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

Larry Goldstein

University of Southern California

IV. Local dependence

   (a) Smooth function bound under local dependence

IV. Nonsmooth functions

   (a) Concentration Inequalities
   (b) Smoothing Inequalities
   (c) Inductive Approach

Local Dependence

Let $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ be a collection of mean zero random variables with $\mathrm{Var}(W) = 1$ where $W = \sum_{\alpha \in \mathcal{A}} X_\alpha$.

For each $\alpha \in \mathcal{A}$ suppose there exists $S_\alpha \subset \mathcal{A}$ such that

$$X_\alpha \quad \text{is independent of} \quad \{X_\beta : \beta \in S_\alpha^c\}.$$

Local Dependence

Let $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ be a collection of mean zero random variables with $\mathrm{Var}(W) = 1$ where $W = \sum_{\alpha \in \mathcal{A}} X_\alpha$. For each $\alpha \in \mathcal{A}$ suppose there exists $S_\alpha \subset \mathcal{A}$ such that

$$X_\alpha \quad \text{is independent of} \quad \{X_\beta : \beta \in S_\alpha^c\}.$$

In particular, note therefore, that

$$\sum_{\alpha \in \mathcal{A}} \sum_{\beta \in S_\alpha} E X_\alpha X_\beta = 1.$$

Local Dependence

For $\alpha \in A$ let

$$W_\alpha = \sum_{\beta \notin S_\alpha} X_\beta.$$

Given h, consider

$$\begin{aligned}
& E[h(W) - Nh] \\
= \;& E[f'(W) - Wf(W)] \\
= \;& E[f'(W) - \sum_{\alpha \in \mathcal{A}} X_\alpha f(W)] \\
= \;& E[f'(W) - \sum_{\alpha \in \mathcal{A}} X_\alpha (f(W) - f(W_\alpha)) + \sum_{\alpha \in \mathcal{A}} X_\alpha f(W_\alpha)] \\
= \;& E[f'(W) - \sum_{\alpha \in \mathcal{A}} X_\alpha (f(W) - f(W_\alpha))].
\end{aligned}$$

Local Dependence

Summands are

$$X_\alpha(f(W) - f(W_\alpha))] = X_\alpha \left(f'(W)(W - W_\alpha)\right) + R_\alpha,$$

so subtracting sum of first term from $f'(W)$ yields

$$\begin{aligned} & |E[f'(W)(1 - \sum_{\alpha \in \mathcal{A}} X_\alpha \sum_{\beta \in S_\alpha} X_\beta]|| \\ \leq \quad & 2||h||_\infty \sqrt{\text{Var}\left(\sum_{\alpha \in \mathcal{A}, \beta \in S_\alpha} X_\alpha X_\beta\right)}. \end{aligned}$$

Local Dependence

For the remainder term, recalling

$$X_\alpha(f(W) - f(W_\alpha))] = X_\alpha\left(f'(W)(W - W_\alpha)\right) + R_\alpha,$$

we see

$$|R_\alpha| \leq \frac{1}{2}||f''||_\infty |X_\alpha(\sum_{\beta\in S_\alpha} X_\beta)^2| \leq ||h'||_\infty |X_\alpha(\sum_{\beta\in S_\alpha} X_\beta)^2|.$$

Local Dependence: Smooth Function Theorem

If $||h'||_\infty < \infty$, then $|Eh(W) - Nh|$ is bounded by

$$2||h||_\infty \sqrt{\mathsf{Var}\left(\sum_{\alpha\in\mathcal{A},\beta\in S_\alpha} X_\alpha X_\beta\right)}$$

$$+ \quad ||h'||_\infty \sum_{\alpha\in A} E|X_\alpha(\sum_{\beta\in S_\alpha} X_\beta)^2|.$$

Local Dependence: Smooth Function Theorem

Application:

> Let each edge in a finite lattice in $\mathbb{Z}^2$ be present with probability $p \in (0,1)$ independent of the presence of all other edges. Let W be the number of squares.

Compute a bound on the normal approximation to W.

