Math/Stat 523, Spring 2020

Jon A. Wellner

Lecture 8

Wednesday April 22 and Friday April 24

Outline

- 1: Stein's method, the basics.
- 2: Stein's method and the zero-bias transformation.
- 3: More on Stein's method: exchangeable pairs.

1. Stein's method, the basics

This material is based on Chen, Goldstein, and Shao (2011), *Normal Approximation by Stein's Method*.

Our goal here is to introduce Stein's method in the context of convergence to Gaussian limit distributions. Our starting point is the following lemma characterizing the standard normal distribution.

Lemma 1. If $W \sim N(0,1)$, then

$$E[f'(W)] = E[Wf(W)] \tag{1}$$

for all absolutely continuous and piecewise differentiable functions f with $E|f'(Z)| < \infty$. Conversely, if (1) holds for all bounded, continuous, and piecewise differentiable functions f with $E|f'(Z)| < \infty$, then $W \sim N(0,1)$.

Lemma 2. Fix $z \in \mathbb{R}$ and let $\Phi(z) \equiv P(Z \le z)$ be the distribution function of a standard normal random variable Z. Then the unique solution $f \equiv f_z$ of

$$f'(w) - wf(w) = 1_{[w \le z]} - \Phi(z) \tag{2}$$

is given by

$$f_z(w) = \begin{cases} \sqrt{2\pi} e^{w^2/2} \Phi(w)(1 - \Phi(z)) & \text{if } w \le z, \\ \sqrt{2\pi} e^{w^2/2} \Phi(z)(1 - \Phi(w)) & \text{if } w > z. \end{cases} \tag{3}$$

$$= (1/\phi(w))\{\Phi(z \wedge w) - \Phi(z)\Phi(w)\}. \tag{4}$$

Suppose that we want to show that some sequence of random variables $\{W_n\}$ satisfies $W_n \to_d Z \sim N(0,1)$, i.e.

$$F_{W_n}(z) - \Phi(z) \to 0 \quad \text{for all } z \in \mathbb{R}.$$

But if f_z satisfies (2) then

$$F_{W_n}(z) - \Phi(z) = E\{1_{[W_n \le z]} - \Phi(z)\} = E\{f_z'(W_n) - W_n f_z(W_n)\}$$

and it might be easier to show that the right side in the last display converges to 0.

Lemma 3. (Properties of the solution). Let $z \in \mathbb{R}$ and let f_z be given by (3). Then $w f_z(w)$ is a non-decreasing function of w. Moreover, for all real w, u, v:

$$|w f_z(w)| \le 1, \qquad |w f_z(w) - u f_z(u)| \le 1,$$

$$|f_z'(w)| \le 1, \qquad |f_z'(w) - f_z'(u)| \le 1$$

$$0 \le f_z(w) \le \min\{\sqrt{2\pi}/4, 1/|z|\}$$

$$|(w+u) f_z(w+u) - (w+v) f_z(w+v)| \le (|w| + \sqrt{2\pi}/4)(|u| + |v|).$$

Proof of Lemma 1: *(Necessity)* Let f be an absolutely continuous function with $E|f'(Z)| < \infty$. If $W \sim N(0,1)$, then

$$\begin{aligned} Ef'(W) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(w) e^{-w^2/2} dw \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(w) \left(\int_{-\infty}^{w} (-x) e^{-x^2/2} dx \right) dw \\ &\quad + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} f'(w) \left(\int_{w}^{\infty} x e^{-x^2/2} dx \right) dw. \end{aligned}$$

By Fubini's theorem this yields

$$\begin{aligned} Ef'(W) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \left(\int_{x}^{0} f'(w) dw \right) (-x e^{-x^2/2}) dx \\ &\quad + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \left(\int_{0}^{x} f'(w) dw \right) (x e^{-x^2/2}) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (f(x) - f(0)) (x e^{-x^2/2}) dx \\ &= E[W f(W)]. \end{aligned}$$

(Sufficiency) The function f_z as given in (3) is clearly continuous and piecewise continuously differentiable. By Lemma 3, f_z is bounded. Hence if (1) holds for all bounded continuous and continuously differentiable functions, then by (2)

$$0 = E[f_z'(W) - Wf_z(W)] = E\{1_{[W \le z]} - \Phi(z)\} = P(W \le z) - \Phi(z).$$

Thus $W \sim N(0,1)$. $\square$

When f is absolutely continuous and bounded, (1) follows for $W \sim N(0,1)$ via integration by parts as follows:

$$\begin{aligned} E[Wf(W)] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} wf(w)e^{-w^2/2}dw \\ &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(w)d(e^{-w^2/2}) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(w)(e^{-w^2/2})dw = Ef'(W). \end{aligned}$$

Proof of Lemma 2: First multiply both sides of (2) by the integrating factor $e^{-w^2/2}$ to find that

$$\left(e^{-w^2/2} f(w)\right)' = e^{-w^2/2}\{1_{[w \le z]} - \Phi(z)\}.$$

Integrating both sides of this identity yields

$$\begin{aligned} f_z(w) &= e^{w^2/2} \int_{-\infty}^{w} \{1_{[x \le z]} - \Phi(z)\} e^{-x^2/2} dx \\ &= -\, e^{w^2/2} \int_{w}^{\infty} \{1_{[x \le z]} - \Phi(z)\} e^{-x^2/2} dx, \end{aligned}$$

which is equivalent to (3). Lemma 3 below shows that $f_z(w)$ is bounded.

