Bounds on the Constant in the Mean Central Limit Theorem

Larry Goldstein

University of Southern California, Ann. Prob. 2010

Classical Berry Esseen Theorem

Let $X, X_1, X_2, \ldots$ be i.i.d. with distribution G having mean zero, variance σ^2 and finite third moment. Then there exists C such that

$$||F_n - \Phi||_\infty \le \frac{CE|X|^3}{\sigma^3\sqrt{n}} \quad \text{for } n \in \mathbb{N}$$

where F_n is the distribution function of

$$S_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i,$$

where for distribution functions F and G

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

Different Metrics

L^∞, type of worse case error:

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

L^1, type of average case error:

$$||F - G||_1 = \int_{-\infty}^{\infty} |F(x) - G(x)| dx$$

L^p Berry Esseen Theorem

For $p \geq 1$ there exists a constant C such that

$$||F_n - \Phi||_p \leq \frac{CE|X|^3}{\sigma^3 \sqrt{n}} \quad \text{for } n \in \mathbb{N}. \tag{1}$$

Let $\mathcal{F}_\sigma$ be the collection of all distributions with mean zero, positive variance σ^2, and finite third moment. The L^p Berry-Esseen constant c_p is given by

$$c_p = \inf\{C : \frac{\sqrt{n}\sigma^3 ||F_n - \Phi||_p}{E|X|^3} \leq C, n \in \mathbb{N}, G \in \mathcal{F}_\sigma\}.$$

Each C in (1) is an upper bound on c_p.

Upper Bounds in the Classical Case

Classical case $p = \infty$,

1. $1.88/7.59$ (Berry/Esseen, 1941/1942)
2. $\ldots$
3. 0.7975 (P. van Beeck, 1972).
4. 0.7655 (I. S. Shiganov in 1986).
5. 0.7056 (I.G. Shevtsova in 2006)

Asymptotic Refinements

Let

$$c_{p,m} = \inf\{C : \frac{\sqrt{n}\sigma^3 ||F_n - \Phi||_p}{E|X|^3} \leq C, n \geq m, G \in \mathcal{F}_\sigma\}$$

Clearly $c_{p,m}$ decreases in m, so we have existence of the limit

$$\lim_{m\to\infty} c_{p,m} = c_{p,\infty}.$$

Asymptotically Correct Constant: $p = 1$

For $G \in \mathcal{F}_\sigma$ Esseen explicitly calculates the limit

$$\lim_{n\to\infty} n^{1/2}||F_n - \Phi||_1 = A_1(G).$$

Zolotarev (1964), using characteristic function techniques, shows that

$$\sup_{G\in\mathcal{F}_\sigma} \frac{\sigma^3 A_1(G)}{E|X|^3} = \frac{1}{2},$$

so

$$c_{1,\infty} = \frac{1}{2}.$$

Stein Functional

A bound on the (non-asymptotic) L_1 constant can be obtained by considering the extremum of a Stein functional.

Extrema of Stein functionals are considered by Utev and Lefévre, 2003, who computed some exact norms of Stein operators.

Bound using Zero Bias

Let W be a mean zero random variable with finite positive variance σ^2. We say W^* has the W zero bias distribution if

$$E[Wf(W)] = \sigma^2 E[f'(W^*)] \quad \text{for all smooth } f.$$

If the distribution F of W has variance 1 and W and W^* are on the same space with W^* having the W zero bias distribution, then

$$||F - \Phi||_1 \leq 2E|W^* - W|.$$

Exchange One Zero Bias Coupling

If $W = X_1 + \cdots + X_n$, independent mean zero variables with variances $\sigma_1^2, \ldots, \sigma_n^2$, and I is an independent index with distribution

$$P(I = i) = \frac{\sigma_i^2}{\sum_{j=1}^n \sigma_j^2},$$

then

$$W^* = \sum_{j \neq I} X_j + X_I^*$$

has the W-zero bias distribution, when for each i, X_i^* has the X_i-zero bias distribution independent of $X_j, j \neq i$.

