Math/Stat 523, Spring 2020

Jon A. Wellner

Lecture 9

Monday April 27 and Wednesday April 29

Outline

- 0: C-G-S (2011), Proposition 2.4
- 1: Stein's method for sums of independent rv's via Lipschitz functions (C-G-S Chapter 3)
  - ▷ Berry-Esseen bounds, bounded case
  - ▷ Berry-Esseen bounds, unbounded case
- 2: Stein's method: L_1 bounds.
  (C-G-S Chapter 4)
  - ▷ Combinatorial CLT's (C-G-S Section 4.4
  - ▷ Sampling without replacement (C-G-S Section 4.5)
- 3: Stein's method: Kolmogorov/L^∞ metric
  (C-G-S Section 6.1)

0. Proposition 2.4

The following two general identities can be exploited to yield normal approximations in the sense of comparison of expectations for Lipschitz functions: The first of these is:

$$E[Wf(W)] \;=\; E\int_{\mathbb{R}} f'(W+t)\widehat{K}(t)dt + E[Rf(W)], \qquad (2.76)$$

The second identity is,

$$E[Wf(W) = Ef'(W+\Delta) \qquad (2.77)$$

where $\Delta \equiv W^* - W$. The following proposition summarizes the general bounds which follow from these two identities:

Proposition 2.4 Let h be an absolutely continuous function with $\|h'\|_\infty < \infty$ and $\mathcal{F}$ any $\sigma-$field containing $\sigma[W]$.
(i) If (2.76) holds, then

$$|Eh(W) - Eh(Z)| \;\leq\; \|h'\| \left\{ \sqrt{\frac{2}{\pi}} E|1 - \widehat{K}_1| + 2E\widehat{K}_2 + 3E|R| \right\}, \qquad (2.78)$$

where

$$\widehat{K}_1 = E\left\{ \int_{\mathbb{R}} \widehat{K}(t)dt \middle| \mathcal{F} \right\} \quad \text{and} \quad \widehat{K}_2 = \int_{\mathbb{R}} |t\widehat{K}(t)|dt.$$

(ii) If (2.77) holds, then, with $\Delta \equiv W^* - W$,

$$|Eh(W) - Eh(Z)| \leq 2\|h'\| E|\Delta|. \qquad (2.79)$$

Proof: Here we will prove (2.79); for the proof of (2.78), see C-G-S (2011).

Proof of (2.79): By (2.77) we have, with $\Delta \equiv W^* - W$,

$$
\begin{aligned}
|Eh(W) - Eh(Z)| &= |E\left(f_h'(W) - W f_h(W)\right)| \\
&= |E\left(f_h'(W) - f_h'(W + \Delta)\right)| \\
&\leq \|f_h''\|_\infty E|\Delta| \\
&\leq 2\|h'\|E|\Delta|
\end{aligned}
$$

where the last inequality follows from (2.13). $\square$

1. Stein's method for sums of independent rv's

This material is based on Chen, Goldstein, and Shao (2011), Chapter 3.

With the basic facts about Stein's method in hand, we can begin to formulate some precise bounds by using the properties of solutions of the Stein equation given by Lemmas 3 and 4 of Lecture 8.

Notation and assumptions: Here we assume that $\xi_1, \ldots, \xi_n$ are independent rv's, $W \equiv \sum_{i=1}^n \xi_i$, and

$$E\xi_i = 0, \quad 1 \le i \le n, \quad \text{and} \quad \sum_{i=1}^n Var(\xi_i) = 1. \tag{1}$$

The supremum, or L^∞, or Kolmogorov distance between two distribution functions F and G is given by

$$\|F - G\|_\infty = \sup_{z \in \mathbb{R}} |F(z) - G(z)| = \sup_{h \in \mathcal{H}} |\int h dF - \int h dG|$$

where $\mathcal{H} = \{h_z(x) = 1_{(-\infty,z]}(x) : \quad z \in \mathbb{R}\}$.

One of our main goals here is to (re-prove) the Berry-Esseen theorem: for $F \equiv F_W$ and $G = \Phi = P(Z \le \cdot)$ where $Z \sim N(0,1)$,

$$\|F - \Phi\|_\infty \le C \sum_{i=1}^{n} E\|\xi_i\|^3.$$

where C is an absolute constant.

Recall that $h : \mathbb{R} \to \mathbb{R}$ is a Lipschitz function if there exists a constant K such that

$$|h(x) - h(y)| \le K|x - y| \quad \text{for all} \quad x, y \in \mathbb{R}.$$

Equivalently, h is Lipschitz - continuous if and only if h is absolutely continuous if and only if h is absolutely continuous with $\|h'\|_\infty < \infty$.

