Multivariate normal approximation: permutation statistics, local dependence and beyond

Martin Raič

University of Ljubljana
and
University of Primorska
Slovenia

Workshop on New Directions in Stein's Method
NUS, IMS, Singapore
May 2015

OBJECTIVE

- W ... a $\mathbb{R}^d$-valued random vector to be approximated
- Z_d ... a standard d-variate normal random vector
- $\mathscr{C}$ ... the family of all measurable convex sets in $\mathbb{R}^d$
- Consider the following type of the error in the approximation of $\mathscr{L}(W)$ with $\mathscr{L}(Z_d)$:

$$\sup_{C \in \mathscr{C}} \left| \mathbb{P}(W \in C) - \mathbb{P}(Z_d \in C) \right| .$$

COMBINATORIAL CENTRAL LIMIT THEOREM (1)

- N ... set with n elements
- $a(i,j) \in \mathbb{R}^d; \quad i,j \in N$
- $\pi\colon\ N \to N$ ... uniformly distributed random permutation
- $W = \sum_{i \in N} a(i, \pi(i))$
- $a(i,j)$ can be chosen so that $\mathbb{E}\, W = \mathbf{0}$ and $\operatorname{var}(W) = \mathbb{E}\, W W^{\mathrm{T}} = \mathbf{I}$.

Theorem

There exists a universal constant K, such that for all $C \in \mathscr{C}$,

$$\big|\mathbb{P}(W \in C) - \mathbb{P}(Z_d \in C)\big| \le K\, d^{1/4}\, \frac{1}{n} \sum_{i \in N} \sum_{j \in N} |a(i,j)|^3\,,$$

where $|\cdot|$ denotes the Euclidean norm.

COMBINATORIAL CENTRAL LIMIT THEOREM (2)

- Bolthausen (1984) proved the univariate version.
- Bolthausen and Götze (1993) claimed to have proved even an extension of the preceding theorem. However …
- Chen and Shao (2007) found a counter-example: the extended result is wrong and the proof does not work even for the basic version.
- Nevertheless, the basic version holds true and can be extended to various other cases: vectors $a(i,j)$ being random, multiply-indexed permutation statistics, i. e.,

$$\sum_{i_1,\dots,i_k \in N} a\big(i_1,\dots,i_k,\pi(i_1),\dots,\pi(i_k)\big) \text{ etc.}$$

COMBINATORIAL CENTRAL LIMIT THEOREM (3)

Theorem

There exists a universal constant K, such that for all $C \in \mathscr{C}$,

$$\left|\mathbb{P}(W \in C) - \mathbb{P}(Z_d \in C)\right| \leq K\, d^{1/4}\, \frac{1}{n} \sum_{i \in N} \sum_{j \in N} |a(i,j)|^3\,,$$

where $|\cdot|$ denotes the Euclidean norm.

LOCAL DEPENDENCE (1)

- N ... index set
- G ... undirected graph with vertex set N
- $(X_i)_{i \in N}$ ... a family of $\mathbb{R}^d$-valued random vectors
- For any disjoint $I, J \subseteq N$, such that there is no edge with one endpoint in I and the other in J, the families $(X_i)_{i \in I}$ and $(X_j)_{j \in J}$ are independent.
- $W = \sum_{i \in N} X_i$; $\quad \mathbb{E}\, W = \mathbf{0}$, $\operatorname{var}(W) = \mathbf{I}$

Theorem

There exists a universal constant K, such that for all $C \in \mathscr{C}$,

$$\left|\mathbb{P}(W \in C) - \mathbb{P}(Z_d \in C)\right| \leq K\, d^{1/4} (D+1)^2 \sum_{i \in N} \mathbb{E}\, |X_i|^3 \,,$$

where D denotes the maximum degree of a vertex of G.

LOCAL DEPENDENCE (2)

- Barbour, Karoński and Ruciński (1989) proved the result for the univariate case and the Wasserstein metric.
- Rinott and Rotar (1996) derived a result for the multivariate case and indicators of measurable convex sets. However, the summands need to be bounded, there is an additional logarithmic factor and the factor arising from the dependence graph is in general larger.
- Chen and Shao (2004) proved a univariate Berry–Esséen type bound, but again, the factor arising from the dependence graph is in general larger.
- Fang and Röllin (2012) proved a multivariate CLT, but the summands essentially need to be bounded and the dependence on dimension is suboptimal. However, their result yields optimal bounds in many special cases.

LOCAL DEPENDENCE (3)

Theorem

There exists a universal constant K, such that for all $C \in \mathscr{C}$,

$$\left|\mathbb{P}(W \in C) - \mathbb{P}(Z_d \in C)\right| \leq K\, d^{1/4}(D+1)^2 \sum_{i \in N} \mathbb{E}\,|X_i|^3\,,$$

where D denotes the maximum degree of a vertex of G.