Functional $B(X)$

For $X \in \mathcal{F}_\sigma$ let

$$B(X) = \frac{2\sigma^2 ||\mathcal{L}(X^*) - \mathcal{L}(X)||_1}{E|X^3|}.$$

For $\mathbb{R}$ (more generally on any Polish space) valued random variables, given distributions F and G, one can construct $X \sim F$ and $Y \sim G$ such that

$$E|X - Y| = ||F - G||_1$$

In fact, let $X = F^{-1}(U), Y = G^{-1}(U)$ for $U \sim \mathcal{U}[0, 1]$.

Exchange One Zero Bias Coupling

Let $X_1, \ldots, X_n$ be independent random variables with distributions $G_i \in \mathcal{F}_{\sigma_i}, i = 1, \ldots, n$ and let F_n be the distribution function of $W = (X_1 + \cdots + X_n)/\sigma$ with $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$. Then with $E|X_i^* - X_i| = ||G_i^* - G_i||_1$,

$$\begin{aligned} E|W^* - W| &= \frac{1}{\sigma} E|X_I^* - X_I| = \frac{1}{\sigma} \sum_{i=1}^n \frac{\sigma_i^2}{\sigma^2} E|X_i^* - X_i| \\ &= \frac{1}{\sigma^3} \sum_{i=1}^n \frac{\sigma_i^2 E|X_i^* - X_i|}{E|X_i|^3} E|X_i^3| \\ &= \frac{1}{2\sigma^3} \sum_{i=1}^n B(X_i) E|X_i^3|. \end{aligned}$$

Exchange One Zero Bias Coupling

If $X_1, \ldots, X_n$ are independent mean zero random variables with distributions $G_1, \ldots, G_n$ having finite variances $\sigma_1^2, \ldots, \sigma_n^2$ and finite third moments, then the distribution function F_n of $(X_1 + \cdots + X_n)/\sigma$ with $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$ obeys

$$||F_n - \Phi||_1 \leq \frac{1}{\sigma^3} \sum_{i=1}^{n} B(G_i) E|X_i|^3$$

where the functional $B(G)$ is given by

$$B(G) = \frac{2\sigma^2 ||G^* - G||_1}{E|X|^3}$$

when X has distribution $G \in \mathcal{F}_\sigma$.

Distribution Specific Constants

In the i.i.d. case,

$$||F_n - \Phi||_1 \le \frac{B(G)E|X^3|}{\sqrt{n}\sigma^3},$$

and e.g.,

1. $B(G) = 1$ for mean zero two point distributions
2. $B(G) = 1/3$ for mean zero uniform distributions
3. $B(G) = 0$ for mean zero normal distributions

Universal Bound

Recall that for $G \in \mathcal{F}_\sigma$

$$B(G) = \frac{2\sigma^2 ||G^* - G||_1}{E|X|^3}.$$

For a collection of distributions $\mathcal{F} \subset \bigcup_{\sigma>0} \mathcal{F}_\sigma$, let

$$B(\mathcal{F}) = \sup_{G \in \mathcal{F}} B(G).$$

Then for $X_1, \ldots, X_n$ i.i.d. with distribution in $\mathcal{F}_\sigma$,

$$||F_n - \Phi||_1 \leq \frac{B(\mathcal{F}_\sigma) E|X^3|}{\sqrt{n}\sigma^3}.$$

Bounds on $B(\mathcal{F}_\sigma)$

Mean zero two point distributions give $B(\mathcal{F}_\sigma) \geq 1$ for all $\sigma > 0$.

Using essentially only

$$E|X^* - X| \leq E|X^*| + E|X|$$

gives $B(\mathcal{F}_\sigma) \leq 3$ for all $\sigma > 0$.

By coupling X and X^* together we improve the value of the constant from 3.

