid W]$
- $\mathbb{E}|W_\varepsilon - W|^3$

For convenience, consider

$$W = X_1 \qquad W_\varepsilon = (X_\varepsilon)_1$$

Now, $$X_\varepsilon - X = [U R_\varepsilon U^T - I] X$$

Let $$K = \begin{bmatrix} u_{11} & u_{12} \\ \vdots & \vdots \\ u_{n1} & u_{n2} \end{bmatrix} \qquad C_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then $$U R_\varepsilon U^T - I = U\left(\begin{bmatrix} \sqrt{1-\varepsilon^2} & \varepsilon \\ -\varepsilon & \sqrt{1-\varepsilon^2} \end{bmatrix} \oplus I_{n-2}\right) U^T - I$$

$$= U\left(\begin{bmatrix} \sqrt{1-\varepsilon^2} - 1 & \varepsilon \\ -\varepsilon & \sqrt{1-\varepsilon^2} - 1 \end{bmatrix} \oplus O_{n-2}\right) U^T$$

$$= \left[-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right] K K^T + \varepsilon K C_2 K^T$$

$$\Rightarrow X_\varepsilon - X = \left[-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right] K K^T X + \varepsilon K C_2 K^T X$$

$$\Rightarrow (X_\varepsilon)_1 - X_1 = \left[-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right] (K K^T X)_1 + \varepsilon (K C_2 K^T X)_1$$

Now,

$$(KK^T)_{ij} = u_{i1}u_{j1} + u_{i2}u_{j2} \Rightarrow \mathbb{E}(KK^T)_{ij} = \frac{2}{n}\delta_{ij}$$

$$(KC_2K^T)_{ij} = u_{i1}u_{j2} - u_{i2}u_{j1}$$

$$\Rightarrow \quad \mathbb{E}[(KC_2K^T)_{ij}] = 0$$

Thus

$$\mathbb{E}[(X_\varepsilon)_1 - (X)_1 \mid X_1] = \left(-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right) \cdot \frac{2}{n} X_1$$

$$\boxed{\lambda = \frac{1}{n}}$$

It's not much more work to show:

$$\mathbb{E}[((X_\varepsilon)_1 - X_1)^2 \mid X_1]$$

$$= \frac{2\varepsilon^2}{n-1} - \frac{2\varepsilon^2}{n(n-1)}\left[X_1^2 + O(\varepsilon)\right].$$

$$\boxed{E = \frac{2}{n(n-1)}(1 - X_1^2)}$$

Finally, from the expression

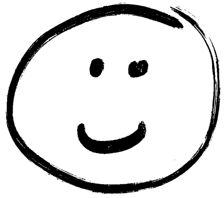
$$X_\varepsilon - X = \left(-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right) KK^T X + \varepsilon K C_2 K^T X$$

makes it clear that

$$\mathbb{E}\,|(X_\varepsilon)_1 - X_1|^3 = O(\varepsilon^3).$$

Plugging into the theorem (and bounding $\mathbb{E}|X_1^2 - 1| \le \mathbb{E}X_1^4 - 1 \le 2$) gives

$$d_{T.V.}(X_1, Z) \le \frac{4}{n-1}.$$



Taking advantage of the continuous symmetries gave a better rate in a stronger metric (and this is the right answer - due to Diaconis / Freedman.)

Application to Random Matrices

Theorem (M): Let $M \in \mathbb{O}_n$ be a Haar-distributed random orthogonal matrix, and let $A \in M_n(\mathbb{R})$ be a fixed $n \times n$ matrix. Assume that $tr(AA^t) = n$, and define

$$W := tr(AM)$$

Then

$$d_{T.V.}(W, Z) \leq \frac{2\sqrt{3}}{n-1}$$

for $Z \sim N(0,1)$.

Intuition: Haar matrices are a lot like Gaussian matrices. W is a rank 1 projection of Haar measure on $\mathbb{O}_n$ and is approximately Gaussian. All lower-dim. projections of Gauss measure on the space of $n \times n$ matrices are exactly Gaussian.

The idea of the proof is very close to the sphere example:

$M \sim \text{Haar}(\mathcal{O}_n)$; $M_\varepsilon := U R_\varepsilon U^T M$

for $U \sim \text{Haar}(\mathcal{O}_n)$, independent of M

and $R_\varepsilon = \begin{bmatrix} \sqrt{1-\varepsilon^2} & \varepsilon \\ -\varepsilon & \sqrt{1-\varepsilon^2} \end{bmatrix} \oplus I_{n-2}$

as before.

There is an analagous result for $W = \mathrm{Re}(\mathrm{tr}(AM))$ for $M \sim \text{Haar}(\mathcal{U}_n)$, $A \in M_n(\mathbb{C})$.

To consider $\mathrm{tr}(AM)$ itself requires multivariate tools.