GENERAL APPROACH

- $W = \int \Psi(d\xi)$, where Ψ is a random $\mathbb{R}^d$-valued measure
- $\psi = \mathbb{E}\,\Psi$ ... its mean measure, i. e., $\mathbb{E}\,W = \int \psi(d\xi)$
- V ... another random variable
- A family $(V_\xi)_\xi$ of random variables fits the conditioning of V given Ψ if for any suitable function F,

  $$\mathbb{E} \int F(\xi, V)\,\Psi(d\xi) = \int \mathbb{E}\,F(\xi, V_\xi)\,\psi(d\xi)\,.$$

  If the domain of Ψ is finite or countable and the conditional distribution of V given $\Psi(\{\xi\}) = x$ is the same for all $x \neq \mathbf{0}$, this can be taken to be the distribution of V_ξ.
- This concept is a generalization of Palm measures and encompasses summation as well as conditioning. It could potentially be used in stochastic geometry.

THE STEIN EXPECTATION (1)

Suppose that $\mathbb{E}\, W = \mathbf{0}$ and $\mathrm{var}(W) = \mathbf{I}$ and that $(W_\xi)_\xi$ fits the conditioning of W given Ψ. Then for any suitable function g,

$$\mathbb{E}\langle \nabla g(W), W\rangle = \mathbb{E}\int \langle \nabla g(W), \Psi(d\xi)\rangle = \int \langle \mathbb{E}\, \nabla g(W_\xi), \psi(d\xi)\rangle\,.$$

Now as $\mathbb{E}\, W = \mathbf{0}$, we also have:

$$\int \langle \mathbb{E}\, \nabla g(W), \psi(d\xi)\rangle = 0\,,$$

so that:

$$\mathbb{E}\langle \nabla g(W), W\rangle = \int \big\langle \mathbb{E}\, \nabla g(W_\xi) - \nabla g(W), \psi(d\xi)\big\rangle =$$

$$= \int \mathbb{E}\big\langle \nabla^2 g(\tau W_\xi + (1-\tau)W), (W_\xi - W)\otimes \psi(d\xi)\big\rangle\,,$$

where $\tau \sim U(0,1)$ is independent of all other variates.

Now let:

- $W_\xi - W = \int \Psi_\xi(d\eta)$;
- $\left(W_{\xi\eta}^{(0)}, W_{\xi\eta}^{(1)}\right)_\eta$ fit the conditioning of (W_ξ, W) given Ψ_ξ.
- ψ_ξ ... the mean measure of Ψ_ξ.

Then:

$$\begin{aligned}
&\mathbb{E}\langle \nabla g(W), W\rangle = \\
&= \int \mathbb{E} \int \left\langle \nabla^2 g(\tau W_\xi + (1-\tau) W), \Psi_\xi(d\eta) \otimes \psi(d\xi) \right\rangle = \\
&= \iint \mathbb{E} \left\langle \nabla^2 g\left(\tau W_{\xi\eta}^{(0)} + (1-\tau) W_{\xi\eta}^{(1)}\right), \psi_\xi(d\eta) \otimes \psi(d\xi) \right\rangle .
\end{aligned}$$

THE STEIN EXPECTATION (3)

Next,

$$\mathbf{I} = \mathbb{E}\, WW^{\mathrm{T}} = \mathbb{E} \int W\, \Psi(d\xi)^{\mathrm{T}} = \int \mathbb{E}\, W_\xi\, \psi(d\xi)^{\mathrm{T}} =$$
$$= \int \mathbb{E}(W_\xi - W)\, \psi(d\xi)^{\mathrm{T}} = \iint \psi_\xi(d\eta)\, \psi(d\xi)^{\mathrm{T}}\,,$$

and consequently,

$$\Delta g(w) = \langle \nabla^2 g(w), \mathbf{I} \rangle = \iint \big\langle \nabla^2 g(w), \psi_\xi(d\eta) \otimes \psi(d\xi) \big\rangle\,.$$

Thus,

$$\mathbb{E}\big[\langle \nabla g(W), W \rangle - \Delta g(W)\big] =$$
$$= \iint \mathbb{E}\big\langle \nabla^2 g\big(\tau W^{(0)}_{\xi\eta} + (1-\tau) W^{(1)}_{\xi\eta}\big) - \nabla^2 g(W),$$
$$\psi_\xi(d\eta) \otimes \psi(d\xi) \big\rangle\,.$$

THE STEIN EXPECTATION (4)

Thus, if g solves the Stein equation:

$$\langle \nabla g(w), w \rangle - \Delta g(w) = f(w) - \mathbb{E}\, f(Z_d)$$

and the third derivatives are bounded, one can estimate the error in terms of:

$$\iint \mathbb{E}\Big[\big|W^{(0)}_{\xi\eta} - W\big| + \big|W^{(1)}_{\xi\eta} - W\big|\big)\Big] |\psi_\xi|(d\eta)\, |\psi|(d\xi)\,.$$

Remarks:

- The right hand side corresponds to third moments.
- The expansion can be continued to arbitrary order.
- The key point is coupling of conditional distributions, e. g., of W_ξ, with the original distribution of W.

