A Gentle Introduction to Stein's Method for Normal Approximation III

Larry Goldstein

University of Southern California

III. Exchangeable Pair, Zero Bias Couplings

IV. Local dependence, Nonsmooth functions

III. Stein Exchangeable Pair

   (a) Linearity Condition
   (b) Mean, Variance, Function Identities
   (c) Smooth function bound

III. Zero Bias Couplings

   (a) Fixed point zero bias rationale for CLT
   (b) Change 1 property
   (c) L^1 bound
   (d) Connection to the Stein pair, K function

Stein Equation

Let W satisfy $EW = 0, \text{Var}(W) = 1$. Recall, for given $h \in \mathcal{H}$,

$$f'(w) - wf(w) = h(w) - Nh \quad \text{where} \quad Nh = Eh(Z).$$

For the given W, we calculate

$$Eh(W) - Nh$$

by computing

$$E[f'(W) - Wf(W)].$$

Stein Exchangeable Pair

[Stein, (1986)]

We say the random variables (W, W') form a λ-Stein pair if (W, W') is exchangeable and satisfy the 'linearity' or 'linear regression' condition

$$E(W'|W) = (1-\lambda)W \quad \text{for some } \lambda \in (0,1).$$

Linearity Condition: Bivariate Normal Connection

Parallel to a property of bivariate normal variables Z_1, Z_2: conditional expectation of Z_1 given Z_2 is linear

$$E(Z_1|Z_2) = \mu_1 + \sigma_1 \rho \left(\frac{Z_2 - \mu_2}{\sigma_2} \right).$$

When Z_1 and Z_2 have mean zero and equal variance,

$$E(Z_1|Z_2) = (1 - \lambda) Z_2 \quad \text{for } \lambda = 1 - \rho.$$

Linearity Condition: Generator Connection

$E(W'|W) = (1 - \lambda)W$ or $E(W' - W|W) = -\lambda W.$

Embed in a sequence, $E(W_{t+1} - W_t|W_t) = -\lambda W_t$,

$$\Delta W_t = -\lambda W_t + \epsilon_t \quad \text{where } E[\epsilon_t|W_t] = 0.$$

Reminiscent of Ornstein Uhlenbeck process

$$dW_t = -\lambda W_t + \sigma dB_t.$$

Linearity Condition: Reversible Markov Chain Connection

If $W_1, W_2, \ldots$ is a reversible Markov Chain in stationarity, then (W_t, W_{t+1}) is exchangeable.

To apply the method for a given distribution W, construct a reversible Markov chain with stationary distribution W.

Anti-Voter Model

On the graph $(\mathcal{V}, \mathcal{E})$, with $|\mathcal{V}| = n$, consider the evolution of the state $\mathbf{X}_t \in \{-1, 1\}^n$ where at each time step a vertex chosen uniformly at random chooses a neighbor at random and adopts the opposite state.

Though $\mathbf{X}_t$ is not reversible, if stationary and the function W satisfies $W(\mathbf{X}_{t+1}) - W(\mathbf{X}_t) \in \{-1, 0, 1\}$ then $(W(\mathbf{X}_t), W(\mathbf{X}_{t+1}))$ is exchangeable.

[Liggett (1985), Rinott and Rotar (1997)]

Anti-Voter Model: Linearity

Let T denote the number of vertices i with $X_i = 1$, and let $U = 2T - n$. Further, let a, b and c be the number of edges whose vertices agree with $1, -1$, or disagree, respectively.

Observe that for a regular graph of degree r

$$T = [2a + c]/r, \qquad n - T = [2b + c]/r.$$

$$P(U' - U = -2 \,|\, \mathbf{X}) = \frac{2a}{rn}, \;\; P(U' - U = 2 \,|\, \mathbf{X}) = \frac{2b}{rn}.$$

Therefore, using $a + b + c = rn/2$,

$$E[(U' - U) \,|\, \mathbf{X}] = \frac{4b - 4a}{rn} = \frac{2(n - 2T)}{n} = -\frac{2U}{n}.$$

