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

Larry Goldstein

University of Southern California

II . Size Bias Couplings

(a) Stein equation with mean and variance

(b) Size biasing

(c) Relation to Stein equation

(d) Smooth function bound

(e) Examples

[Baldi, Rinott and Stein (1989)]

Stein Equation

For given $h \in \mathcal{H}$, solve for f in

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

For W satisfying $EW = 0, \mathsf{Var}(W) = 1$ we calculate

$$Eh(W) - Nh$$

by computing

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

Stein Equation, Mean and Variance

If W has mean zero and variance 1, consider

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

If Y has mean μ and variance σ^2, letting $w = (y-\mu)/\sigma$,

$$g'\left(\frac{y-\mu}{\sigma}\right) - \left(\frac{y-\mu}{\sigma}\right) g\left(\frac{y-\mu}{\sigma}\right)$$

and $f(y) = \sigma g((y-\mu)/\sigma)$ gives

$$f'(y) - \left(\frac{y-\mu}{\sigma^2}\right) f(y) = h\left(\frac{y-\mu}{\sigma}\right) - Nh.$$

Stein Equation, Scaling and Bounds

When

$$f(y) = \sigma g((y-\mu)/\sigma)$$

then

$$||f^{(k)}||_\infty = \sigma^{-k+1}||g^{(k)}||_\infty.$$

In particular,

$$||f'||_\infty = ||g'||_\infty \leq 2||h - Nh||_\infty$$

and

$$||f''||_\infty = \sigma^{-1}||g''||_\infty \leq 2\sigma^{-1}||h'||_\infty.$$

Size Biasing

Let $Y \geq 0$ have nonzero finite mean $EY = \mu$. We say Y^s has the Y-size bias distribution if

$$\frac{dF^s}{dF} = \frac{y}{\mu}$$

where F and F^s are the distributions of Y and Y^s, respectively. Alternatively, the distribution of Y^s is characterized by

$$E[Yf(Y)] = \mu E[f(Y^s)]$$

for all functions for which these expectations exist.

Size Biased Sampling

Oil exploration, find large reserves first.

For Y nonnegative integer valued with finite nonzero mean,

$$P(Y^s = k) = \frac{kP(Y = k)}{EY}, \quad k = 0, 1, \ldots.$$

Random Digit Dialing, Sampling (zero mass at zero).

Note if Y is Bernoulli $p \in (0, 1)$, then

$$Y^s = 1.$$

Waiting Time Paradox

Consider $\Gamma(\alpha, 1/\lambda)$ distribution,

$$g(y; \alpha, 1/\lambda) = \frac{\lambda^\alpha y^{\alpha-1} e^{-\lambda y}}{\Gamma(\alpha)}.$$

Poisson Process with exponential $Y_i \sim \Gamma(1, 1/\lambda)$ interarrival times. The memoryless property says one lands in interval of length with distribution $Y_1 + Y_2 \sim \Gamma(2, 1/\lambda)$. Note $EY_1 = 1/\lambda$, and

$$\lambda^2 y e^{-\lambda y} \quad \text{is the size biased density of} \quad \lambda e^{-\lambda y}.$$

Single Summand Property

Let $X_1, \ldots, X_n$ be nonnegative independent random variables with finite means $\mu_1, \ldots, \mu_n$, respectively, and

$$Y = \sum_{i=1}^{n} X_i.$$

Let I be an index with distribution $P(I = i) = \mu_i / \sum_{j=1}^{n} \mu_j$, and for $i = 1, 2 \ldots, n$ let X_i^i have the X_i size biased distribution, and be independent of $X_j, j \neq i$. Then with

$$Y^j = \sum_{i \neq j} X_i + X_j^j,$$

the variable Y^I has the Y-size biased distribution.

Size Biasing under Dependence

For $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n$ with nonnegative components and positive means $\mu_1, \ldots, \mu_n$ we say $\mathbf{X}^i$ has the $\mathbf{X}$ distribution biased in direction i if

$$EX_i g(\mathbf{X}) = \mu_i \, Eg(\mathbf{X}^i) \quad \text{or} \quad dF^i(\mathbf{x}) = \frac{x_i dF(\mathbf{x})}{\mu_i}.$$

Then letting I be a random index independent of $\mathbf{X}, \mathbf{X}^i, i = 1, \ldots, n$ with distribution

$$P(I = i) = \frac{\mu_i}{\sum_{j=1}^n \mu_j}, \quad \text{the variable} \quad Y^I = \sum_{i=1}^n X_i^I$$

has the Y size biased distribution.

