Berry Esseen Bounds for Local Extremes and Combinatorial Central Limit Theorems, using Size and Zero Biasing

Larry Goldstein

http://math.usc.edu/∼larry

Local Maxima of Finite Graphs

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a finite graph, and for $v \in \mathcal{V}$ let N_v be the set of neighbors of v,

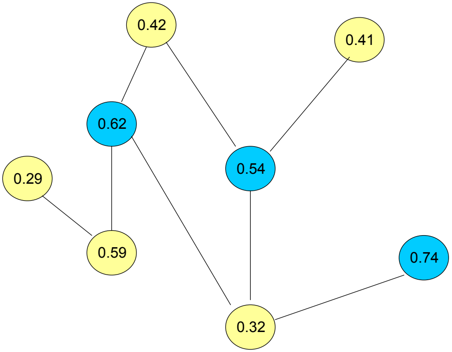
$$N_v = \{w : \{w, v\} \in \mathcal{E}\}.$$

Let $U_v, v \in \mathcal{V}$ be iid with any continuous distribution. Let X_v be the indicator that $v \in \mathcal{V}$ is a local maximum, that is, that

$$U_v > U_w \quad \text{for all } w \in N_v.$$

Describe the distribution of the total number of local maxima

$$W = \sum_{v \in \mathcal{V}} X_v.$$


0.42
0.41
0.62
0.54
0.29
0.59
0.74
0.32

Combinatorial Central Limit Theorem

For $\{a_{ij}\}_{1 \le i,j \le n}$ an array of numbers and π a random permutation of $\{1, \ldots, n\}$, find the distribution of

$$W = \sum_{i=1}^{n} a_{i,\pi(i)}.$$

When π is uniform over $\mathcal{S}_n$, $e_i = \mathbf{1}(1 \le i \le k)$ and

$$a_{ij} = d_i e_j$$

then W is the sum of the characteristics d_i in a simple random sample of size k from a population of size n.

Size Biasing

For $X \in \{0, 1, 2, \ldots\}$ with $EX = \mu < \infty$, consider the size biased distribution

$$P(X^s = k) = \frac{kP(X = k)}{\mu}.$$

Appears in sampling, generates the waiting time paradox.

The distribution is also characterized by

$$EXf(X) = \mu Ef(X^s) \quad \text{all } f,$$

and can be applied to any $X \geq 0$ with finite mean μ.

Size Bias Coupling

If $X_1, \ldots, X_n$ are non-negative independent variables with finite means $\mu_1, \ldots, \mu_n$, then with $W = X_1 + \cdots + X_n$,

$$W^s = W - X_I + X_I^s,$$

where

$$P(I = i) = \frac{\mu_i}{\sum_{j=1}^n \mu_j} = \frac{\mu_i}{\mu}.$$

The sum is size biased by replacing one summand, chosen with probability proportional to its expectation, by an independent variable having that summand's size biased distribution.

Coupling

$$
\begin{aligned}
\mu E f(W^s) &= \mu E f(W - X_I + X_I^s) \\
&= \mu \sum_{i=1}^{n} E f(W - X_i + X_i^s) \frac{\mu_i}{\mu} \\
&= \sum_{i=1}^{n} \mu_i E f(\sum_{t \neq i} X_t + X_i^s) \\
&= \sum_{i=1}^{n} E X_i f(\sum_{t \neq i} X_t + X_i) \\
&= \sum_{i=1}^{n} E X_i f(W) \\
&= E W f(W)
\end{aligned}
$$

Stein's Method for Normal

Characterization of the Normal:

$$Z \sim \mathcal{N}(0, \sigma^2)$$

if and only if

$$EZf(Z) = \sigma^2 Ef'(Z) \quad \text{for all } f.$$

For $EW = 0, EW^2 = \sigma^2$, if

$$E[Wf(W) - \sigma^2 f'(W)]$$

is close to zero for many functions f, then W should be close in distribution to Z.

Stein Differential Equation

Given a test function h, let $Nh = Eh(Z/\sigma)$, solve for f in

$$\sigma^2 f'(w) - wf(w) = h(w/\sigma) - Nh,$$

and evaluate expectation of RHS by expectation of LHS.
Bounds for $\sigma^2 = 1$,

$$||f''|| \leq 2||h'||.$$

Note the appearance of the term $wf(w)$ both in this Stein equation, and also in the characterization of the size bias distribution.

Zero Bias Transformation

Goldstein and Reinert 1997: For W a mean zero variance σ^2 random variable, there exists W^* such that for all absolutely continuous f with $E|Wf(W)| < \infty$,

$$EWf(W) = \sigma^2 Ef'(W^*).$$

From Stein's characterization,

$$EZf(Z) = \sigma^2 Ef'(Z) \quad \text{if and only if} \quad Z \sim \mathcal{N}(0, \sigma^2).$$

Hence:

$$W^* =_d W \quad \text{if and only if} \quad W \sim \mathcal{N}(0, \sigma^2).$$

Zero and Size Biasing

To zero (size) bias a sum

$$W = \sum_{i=1}^{n} X_i$$

of mean zero (non-negative) independent variables, pick one proportional to its variance (mean) and replace with biased version. Size biasing applied to normal in Goldstein and Rinott, 1996. Zero biasing?