Local Dependence: Smooth Function Theorem

Application:

> Let each edge in a finite lattice in $\mathbb{Z}^2$ be present with probability $p \in (0, 1)$ independent of the presence of all other edges. Let W be the number of squares.
>
> Assign an independent uniform random variable to each vertex of a fixed graph. Let W be the number of local maxima.

Compute a bound on the normal approximation to W.

Kolmogorov, or L^∞ Bounds

So far we have dealt with expectations of smooth functions. Now we consider deriving bounds on

$$||F - \Phi||_\infty = \sup_{-\infty < x < \infty} |F(x) - \Phi(x)|.$$

As $F(x) = E\mathbf{1}(X \le x)$, an expectation of a nonsmooth function, we need some new ideas to obtain a bound using Stein's method.

Concentration Inequalities

[Chen and Shao (2004)]

A bound on the probability that a random variable takes values in a (small) interval.

Let $z \in \mathbb{R}$ and $\delta > 0$. Now let

$$g(w) = \begin{cases} -\delta/2 & w \in (-\infty, z] \\ \text{linear} & w \in (z, z+\delta] \\ \delta/2 & w \in (z+\delta, \infty) \end{cases}$$

Then for W with mean zero and variance 1,

$$\begin{aligned} P(z \le W^* \le z+\delta) &= Eg'(W^*) \\ &= EWg(W) \le (\delta/2)E|W| \le \delta/2. \end{aligned}$$

Concentration Inequalities

Suppose for the given W we can find W^* such that $|W^* - W| \leq \delta$. For example, if $X_1, \ldots, X_n$ are i.i.d. mean zero with $|X_1| \leq K$,

$$W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \quad \text{then} \quad W^* = W - \frac{1}{\sqrt{n}} \left(X_I - X_I^* \right),$$

where $P(I = i) = 1/n$, so in particular

$$|W^* - W| \leq \delta \quad \text{where} \quad \delta = \frac{2K}{\sqrt{n}}.$$

Concentration Inequalities

When $|W^* - W| \le \delta$,

$$\begin{aligned} & P(W \le z) - P(W^* \le z) \\ = \; & P(W^* \le z + W^* - W) - P(W^* \le z) \\ \le \; & P(W^* \le z + \delta) - P(W^* \le z) \\ = \; & P(z \le W^* \le z + \delta) \\ \le \; & \delta/2. \end{aligned}$$

Interchanging the roles of W and W^*, we have

$$P(W^* \le z) - P(W \le z) \le P(z \le W \le z + \delta)$$

Concentration Inequalities

From the concentration inequality for W^*, we have likewise that

$$\begin{aligned}
& P(W^* \leq z) - P(W \leq z) \\
\leq\ & P(z \leq W \leq z + \delta) \\
=\ & P(z \leq W^* - (W^* - W) \leq z + \delta) \\
=\ & P(z + (W^* - W) \leq W^* \leq z + \delta + (W^* - W)) \\
\leq\ & P(z - \delta \leq W^* \leq z + 2\delta) \\
\leq\ & \frac{3}{2}\delta.
\end{aligned}$$

Hence

$$|P(W \leq z) - P(W^* \leq z)| \leq \frac{3}{2}\delta.$$

Concentration Inequalities

Letting

$$f'(w) - wf(w) = \mathbf{1}_{(-\infty,z]}(w) - \Phi(z)$$

we have

$$\begin{aligned} & |P(W \le z) - P(Z \le z)| = |E[f'(W) - Wf(W)]| \\ = \quad & |E[f'(W) - f'(W^*)]| \\ = \quad & |E[Wf(W) - W^*f(W^*)] + P(W \le z) - P(W^* \le z)| \\ \le \quad & |E[Wf(W) - W^*f(W^*)]| + \frac{3}{2}\delta. \end{aligned}$$

Concentration Inequalities

For the first term $E\left(Wf(W) - W^* f(W^*)\right)$ write

$$
\begin{aligned}
& |E[(W(f(W) - f(W^*)) - (W^* - W)f(W^*)]| \\
\leq \; & ||f'||_\infty E|W(W - W^*)| + ||f||_\infty E|W^* - W| \\
\leq \; & E|W(W - W^*)| + \frac{\sqrt{2\pi}}{4} E|W^* - W| \\
\leq \; & \delta(E|W| + \frac{\sqrt{2\pi}}{4}) \\
\leq \; & \delta(1 + \frac{\sqrt{2\pi}}{4})
\end{aligned}
$$