Value of Supremum

Theorem 1 *For all $\sigma \in (0, \infty)$,*

$$B(\mathcal{F}_\sigma) = 1.$$

Hence, when $X_1, \ldots, X_n$ are independent with distributions in $\mathcal{F}_{\sigma_i}, i = 1, \ldots, n$ and $\sum_{i=1}^n \sigma_i^2 = \sigma^2$,

$$||F_n - \Phi||_1 \leq \frac{1}{\sigma^3} \sum_{i=1}^n E|X_i|^3,$$

and when these variables are identically distributed with variances σ^2,

$$||F_n - \Phi||_1 \leq \frac{E|X_i|^3}{\sqrt{n}\sigma^3}.$$

Bounds on the Constant c_1

We can also prove the lower bound

$$c_1 \quad \geq \quad \frac{2\sqrt{\pi}(2\Phi(1)-1) - (\sqrt{\pi}+\sqrt{2}) + 2e^{-1/2}\sqrt{2}}{\sqrt{\pi}}$$

Supremum of $B(\mathcal{F}_\sigma)$

Want to compute

$$\sup_{G \in \mathcal{F}_\sigma} B(G) \quad \text{where} \quad B(G) = \frac{2\sigma^2 ||G^* - G||_1}{E|X|^3}.$$

Successively reduce, in four steps, the computation of the supreumum of $B(G)$ on $\mathcal{F}_\sigma$ to computations over smaller collections of distributions.

First Reduction: $\sigma = 1$

Recall

$$B(G) = \frac{2\sigma^2 ||G^* - G||_1}{E|X|^3}.$$

By the scaling property

$$B(aX) = B(X) \quad \text{for all } a \neq 0$$

it suffices to consider $\mathcal{F}_1$.

Second Reduction: compact support

For $X \in \mathcal{F}_1$, show that there exists $X_n, n = 1, 2, \ldots$, each in $\mathcal{F}_1$ and having compact support, such that $B(X_n) \to B(X)$.

Hence it suffices to consider the class of distributions $\mathcal{M} \subset \mathcal{F}_1$ with compact support.

Third Reduction: finite support

For $X \in \mathcal{M}$ show that there exists $X_n, n = 1, 2, \ldots$ in $\mathcal{M}$, finitely supported, such that $B(X_n) \to B(X)$.

Hence it suffices to consider $\bigcup_{m \geq 3} D_m$, where D_m are mean zero variance one distributions supported on at most m points.

Fourth Reduction: three point support

Use a convexity type property of $B(G)$, which depends on the behavior of the zero bias transformation on a mixture, to obtain

$$B(D_3) = B(\bigcup_{m \geq 3} D_m).$$

Hence it suffices to consider D_3.

Lastly

Show

$$B(D_3) = 1.$$

Finding Extremes of Expectations

Arguments along these lines were first considered by Hoeffding for the calculation of the extremes of $EK(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are independent.

Though $B(G)$ is not of this form, the reasoning of Hoeffding applies.

In some cases the final result obtained is not in closed form.

Reduction to Compact Support and Finite Support

Continuity of the zero bias transformation: If

$$X_n \Rightarrow_d X, \quad \text{and} \quad \lim_{n\to\infty} EX_n^2 = EX^2$$

then

$$X_n^* \Rightarrow_d X^* \quad \text{as } n \to \infty.$$

Leads to continuity of $B(G)$: If

$$X_n \Rightarrow_d X, \quad \lim_{n\to\infty} EX_n^2 = EX^2 \quad \text{and} \quad \lim_{n\to\infty} E|X_n^3| = E|X^3|$$

then

$$B(X_n) \to B(X) \quad \text{as } n \to \infty.$$

From $\bigcup_{m \geq 3} D_m$ to D_3

If X_μ be the μ mixture of a collection $\{X_s, s \in S\}$ of mean zero, variance 1 random variables satisfying $E|X_\mu^3| < \infty$. Then