The general solution of (2) is given by $f_z(w)$ plus some constant multiple, say $ce^{w^2/2}$, of the solution to the homogeneous equation. Hence the only bounded solution is the one with $c = 0$.
□

For a given real-valued and measurable function h with $E|h(Z)| < \infty$, let $Eh(Z)$ be denoted by Nh, and call

$$f'(w) - wf(w) = h(w) - Nh \tag{5}$$

the Stein equation for h, or simply the Stein equation. Note that (2) is the special case of (5) for $h(w) = 1_{[w \le z]}$. By the method of integrating factors that produced (3), it can be shown that the unique bounded solution of (5) is given by

$$\begin{aligned} f_h(w) &= e^{w^2/2} \int_{-\infty}^{w} (h(x) - Nh) e^{-x^2/2} dx \\ &= -e^{w^2/2} \int_{w}^{\infty} (h(x) - Nh) e^{-x^2/2} dx. \end{aligned} \tag{6}$$

Stein (1986) showed that when h is a bounded differentiable function with bounded derivative h', the solution f is twice differentiable and satisfies

$$\|f'\|_\infty \le 2\|h\|_\infty \quad \text{and} \quad \|f''\|_\infty \le 2\|h'\|_\infty. \tag{7}$$

2. Stein's method: the zero-bias transform

The material in this section is based primarily on Goldstein's (2009) paper. For the detailed reference see the last page of Lecture 8.

The zero-bias transformation: Let X be a random variable with $E(X) = 0$ and $Var(X) = E(X^2) < \infty$. Then X^* has the $X-$zero biased distribution F^* satisfying

$$E(Xf(X)) = \sigma^2 E(f'(X^*)) \tag{8}$$

for all absolutely continuous functions f for which these expectations exist.

Lemma 1. (Goldstein and Reinert (1997)) $X^* \stackrel{d}{=} X$ if and only if $X \sim N(0, \sigma^2)$.

Proof: Suppose that $X \sim N(0, \sigma^2)$. Note that the density of $X \sim N(0, \sigma^2)$ is $\varphi_{\sigma^2}(x) = \phi(x/\sigma)/\sigma$. It satisfies

$$\sigma^2 \varphi'_{\sigma^2}(x) = -x\varphi_{\sigma^2}(x).$$

Then

$$\begin{aligned} E[Xf(X)] &= \int x f(x)\varphi_{\sigma^2}(x)dx = -\sigma^2 \int f(x)\varphi'_{\sigma^2}(x)dx \\ &= \sigma^2 \int f'(x)\varphi_{\sigma^2}(x)dx \end{aligned}$$

by integration by parts. Thus (8) holds. □

Example: Suppose that $Y \sim$ Bernoulli(p). Then $X \equiv Y - p$ has $E(X) = 0$ and $Var(Y) = p(1-p)$. Then the left side of (8) becomes

$$\begin{aligned} E\{Xf(X)\} &= E\{(Y-p)f(Y-p)\} \\ &= p(1-p)f(1-p) + (-p)(1-p)f(-p) \\ &= \sigma^2\{f(1-p) - f(-p)\} \\ &= \sigma^2 \int_{-p}^{1-p} f'(u)du = \sigma^2 E\{f'(U)\} \end{aligned}$$

where $U \sim$ Uniform$(-p, 1-p)$. Thus $X^* = (Y-p)^* \stackrel{d}{=} U \sim$ Uniform (-p,1-p) . Note that while $F = \mathcal{L}(X)$ is discrete, $F^* = \mathcal{L}(X^*)$ is absolutely continuous. The following proposition shows that this is a general phenomenon.

Proposition 1: Suppose that X is a random variable with mean zero and finite positive variance σ^2. Then there exists a unique distribution for X^* such that

$$Ef'(X^*) = \sigma^2 E\{Xf(X)\}$$

for every absolutely continuous function f for which $E|Xf(X)| < \infty$. Moreover, the distribution of X^* is absolutely continuous with density

$$p^*(x) = E\left\{X1_{[X>x]}\right\}/\sigma^2 = -E\left\{X1_{[X\leq x]}\right\}/\sigma^2.$$

and distribution function

$$G^*(x) = E\{X(X-x)1_{[X\leq x]}\}/\sigma^2.$$

Proof: See Chen, Goldstein, and Shao (2011), page 27.

The characterization (8) specifies a relationship between the moments of X and the moments of X^*. Here is an interesting example: if $f(x) = (1/2)x^2\text{sign}(x)$, then $f'(x) = |x|$, and we therefore find that

$$\sigma^2 E|X^*| = 2^{-1} E|X|^3 \quad \text{where} \quad \sigma^2 = Var(X).$$

Thus $E|X|^3 < \infty$ if and only if $E|X^*| < \infty$. The most important property of the zero-bias transform is given in the following proposition.

Lemma 2. Let $X_{n,i}$, $i = 1, \ldots, n$, be independent mean zero random variables with $Var(X_{n,i}) = \sigma^2_{n,i}$ summing to 1. Let $X^*_{n,i}$ have the $X_{n,i}$-zero bias distribution with $X^*_{n,i}$ mutually independent and $X^*_{n,j}$ independent of the $X_{n,i}$'s. Furthermore, let I_n be a random index, independent of $X_{n,i}$, $X^*_{n,j}$, $i = 1, \ldots, n$, with distribution

$$P(I_n = i) = \sigma^2_{n,i}.$$

Then

$$W_n^* \stackrel{d}{=} W_n - X_{n,I_n} + X^*_{n,I_n}, \tag{9}$$

where W_n^* has the W_n-zero bias distribution.