Stein Exchangeable Pair: Mean

When expectations exist they must equal zero, as

$$EW = EW' = E(E(W'|W)) = E(1-\lambda)W = (1-\lambda)EW.$$

As $1 - \lambda \neq 0$,

$$EW = 0.$$

Stein Exchangeable Pair: Variance Identity

$$\begin{aligned} E[W'W] &= E\left(E(W'W|W)\right) \\ &= E\left(WE(W'|W)\right) \\ &= (1-\lambda)E\left(W^2\right) \\ &= (1-\lambda)\sigma^2 \end{aligned}$$

gives

$$\begin{aligned} E(W'-W)^2 &= 2(EW^2 - EW'W) \\ &= 2(\sigma^2 - (1-\lambda)\sigma^2) \\ &= 2\lambda\sigma^2. \end{aligned}$$

Stein Exchangeable Pair: Function Identity

Linearity condition gives

$$E[W'f(W)] = E[f(W)E(W'|W)] = (1-\lambda)E[Wf(W)],$$

so

$$\begin{aligned} E(W'-W)(f(W')-f(W)) &= 2E(Wf(W)-W'f(W)) \\ &= 2\lambda E[Wf(W)] \end{aligned}$$

or

$$E[Wf(W)] = \frac{E(W'-W)(f(W')-f(W))}{2\lambda}.$$

Exchangeable Pair and the Stein Equation

If W, W' is Stein pair with variance 1, then

$$E\left(\frac{(W'-W)(f(W')-f(W))}{2\lambda}\right) \quad = \quad E[Wf(W)].$$

Taylor expansion

$$f(W')-f(W) = (W'-W)f'(W) + \int_W^{W'} (W'-s)f''(s)ds.$$

Multiplying by $(W'-W)/(2\lambda)$ results in two terms, the first of which is

$$\frac{1}{2\lambda}(W'-W)^2 f'(W).$$

Exchangeable Pair: First Term

First term of the difference $f'(W) - Wf(W)$ is

$$E\left(f'(W)[1 - \frac{(W' - W)^2}{2\lambda}]\right)$$

Since $E(W' - W)^2/(2\lambda) = 1$, conditioning on W, applying the Cauchy Schwarz inequality and that $||f'||_\infty \le 4||h||_\infty$ yields the bound

$$R_1 = \frac{2||h||_\infty}{\lambda}\sqrt{\mathsf{Var}(E((W' - W)^2|W))}.$$

Exchangeable Pair: Second Term

Expectation of

$$\frac{1}{2\lambda}|(W'-W)\int_W^{W'}(W'-s)f''(s)| \leq \frac{1}{4\lambda}||f''||_\infty|W'-W|^3$$

so, applying the bound $||f''||_\infty \leq 2||h'||_\infty$, the second term is bounded by

$$R_2 = \frac{||h'||_\infty}{2\lambda}E|W'-W|^3.$$

Exchangeable Pair: Smooth Functions

Let h be bounded and have bounded derivative, and let W, W' be a mean zero, variance 1, λ-Stein pair. Then

$$|Eh(W) - Nh| \leq R_1 + R_2$$

where

$$R_1 = \frac{2||h||_\infty}{\lambda} \sqrt{\mathsf{Var}(E((W' - W)^2|W))}$$

and

$$R_2 = \frac{||h'||_\infty}{2\lambda} E|W' - W|^3.$$

Exchangeable Pair: Example

Let π be uniform over $\Pi_n \subset \mathcal{S}_n$, the collection of fixed point free ($\pi(i) \neq i$) involutions ($\pi^2(i) = i$) of $\{1, \ldots, n\}$. Special case of a distribution on $\mathcal{S}_n$ constant on cycle type, that is, one satisfying

$$P(\pi) = P(\rho^{-1}\pi\rho) \quad \text{for all } \pi, \rho \in \mathcal{S}_n.$$

Let $\{a_{ij}\}_{i,j}$ be a collection of n^2 real numbers. Approximate the distribution of

$$W = \sum_{i=1}^{n} a_{i,\pi(i)}.$$

May assume $a_{ij} = a_{ji}$ and $a_{ii} = 0$ without loss of generality.