Theorem: Let $W = \sum_1^n \xi_i$ be the sum of mean zero independent random variables ξ_i, $1 \le i \le n$ with $\sum_1^n Var(\xi_i) = 1$, and let h be a Lipschitz continuous function. If $E|\xi_i|^3 < \infty$ for $i = 1, \ldots, n$, then

$$|Eh(W) - Eh(Z)| \le 3\gamma \|h'\|_\infty \tag{2}$$

where

$$\gamma \equiv \sum_1^n E|\xi_i|^3. \tag{3}$$

Proof: By Lemma 2.8,

$$E\{Wf(W)\} = E\{f'(W + \Delta) \quad \text{with} \quad \Delta \equiv \xi_I^* - \xi_I$$

where ξ_i^* has the ξ_i−zero bias distribution and is independent of ξ_j, $j \neq i$, and I is a random index with distribution $P(I = i) = E\xi_i^2$ independent of all other variables. It then follows from (2.79) of Proposition 2.4 that

$$
\begin{aligned}
|Eh(W) - Eh(Z)| &\le 2\|h'\| E|\xi_I^* - \xi_I| \\
&= 2\|h'\| \sum_1^n E\xi_i^2 E|\xi_i^* - \xi_i| \\
&\le 2\|h'\| \sum_1^n \left(E|\xi_i^2 E|\xi_i^*| + E|\xi_i| E\xi_i^2 \right) \\
&= 2\|h'\| \sum_1^n \left(\frac{1}{2} E|\xi_i|^3 + E|\xi_i| E\xi_i^2 \right) \\
&\le 3\|h'\| \sum_1^n E|\xi_i|^3.
\end{aligned}
$$

where we used $\sigma^2 E|X^*| = 2^{-1} E|X|^3$ (this is (2.57)) and Liapunov's inequality: $E|\xi_i| \le (E|\xi_i|^3)^{1/3}$ and $E|\xi_i|^2 \le (E|\xi_i|^3)^{2/3}$.

The constant 3 can be improved to 1; see Corollary 4.2.

The following theorem provides a bound for $|Eh(W) - Eh(Z)|$ in terms of the truncated second and third moments, thereby avoiding the existence of 3rd moments,

$$\beta_2 \equiv \sum_{i=1}^{n} E\xi_i^2 1_{[|\xi_i|>1]} \quad \text{and} \quad \beta_3 \equiv \sum_{i=1}^{n} E|\xi_i|^3 1_{[|\xi_i|\le 1]}.$$

Theorem 3.2: If $W = \sum_1^n \xi_i$ is the sum of mean zero independent random variables, $1 \le i \le n$ with $\sum_1^n Var(\xi_i) = 1$, then for any Lipschitz function h

$$|Eh(W) - Eh(Z)| \le \|h'\|(4\beta_2 + 3\beta_3).$$

Proof: Use the same notation as in the previous theorem. Then follow the proof of (2.79) in Proposition 2.4, but rather than applying $|f_h'(W) - f_h'(W + \Delta)| \le 2\|h'\||\Delta|$, we instead use

$$|f_h'(W) - f_h'(W + \Delta)| \le \min\left\{2\|f_h'\|, \|f_h''\||\Delta|\right) \le 2\|h'\|(1 \wedge |\Delta|)$$

which holds by (2.13) where $a \wedge b \equiv \min\{a, b\}$. Hence

$$\begin{aligned} |Eh(W) - Eh(Z)| &\leq 2\|h'\| E\left(1 \wedge |\xi_I^* - \xi_I|\right) \\ &\leq 2\|h'\| E\left(1 \wedge (|\xi_I^*| + |\xi_I|)\right) \\ &\leq 2\|h'\| \left(E(1 \wedge |\xi_I^*|) + E(1 \wedge |\xi_I|)\right). \end{aligned}$$

Now let

$$f(x) = x 1_{[|x|>1]} + \frac{1}{2} x^2 \mathrm{sign}(x) 1_{[|x| \leq 1]},$$

so that $f'(x) = 1 \wedge |x|$. Hence by (2.60) and (2.51)

$$\begin{aligned} E(1 \wedge |\xi_I^*|) &= \sum_{i=1}^n E(1 \wedge |\xi_i^*|) E\xi_i^2 \\ &= \sum_{i=1}^n E\left(\xi_i^2 1_{[|\xi_i|>1]} + \frac{1}{2}|\xi_i|^3 1_{[|\xi_i| \leq 1]}\right) \\ &\equiv \beta_2 + \beta_3. \end{aligned}$$