Application: Eigenfunctions of Δ

Theorem (M): Let (M, g) be a compact manifold without boundary, and let f be an eigenfunction of Δ_g:

$$\Delta_g f = -\lambda f \qquad \lambda > 0$$

Assume that $\int_M f^2 \, d\overline{vol} = 1$

(normalized volume – $\int_M 1 \, d\overline{vol} = 1$)

Then if X is a random point of M,

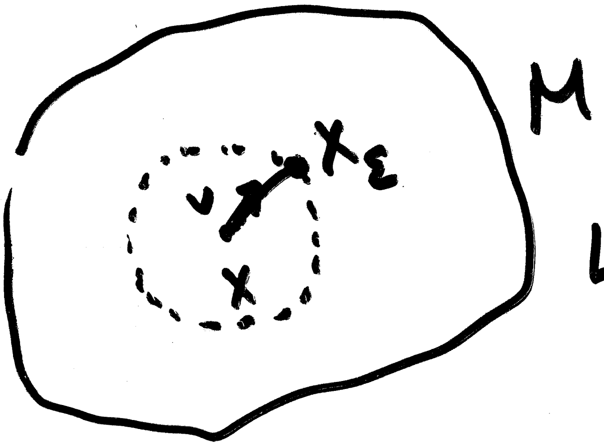
$$d_{TV}(f(X), Z) \le \frac{1}{\lambda} \sqrt{\mathrm{Var}(\|\nabla f\|^2)}.$$

Exchangeable pair:

Fix $\varepsilon > 0$ smaller than the injectivity radius of the exponential map at each point.

Let X be a random point of M.

Pick $v \in S_X M$ (unit vectors at X),



s.t. v is distributed uniformly on $S_X M$.

Let

$$X_\varepsilon = \exp_X(\varepsilon v).$$

Then (X, X_ε) is an exchangeable pair.

So if $W = f(X)$, $W_\varepsilon = f(X_\varepsilon)$, then (W, W_ε) is exchangeable as well.

Why eigenfunctions of Δ_g?

Recall that to apply the theorem, it is necessary that

$$\mathbb{E}[W_\varepsilon - W \mid W] \approx -\varepsilon^2 \lambda W.$$

$$W_\varepsilon - W = f(X_\varepsilon) - f(X)$$

<u>Claim:</u>

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \mathbb{E}[f(X_\varepsilon) - f(X) \mid X]$$
$$= \frac{1}{2n} \Delta_g f(X).$$

So for this to be proportional to $W = f(X)$, need

$$\Delta_g f(X) = -\lambda f(X).$$

Degree ℓ spherical harmonics

(ℓ odd)

Let $C_\ell^{\frac{n-2}{2}}(t)$ denote the usual Gegenbauer polynomial of degree ℓ (if ℓ is odd, all powers of t appearing are odd)

$C_\ell^{\frac{n-2}{2}}(x_k)$ is an eigenfunction of $\Delta_{S^{n-1}}$ for each $k \in \{1, \dots, n\}$, with eigenvalue $-\ell(n+\ell-2)$. For ℓ odd

$$\mathbb{E}\left[C_\ell^{\frac{n-2}{2}}(x_k)\, C_\ell^{\frac{n-2}{2}}(x_j)\right] = 0, \quad k \neq j$$

Pick A s.t. $\mathbb{E}\left[A^2 \left(C_\ell^{n-2/2}(x_k)\right)^2\right] = 1 \quad \forall k.$

Let $\{a_k\}$ s.t. $\sum a_k^2 = 1$. Then the theorem applies to $p(x) = \sum_{i=1}^n a_i \cdot A \cdot C_\ell^{\frac{n-2}{2}}(x_i)$ and if $X \sim \text{unif}(S^{n-1})$, $\exists$ an absolute constant c s.t.

$$d_{TV}(p(X), Z) \leq c\, \|a\|_4^2$$

A multi-variate extension

The following well-known result gives a characterization of multi-variate Gaussian.

Lemma: Let $Z \in \mathbb{R}^k$ with $\{Z_i\}$ i.i.d. standard Gaussians.

1. If $f \in C_c^2(\mathbb{R}^k)$,

$$\mathbb{E}\left[\Delta f(Z) - Z \cdot \nabla f(Z)\right] = 0.$$

2. If Y is a random vector with

$$\mathbb{E}\left[\Delta f(Y) - Y \cdot \nabla f(Y)\right] = 0$$

for all $f \in C_b^2(\mathbb{R}^k)$, then $Y \sim Z$.