Easy Smooth Function Zero Bias CLT

Take $X_1, \ldots, X_n$ iid mean zero, variance 1, let

$$W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \quad \text{having variance 1.}$$

Given h, solve for f, and write

$$\begin{aligned} Eh(W) - Nh &= E\left[f'(W) - Wf(W)\right] \\ &= E\left[f'(W) - f'(W^*)\right] \\ &\leq ||f''|| E|W - W^*| \\ &\leq \frac{2||h'||}{\sqrt{n}} E|X_I - X_I^*|. \end{aligned}$$

Berry Esseen Bounds

Bounds over collection of non smooth (e.g.indicator) functions $\mathcal{H}$;

$$\delta = \sup_{h \in \mathcal{H}} |Eh(W) - Eh(Z)|.$$

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

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

Functions $\mathcal{H}$

3. For any $\epsilon > 0$ and any $h \in \mathcal{H}$, the functions

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

are in $\mathcal{H}$, and

$$\int \tilde{h}(x;\epsilon)\phi(x)ds \le a\epsilon$$

for some constant a which depends only on the class $\mathcal{H}$, where

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

and ϕ is the standard normal density.

Classical Berry Esseen

The collection of indicators of all half lines, and indicators of all intervals, for example, each form classes $\mathcal{H}$ which satisfy 1,2 and 3 with $a = \sqrt{2/\pi}$ and $a = 2\sqrt{2/\pi}$ respectively (see Bolthausen, Goetze, Rinott-Rotar).

Smoothing Lemma

Define the t-smoothed version of h,

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

and the t smoothed bound

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

Lemma 1 *For W a random variable on $\mathbb{R}$ and $\mathcal{H}$ a class of measurable functions satisfying properties 1,2, and 3,*

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

Size Bias Theorem

Theorem 1 *Let $W \geq 0$ be a mean μ, variance σ^2 random variable and suppose $|W^s - W| \leq B$ for some B. Then*

$$\delta \leq aA + \frac{\mu}{\sigma}\left((19 + 30a)A^2 + 4A^3\right) + \frac{23\Delta\mu}{\sigma^2},$$

where

$$\Delta = \sqrt{\textit{Var}(E(W^s - W|W))}$$

and $A = B/\sigma$.

Classical Berry Esseen

For indicators of all half lines, and indicators of all intervals, using $a = \sqrt{2/\pi}$ and $a = 2\sqrt{2/\pi}$, we have respectively

$$\delta \leq 0.8A + \frac{\mu}{\sigma}\left(43A^2 + 4A^3\right) + \frac{23\Delta\mu}{\sigma^2}$$

and

$$\delta \leq 1.6A + \frac{\mu}{\sigma}\left(67A^2 + 4A^3\right) + \frac{23\Delta\mu}{\sigma^2}.$$

Gives (correct) order σ^{-1}.

Zero Bias Theorem

Theorem 2 *Let W be a mean zero, variance σ^2 random variable. If*

$$|W^* - W| \leq 3B$$

then

$$\delta \leq A\left(37 + 12A + 60a\right),$$

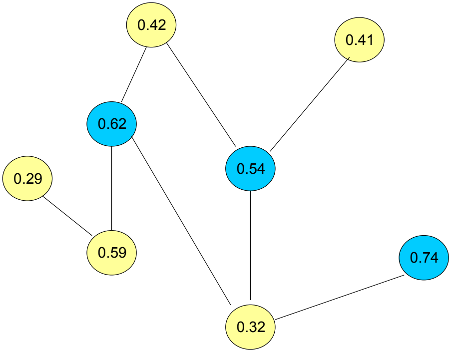
for $A = 3B/\sigma$. For indicators of all half lines, and the indicators of all intervals, using $a = \sqrt{2/\pi}$ and $a = 2\sqrt{2/\pi}$, we have respectively

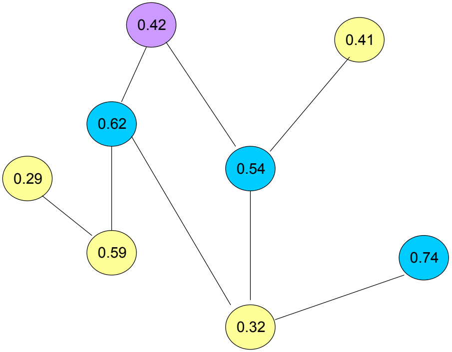
$$\delta \leq A\left(85 + 12A\right) \quad \textit{and} \quad \delta \leq A\left(133 + 12A\right).$$