$$B(X_\mu) \leq \sup_{s \in S} B(X_s).$$

In particular, if $\mathcal{C}$ is a collection of mean zero, variance 1 random variables with finite absolute third moments and $\mathcal{C} \supset \mathcal{D}$ such that every distribution in $\mathcal{C}$ can be represented as a mixture of distributions in $\mathcal{D}$, then

$$B(\mathcal{C}) = B(\mathcal{D}).$$

Zero Biasing a Mixture

Theorem 2 *Let* $\{m_s, s \in S\}$ *be a collection of mean zero distributions on* $\mathbb{R}$ *and* μ *a probability measure on* S *such that the variance* σ_μ^2 *of the mixture distribution is positive and finite. Then* m_μ^*, *the* m_μ *zero bias distribution exists and is given by the mixture*

$$m_\mu^* = \int m_s^* d\nu \quad \textit{where} \quad \frac{d\nu}{d\mu} = \frac{\sigma_s^2}{\sigma_\mu^2}.$$

In particular, $\nu = \mu$ *if and only if* σ_s^2 *is a constant* μ *a.s.*

Mixture of Constant Variance: $\nu = \mu$

$$
\begin{aligned}
||\mathcal{L}(X_\mu^*) - \mathcal{L}(X_\mu)||_1 &= \sup_{f \in L} |Ef(X_\mu^*) - Ef(X_\mu)| \\
&= \sup_{f \in L} |\int_S Ef(X_s^*) d\mu - \int_S Ef(X_s) d\mu| \\
&\leq \sup_{f \in L} \int_S |Ef(X_s^*) - Ef(X_s)| \, d\mu \\
&\leq \sup_{f \in L} \int_S ||\mathcal{L}(X_s^*) - \mathcal{L}(X_s)||_1 d\mu \\
&= \int_S ||\mathcal{L}(X_s^*) - \mathcal{L}(X_s)||_1 d\mu.
\end{aligned}
$$

$B(X_\mu) \leq \sup_s B(X_s)$

The relation

$$\frac{d\tau}{d\mu} = \frac{E|X_s^3|}{E|X_\mu^3|}. \tag{2}$$

defines a probability measure, as $E|X_\mu^3| = \int E|X_s^3| ds$.

$$B(X_\mu) \le \sup_s B(X_s)$$

Then

$$\begin{aligned} B(X_\mu) &= \frac{2||\mathcal{L}(X_\mu^*) - \mathcal{L}(X_\mu)||_1}{E|X_\mu^3|} \\ &\le \frac{\int_S 2||\mathcal{L}(X_s^*) - \mathcal{L}(X_s)||_1 d\mu}{E|X_\mu^3|} \\ &= \frac{\int_S B(X_s) E|X_s^3| d\mu}{E|X_\mu^3|} \\ &= \int_S B(X_s) d\tau \\ &\le \sup_{s \in S} B(X_s) \end{aligned}$$

Reduction of $D_m, m > 3$

The distribution of any $X \in D_m$ is determined by the values $a_1 < \cdots < a_m$ and probabilities $\mathbf{p} = (p_1, \ldots, p_m)'$, all positive. The vector $\mathbf{p}$ must satisfy $A\mathbf{p} = \mathbf{c}$ where

$$A = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ a_1 & a_2 & \ldots & a_m \\ a_1^2 & a_2^2 & \ldots & a_m^2 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

For $m > 3$ there exists vector $v \neq 0$ satisfying $Av = 0$. Hence $X \in D_m$ can be represented of a mixture of two distributions in D_{m-1}.

Reduction to D_3

For every $m > 3$, every $G \in D_m$ can be represented as a finite mixture of distributions in D_{m-1}. Hence

$$B(D_3) = B(\bigcup_{m \geq 3} D_m).$$

Every distribution D_3 with support points, say $x < y < 0 < z$, can be written as

$$m_\alpha = \alpha m_1 + (1 - \alpha) m_0,$$

a mixture of the (unequal variance) mean zero distributions m_1 and m_0 supported on $\{x, z\}$ and $\{y, z\}$, respectively.