Proof: For simplicity, we drop the n in the double indexing of the triangular array(s). First note that we can write

$$X_I = \sum_{i=1}^n 1_{[I=i]} X_i \qquad \text{and} \qquad X_I^* = \sum_{i=1}^n 1_{[I=i]} X_i^*.$$

Thus it is clear that

$$\mathcal{L}(X_I) = \sum_{i=1}^n \mathcal{L}(X_i)\sigma_i^2, \quad \text{and} \quad \mathcal{L}(X_I^*) = \sum_{i=1}^n \mathcal{L}(X_i^*)\sigma_i^2.$$

Now suppose that W^* has the $W-$zero bias distribution. Then for all absolutely continuous functions f for which the following expectations exist,

$$
\begin{aligned}
E\{f'(W^*)\} &= E\{Wf(W)\} = \sum_{i=1}^n E\{X_i f(W)\} \\
&= \sum_{i=1}^n E[X_i f(W - X_i + X_i)] \\
&= \sum_{i=1}^n \sigma_i^2 E\{f'(W - X_i + X_i^*)\} \\
&= \sum_{i=1}^n E\{f'(W - X_i + X_i^*) 1_{[I=i]} \\
&= E\{f'(W - X_I + X_I^*)\}.
\end{aligned}
$$

Thus $W^* \stackrel{d}{=} W - X_I + X_I^*$. □

The swapping identity (9) suggests that the CLT should hold when the random variables X_{n,I_n} and X^*_{n,I_n} are both small asymptotically, since then the distributions of W_n and W_n^* are close, and then W_n becomes an approximate fixed point of the zero-bias transformation. The following theorem makes this heuristic argument precise.

Theorem 1: Suppose that $\{X_{n,i} : 1 \le i \le n\}$ is a triangular array of row-independent random variables satisfying the hypotheses of Lemma 2: (Thus they satisfy "Condition 1.1" of Goldstein (2009).) Then the small zero-bias condition

$$X^*_{n,I_n} \to_p 0 \tag{10}$$

is equivalent to the Lindeberg condition: for every $\epsilon > 0$

$$L_n(\epsilon) \equiv \sum_{i=1}^{n} E\{X^2_{n,i} 1_{[|X_{ni}|>\epsilon]}\} \to 0.$$

Note: This is Goldstein's Theorem 1.1.

Goldstein's proof of this theorem proceeds by showing that the small zero-bias condition (10) implies that

$$X_{n,I_n} \to_p 0,$$

and hence

$$W_n^* - W_n = X_{n,I_n}^* - X_{n,I_n} \to_p 0.$$

Theorem 2: Suppose that $\{X_{n,i} : \ 1 \le i \le n\}$ is a triangular array of row-independent random variables satisfying the hypotheses of Lemma 2; thus they satisfy "Condition 1.1" of Goldstein. Suppose that the small zero bias condition (10) holds. Then $W_n \to_d Z \sim N(0,1)$.

Thus the Lindeberg condition (or, equivalently, the small zero bias condition) is sufficient for convergence to Z, but it is not necessary without some "uan condition" to insure that the individual $X_{n,i}$'s are "uniformly asymptotically negligible". Here is the resulting (partial) converse of Theorem 2.

Theorem 3: Suppose that $\{X_{n,i}: \; 1 \le i \le n\}$ is a triangular array satisfying "Condition 1.1" and

$$m_n \equiv \max_{1 \le i \le n} \sigma^2_{n,i} \to 0.$$

Then $W_n \to_d Z$ implies that $X^*_{n,I_n} \to_p 0$.

Here is an outline of the proof of Theorem 3:

- First show that $W_n \to_d Z$ implies that $W^*_n \to_d Z$.
- Next, show that $W^*_n \to_d Z$ and $m_n \to 0$ implies $X_{n,I_n} \to_p 0$.
- Then it follows that $W_n + X^*_{n,I_n} = W^*_n + X_{n,I_n} \to_d Z$.
- Finally, $W_n \to_d Z$ and $W_n + X^*_{n,I_n} \to_d Z$ imply $X^*_{n,I_n} \to_p 0$.

Proof of Theorem 1: Since the random index I_n is independent of the X_{ni}'s and the $X^*_{n,i}$'s,

$$Ef(X_{n,I_n}) = \sum_{i=1}^n \sigma^2_{n,i} Ef(X_{n,i}) \quad \text{and} \quad Ef(X^*_{n,I_n}) = \sum_{i=1}^n \sigma^2_{n,i} Ef(X^*_{n,i}). \tag{1.11}$$

Let $\epsilon > 0$ and set

$$f(x) = (x+\epsilon)1_{[x \le -\epsilon]} + 0 \cdot 1_{[|x|<\epsilon]} + (x-\epsilon)1_{[x \ge \epsilon]}.$$

and note that

$$xf(x) = (x^2 - \epsilon x)1_{[x \ge \epsilon]} + (x^2 - \epsilon|x|)1_{[-x \ge \epsilon]}.$$

The function f is absolutely continuous with $f'(x) = 1_{[|x| \ge \epsilon]}$ almost everywhere. Hence, using the definition of the zero bias transform in the second equality

$$\begin{aligned}\sigma_{n,i}^2 P(|X_{n,i}^*| \geq \epsilon) &= \sigma_{n,i}^2 E f'(X_{n,i}^*) \\ &= E[(X_{n,i}^2 - \epsilon|X_{n,i}|)1_{[|X_{n,i}|\geq\epsilon]}] \\ &\leq E[(X_{n,i}^2 + \epsilon|X_{n,i}|)1_{[|X_{n,i}|\geq\epsilon]}] \\ &\leq 2E[X_{n,i}^2 1_{[|X_{n,i}|\geq\epsilon]}].\end{aligned}$$

Summing over $i = 1, \ldots, n$ and applying (11) for the indicator function $1_{[|x|\geq\epsilon]}$ yields

$$P(|X_{n,I_n}^*| \geq \epsilon) = \sum_{i=1}^{n} \sigma_{n,i}^2 P(|X_{n,i}^*| \geq \epsilon) \leq 2L_n(\epsilon)$$

Thus the Lindeberg condition implies the small zero-bias condition $X_{n,I_n}^* \to_p 0$.