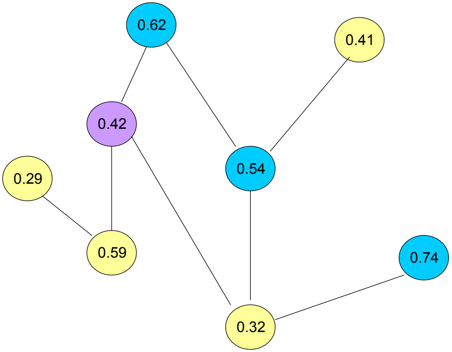
Size Bias Construction, Dependent Variables

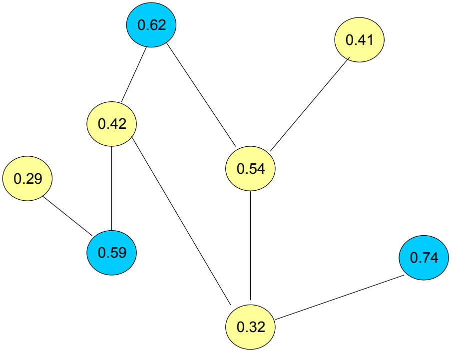
Pick a summand with probability proportional to its expectation, replace with one from that summands size bias distribution, and adjust other variables to have correct conditional distribution given the new value of the selected variable.

Though we apply here to the local maxima problem, having local dependence, the construction can also be used in cases of global dependence.


0.42
0.41
0.62
0.54
0.29
0.59
0.74
0.32


0.42
0.41
0.62
0.54
0.29
0.59
0.74
0.32


0.62
0.41
0.42
0.54
0.29
0.59
0.74
0.32


0.62
0.41
0.42
0.54
0.29
0.59
0.74
0.32

Local Maxima on Graphs

For bounded degree graphs, we can couple such that $|W - W^s|$ is bounded. For example, Baldi, Rinott and Stein, consider hypercube $\mathcal{V} = \{0,1\}^d$, for which we find, for say half lines,

$$\delta \leq 0.8A + \frac{\mu}{\sigma}\left(43A^2 + 4A^3\right) + \frac{23\Delta\mu}{\sigma^2}$$

with

$$\mu = EY, \sigma^2 = \mathsf{Var}(Y), \quad A = (d+1)/\sigma$$

and

$$\Delta \leq 2^{-d/2}(d+1)\sqrt{\binom{d}{3} + \binom{d}{2} + d + 1}.$$

Zero Bias by Square Bias Interpolation

If $(W', W'') \sim dF(w', w'')$ are exchangeable with

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

then taking $(W^\dagger, W^\ddagger)$ according to

$$dG(w^\dagger, w^\ddagger) = \frac{(w^\dagger - w^\ddagger)^2 dF(w^\dagger, w^\ddagger)}{E(W' - W'')^2}$$

and $U \sim \mathcal{U}[0,1]$, we have

$$W^* = UW^\dagger + (1-U)W^\ddagger.$$

Combinatorial CLT

$$W = \sum_{i=1}^{n} a_{i,\pi(i)},$$

Uniform π, von Bahr, Bothausen. Non uniform, Kolchin single cycle n^1 is special case where permutation distribution depends only on cycle type

$$1^{c_1} 2^{c_2} \cdots n^{c_n} \quad \text{where } \sum_{j=1}^{n} jc_j = n.$$

To construct exchangeable pair W', W'' with distribution W use exchangeable pair π', π''.

Exchangeable Pair of Permutations

When distribution depends on cycle type, given π', construct π'' by taking indices $I \neq J$ uniformly and independently of π, and let π' be given by the interchange of I with J in the cycle representation of π. If in the cycle representation of π we have

$$\cdots \pi^{-1}(I) \to I \to \pi(I) \cdots \quad \text{and} \quad \cdots \pi^{-1}(J) \to J \to \pi(J) \cdots$$

then in the cycle representation of π' we put

$$\cdots \pi^{-1}(I) \to J \to \pi(I) \cdots \quad \text{and} \quad \cdots \pi^{-1}(J) \to I \to \pi(J) \cdots$$

Square Bias and Interpolate

Generate $W^\dagger, W^\ddagger$ proportional to $(W' - W'')^2$ and set

$$W^* = UW^\dagger + (1-U)W^\ddagger.$$

Then, even though the dependence is global,

$$|W^* - W| \le 14C \quad \text{where } C = \max_{ij} |a_{ij}|.$$

Dependent CLT and δ

Combinatorial Central Limit Theorem:

$$Y = \sum_{i=1}^{n} a_{i,\pi(i)}$$

for an array of real numbers $\{a_{i,j}\}_{i,j=1}^{n}$ and a random permutation $\pi \in S_n$. For permutation distributions constant over cycle type with no fixed points and $a_{i,j} = a_{j,i}$,

$$\delta \leq A\,(85 + 12A)$$

where $A = 42C/\sigma$ with $\sigma^2 = \mathrm{Var}(Y)$ and $C = \max_{i,j} |a_{i,j}|$.

Directions

Applications: Geometric Functionals, e.g. edge length of nearest neighbor graphs.

Extension: Multivariate Versions