Size Biasing under Dependence

Let $Y = \sum_{j=1}^n X_j$ and $Y^i = \sum_{j=1}^n X_j^i$. Since

$$EX_i g(\mathbf{X}) = \mu_i Eg(\mathbf{X}^i)$$

for $g(\mathbf{x}) = f(x_1 + \cdots + x_n)$ we have

$$EX_i f(Y) = \mu_i Ef(Y^i).$$

Summing over i yields

$$\begin{aligned} E[Yf(Y)] &= \sum_{i=1}^n \mu_i Ef(Y^i) \\ &= \mu \sum_{i=1}^n P(I=i) Ef(Y^i) \\ &= \mu Ef(Y^I). \end{aligned}$$

Dependent Bernoullis

If $X_1, \ldots, X_n$ are independent non trivial Bernoulli random variables, then $X_i^i = 1$ so

$$\mathbf{X}^i = \mathcal{L}(\mathbf{X}|X_i = 1).$$

For instance, if $\mathbf{X} \in \mathbb{R}^N$ are the inclusion indicator variables for individuals in a simple random sample of size n, $\mathbf{X}^i$ are the inclusion indicators when $X_i^i = 1$ and the remaining $X_j^i, j \neq i$ are indicators for a simple random sample of size $n-1$.

Include individual i, then sample $n-1$ individuals from those that remain. Will need coupling.

Size Biasing: Mean and Variance Relation

With $\mu = EY$ recall

$$\mu E f(Y^s) = E[Y f(Y)].$$

If Y and Y^s are defined on the same space,

$$\begin{aligned} \mu E(Y^s - Y) &= \mu E Y^s - \mu E Y \\ &= EY^2 - \mu^2 \\ &= \sigma^2 \end{aligned}$$

where $\sigma^2 = \mathrm{Var}(Y)$.

Size Biasing and the Stein Equation

Recall (μ, σ^2) Stein's Lemma: If

$$E\left[\left(\frac{X-\mu}{\sigma^2}\right) f(X)\right] = Ef'(X) \quad \text{for } f \in \mathcal{F}$$

then $X \sim \mathcal{N}(\mu, \sigma^2)$.

Hence, if

$$E\left[\left(\frac{Y-\mu}{\sigma^2}\right) f(Y)\right] \approx Ef'(Y) \quad \text{for } f \in \mathcal{F}$$

then $Y \approx \mathcal{N}(\mu, \sigma^2)$.

Size Bias Coupling

Suppose Y and Y^s are defined on a common space, with Y^s having the Y size bias distribution. Then for a twice differentiable function f,

$$
\begin{aligned}
E[\left(\frac{Y-\mu}{\sigma^2}\right) f(Y)] &= E[\frac{\mu}{\sigma^2}(f(Y^s)-f(Y))] \\
&= \frac{\mu}{\sigma^2} E(Y^s - Y) f'(Y) + R
\end{aligned}
$$

where

$$
R = \frac{\mu}{\sigma^2} E \int_Y^{Y^s} (Y^s - t) f''(t) dt.
$$

Size Bias Coupling

Taking the difference,

$$
\begin{aligned}
E&\left[f'(Y) - \left(\frac{Y-\mu}{\sigma^2}\right) f(Y)\right] \\
&= E\left[\left(1 - \frac{\mu}{\sigma^2} E(Y^s - Y)\right) f'(Y)\right] + R \\
&= E\left[f'(Y) E[\left(1 - \frac{\mu}{\sigma^2}(Y^s - Y)\right) |Y]\right] + R.
\end{aligned}
$$

Main Term

$$
\begin{aligned}
& |E\left[f'(Y)E[\Big(1-\frac{\mu}{\sigma^2}(Y^s-Y)\Big)|Y]\right]| \\
\leq\ & \frac{\mu}{\sigma^2}\sqrt{E[f'(Y)]^2}\sqrt{\mathsf{Var}(E(Y^s-Y|Y))} \\
\leq\ & \frac{2\mu}{\sigma^2}||h-Nh||_\infty\sqrt{\mathsf{Var}(E(Y^s-Y|Y))}.
\end{aligned}
$$

When $Y=\sum_{i=1}^n X_i$, sum of nonnegative variables, typically we have μ and σ^2 of $O(n)$. Hence if the variance term is $O(1/n)$, this term has order $n^{-1/2}$.