Concentration Inequalities

Bounded zero bias coupling theorem: If $|W^* - W| \leq \delta$ then

$$\begin{aligned} |P(W \leq z) - P(Z \leq z)| &\leq \frac{3}{2}\delta + \delta(1 + \frac{\sqrt{2\pi}}{4}) \\ &= \delta\left(\frac{5}{2} + \frac{\sqrt{2\pi}}{4}\right). \end{aligned}$$

Concentration Inequalities

Bounded zero bias coupling theorem: If $|W^* - W| \leq \delta$ then

$$\begin{aligned} |P(W \leq z) - P(Z \leq z)| &\leq \frac{3}{2}\delta + \delta(1 + \frac{\sqrt{2\pi}}{4}) \\ &= \delta\left(\frac{5}{2} + \frac{\sqrt{2\pi}}{4}\right). \end{aligned}$$

When $W = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ with $X_1, \ldots, X_n$ i.i.d mean zero variance 1 and $|X_i| \leq K$, may take $\delta = 2K/\sqrt{n}$ so

$$|P(W \leq z) - P(Z \leq z)| \leq n^{-1/2}K\left(5 + \sqrt{\frac{\pi}{2}}\right).$$

Smoothing Inequalities

[Bhattacharya and Ranga Rao (1986), Göetze (1991), Rinott and Rotar (1996)]

Works more generally than on indicators, and also in higher dimension, though will illustrate in $\mathbb{R}$.

For a function $h : \mathbb{R} \to \mathbb{R}$ let

$$h_\epsilon^+(x) = \sup_{|y| \le \epsilon} h(x+y), \quad h_\epsilon^-(x) = \inf_{|y| \le \epsilon} h(x+y)$$

and

$$\tilde{h}_\epsilon(x) = h_\epsilon^+(x) - h_\epsilon^-(x).$$

Smoothing Inequalities

Let $\mathcal{H}$ be a class of measurable functions on $\mathbb{R}$ such that

(i) The functions $h \in \mathcal{H}$ are uniformly bounded in absolute value by a constant, which can be taken to be 1 without loss of generality,

(ii) For any real numbers c and d, and for any $h \in \mathcal{H}$, the function $h(cx + d) \in \mathcal{H}$,

(iii) For any $\epsilon > 0$ and $h \in \mathcal{H}$, the functions h_ϵ^+, h_ϵ^- are also in $\mathcal{H}$, and there exists a, depending only on $\mathcal{H}$, such that

$$E\tilde{h}_\epsilon(Z) \leq a\epsilon.$$

Smoothing Inequalities

Let $\mathcal{H}$ denote a class of measurable functions satisfying (i),(ii), and (iii) and let $h \in \mathcal{H}$. let $\phi(t)$ denote the standard normal density, and, for $t \in (0,1)$, define

$$h_t(x) = \int h(x+ty)\phi(y)dy$$

and

$$\delta_t = \sup\{|Eh_t(W) - Nh_t| : h \in \mathcal{H}\}.$$

Lemma 1 *For any random variable W on $\mathbb{R}$*

$$\delta \leq 2.8\delta_t + 4.7at \quad \textit{for all } t \in (0,1).$$

Smoothing Inequalities

Use techniques for smooth functions to obtain a bound on δ_t, which may involve the original δ, e.g. for some constants c_1, c_2 and c_3,

$$\delta_t \leq c_1 + \frac{1}{t}\left(c_2\delta + c_3\right) \quad \text{for all } t \in (0,1).$$

Substitution of δ_t into the smoothing inequality

$$\delta \leq 2.8\delta_t + 4.7at$$

yields a minimization problem over t, and a bound on δ.