To bound $E(1 \wedge |\xi_I|)$, recall the fact that if g and h are increasing functions from $\mathbb{R}$ to $\mathbb{R}$, then $Eg(\xi)Eh(\xi) \leq Eg(\xi)h(\xi)$

(so that $Cov[g(\xi), h(\xi)] \geq 0$). Thus, since both $1 \wedge |x|$ and x^2 are increasing functions of $|x|$, we see that (via (2.60))

$$\begin{aligned} E(1 \wedge |\xi_I|) &= \sum_{i=1}^n E(1 \wedge |\xi_i|) E\xi_i^2 \leq E \sum_1^n (1 \wedge |\xi_i|)\, \xi_i^2 \\ &\leq E \left\{ \sum_1^n \xi_i^2 1_{[|\xi_i|>1]} + |\xi_i|^3 1_{[|\xi_i|\leq 1]} \right\} = \beta_2 + \beta_3. \end{aligned}$$

Using these two bounds for the right side of (3.8) yields the claim. $\square$

Remarks:

(1) Note that for $\delta \in (0, 1]$ we have

$$\begin{aligned} \beta_2 &= \sum_1^n E\left\{ \xi_i^2 1_{[|\xi_i|>1]} \right\} \leq \sum_1^n E\left\{ |\xi_i|^{2+\delta} 1_{[|\xi_i|>1]} \right\}, \quad \text{and} \\ \beta_3 &= \sum_1^n E\left\{ |\xi_i|^3 1_{[|\xi_i|\leq 1]} \right\} \leq \sum_1^n E\left\{ |\xi_i|^{2+\delta} 1_{[|\xi_i|\leq 1]} \right\}, \end{aligned}$$

Thus $\beta_2 + \beta_3 \le \sum_1^n E\left\{|\xi_i|^{2+\delta}\right\}$.
(2) When $\xi_i \equiv X_i/\sigma_n$ with the X_i's i.i.d. with $\sigma^2 = Var(X_1)$, then the bound becomes $E|X_1/\sigma|^{2+\delta}/n^{\delta/2}$, so we recover the classical Berry-Esseen rate $n^{-1/2}$ when $\delta = 1$ (so $E|X_1|^3 < \infty$).
(3) Furthermore, for any $0 < \epsilon < 1$,

$$\begin{aligned}
\beta_2 + \beta_3 &= \sum_1^n E\left\{\xi_i^2 1_{[|\xi_i|>1]}\right\} + \sum_1^n E\left\{|\xi_i|^3 1_{[|\xi_i|\le 1]}\right\} \\
&\le \sum_1^n E\left\{\xi_i^2 1_{[|\xi_i|>1]}\right\} + \sum_1^n E\left\{\xi_i^2 1_{[\epsilon<|\xi_i|\le 1]}\right\} + \epsilon \sum_1^n E\left\{\xi_i^2 1_{[|\xi_i|\le\epsilon]}\right\} \\
&\le \sum_1^n E\left\{\xi_i^2 1_{[|\xi_i|>\epsilon]}\right\} + \epsilon.
\end{aligned}$$

The smooth (Lipschitz) function bounds (3.3) or (3.6) of the previous two theorems do not directly yield Berry-Esseen type results with optimal rate of convergence (for the Kolmogorov metric). But they do imply a weak L^∞ bound as follows:

Theorem 3.3: (Erickson, 1974). Suppose that there exists a δ such that for any Lipschitz function h

$$|Eh(W) - Eh(Z)| \le \delta \|h'\|_\infty. \qquad (3.11) \qquad (4)$$

Then

$$\sup_{z \in \mathbb{R}} |P(W \le z) - \Phi(z)| \le 2\delta^{1/2} \qquad (3.12)$$

Proof: Without loss of generality we can assume that $\delta \le 1/4$; if $\delta > 1/4$ the bound claimed in (3.12) is trivial. Let $c \equiv \delta^{1/2}(2\pi)^{1/4}$, and for fixed $z \in \mathbb{R}$ define

$$h_c(w) = \begin{cases} 1 & \text{if } \ w \le z, \\ 0 & \text{if } \ w > z + c, \\ \text{linear} & \text{if } z < w < z + c. \end{cases}$$

Then h_c is Lipschitz continuous with $\|h'\|_\infty = 1/c$, and hence by (3.11)

$$\begin{aligned} P(W \le z) - \Phi(z) &\le Eh_c(W) - Eh_c(Z) + Eh_c(Z) - \Phi(z) \\ &\le \frac{\delta}{c} + \frac{c}{\sqrt{2\pi}} = 2(2\pi)^{-1/4}\delta^{1/2} \le 2\delta^{1/2}. \end{aligned}$$