Combinatorial CLT: Involutions

Let for a, b, c distinct, let $A = \{\pi : \pi(a) = c\}$, and $B = \{\pi : \pi(b) = c\}$, and let τ_{ab} be the transposition of a and b. Then

$$\pi \in A \quad \text{if and only if} \quad \tau_{ab}^{-1} \pi \tau_{ab} \in B$$

so $P(A) = P(\tau_{ab}^{-1} A \tau_{ab}) = P(B)$ and therefore

$$Ea_{i,\pi(i)} = \frac{1}{n-1} \sum_{j \neq i} a_{i,j} = \frac{1}{n-1} \sum_{j=1}^{n} a_{i,j}.$$

When considering $\mathcal{L}((W - EW)/\sigma_W)$ we may assume $\sum_j a_{i,j} = 0$ for all i without loss of generality.

Coupling: Involutions

Let I, J with $I \neq J$ be chosen uniformly from $\{1, \ldots, n\}$, and set

$$\pi' = \pi\alpha_{IJ} \quad \text{where} \quad \alpha_{ij} = \tau_{i,\pi(j)}\tau_{j,\pi(i)}.$$

For $\pi \in \Pi_n$ and $i \neq j$, whereas π has the cycle(s)

$$(i, \pi(i)), (j, \pi(j))$$

π' has the cycle(s)

$$(i, j), (\pi(i), \pi(j)).$$

Linearity Condition

Recalling $W = \sum_i a_{i,\pi(i)}$ and letting $W' = \sum_i a_{i,\pi'(i)}$, we have

$$W' - W = 2\left(a_{I,J} + a_{\pi(I),\pi(J)} - (a_{I,\pi(I)} + a_{J,\pi(J)})\right).$$

$$E[a_{I,J}|\pi] = E[a_{I,J}] = \frac{1}{n(n-1)}\sum_{i,j} a_{ij} = 0$$

and

$$E[a_{I,\pi(I)}|\pi] = \frac{1}{n}\sum_{i=1}^{n} a_{i,\pi(i)} = \frac{1}{n}W,$$

and so, since the resulting expression is W measurable,

$$E[W'|W] = (1 - \frac{4}{n})W.$$

Involutions: Calculating the Bound

Need to compute

$$R_1 = \frac{2||h||_\infty}{\lambda}\sqrt{\mathsf{Var}(E((W' - W)^2|W))}$$

and

$$R_2 = \frac{||h'||_\infty}{2\lambda} E|W' - W|^3$$

for

$$W' - W = 2\left(a_{I,J} + a_{\pi(I),\pi(J)} - (a_{I,\pi(I)} + a_{J,\pi(J)})\right).$$

Under the usual asymptotic $R_2 = O(n^{-1/2})$.

Involutions: Calculating the Bound

Recall

$$W' - W = 2\left(a_{I,J} + a_{\pi(I),\pi(J)} - (a_{I,\pi(I)} + a_{J,\pi(J)})\right).$$

To show $R_1 = O(n^{-1/2})$ use

$$\text{Var}(E((W'-W)^2|W)) \leq \text{Var}(E((W'-W)^2|\pi)).$$

Requires calculation of the variance of a sum of terms such as

$$E(a_{I,\pi(I)}^2|\pi) = \frac{1}{n}\sum_{i=1}^{n} a_{i,\pi(i)}^2.$$

Zero Bias Coupling

[Goldstein and Reinert (1997)]

Stein identity: $Z \sim \mathcal{N}(0, \sigma^2)$ if and only if

$$E[Zf(Z)] = \sigma^2 E[f'(Z)] \quad \text{for all smooth } f.$$

For any mean zero, variance σ^2 distribution $\mathcal{L}(W)$ there exists $\mathcal{L}(W^*)$ satisfying

$$E[Wf(W)] = \sigma^2 E[f'(W^*)].$$

Distributional transformation $W \to W^*$, of which $\mathcal{N}(0, \sigma^2)$ is the unique fixed point.

Absolutely continuous, support$(W^*) \subset$ co(support(W)).

Zero Bias Distribution

Density of W^* is given by

$$p^*(t) = E[X; X > t]/\sigma^2.$$

Distribution can also be specified as 'square biasing' followed by multiplication by an independent uniform,

$$X^* =_d UY$$

where

$$\frac{dF_Y}{dF_X} = \frac{x^2}{\sigma^2}.$$

Fixed Point Proximity

If W is close to W^*, then W is close to being a fixed point of the zero bias transformation, so close to the unique fixed point, so close to normal.

