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

$$f'(w) - wf(w) = h(w) - Nh$$

where $Nh = Eh(Z)$.

For $W$ satisfying $EW = 0$, $\text{Var}(W) = 1$ 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
\[ g'(w) - wg(w) = h(w) - Nh \]
where \( Nh = Eh(Z) \).

If $Y$ has mean $\mu$ and variance $\sigma^2$, 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

$$Y^s = 1.$$
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

$$\lambda^2 ye^{-\lambda y} \quad \text{is the size biased density of} \quad \lambda e^{-\lambda y}.$$
Let $X_1, \ldots, X_n$ be nonnegative independent random variables with finite means $\mu_1, \ldots, \mu_n$, respectively, and let $Y = \sum_{i=1}^{n} X_i$.

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 $X_i^j$ 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 \( X^i \) has the \( \mathbf{X} \) distribution biased in direction \( i \) if

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

Then letting \( I \) be a random index independent of \( \mathbf{X}, X^i, i = 1, \ldots, n \) 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 $Y^i = \sum_{j=1}^{n} X^i_j$. Since

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

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

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

Summing over $i$ yields

$$E[Y f(Y)] = \sum_{i=1}^{n} \mu_i Ef(Y^i)$$

$$= \mu \sum_{i=1}^{n} P(I = i) Ef(Y^i)$$

$$= \mu Ef(Y^I).$$
If $X_1, \ldots, X_n$ are independent non trivial Bernoulli random variables, then $X_i^i = 1$ so

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

For instance, if $X \in \mathbb{R}^N$ are the inclusion indicator variables for individuals in a simple random sample of size $n$, $X^i$ 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

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

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

$$\mu E(Y^s - Y) = \mu EY^s - \mu EY$$
$$= EY^2 - \mu^2$$
$$= \sigma^2$$

where $\sigma^2 = \text{Var}(Y)$. 
Recall \((\mu, \sigma^2)\) 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 $f$,

$$E\left[ \left( \frac{Y - \mu}{\sigma^2} \right) f(Y) \right] = E\left[ \frac{\mu}{\sigma^2} (f(Y^s) - f(Y)) \right]$$

$$= \frac{\mu}{\sigma^2} E(Y^s - Y) f'(Y) + R$$

where

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

Taking the difference,

\[
E \left[ f'(Y) - \left( \frac{Y - \mu}{\sigma^2} \right) f(Y) \right] \\
= E \left[ \left( 1 - \frac{\mu}{\sigma^2} E(Y^s - Y) \right) f'(Y) \right] + R \\
= E \left[ f'(Y) E\left[ \left( 1 - \frac{\mu}{\sigma^2} (Y^s - Y) \right) |Y| \right] \right] + R.
\]
\[
|E \left[ f'(Y) E\left[ \left( 1 - \frac{\mu}{\sigma^2} (Y^s - Y) \right) |Y| \right] \right] |
\leq \frac{\mu}{\sigma^2} \sqrt{E[f'(Y)]^2} \sqrt{\text{Var}(E(Y^s - Y|Y))}
\leq \frac{2\mu}{\sigma^2} \|h - Nh\|_{\infty} \sqrt{\text{Var}(E(Y^s - Y|Y))}.
\]

When \( Y = \sum_{i=1}^{n} X_i \), sum of nonnegative variables, typically we have \( \mu \) and \( \sigma^2 \) 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 \( \|f''\|_\infty \leq (2/\sigma) \|h'\|_\infty \), may be bounded by

\[ |R| \leq \|f''\|_\infty \frac{\mu}{\sigma^2} E(Y_s - Y)^2 \leq \|h'\|_\infty \frac{\mu}{\sigma^3} E(Y_s - Y)^2. \]

If \( \mu \) and \( \sigma^2 \) are both order \( O(n) \) then if \( E(Y_s - Y)^2 \) is bounded the remainder term \( R \) has order \( n^{-1/2} \).
The remainder term depends on $E(Y^s - Y)^2$. Berry-Esseen bounds depend on third moments.

Note

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

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,

\[
|Eh((Y - \mu)/\sigma) - Nh| \leq R_1 + R_2
\]

where

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

and

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

Typically \( \mu \) and \( \sigma^2 \) are \( O(n) \), so we want

\[
\text{Var}(E(Y^s - Y|Y)) = O(n^{-1}) \quad \text{and} \quad E(Y^s - Y)^2 = O(1).
\]
When $Y = X_1 + \cdots + X_n$, a sum of nonnegative i.i.d. variables with variances $\sigma^2$ 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

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

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 $J = (J_1, \ldots, J_N) \in \{0, 1\}^N$ with $\sum_{i=1}^{N} J_i = n$ are equally likely.
Given $J$, let $K$ be chosen uniformly from the collection of $k$ for which $J_k = 1$. For each $i$ let

$$J^i_j = \begin{cases} J_j & j \not\in \{i, K\} \\ J_i & j = K \\ 1 & j = i. \end{cases}$$

Interchanging the sampling indicators of $i$ and the sampled unit $K$ gives $J^i$ indicators with

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

on the same space as, and close to, $J$. 
Simple Random Sampling Coupling

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

$$\overline{Y} = \sum_{i \not\in \{I,K\}} a_i J_i$$

when $I \neq K$ we have

$$Y = \overline{Y} + a_I J_I + a_K J_K \quad \text{and} \quad Y^I = \overline{Y} + a_I J_K + a_K J_I,$$

and then, in all cases,

$$Y^I - Y = a_I J_K + a_K J_I - a_I J_I - a_K J_K = (1 - J_I)(a_I - a_K).$$
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

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

and

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

Hence, conditioning on more yields an upper bound,

\[ \text{Var}[E(\Delta | \mathcal{F})] \geq \text{Var}[E(\Delta | Y)]. \]
Conditional Expectation of Difference

Condition on more:

\[ \text{Var}(E((1-J_I)(a_I-a_K)|Y)) \leq \text{Var}(E((1-J_I)(a_I-a_K)|J)) \].

Tractable conditional expectation:

\[
E((1 - J_I)(a_I - a_K)|J) \\
= \sum_{i,k} (1 - J_i)(a_i - a_k)P(I = i, K = k|J) \\
= \sum_{i,k} (1 - J_i)(a_i - a_k) \frac{a_i}{N\bar{a}} \frac{J_k}{n}
\]

Under ‘typical’ conditions [Luk (1994)] \( n/N \rightarrow f \in (0, 1) \) and \( a_i = O(1) \), the variance will be \( O(1/n) \), as desired
Graph Degree Problem on $G_n$

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

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

$$Y = \sum_{v \in \mathcal{V}} X_v \quad \text{where} \quad X_v = 1(D(v) = d).$$
For every pair of vertices in the set \( V \) of size \( n \), draw an edge, independently of all other edges, with probability \( \pi_n \). For \( d \) a nonnegative integer, let \( Y \) be the number of edges of the resulting graph \( G_n \) which has degree \( d \), that is,

\[
Y = \sum_{v \in V} X_v \quad \text{where} \quad X_v = 1(D(v) = d).
\]

To size bias, select \( V \) uniformly over \( V \). 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 \( G_n \), 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 $G_n$

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

$$Y = \sum_{v \in \mathcal{V}} X_v$$

where $X_v = 1(D(v) = d)$.

See Jay Bartroff’s talk.
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 $\rho$ 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 $\rho$ 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.
III. Exchangeable Pair, Zero Bias Couplings

IV. Local dependence, Nonsmooth functions