Size Bias Coupling, Remainder Term

$$R = \frac{\mu}{\sigma^2} E \int_Y^{Y^s} (Y^s - t) f''(t) dt.$$

Recalling $||f''||_\infty \le (2/\sigma)||h'||_\infty$, may be bounded by

$$|R| \le ||f''||_\infty \frac{\mu}{\sigma^2} \frac{1}{2} E(Y^s - Y)^2 \le ||h'||_\infty \frac{\mu}{\sigma^3} E(Y^s - Y)^2.$$

If μ and σ^2 are both order $O(n)$ then if $E(Y^s - Y)^2$ is bounded the remainder term R has order $n^{-1/2}$.

Error Bound

The remainder term depends on $E(Y^s - Y)^2$. Berry-Esseen bounds depend on third moments.

Note

$$E[Yf(Y)] = \mu E f(Y^s)$$

applied with $f(w) = w^2$ gives $\mu E(Y^s)^2 = EY^3$.

Putting terms together

Smooth function bound: If h' exists and is bounded,

$$|Eh((Y-\mu)/\sigma) - Nh| \le R_1 + R_2$$

where

$$R_1 = \frac{2\mu}{\sigma^2}||h - Nh||_\infty \sqrt{\mathsf{Var}(E(Y^s - Y|Y))}$$

and

$$R_2 = ||h'||_\infty \frac{\mu}{\sigma^3} E(Y^s - Y)^2.$$

Typically μ and σ^2 are $O(n)$, so we want

$\mathsf{Var}(E(Y^s - Y|Y)) = O(n^{-1})$ and $E(Y^s - Y)^2 = O(1)$.

Independent Identically Distributed Variables

When $Y = X_1 + \cdots + X_n$, a sum of nonnegative i.i.d. variables with variances σ^2 and finite third moments, then with $P(I = i) = 1/n$ and X_i^s independent of all other variables

$$Y^I - Y = X_I^s - X_I.$$

Hence $E[X_I^s - X_I | Y] = EX_I^s - Y/n$ and therefore

$$\mathrm{Var}(E[X_I^s - X_I|Y]) = \mathrm{Var}(Y)/n^2 = \sigma^2/n = O(n^{-1}),$$

and since $E(X_i^s)^2 = EX_i^3/EX_i$,

$$E(Y^s - Y)^2 = E(X_I^s - X_I)^2 \leq 2E((X_I^s)^2 + X_I^2) = O(1).$$

Example: Simple Random Sampling n of N

Population $\mathcal{A} = \{a_1, \ldots, a_N\} \subset (0, \infty)$. Want to approximate the standardized distribution of

$$Y = \sum_{i=1}^{N} a_i J_i,$$

where all $\mathbf{J} = (J_1, \ldots, J_N) \in \{0,1\}^N$ with $\sum_{i=1}^{N} J_i = n$ are equally likely.

Coupling Y and Y^s

Given $\mathbf{J}$, let K be chosen uniformly from the collection of k for which $J_k = 1$. For each i let

$$J_j^i = \begin{cases} J_j & j \notin \{i, K\} \\ J_i & j = K \\ 1 & j = i. \end{cases}$$

Interchanging the sampling indicators of i and the sampled unit K gives $\mathbf{J}^i$ indicators with

$$\mathcal{L}(\mathbf{J}^i) = \mathcal{L}(J_1, \ldots, J_N | J_i = 1)$$

on the same space as, and close to, $\mathbf{J}$.

Simple Random Sampling Coupling

As $Ea_iJ_i = a_in/N$, upon picking $P(I=i) \propto a_i$, Y^I has the Y-size biased distribution, where $Y^i = \sum_{j=1}^N a_jJ_j^i$. Letting

$$\overline{Y} = \sum_{i \notin \{I,K\}} a_iJ_i$$

when $I \neq K$ we have

$$Y = \overline{Y} + a_IJ_I + a_KJ_K \quad \text{and} \quad Y^I = \overline{Y} + a_IJ_K + a_KJ_I,$$

and then, in all cases,

$$Y^I - Y = a_IJ_K + a_KJ_I - a_IJ_I - a_KJ_K = (1-J_I)(a_I - a_K).$$

Conditional Expectation of Difference

May be difficult to calculate the conditional expectation

$$E(Y^I - Y|Y) = E((1 - J_I)(a_I - a_K)|Y).$$

Let $X = E(\Delta|\mathcal{F})$ where Y is $\mathcal{F}$ measurable. By the conditional variance formula

$$\mathsf{Var}(X) = E[\mathsf{Var}(X|Y)] + \mathsf{Var}[E(X|Y)] \geq \mathsf{Var}[E(X|Y)],$$

and

$$E(X|Y) = E(E(\Delta|\mathcal{F})|Y) = E(\Delta|Y).$$

Hence, conditioning on more yields an upper bound,

$$\mathsf{Var}[E(\Delta|\mathcal{F})] \geq \mathsf{Var}[E(\Delta|Y)].$$

Conditional Expectation of Difference

Condition on more:

$\mathsf{Var}(E((1-J_I)(a_I-a_K)|Y)) \leq \mathsf{Var}(E((1-J_I)(a_I-a_K)|\mathbf{J}))$.