COMBINATORIAL CLT: THE BASIC COUPLING

- $W = \sum_{i \in N} a(i, \pi(i)) = \int d\Psi(\xi)$, where
- the domain of Ψ is $N \times N$ and $\Psi = \sum_{i \in N} a(i, \pi(i)) \delta_{(i, \pi(i))}$.
- For $x \neq 0$, either $\{\Psi(\{(i,j)\}) = x\} = \emptyset$ or $\{\Psi(\{(i,j)\}) = x\} = \{\pi(i) = j\}$.
- $\pi_{i \to j}(k) = \begin{cases} j & ;\ k = i \\ \pi(i) & ;\ k = \pi^{-1}(j) \\ \pi(k) & ;\ \text{otherwise} \end{cases}$ is close to π.
- $W_{(i,j)} = \sum_{k \in N} a(k, \pi_{i \to j}(k))$ is close to W.

- $W = \sum_{i \in N} X_i$ with X_i being locally dependent
- For the domain of Ψ, take $N \times \mathbb{R}^d$.
- $\Psi = \sum_{i \in N} X_i \delta_{(i, X_i)}$.
- Conditioning at the point (i, x) is equivalent to conditioning on $\{X_i = x\}$.
- X_i is independent of the family $(X_j)_{\substack{j \neq i \\ j \nsim i}}$, where $\sim$ denotes adjacency.
- For $x \in \mathbb{R}^d$, choose $(X_{j|(i,x)})_{\substack{j \neq i \\ j \sim i}}$ so that its conditional distribution given $(X_j)_{\substack{j \neq i \\ j \nsim i}}$ matches the conditional distribution of $(X_j)_{\substack{j \neq i \\ j \sim i}}$ given $X_i = x$ and $(X_j)_{\substack{j \neq i \\ j \nsim i}}$.
- $W_{(i,x)} = x + \sum_{\substack{j \neq i \\ j \sim i}} X_{j|(i,x)} + \sum_{\substack{j \neq i \\ j \nsim i}} X_j$ is close to W.

Loosely speaking, the Stein expectation can be expressed in the form:

$$\mathbb{E}\big[\langle \nabla g(W), W\rangle - \Delta g(W)\big] = \iiint \mathbb{E}\langle \nabla^3 g(\tilde{W}), XXX\rangle\,,$$

where X stands for $\psi(d\xi)$, $\psi_\xi(d\eta)$ or $\psi^{(r)}_{\xi\eta}(d\zeta)$ and where $W^{(r)}_{\xi\eta} - W = \int \Psi^{(r)}_{\xi\eta}(d\zeta)$. In the combinatorial limit theorem, we have:

$$\iiint |XXX| \le \frac{K}{n} \sum_{i\in N} \sum_{j\in N} |a(i,j)|^3\,,$$

while for local dependence, we have:

$$\iiint |XXX| \le K(D+1)^2 \sum_{i\in N} \mathbb{E}\,|X_i|^3\,.$$

This allows us to bound $|\,\mathbb{E}\,f(W) - \mathbb{E}\,f(Z_d)|$, provided that $\nabla^3 g(W)$ is bounded. This happens if ∇f is Lipschitz.

TOWARDS THE MAIN RESULT (1)

Ball (1993), Bentkus (2003): for each $\varepsilon > 0$, there exist smooth functions f_ε^+ and f_ε^-, such that:

$$\mathbb{P}(W \in C) - \mathbb{P}(Z_d \in C) \le \mathbb{E}\, f_\varepsilon^+(W) - \mathbb{E}\, f_\varepsilon^+(Z_d)) + 4\varepsilon d^{1/4}$$
$$\mathbb{P}(W \in C) - \mathbb{P}(Z_d \in C) \ge \mathbb{E}\, f_\varepsilon^-(W) - \mathbb{E}\, f_\varepsilon^-(Z_d)) - 4\varepsilon d^{1/4}$$

and that $|\nabla^2 f_\varepsilon^{+/-}| = O(\varepsilon^{-2})$.