Inductive Approach

For $n \in \mathbb{N}$ and $\gamma \geq 1$ let $\mathcal{L}(n, \gamma)$ be collection of distributions on $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ with i.i.d. mean zero, variance 1 components, and $1 \leq \gamma = E|X_1^3| < \infty$. Let

$$S_n = n^{-1/2}(X_1 + \cdots + X_n)$$

and for $\lambda \geq 0$,

$$\delta(\lambda, \gamma, n) = \sup\{|E(h_{z,\lambda}(S_n) - h_{z,\lambda}(Z))| : z \in \mathbb{R}, \mathbf{X} \in \mathcal{L}(n, \gamma)\}$$

where $h_{z,0}(x) = \mathbf{1}(x \leq z)$ and for $\lambda > 0$

$$h_{z,\lambda}(x) = \begin{cases} 1 & x < z \\ 1 + \frac{z-x}{\lambda} & x \in [z, z + 1/\lambda) \\ 0 & x \geq z + 1/\lambda \end{cases}$$

Inductive Approach

Letting $\delta(\gamma, n) = \delta(0, \gamma, n)$, the Berry-Esseen Theorem is

$$\sup\{\sqrt{n}\delta(\gamma, n)/\gamma : \gamma \geq 1, n \in \mathbb{N}\} < \infty.$$

By $h_{z,0}(x) \leq h_{z,\lambda}(x) \leq h_{z+\lambda,0}(x)$ we have

$$\delta(\gamma, n) \leq \delta(\lambda, \gamma, n) + \lambda/\sqrt{2\pi}.$$

When f solves the Stein equation for $h = h_{z,\lambda}$, Chen & Shao '04 give

$$|f'(x+s) - f'(x+t)| \leq$$

$$(|x|+1)\min(|s|+|t|, 1) + \lambda^{-1}\left|\int_s^t \mathbf{1}(z \leq x+u \leq z+\alpha)du\right|$$

Inductive Approach

We have

$$\begin{aligned} Eh(S_n) - h(Z) &= E[f'(S_n) - S_n f(S_n)] \\ &= E[f'(S_n) - f'(S_n^*)] \end{aligned}$$

which is

$$E[f'(S_{n-1} + X_n/\sqrt{n}) - f'(S_{n-1} + X_n^*/\sqrt{n})].$$

Need to bound the expectation of

$$(|S_{n-1}| + 1)\min((|X_n| + |X_n^*|)/\sqrt{n}, 1) +$$

$$\lambda^{-1} \left| \int_{X_n/\sqrt{n}}^{X_n^*/\sqrt{n}} \mathbf{1}(z \le S_{n-1} + u \le z + \lambda) du \right|.$$

Inductive Approach

Regarding the first term, we have

$$E|X_n| \le (E|X_n|^2)^{1/2} = 1 \le (E|X_n^3|)^{1/3} \le E|X_n^3|.$$

From the zero bias identity with function $\text{sign}(x)x^2$ we have $E|X_n^*| = (1/2)E|X_n^3|$. Note, therefore, that

$$c_X = E|X_n^* - X_n| \le E|X_n^*| + E|X_n| \le (3/2)\gamma.$$

Since $E(|S_{n-1}| + 1) \le 2$,

$$E\left((|S_{n-1}| + 1)\min((|X_n| + |X_n^*|)/\sqrt{n}, 1)\right) \le 3n^{-1/2}\gamma.$$

Inductive Approach

Second term: expectation of

$$\lambda^{-1} \left| \int_{X_n/\sqrt{n}}^{X_n^*/\sqrt{n}} \mathbf{1}(z \leq S_{n-1} + u \leq z + \lambda) du \right|$$

which we decompose as λ^{-1} times

$$\begin{aligned} & E\left(\int_{X_n/\sqrt{n}}^{X_n^*/\sqrt{n}} \mathbf{1}(z \leq S_{n-1} + u \leq z + \lambda) du; X_n \leq X_n^* \right) \\ + \quad & E\left(\int_{X_n^*/\sqrt{n}}^{X_n/\sqrt{n}} \mathbf{1}(z \leq S_{n-1} + u \leq z + \lambda) du; X_n > X_n^* \right) \end{aligned}$$