Similarly, $P(W \le z) - \Phi(z) \ge -2\delta^{1/2}$, and thus (3.12) holds. $\square$

By Remark (3) above we see that if the Lindeberg condition holds, then $\beta_2 + \beta_3 \to 0$ since ϵ is arbitrary. By Theorems 3.2 and 3.3 it follows that

$$\sup_{z \in \mathbb{R}} |P(W \leq z) - \Phi(z)| \leq 4(\beta_2 + \beta_3)^{1/2} \to 0$$

as $n \to \infty$. Thus the forward part of the Lindeberg - Feller CLT holds (again)!

A further beautiful result in C-G-S section 3.5 is the following result of Hall and Barbour (1984):

Theorem 3.8: Suppose that $\xi_1, \ldots, \xi_n$ are independent random variables with zero means and finite variances $E\xi_i^2 = \sigma_i^2$, $1 \le i \le n$, satisfying $\sum_{i=1}^n \sigma_i^2 = 1$, and let $W \equiv \sum_{i=1}^n \xi_i$. Then there exists an absolute constant $C \le 43$ such that for all $\epsilon > 0$

$$\begin{aligned}
&\left(1 - e^{-\epsilon^2/4}\right) \sum_{i=1}^n E\{\xi_i^2 1_{[|\xi_i|>\epsilon]}\} \\
&\quad \le \ C \left(\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| + \sum_{i=1}^n \sigma_i^4 \right). \qquad (3.44)
\end{aligned}$$

Recall that $\sum_1^n \sigma_i^4 \le \max_{1\le i\le n} \sigma_{ni}^2 \cdot 1 \to 0$ is the "uan" condition in the if and only if formulation of the Lindeberg-Feller CLT.
s This inequality implies the converse part of the Lindeberg-Feller theorem .... and indeed goes a considerable way toward quantifying it. (It is interesting that C-G-S (2011) do not reference Goldstein (2009), while Goldstein (2009) does not reference Hall and Barbour (1984).

2. Stein's method: L_1 bounds. C-G-S chapter 4

In this chapter the focus is on L_1 bounds. Let

$$\mathcal{L} \equiv \{h : \mathbb{R} \to \mathbb{R} |\ |h(y) - h(x)| \le |y - x|\}.$$

Recall that

$$\begin{aligned} \|F - G\|_1 &\equiv \int_{\mathbb{R}} |F(x) - G(x)| dx \\ &= \text{inf}\{E_\pi |X - Y|\} \\ &= \sup_{h \in \mathcal{L}} |E_F h(X) - E_G h(X)| \\ &= E|F^{-1}(U) - G^{-1}(U)| \quad \text{where} \quad U \sim \text{Uniform}(0,1) \end{aligned}$$

and where the infimum in the second equality is over all joint laws with marginals F and G respectively. See pages 63 - 65 of C-G-S for a discussion.

Suppose that $\{a_{i,j} :\ 1 \le i \le N,\ 1 \le j \le N\}$ is an array of real numbers, and define

$$Y \equiv \sum_{i=1}^{N} a_{i,\pi(i)}. \qquad (4.104)$$

Elementary calculations show that $\mu \equiv E(Y) = Na_{\cdot,\cdot}$ where we write

$$a_{\cdot\cdot} \equiv \frac{1}{N^2} \sum_{i,j=1}^{N} a_{i,j}, \qquad a_{i,\cdot} \equiv \frac{1}{N} \sum_{j=1}^{N} a_{i,j} \qquad a_{\cdot,j} \equiv \frac{1}{N} \sum_{i=1}^{N} a_{i,j}$$

Moreover,

$$\begin{aligned} Var(Y) &= \frac{1}{N-1} \sum_{i,j} (a_{i,j} - a_{i,\cdot} - a_{\cdot,j} + a_{\cdot,\cdot})^2 \\ &= \frac{1}{N-1} \sum_{i,j} (a_{i,j}^2 - a_{i,\cdot}^2 - a_{\cdot,j}^2 + a_{\cdot,\cdot}^2) \\ &= \frac{1}{4N^2(N-1)} \sum_{i,j,k,l} [(a_{i,k} + a_{j,l}) - (a_{i,l} + a_{j,k})]^2. \qquad (4.105) \end{aligned}$$

and

$$\gamma \equiv \gamma_A = \sum_{i,j=1}^{N} |a_{i,j} - a_{i\cdot} - a_{\cdot,j} + a_{\cdot,\cdot}|^3. \quad (4.107)$$