3. If $g \in C_c^\infty(\mathbb{R}^k)$, the function

$$U_0 g(x) = \int_0^1 \frac{1}{2t}\left[\mathbb{E}\, g(\sqrt{t}\, x + \sqrt{1-t}\, Z) - \mathbb{E} g(Z)\right] dt$$

is a solution to the differential equation

$$\Delta h(x) - x \cdot \nabla h(x) = g(x) - \mathbb{E} g(Z)$$

Recall that the main ingredients needed for the univariate theorem were

1. Make exchangeable pair.
2. Use pair to approximate derivative by difference.
3. Bound $U_0 g$, its derivatives in terms of g and its derivatives.

Here, step 3 doesn't work as easily, since

$$f \longmapsto \Delta f(x) - x \cdot \nabla f(x)$$

is a second order operator

$\rightsquigarrow$ more smoothness of test functions needed $\rightsquigarrow$ weaker notion of distance

The same ideas as in the univariate case lead to the following.

Theorem (S. Chatterjee, E.M.):

Let X, X_ε be two random vectors with $\mathcal{L}(X) = \mathcal{L}(X_\varepsilon)$ s.t. $\lim_{\varepsilon \to 0} X_\varepsilon = X$ almost surely.

Suppose there is a constant λ, functions $E_{ij}(X)$, $\alpha(X)$, $\beta(X)$ with $\mathbb{E}|\alpha(X)|$, $\mathbb{E}|\beta(X)| < \infty$, such that

1. $\mathbb{E}[(X_\varepsilon - X)_i \mid X] = -\lambda \varepsilon^2 X_i + o(\varepsilon^2)\alpha(X)$

2. $\mathbb{E}[(X_\varepsilon - X)_i (X_\varepsilon - X)_j \mid X] = 2\lambda\varepsilon^2 \delta_{ij} + \varepsilon^2 E_{ij} + o(\varepsilon^2)\beta(X)$

3. $\mathbb{E}|X_\varepsilon - X|^3 = o(\varepsilon^2)$.

Then

$$d_{L^*}(X, Z) \le \min\left\{ \frac{1}{2\lambda} \sum_{i,j} \mathbb{E}|E_{ij}|, \ \frac{\sqrt{k}}{2\lambda} \mathbb{E}\left(\sum_{i,j} E_{ij}^2 \right)^{\frac{1}{2}} \right\}.$$

If, in addition, $E_{ij} = E_i F_j + \delta_{ij} R_i$, then

$$d_{L^*}(X, Z) \le \frac{1}{\lambda\sqrt{2\pi}} \mathbb{E}\left[\|\vec{E}\|_2 \|\vec{F}\|_2 \right] + \frac{1}{\lambda} \sum_{i=1}^{k} \mathbb{E}|R_i| .$$

Application: Rank k projections of spherically symmetric distributions on $\mathbb{R}^n$.

<u>Theorem</u> (M): Let X be a random vector in $\mathbb{R}^n$ whose distribution is spherically symmetric, with $\mathbb{E}X_i = 0 \ \forall i$, $\mathbb{E}X_iX_j = \delta_{ij}$. Suppose there is a constant a s.t.

$$\mathrm{Var}(|X|^2) \leq a$$

Let $P_k : \mathbb{R}^n \to \mathbb{R}^k$, $P_k(x_1, \ldots, x_n) = (x_1, \ldots, x_k)$, and let $Y = P_k(X)$. Then

$$d_{L^*}(Y, Z) \leq \frac{ck}{n}$$

for a constant c depending only on a, and $Z = \{Z_i\}_{i=1}^k$ with Z_i i.i.d. standard Gaussians.

Rank k projections of Haar measure

Theorem (M): Let $B_1, \ldots, B_k$ be linearly independent $n \times n$ matrices over $\mathbb{R}$; assume $\mathrm{tr}(B_i B_i^t) = n$ for each i. Let $b_{ij} = \mathrm{tr}(B_i B_j^t)$. Let M be a random orthogonal matrix and

$$W = (\mathrm{tr}(B_1 M), \mathrm{tr}(B_2 M), \ldots, \mathrm{tr}(B_k M)).$$

Let $Y = (Y_1, \ldots, Y_k)$ be a vector of standard Gaussian r.v.'s with covariance matrix $\frac{1}{n}(b_{ij})$. Then

$$d_{L^*}(W, Y) \le \frac{2\sqrt{\lambda}\, k^{3/2}}{n-1}$$

where λ is the largest eigenvalue of $\frac{1}{n}(b_{ij})$.

Complex-linear functions on $\mathcal{U}_n$

Theorem (M): Let $M \in \mathcal{U}_n$ be distributed according to Haar measure and let $A \in M_n(\mathbb{C})$ be such that $\operatorname{tr}(AA^*) = n$. Let

$$W = \operatorname{tr}(AM),$$

and Z a standard complex Gaussian. Then

$$d_{L^*}(W, Z) \leq \frac{c}{n}$$

for a constant c which is asymptotically $3\sqrt{2}$ (c can be taken to be 6 if $n \geq 6$.)