Tractable conditional expectation:

$$
\begin{aligned}
& E((1-J_I)(a_I-a_K)|\mathbf{J}) \\
= & \sum_{i,k}(1-J_i)(a_i-a_k)P(I=i,K=k|\mathbf{J}) \\
= & \sum_{i,k}(1-J_i)(a_i-a_k)\frac{a_i}{N\overline{a}}\frac{J_k}{n}
\end{aligned}
$$

Under 'typical' conditions [Luk (1994)] $n/N \to f \in (0,1)$ and $a_i = O(1)$, the variance will be $O(1/n)$, as desired

Advertisements

Graph Degree Problem on $\mathcal{G}_n$

For every pair of vertices in the set $\mathcal{V}$ of size n, draw an edge, independently of all other edges, with probability π_n. For d a nonnegative integer, let Y be the number of edges of the resulting graph $\mathcal{G}_n$ which has degree d, that is,

$$Y = \sum_{v \in \mathcal{V}} X_v \quad \text{where} \quad X_v = \mathbf{1}(D(v) = d).$$

Graph Degree Problem on $\mathcal{G}_n$

For every pair of vertices in the set $\mathcal{V}$ of size n, draw an edge, independently of all other edges, with probability π_n. For d a nonnegative integer, let Y be the number of edges of the resulting graph $\mathcal{G}_n$ which has degree d, that is,

$$Y = \sum_{v \in \mathcal{V}} X_v \quad \text{where} \quad X_v = \mathbf{1}(D(v) = d).$$

Graph Degree Problem on $\mathcal{G}_n$

For every pair of vertices in the set $\mathcal{V}$ of size n, draw an edge, independently of all other edges, with probability π_n. For d a nonnegative integer, let Y be the number of edges of the resulting graph $\mathcal{G}_n$ which has degree d, that is,

$$Y = \sum_{v \in \mathcal{V}} X_v \quad \text{where} \quad X_v = \mathbf{1}(D(v) = d).$$

To size bias, select V uniformly over $\mathcal{V}$. Conditional on $D(V) = d$, the d edges of V are uniform over all possible $\binom{n-1}{d}$ choices. Edges not involving V are independent. Hence, a coupling can be achieved by first generating $\mathcal{G}_n$, selecting V, and then adding or removing edges from V as needed for the cases $D(V) < d$ and $D(V) > d$, respectively.

Graph Degree Problem on $\mathcal{G}_n$

For every pair of vertices in the set $\mathcal{V}$ of size n, draw an edge, independently of all other edges, with probability π_n. For d a nonnegative integer, let Y be the number of edges of the resulting graph $\mathcal{G}_n$ which has degree d, that is,

$$Y = \sum_{v \in \mathcal{V}} X_v \quad \text{where} \quad X_v = \mathbf{1}(D(v) = d).$$

See Jay Bartroff's talk.

Covered Volume of Balls around Randomly Placed Points

Let $U_1, \ldots, U_n$ be i.i.d. in $C_n = [0, n^{1/d})^d$ with periodic boundary conditions, and let $B_{i,\rho}$ be the ball of radius ρ centered around U_i. Let V be the volume of their union,

$$V = \text{Volume}(\bigcup_{i=1}^{n} B_{i,\rho}).$$

Unlike previous examples, there are no obvious indicators to 'set to 1'; in fact, V is continuous.

Q: So, how to size bias V?

Covered Volume of Balls around Randomly Placed Points

Let $U_1, \ldots, U_n$ be i.i.d. in $C_n = [0, n^{1/d})^d$ with periodic boundary conditions, and let $B_{i,\rho}$ be the ball of radius ρ centered around U_i. Let V be the volume of the union

$$V = \text{Volume}(\bigcup_{i=1}^{n} B_{i,\rho}).$$

Unlike previous examples, there are no obvious indicators to 'set to 1'; in fact, V is continuous.

Q: So, how to size bias V?

A: See Mathew Penrose's talk.

Coming Attractions

III. Exchangeable Pair, Zero Bias Couplings

IV. Local dependence, Nonsmooth functions