Theorem 4.8: For $N \geq 3$, let $\{a_{i,j}\}_{i,j=1}^N$ be the components of a matrix $A \in \mathbb{R}^{N\times N}$, let π be random permutation uniformly distributed over $\mathcal{S}_N$, and let Y be given by (4.104). Then with σ^2 as given above and γ given in

$$\|F_W - \Phi\|_1 \leq \frac{\gamma}{(N-1)\sigma^3}\left(16 + \frac{56}{N-1} + \frac{8}{N-1)^2}\right).$$

Remark: When the elements of A are all of comparable order, σ^2 is of order N, and γ is of order N^2, and hence the bound is of order $N^{-1/2}$.

3. Stein's method: L^∞ bounds. C-G-S chapter 6

Theorem 6.1: Let $\{a_{i,j}\}_{i,j=1}^n$ be an array of real numbers and let π be a random permutation with uniform distribution over $\mathcal{S}_n$. Then with Y as before and $W \equiv (Y-\mu)/\sigma$,

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \le 16.3C/\sigma \quad \text{for} \quad n \ge 3,$$

where $\mu = E(Y)$ and $\sigma^2 = Var(Y)$ are given by (4.105) and

$$C \equiv \max_{1\le i,j\le n} |a_{i,j} - a_{i\cdot} - a_{\cdot,j} + a_{\cdot\cdot}|.$$

Theorem 6.2: (Bolthausen, 1984) Let $\{a_{i,j}\}_{i,j=1}^{n}$ be an array of real numbers and let π be a random permutation with uniform distribution over $\mathcal{S}_n$. Then with Y as before and $W \equiv (Y-\mu)/\sigma$, there exists a constant c such that

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \le c\gamma_A/(n\sigma_A^3) \quad \text{for} \quad n \ge 2,$$

and γ_A is as before.

Now suppose that $a_{ij} = b_i c_j$ where $\{b_{N1}, \ldots, b_{NN}\}$ and $\{c_{N1}, \ldots, c_{NN}\}$ are real numbers. Suppose that $\pi = (\pi_1, \ldots, \pi_N)$ is uniformly distributed on the $N!$ permutations of $\{1, \ldots, N\}$. Then the statistic Y becomes

$$Y \equiv Y_N = \sum_{i=1}^{N} b_{Ni} c_{N\pi_i} = \sum_{i=1}^{N} b_i c_{\pi(i)}.$$

We easily find that $E(Y_N) = N \cdot \bar{b}_N \bar{c}_N$ and
$Var(Y_N) = B_N^2 C_N^2/(N-1)$ where

$$B_N^2 = \sum_{i=1}^{N} (b_{Ni} - \bar{b}_N)^2, \quad \text{and} \quad C_N = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)^2.$$

The following theorem is due to Hájek (1961).

Theorem: (Hájek, 1961). Suppose that

$$\max_{1\le i\le N} \frac{|b_{Ni}-\bar{b}_N|}{B_N} \to 0, \quad \text{and} \quad \max_{1\le i\le N} \frac{|c_{Ni}-\bar{c}_N|}{C_N} \to 0.$$

Then the sequence $(Y_N - E(Y_N))/\sigma(Y_N) \to_d Z \sim N(0,1)$ if and only if

$$\sum_{(i,j):\sqrt{N}|b_{Ni}-\bar{b}_N|\cdot|c_{Ni}-\bar{c}_N|>\epsilon B_N C_N} \frac{|b_{Ni}-\bar{b}_N|^2|c_{Nj}-\bar{c}_N|^2}{B_N^2 C_N^2} \to 0$$

for every $\epsilon > 0$.

Also see Proposition A.5.3, van der Vaart & W, page 458. Note that Hájek (1961) is not cited by C-G-S (2011).

Questions:

- Is there a corresponding upper bound version of the Hall-Barbour theorem?
- Is there an analogue of the Lindeberg-Feller CLT for Y based on a general array $\{a_{i,j}\}$?
- Is there an analogue of the Hall - Barbour theorem for Y based on a general array $\{a_{i,j}\}$?
- Is there an analogue of the Lindeberg - Feller CLT for Y in the case of simple random sampling? That is, when $a_{i,j} = b_i c_j$ where n of the c_j's are 1 and $N - n$ are 0's?
- Is there an analogue of the Hall - Barbour theorem for Y in the case of simple random sampling? That is, when $a_{i,j} = b_i c_j$ where n of the c_j's are 1 and $N - n$ are 0's?