For the argument in the other direction, first note that for all x,

$$x^2 1_{[|x|\geq\epsilon]} \leq 2\left(x^2 - \frac{\epsilon}{2}|x|\right) 1_{[|x|\geq\epsilon/2]}.$$

See Figure 1. This yields

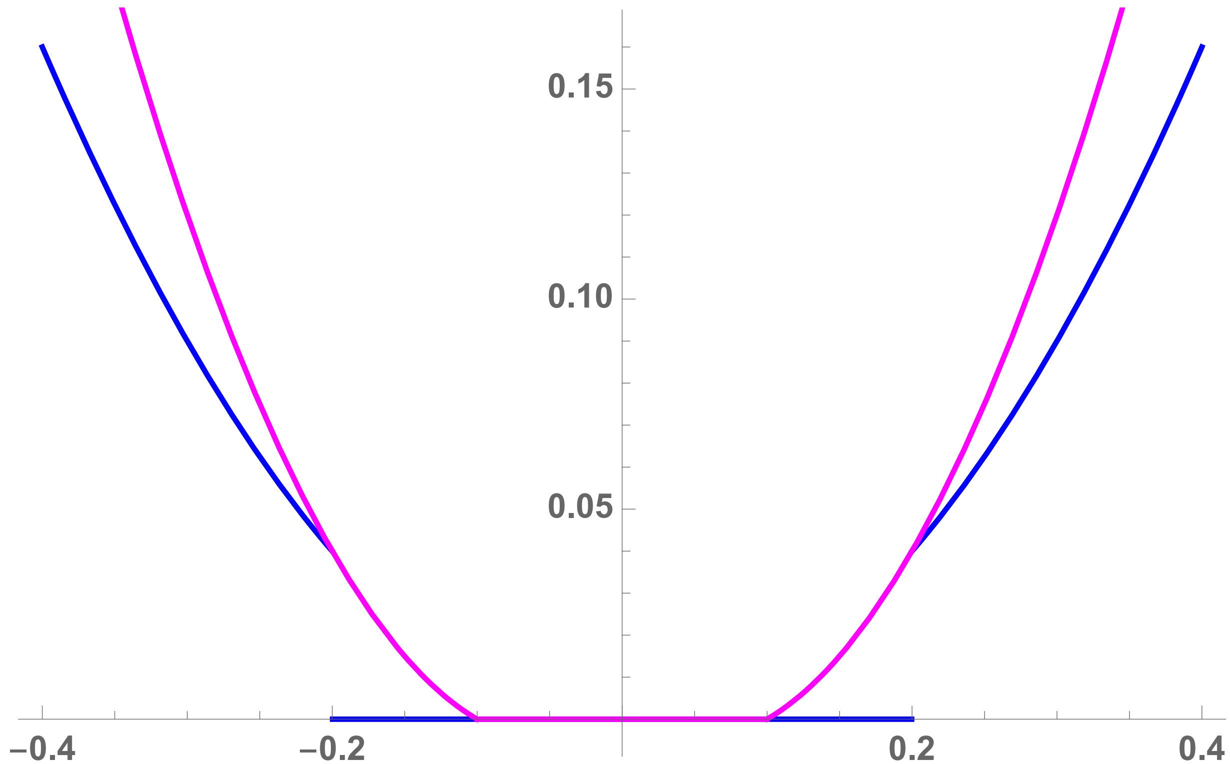

Figure 1: Truncated quadratic (blue) and bound (magenta)

$$
\begin{aligned}
L_n(\epsilon) &= \sum_{i=1}^{n} E\{X_{n,i}^2 1_{[|X_{n,i}|\ge\epsilon]}\} \\
&\le 2\sum_{i=1}^{n} E\left(X_{n,i}^2 - \frac{\epsilon}{2}|X_{n,i}|\right) 1_{[|X_{n,i}|\ge\epsilon/2]} \\
&= 2\sum_{i=1}^{n} \sigma_{n,i}^2 P\left(|X_{n,i}^*| \ge \frac{\epsilon}{2}\right) \\
&= 2P\left(|X_{n,I_n}^*| \ge \frac{\epsilon}{2}\right)
\end{aligned}
$$

and hence the small zero-bias condition implies the Lindeberg condition. □

In preparation for the forward part of the Lindeberg-Feller CLT, we need the following two lemmas:

Lemma 1. (This is Lemma 4.1 of Goldstein (2009).) Suppose that $\{X_{n,i} : 1 \le i \le n\}$ satisfies Condition 1.1 and let $m_n \equiv \max_{i\le n} \sigma_{n,i}^2$. Then $m_n \to 0$ implies that $X_{n,I_n} \to_p 0$.

Lemma 2. (This is Lemma 4.2 of Goldstein (2009).) Suppose that $\{X_{n,i} : 1 \le i \le n\}$ satisfies Condition 1.1. If the small zero bias condition $X^*_{n,I_n} \to_p 0$ holds, then $X_{n,I_n} \to_p 0$.