Change One Property

Parallel to the size biasing: If $X_1, \ldots, X_n$ are independent nonnegative (mean zero) random variables with finite nonzero mean (variance), then

$$W = \sum_{i=1}^{n} X_i$$

can be size (zero) biased by replacing a single summand, chosen with probability proportional to its mean (variance) and replacing it with an independent random variable having that summands size (zero) biased distribution, e.g.

$$W^* - W = X_I^* - X_I.$$

Zero Bias Rationale for CLT

When W is the sum of many comparable variables, W^* differs from W by only a single summand. Hence the distributions of W and W^* are close, so $\mathcal{L}(W)$ is close to being a fixed point of the zero bias transformation, and so must be close to the normal.

One way to make this statement precise is with the following L^1 bound: For any coupling of W, having variance 1, to W^*,

$$||\mathcal{L}(W) - \mathcal{L}(Z)||_1 \leq 2E|W^* - W|.$$

Proof of L^1 Bound

Let W^* have the W-zero bias distribution, and be defined on the same space as W. Then when $||h'||_\infty \leq 1$,

$$\begin{aligned} |Eh(W) - Nh| &= |E[f'(W) - Wf(W)]| \\ &= |Ef'(W) - Ef'(W^*)| \\ &\leq ||f''||_\infty E|W - W^*| \\ &\leq 2||h'||_\infty E|W - W^*| \\ &\leq 2E|W^* - W|. \end{aligned}$$

Taking supremum over all h with $||h'||_\infty \leq 1$ yields

$$||\mathcal{L}(W) - \mathcal{L}(Z)||_1 \leq 2E|W^* - W|.$$

Connection to Exchangeable Pair

If $dF(w', w'')$ is the distribution of the Stein pair W', W'' let $W^\dagger, W^\ddagger$ have distribution

$$dF^\dagger(w', w'') = \frac{(w' - w'')^2}{2\lambda\sigma^2} dF(w', w'').$$

Then if U is a uniform variable, independent of $W^\dagger, W^\ddagger$,

$$W^* = UW^\dagger + (1 - U)W^\ddagger$$

has the W-zero biased distribution.

Square biasing under symmetry

Let $\mathbf{Y} = (Y_1, \ldots, Y_n) =_d (\pm Y_1, \ldots, \pm Y_n)$ with $\mathsf{Var}(Y_i) = \sigma_i^2 \in (0, \infty)$ and $W = \sum_{i=1}^n Y_i$. Let $\mathbf{Y}^i \sim y_i^2 dF(\mathbf{y})/\sigma_i^2$, I a random index independent of $\mathbf{Y}$ and $\{\mathbf{Y}^i, i = 1, \ldots, n\}$ with distribution

$$P(I = i) = \frac{\sigma_i^2}{\sum_{j=1}^n \sigma_j^2},$$

and $U \sim \mathcal{U}[-1, 1]$ independent of all other variables. Then

$$W^* = UY_I^I + \sum_{j \neq I} Y_j^I$$

has the W-zero bias distribution.

Connection to K function

For a mean zero random variable X, Chen and Shao let

$$K(t) = E(X\mathbf{1}_{0 \le t \le X} - X\mathbf{1}_{X \le t < 0}) = E(X\mathbf{1}_{X>t}) \quad \text{a.e.},$$

so $K(t)/\sigma^2$ is the zero bias density. For a sum W of independent variables, letting $W = W^{(i)} + X_i$, they write

$$E[Wf(W)] = \sum_{i=1}^{n} \int_{-\infty}^{\infty} E[f'(W^{(i)} + t)]K_i(t)dt,$$

which is $\sigma^2 Ef'(W^*)$, as the expression above equals

$$\sigma^2 \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma^2} \int_{-\infty}^{\infty} E[f'(W^{(i)}+t)]\frac{K_i(t)}{\sigma_i^2}dt = \sigma^2 Ef'(W_I+X_I^*).$$

Comparison of three couplings

1. Exchangeable pair: linearity condition, evaluation of the variance of a conditional expectation.
2. Size bias: no linearity condition, evaluation of the variance of a conditional expectation.
3. Zero bias:
   (a) Construction through exchangeable pair: linearity condition, no variance of conditional expectation.
   (b) Construction through square biasing: symmetry condition, no variance of conditional expectation.

IV. Local dependence, Nonsmooth functions