Mixture with Unequal Variance

For $\alpha \in [0, 1]$, with

$$m_\alpha = \alpha m_1 + (1-\alpha) m_0$$

we have

$$m_\alpha^* = \beta m_1^* + (1-\beta) m_0^* \quad \text{where} \quad \beta = \frac{\alpha x}{\alpha x + (1-\alpha) y}.$$

Since $x < y < 0$,

$$\frac{\beta}{1-\beta} = \frac{\alpha}{1-\alpha}\frac{x}{y} > \frac{\alpha}{1-\alpha} \quad \text{and therefore} \quad \beta > \alpha.$$

Calculating $G(D_3)$

Write $m \in D_3$ as

$$m_\alpha = \alpha m_1 + (1-\alpha) m_0$$

where m_1 and m_0 are mean zero two point distributions on $\{x, z\}$ and $\{y, z\}$, respectively, $x < y < 0 < z$.

Need to bound

$$||m_\alpha^* - m_\alpha||_1. \qquad (3)$$

Any coupling of variables Y_α^* and Y_α with distributions m_α^* and m_α, respectively, gives an upper bound to (3). Let F_0, F_1, F_0^*, F_1^* be the distribution functions of m_0, m_1, m_0^* and m_1^*, respectively.

Bound by Coupling

Set (Y_1, Y_0, Y_1^*, Y_0^*) equal to

$$(F_1^{-1}(U), F_0^{-1}(U), (F_1^*)^{-1}(U), (F_0^*)^{-1}(U))$$

and let $\mathcal{L}(Y_\alpha, Y_\alpha^*)$ be

$$\alpha\mathcal{L}(Y_1, Y_1^*) + (1-\beta)\mathcal{L}(Y_0, Y_0^*) + (\beta-\alpha)\mathcal{L}(Y_0, Y_1^*).$$

Then (Y_α, Y_α^*) has marginals $Y_\alpha =_d X_\alpha$ and $Y_\alpha^* =_d X_\alpha^*$, and therefore $||m_\alpha^* - m_\alpha||_1$ is upper bounded by

$$\alpha||m_1^* - m_1||_1 + (1-\beta)||m_0^* - m_0||_1 + (\beta-\alpha)||m_1^* - m_0||_1.$$

Bound on D_3

Goal is to have $||m_\alpha - m_\alpha^*||_1$, or its upper bound

$$\alpha||m_1^* - m_1||_1 + (1-\beta)||m_0^* - m_0||_1 + (\beta - \alpha)||m_1^* - m_0||_1,$$

bounded by

$$E|X_\alpha^3|/(2EX_\alpha^2) = \beta||m_1^* - m_1||_1 + (1-\beta)||m_0^* - m_0||_1.$$

Hence it suffices to show

$$||m_1^* - m_0||_1 \leq ||m_1^* - m_1||_1.$$

'Reduces' to computation of L^1 distances between uniform distribution on $[x, z]$ and two point distributions on $\{y, z\}$ and $\{x, z\}$.

$$||m_1^* - m_0||_1 \leq ||m_1^* - m_1||_1$$

m_0 on $\{y, z\}$, m_1 on $\{x, z\}$ with $x < y < 0 < z$. Right hand side is

$$||m_1^* - m_1||_1 = \frac{z^2 + x^2}{2(z - x)}.$$

Left hand side, under case where $F_1^*(y) \leq F_0(y)$, is

$$[2(z-x)(z-y)^2]^{-1} \left(z^4 - 2yz^3 + x^2z^2 - 2x^2yz\right.$$

$$\left.+5y^2z^2 + 3x^2y^2 - 4xy^3 + 4xy^2z - 4xyz^2 + 2y^4 - 4y^3z\right).$$

Using Mathematica

Taking the difference, after much cancelation $||m_1^* - m_1||_1 - ||m_1^* - m_0||_1$ is seen to equal

$$\frac{-4y^2z^2 - 2x^2y^2 + 4xy^3 - 4xy^2z + 4xyz^2 - 2y^4 + 4y^3z}{2(z-x)(z-y)^2},$$

which factors as

$$\frac{-y(y-x)(y^2 + 2z^2 - y(x+2z))}{(z-x)(z-y)^2}$$

and is positive, due to being in case $F_1^*(y) \leq F_0(y)$.