Proof of Lemma 1: From (11) with $f(x) = x$ we find that $EX_{n,I_n} = 0$, and hence $Var(X_{n,I_n}) = EX^2_{n,I_n}$. By (11) again but now with with $f(x) = x^2$ yields

$$Var(X_{n,I_n}) = \sum_{i=1}^{n} \sigma^4_{n,i}.$$

But now note that for all $i \le n$, $\sigma^4_{n,i} \le \sigma^2_{n,i} \max_{1\le j\le n} \sigma^2_{n,j} = \sigma^2_{n,i} m_n$. Hence for any $\epsilon > 0$, Chebyshev's inequality and Condition 1.1 give

$$\begin{aligned} P(|X_{n,I_n}| \ge \epsilon) &\le \frac{Var(X_{n,I_n})}{\epsilon^2} \le \frac{m_n}{\epsilon^2} \sum_{i=1}^{n} \sigma^2_{n,i} \\ &= \frac{1}{\epsilon^2} m_n \to 0 \end{aligned}$$

and hence $X_{n,I_n} \to_p 0$. $\square$

Proof of Lemma 2: For $n \geq 1$, $1 \leq i \leq n$, and $\epsilon > 0$,

$$\begin{aligned} \sigma_{n,i}^2 &= E(X_{n,i}^2 1_{[|X_{n,i}|<\epsilon]}) + E(X_{n,i}^2 1_{[|X_{n,i}|\geq\epsilon]}) \\ &\leq \epsilon^2 + L_n(\epsilon). \end{aligned}$$

It follows that

$$m_n = \max_{1\leq i\leq n} \sigma_{n,i}^2 \leq \epsilon^2 + L_n(\epsilon).$$

Since $\{X_{ni} : \ 1 \leq i \leq n\}$ satisfies the small zero bias condition, theorem 1 implies that

$$\limsup_{n\to\infty} m_n \leq \epsilon^2,$$

and hence $m_n \to 0$. Thus the claim follows from Lemma 1. $\square$

Proof of Theorem 1.2 Let $h \in C_{c,0}^{\infty}$, the set of all functions with compact support which integrate to zero and have derivatives of all orders, since $C_{c,0}^{\infty} \subset C_b$, the set of all bounded and continuous functions on $\mathbb{R}$, and let $f \equiv f_h$ be the solution to the Stein equation for h given by (6). Taking $w = W_n$ in (5), taking expectations, and using (8) we obtain

$$\begin{aligned} E[h(W_n) - Eh(Z)] &= E[f'(W_n) - W_n f(W_n)] \\ &= E[f'(W_n) - f'(W_n^*)] \end{aligned} \tag{12}$$

with W_n^* given by (9). Since

$$W_n^* - W_n = X_{n,I_n}^* - X_{n,I_n},$$

the small zero bias condition and Lemma 2 imply that

$$W_n^* - W_n \to_p 0.$$

By (7) f' is bounded with a bounded derivative f'', hence its global modulus of continuity

$$\eta(\delta) \equiv \sup_{|y-x| \le \delta} |f'(y) - f'(x)|$$

is bounded and satisfies $\lim_{\delta \searrow 0} \eta(\delta) = 0$. Now by (13)

$$\eta(|W_n^* - W_n|) \to_p 0, \tag{13}$$

and by (12) and the triangle inequality

$$\begin{aligned} |Eh(W_n) - Eh(Z)| &= |E(f'(W_n) - f'(W_n^*)| \\ &\leq E|f'(W_n) - f'(W_n^*)| \\ &\leq E\eta(|W_n - W_n^*|). \end{aligned}$$

Therefore

$$Eh(W_n) - Eh(Z) \to 0 \tag{14}$$

by (13) and the bounded convergence theorem. Thus an application of the Helly-Bray theorem completes the proof. □

Now our goal is to prove the converse half of the Lindeberg-Feller theorem. To do this we need two lemmas as follows:

Lemma 5.1. Let U_n and V_n be two sequences of random variables such that U_n and V_n are independent for every n. Then $U_n \to_d U$ and $U_n + V_n \to_d U$ implies $V_n \to_p 0$.

Lemma 5.2. Let Y and Y_n be mean zero random variables with finite non-zero variances $\sigma^2 = Var(Y)$ and $\sigma_n^2 = Var(Y_n)$. If $Y_n \to_d Y$ and $\sigma_n^2 \to \sigma^2$, then $Y_n^* \to_d Y^*$.

Remark: The hypotheses of Lemma 5.2 also imply that $W_2(F_n, F) \to 0$ where d_2 denotes the Wasserstein metric of order 2 given by

$$\begin{aligned} d_2^2(F,G) &= \inf\Big\{E|X-Y|^2 : \text{ all joint laws of } (X,Y) \\ &\qquad \text{with marginals } F \text{ and } G\} \\ &= \int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt. \end{aligned}$$

s

Proof of Lemma 5.2: Let $f \in C_{c,0}^{\infty}$ and $F(y) = \int_{-\infty}^{y} f(t)dt$. Since Y and Y_n both have mean zero and finite variances, their zero bias distributions exists; in particular

$$\sigma_n^2 Ef(Y_n^*) = E\Big[Y_n F(Y_n)\Big] \quad \text{for all} \quad n.$$

By the Helly-Bray theorem, since $yF(y) \in C_b$ it follows that

$$\begin{aligned} \sigma^2 \lim_n Ef(Y_n^*) &= \lim_n Ef(Y_n^*) \\ &= \lim_n E[Y_n F(Y_n)] = E[YF(Y)] = \sigma^2 Ef(Y^*). \end{aligned}$$

Hence $Ef(Y_n^*) \to Ef(Y^*)$ for all $f \in C_{c,0}^{\infty}$, so $Y_n^* \to_d Y^*$ by Theorem 2.1. (Appendix) □

Now we are ready for the proof of Theorem 1.3.