Inductive Approach

For any $u \in \mathbb{R}$,

$$P(z \leq S_{n-1} + u \leq z + \lambda)$$

can be no more than

$$\begin{aligned} &|P(z \leq S_{n-1} + u \leq z + \lambda) - P(z \leq Z + u \leq z + \lambda)| \\ &\qquad + |P(z \leq Z + u \leq z + \lambda)|, \end{aligned}$$

which is bounded by

$$2\delta(\gamma, n-1) + \lambda/\sqrt{2\pi}.$$

Inductive Approach

Hence, with $c_X = E|X_n^* - X_n| \le (3/2)\gamma$, conditioning on X_n and X_n^*,

$$\lambda^{-1}\left|\int_{X_n/\sqrt{n}}^{X_n^*/\sqrt{n}} \mathbf{1}(z \le S_{n-1} + u \le z + \lambda)du\right|$$

is bounded by

$$\begin{aligned} & n^{-1/2}c_X\left(2\delta(\gamma, n-1)/\lambda + 1/\sqrt{2\pi}\right) \\ \le \quad & n^{-1/2}\gamma\left(3\delta(\gamma, n-1)/\lambda + 3/\sqrt{8\pi}\right). \end{aligned}$$

Putting everything together, $\delta(\gamma, n)$ is bounded by

$$\lambda/\sqrt{2\pi} + n^{-1/2}\gamma\left(3\delta(\gamma, n-1)/\lambda + 3/\sqrt{8\pi} + 3\right)$$

Inductive Approach

Choosing $\lambda = 3\beta\gamma/\sqrt{n}$ in

$$\lambda/\sqrt{2\pi} + n^{-1/2}\gamma\left(3\delta(\gamma, n-1)/\lambda + 3/\sqrt{8\pi} + 3\right)$$

yields, with $c = 3\beta/\sqrt{2\pi} + 3/\sqrt{8\pi} + 3$,

$$\delta(\gamma, n) \leq \delta(\gamma, n-1)/\beta + n^{-1/2}\gamma c.$$

Letting $a_n = \sqrt{n}\delta(\gamma, n)/\gamma$, and using that $\sqrt{n/(n-1)} \leq \sqrt{2}$ for all $n \geq 2$ we have

$$a_1 \leq 1 \quad \text{and} \quad a_n \leq \frac{\sqrt{2}}{\beta}a_{n-1} + c \quad \text{for all } n \geq 2.$$

Inductive Approach

With $\beta > 1$, choosing $\lambda = 3\beta\gamma/\sqrt{n}$ in

$$\lambda/\sqrt{2\pi} + n^{-1/2}\gamma\left(3\delta(\gamma, n-1)/\lambda + 3/\sqrt{8\pi} + 3\right)$$

yields, with $c = 3\beta/\sqrt{2\pi} + 3/\sqrt{8\pi} + 3 > 1$,

$$\delta(\gamma, n) \leq \delta(\gamma, n-1)/\beta + n^{-1/2}\gamma c.$$

Letting $a_n = \sqrt{n}\delta(\gamma, n)/\gamma$, and using that $\sqrt{n/(n-1)} \leq \sqrt{2}$ for all $n \geq 2$ we have

$$a_1 \leq 1 \quad \text{and} \quad a_n \leq \frac{\sqrt{2}}{\beta} a_{n-1} + c \quad \text{for all } n \geq 2.$$

Inductive Approach

Now taking $\beta > \sqrt{2}$ so that $\alpha = \sqrt{2}/\beta < 1$, we can show that if

$$a_n \leq \alpha a_{n-1} + c \quad \text{and} \quad a_1 \leq 1,$$

then

$$a_n \leq c_1 + c_2 \alpha^n \quad \text{where} \quad c_1 = \frac{c}{1-\alpha}, c_2 = \frac{1-\alpha-c}{\alpha(1-\alpha)} < 0$$

so that

$$a_n \uparrow \frac{c}{1-\alpha} = \frac{3\beta/\sqrt{2\pi} + 3/\sqrt{8\pi} + 3}{1-\sqrt{2}/\beta},$$

which has minimal value 12.969.

Recap

I. Background, Stein Identity, Equation, Bounds

II. Size Bias Couplings

III. Exchangeable Pair, Zero Bias Couplings

IV. Local dependence, Nonsmooth functions

Stein Mart
GRAND OPENING