Bound over D_3

Since $||m_1^* - m_0||_1 \leq ||m_1^* - m_1||_1$ we have

$$||m_\alpha - m_\alpha^*||_1 \leq E|X_\alpha^3|/(2EX_\alpha^2),$$

and therefore $B(D_3) \leq 1$.

Bound over D_3

Since $||m_1^* - m_0||_1 \leq ||m_1^* - m_1||_1$ we have

$$||m_\alpha - m_\alpha^*||_1 \leq E|X_\alpha^3|/(2EX_\alpha^2),$$

and therefore $B(D_3) \leq 1$.

Hence

$$1 \geq B(D_3) = B(\bigcup_{m \geq 3} D_m) = B(\mathcal{M}) = B(\mathcal{F}_1) \geq 1.$$

The Anti-Normal Distributions

$G \in \mathcal{F}_1$ is normal if and only if $B(G) = 0$; small $B(G)$ close to normal.

G, a mean zero two point distribution on $x < 0 < y$ achieves $\sup_{G \in \mathcal{F}_1} B(G)$, the worst case for $B(G)$, so 'anti-normal'.

Lower Bound

For $\mathcal{L}(X) = G \in \mathcal{F}_1$,

$$||F_n - \Phi||_1 \leq \frac{c_1 E|X^3|}{\sqrt{n}} \quad \text{for all } n \in \mathbb{N},$$

and in particular for $n = 1$

$$c_1 \geq \frac{||F_1 - \Phi||_1}{E|X^3|} = \frac{||G - \Phi||_1}{E|X^3|}.$$

Lower Bound: $0.535377\ldots$

For $B \sim \mathcal{B}(p)$ for $p \in (0,1)$ let G_p be the distribution function of $X = (B-p)/\sqrt{pq}$. Then $||G_p - \Phi||_1$ equals

$$\int_{-\infty}^{-\sqrt{\frac{p}{q}}} \Phi(x)dx + \int_{-\sqrt{\frac{p}{q}}}^{\sqrt{\frac{q}{p}}} |\Phi(x) - q|dx + \int_{\sqrt{\frac{q}{p}}}^{\infty} |\Phi(x) - 1|dx,$$

and letting

$$\psi(p) = \frac{\sqrt{pq}}{p^2+q^2}||G_p - \Phi||_1 \quad \text{for } p \in (0,1)$$

$$\psi(1/2) = \frac{2\sqrt{\pi}(2\Phi(1)-1) - (\sqrt{\pi}+\sqrt{2}) + 2e^{-1/2}\sqrt{2}}{\sqrt{\pi}}.$$

Upper Bounds in the Classical Case

Classical case $p = \infty$,

1. $1.88/7.59$ (Berry/Esseen, 1941/1942)
2. $\ldots$
3. 0.7975 (P. van Beeck, 1972).
4. 0.7655 (I. S. Shiganov in 1986).
5. 0.7056 (I.G. Shevtsova in 2006)
6. 0.4785 (I. Tyurin in 2010)

Higher Order Hermite Functionals

Letting $H_k(x)$ be the k^{th} Hermite Polynomial, if the moments of X match those of the standard normal up to order $2k$, then there exists $X^{(k)}$ such that

$$EH_k(X)f(X) = Ef^{(k)}(X^{(k)}).$$

Can one compute extreme values of the natural generalizations of $B(G)$ such as

$$B_k(G) = \frac{\sigma^{2k}||X^{(k)} - X||_1}{E|X|^{2k+1}}$$

which might be the values of like constants when higher moments match the normal.