Proof of Theorem 1.3: Since $W_n \to Z$ and $Var(W_n) \to Var(Z) = 1$ (where the sequence of variances and the limit are both identically 1), Lemma 5.2 implies $W_n^* \to_d Z^* \overset{d}{=} Z$. But Z is a fixed point of the zero bias transformation, hence $W_n^* \to_d Z$.

Since $m_n \to 0$, Lemma 4.1 shows that $X_{n,I_n} \to_p 0$, and Slutsky's lemma implies that

$$W_n + X_{n,I_n}^* = W_n^* + X_{n,I_n}^* \to_d Z.$$

Hence

$$W_n \to_d Z \quad \text{and} \quad W_n + X_{n,I_n}^* \to_d Z.$$

Since W_n is a function of the $X_{n,i}$'s which is independent of I_n and $\{X_{n,i}^* : \ 1 \le i \le n\}$ and therefore is independent of X_{n,I_n}^*, Lemma 5.1 yields $X_{n,I_n}^* \to_p 0$. □

3. More on Stein's method: exchangeable pairs

Now suppose that W is an arbitrary random variable, not necessarily a sum. Several variants of Stein's method depend on introduction of an auxiliary random variable coupled to W possessing certain properties. The exchangeable pair approach introduced by Stein (1972, 1986) involves another random variable W' on the same probability space as W in such a way that (W, W') is an exchangeable pair: that is, such that $(W, W') \stackrel{d}{=} (W', W)$. The exchangeable pair approach then exploits the easy fact that if (W, W') is an exchangeable pair, then

$$E\{g(W, W')\} = 0 \qquad (15)$$

for all antisymmetric measurable functions $g(x, y)$ such that the expected value exists. [Recall that $g : \mathbb{R}^2 \to \mathbb{R}$ is antisymmetric if $g(v, u) = -g(u, v)$. Thus if (W, W') is exchangeable, $Eg(W, W') = E\{-g(W', W)\} = -Eg(W', W) = -Eg(W, W')$.]

Definition: If the pair (W, W') is exchangeable and satisfies the "linear regression condition"

$$E(W'|W) = (1 - \lambda)W \quad (16)$$

with $\lambda \in (0, 1)$, then we call (W, W') a $\lambda-$Stein pair, or simply a Stein pair.

This is analogous to the conditional expectation property that holds for the bivariate normal distribution, namely if $(Z, Z') \sim N_2(\underline{\mu}, \Sigma)$, then

$$E(Z'|Z) = \mu_1 + \sigma_1 \rho \left(\frac{Z - \mu_2}{\sigma_2} \right)$$

where σ_1^2, σ_2^2 are the variances of Z' and Z respectively and ρ is the correlation coefficient. When the means are zero and the variances are equal this becomes

$$E(Z'|Z) = (1-\lambda)Z$$

with $\lambda = 1-\rho$. Here is a key lemma concerning exchangeable pairs:

Lemma 2.7 Let (W, W') be a Stein pair and $\Delta \equiv W - W'$. Then

$$EW = 0 \quad \text{and} \quad E\Delta^2 = 2\lambda E(W^2) \ \text{ if } E(W^2) < \infty. \tag{17}$$

Furthermore, when $EW^2 < \infty$, for every absolutely continuous function f satisfying $|f(w)| \le C(1 + |w|)$, we have

$$E[Wf(W)] = \frac{1}{2\lambda} E\left\{(W - W')(f(W) - f(W'))\right\}, \tag{18}$$

$$E[Wf(W)] = E\left\{\int_{-\infty}^{\infty} f'(W+t)\widehat{K}(t)dt\right\} \tag{19}$$

and

$$E[f'(W) - E(Wf(W))]$$
$$= \quad Ef'(W)\left(1 - \frac{\Delta^2}{2\lambda}\right) + E\int_{-\infty}^{\infty}(f'(W) - f'(W+t))\widehat{K}(t)dt \tag{20}$$

where

$$\widehat{K}(t) = \frac{\Delta}{2\lambda}\left\{1_{[-\Delta \le t \le 0]} - 1_{[0 < t \le -\Delta]}\right\} \tag{21}$$

satisfies

$$\int_{-\infty}^{\infty}\widehat{K}(t)dt = \frac{\Delta^2}{2\lambda}. \tag{22}$$

Proof: Taking the expectation in (16) yields, by exchangeability, $EW = EW' = (1-\lambda)EW$, so $EW = 0$. Moreover, since

$$E(WW') = E\left\{E(W'W|W)\right\} = E\left\{WE(W'|W)\right\} = (1-\lambda)EW^2,$$

we have

$$E(W' - W)^2 = 2EW^2 - 2EW'W = 2\lambda EW^2.$$

Now we can use (15) with the antisymmetric function $g(x,y) = (x-y)(f(y)+f(x))$, noting that $Eg(W,W')$ exists because of the growth condition on f. Then (15) yields

$$\begin{aligned} 0 &= E\left\{(W-W')(f(W')+f(W))\right\} \\ &= E\left\{(W-W')(f(W')-f(W))\right\} + 2E\{f(W)(W-W')\} \\ &= E\left\{(W-W')(f(W')-f(W))\right\} + 2E\{f(W)E(W-W'|W)\} \\ &= E\left\{(W-W')(f(W')-f(W))\right\} + 2\lambda E\{Wf(W)\} \end{aligned}$$

where the last equality holds by the linear regression condition (16). Rearranging this equality yields (18), and then

$$
\begin{aligned}
E[Wf(W)] &= \frac{1}{2\lambda}E\left\{(W-W')(f(W)-f(W'))\right\} \\
&= \frac{1}{2\lambda}E\left\{\Delta(f(W)-\Delta))\right\} \\
&= \frac{1}{2\lambda}E\int_{-\Delta}^{0}\Delta f'(W+t)dt \\
&= E\int_{-\infty}^{\infty} f'(W+t)\widehat{K}(t)dt. \qquad (23)
\end{aligned}
$$

This proves (19). Note that integrating (21) yields (22); so to prove (20) we just need to note that

$$
Ef'(W) = E\left\{f'(W)\left(1-\frac{\Delta^2}{2\lambda}\right)\right\} + E\left\{\int_{-\infty}^{\infty} f'(W)\widehat{K}(t)dt\right\}
$$

and combine this with (23). □

Example 1: (Independent random variables). Let $\{\xi_i,\ 1 \le i \le n\}$ be independent rv's with 0 means and $\sum_1^n E\xi_i^2 = 1$. Let $W \equiv \sum_1^n \xi_i$. Let $\{\xi_i',\ i = 1,\dots,n\}$ be an independent copy of $\{\xi_i,\ i = 1,\dots,n\}$, and let I have a uniform distribution on $\{1,\dots,n\}$, independent of $\{\xi_i, \xi_i' :\ i = 1,\dots,n\}$. Define $W' = W - \xi_I + \xi_I'$. Then (W, W') is an exchangeable pair, and a straight-forward computation yields

$$E(W'|W) = \left(1 - \frac{1}{n}\right) W.$$

Thus (16) holds with $\lambda = 1/n$.

Example 2: (Exchangeable pair by substitution). Let $W \equiv g(\xi_1,\dots,\xi_n)$ and let ξ_i' have the conditional distribution of ξ_i given ξ_j, $1 \le j \ne i \le n$. Let I be a random index uniformly distributed over $\{1,\dots,n\}$, independent of $\{\xi_i, \xi_i' :\ i = 1,\dots,n\}$. Define $W' \equiv g(\xi_1,\dots,\xi_{I-1},\xi_I',\xi_{I+1},\dots,\xi_n)$. Then $W, W')$ is an exchangeable pair. We note that unlike Example 1, the linear regression condition (16) is not satisfied automatically.

Example 3: (Combinatorial CLT) Let $\{a_{i,j}\}_{1\le i,j\le n}$ be a given array of real numbers, and let $\pi = \pi'$ be a random permutation of $\{1, \ldots, n\}$. Then let

$$Y' \equiv \sum_{i=1}^{n} a_{i,\pi'(i)}. \tag{24}$$

Here we assume that π has a uniform distribution on the symmetric group $\mathcal{S}_n$. Let

$$a_{\cdot,\cdot} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{ij}, \quad a_{i,\cdot} = \frac{1}{n}\sum_{j=1}^{n} a_{ij}, \quad \text{and} \quad a_{\cdot,j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij}. \tag{25}$$

Since π' is uniformly distributed, it follows easily that $EY' = na_{\cdot,\cdot} = \sum_i a_{i\cdot} = \sum_j a_{\cdot j}$, and therefore

$$Y' - EY' = \sum_{i=1}^{n} (a_{i,\pi(i)} - a_{\cdot\cdot}) = \sum_{i=1}^{n} (a_{i,\pi(i)} - a_{i,\cdot} - a_{\cdot,\pi(i)} + a_{\cdot\cdot}). \quad (26)$$

Since our goal is to derive bounds for convergence of the distribution of the standardized variable $(Y' - E(Y'))/\sqrt{Var(Y')}$ to a Gaussian limit law, with loss of generality we may replace $a_{i,j}$ by $a_{i,j} - a_{i\cdot} - a_{\cdot,j} + a_{\cdot,\cdot}$ and assume

$$a_{i,\cdot} = a_{\cdot,j} = a_{\cdot,\cdot} = 0.$$

Let $\tau_{i,j}$ which transposes i and j, $\pi'' = \pi'\tau_{i,j}$, and let Y'' be given by (24) with π' replaced by π''. Since $\pi''(k) = \pi'(k)$ for $k \notin \{i, j\}$ while $\pi''(i) = \pi'(j)$, and $\pi''(j) = \pi'(i)$, we have

$$Y'' - Y' = (b(i, j, \pi(i), \pi(j)),$$

where $b(i, j, k, l) = a_{i,l} + a_{j,k} - (a_{i,k} + a_{j.l})$.

Now take (I, J) to be independent of π' with the uniform distribution over all pairs satisfying $1 \leq I \neq J \leq n$, the permutations π' and $\pi'' = \tau_{I,J}\pi'$ are exchangeable, and hence so are Y and Y'.

To prove that the linear regression property (16) is satisfied, note that

$$Y'' - Y' = (a_{I,\pi'(J)} + a_{J,\pi'(I)}) - (a_{I,\pi'(I)} + a_{J,\pi'(J)}). \qquad (27)$$

Taking the conditional expectation given π' and using (27), we find

$$
\begin{aligned}
E(Y'' - Y'|\pi') &= 2\left(-\frac{1}{n}\sum_{i=1}^{n} a_{i,\pi'(i)} + \frac{1}{n(n-1)}\sum_{i\neq j} a_{i,\pi'(j)}\right)\\
&= -2\left(\frac{1}{n}\sum_{i=1}^{n} a_{i,\pi'(i)} + \frac{1}{n(n-1)}\sum_{i=1}^{n} a_{i,\pi'(j)}\right)\\
&= -\frac{2}{n-1}Y'.
\end{aligned}
$$

Since the right side is measurable with respect to Y', we conclude that

$$E(Y''|Y') = \left(1 - \frac{2}{n-1}\right)Y',$$

so Y', Y'' is a $2/(n-1)$-Stein pair.

A Special Case: If $a_{ij} = b_i c_j$ where $b_1, \ldots, b_n$ are any real numbers and $c_j \in \{0, 1\}$ satisfy $\sum_{j=1}^n c_j = m$, then Y represents the sum in simple random sampling (without replacement) with sample size m from a population consisting of $\{b_1, \ldots, b_n\}$.

Return to Example 1: (Sums of independent rv's again). Suppose that $\xi_1, \ldots, \xi_n$ are independent rv's with $E\xi_i = 0$ and $\sum_{i=1}^n E\xi_i^2 = 1$. Let

$$W_n \equiv \sum_1^n \xi_i, \quad \text{and} \quad W^{(i)} \equiv W - \xi_i.$$

and define

$$K_i(t) \equiv E\{\xi_i(1_{[0 \le t \le \xi_i]} - 1_{[\xi_i \le t < 0]})\}. \tag{28}$$

Note that $K_i(t) \ge 0$ for all $t \in \mathbb{R}$ and that

$$\int_{-\infty}^{\infty} K_i(t)dt = E(\xi_i^2) \quad \text{and} \quad \int_{-\infty}^{\infty} |t| K_i(t)dt = 2^{-1} E|\xi_i|^3. \tag{29}$$

Now let h be a measurable function with $E|h(Z)| < \infty$ and let $f = f_h$ be the corresponding solution of the Stein equation. The goal is to estimate (or bound, or show convergence of)

$$Eh(W) - Eh(Z) = E\{f'(W) - Wf(W)\}.$$

The following argument is key to the $K-$function approach.

Since ξ_i and $W^{(i)}$ are independent for each $1 \le i \le n$, it follows that

$$E[Wf(W)] = \sum_{i=1}^{n} E[\xi_i f(W)] = \sum_{i=1}^{n} E\{\xi_i[f(W) - f(W^{(i)})]\}$$

where the second equality follows because $E\xi_i = 0$. Writing the last expression in integral form yields

$$
\begin{aligned}
E[Wf(W)] &= \sum_{i=1}^{n} E\left\{\xi_i \int_0^{\xi_i} f'(W^{(i)}+t)dt\right\} \\
&= \sum_{i=1}^{n} E\left\{\int_{-\infty}^{\infty} f'(W^{(i)}+t)\xi_i(1_{[0\le t\le \xi_i]} - 1_{[\xi_i \le t<0]})dt\right\} \\
&= \sum_{i=1}^{n} \int_{-\infty}^{\infty} E\left\{f'(W^{(i)}+t)\right\} K_i(t)dt, \qquad (30)
\end{aligned}
$$

from the definition of K_i and again using independence. But from

$$
\sum_{i=1}^{n} \int_{-\infty}^{\infty} K_i(t)dt = \sum_{i=1}^{n} E\xi_i^2 = 1, \qquad (31)
$$

it follows that

$$
Ef'(W) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} E\{f'(W)\}K_i(t)dt. \qquad (32)
$$

Combining (30) and (32) yields

$$
\begin{aligned}
& E[f'(W) - Wf(W)] \\
&= \sum_{i=1}^{n} \int_{-\infty}^{\infty} E\left\{f'(W) - f'(W^{(i)} + t)\right\} K_i(t)dt \qquad (33)
\end{aligned}
$$

Since $K_i(t)$ is non-negative and $\int_{\mathbb{R}} K_i(t)dt = E\xi_i^2$, the ratio $K_i(t)/E(\xi_i^2) \equiv g_i(t)$ can be regarded as a probability density function. Let ξ_i^*, $i = 1, \ldots, n$ be independent rv's, independent of ξ_j^2 for $j \neq i$ with density g_i for each i. Let I be a random index, independent of $\{\xi_i, \xi_i^* : \; i = 1, \ldots, n\}$ with distribution $P(I = i) = E\xi_i^2$. Then (30) can be written as

$$E[Wf(W)] = Ef'(W^{(I)} + \xi_I^*)$$

and (33) can be written as

$$E[f'(W) - Wf(W)] = E\left\{f'(W) - f'(W^{(I)} + \xi_I^*)\right\}.$$

Further results: (Sections in Chen, Goldstein, and Shao).

- Section 2.2, Multivariate Stein equation
- Section 2.3.2, last two paragraphs
- Section 4.4, Combinatorial CLTs
- Section 4.5, Simple random sampling
- Section 6.1, Combinatorial CLTs .

References and further reading:

- Goldstein, L. (2009). A probabilistic proof of the Lindeberg-Feller central limit theorem. *American Math. Monthly* **116**, 45 - 60.

- Ledoux, M., Nourdin, I., and Pecatti, G. (2015). Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25**, 256 - 306.

- Saumard, A. (2019). Weighted Poincaré inequalities, concentration inequalities, and tail bounds related to Stein kernels in dimension one. *Bernoulli* **25**, 3978 - 4006.

- Stein, C. (1986). Approximate Computation of Expectations, IMS Lecture Notes / Monograph Series. Hayward.