Next, the solution g_ε to the Stein equation for $f_\varepsilon^{+/-}$ can be expressed as:

$$g_\varepsilon = \int_0^{\pi/2} g_{\varepsilon,\alpha}\, d\alpha\,,$$

where:

$$g_{\varepsilon,\alpha}(w) = \mathbb{E}\, f^{+/-}(\cos\alpha\, w + \sin\alpha\, Z_d)\tan\alpha\,.$$

Then we have:

$$|\nabla^3 g_{\varepsilon,\alpha}| = O(\min\{\varepsilon^{-2}, \cot^2\alpha\})\,.$$

Thus, one can bound:

$$\left| \iiint \mathbb{E}\langle \nabla^3 g(\tilde{W}), XXX \rangle \right| = O(\min\{\varepsilon^{-2}, \cot^2 \alpha\}) \iiint |XXX| \,,$$

but this leads to suboptimal bounds. However, it can be shown that:

$$\left| \langle \mathbb{E}\, \nabla^3 g_{\varepsilon,\alpha}(Z_d), XXX \rangle \right| \leq \cos^2 \alpha \sin \alpha |XXX|$$

and therefore:

$$\left| \langle \mathbb{E}\, \nabla^3 g_{\varepsilon}(Z_d), XXX \rangle \right| \leq \frac{1}{6} |XXX| \,.$$

TOWARDS THE MAIN RESULT (3)

If $\tilde{W}$ is approximately normal and the error for the indicators of convex sets can be estimated by δ, we have:

$$\left|\mathbb{E}\langle\nabla^3 g_{\varepsilon,\alpha}(\tilde{W}) - \nabla^3 g_{\varepsilon,\alpha}(Z), XXX\rangle\right| \leq K\,\delta \min\{\varepsilon^{-2}, \cot^2\alpha\}|XXX|$$

Applying the estimate:

$$\int_0^{\pi/2} \min\{a, b\cot^2\alpha\}\, d\alpha \leq 2\sqrt{ab}$$

we find that:

$$|\,\mathbb{P}(W \in C) - \mathbb{P}(Z_d \in C)| \leq$$
$$\leq K\left(\varepsilon d^{1/4} + \left(1 + \frac{\delta}{\varepsilon}\right)\iiint |XXX|\right).$$

If $\delta = O(d^{1/4} \iiint |XXX|)$, one can choose $\varepsilon = \delta/d^{1/4}$ to obtain:

$$|\,\mathbb{P}(W \in C) - \mathbb{P}(Z_d \in C)| \leq K\,d^{1/4}\iiint |XXX|\,.$$

TOWARDS THE MAIN RESULT (4)

How do we know that $\tilde{W}$ is approximately normal?

- One reason can be that $\tilde{W}$ is actually close to W. The argument is most effective if the difference is bounded. This argument has been used several times (Rinott and Rotar (1996), Fang and Röllin (2012), Fang (2014)).
- Bolthausen and Götze (1993) used this argument in a wrong way.
- Götze (1991) uses induction: approximate normality of $\tilde{W}$ is the induction hypothesis. However, for sums of independent random vectors, this is easy.

TOWARDS THE MAIN RESULT (5)

- Take an index set Λ.
- For each $\lambda \in \Lambda$, take a random variable W_λ with $\mathbb{E}\, W_\lambda = 0$ and $\mathrm{var}(W_\lambda)) = \mathbf{I}$ along with all necessary decompositions.
- For each $\lambda \in \Lambda$, specify $\beta(\lambda)$, which is a suspected upper bound on the error in the normal approximation of W_λ (typically, one takes $\beta(\lambda) = \int\int\int |XXX|$).
- The goal is to prove that there exists a constant K, such that:
$$\left|\mathbb{P}(W_\lambda \in C) - \mathbb{P}(Z_d \in C)\right| \leq K\, \beta(\lambda)$$
for all $\lambda \in \Lambda$ and for all $C \in \mathscr{C}$.
- Typically, there are several random variables $\tilde{W}_\lambda$ arising from decompositions. Assume that each one of them has the same distribution as $\tilde{\mathbf{Q}}_\lambda W_{\tilde{\lambda}} + \tilde{w}_\lambda$.
- The induction gets through if $\tilde{\mathbf{Q}}_\lambda \tilde{\mathbf{Q}}_\lambda^{\mathrm{T}}$ are not too far from $\mathbf{I}$ and if $\beta(\tilde{\lambda})$ are not too different from $\beta(\lambda)$.

The differences between $\tilde{\mathbf{Q}}_\lambda \tilde{\mathbf{Q}}_\lambda^{\mathrm{T}}$ and $\mathbf{I}$ and between $\beta(\tilde{\lambda})$ and $\beta(\lambda)$ can give rise to higher moments. To avoid this, one needs to compare the first- and the third-order expansions:

$$\begin{aligned} \langle \mathbb{E}\, \nabla g(\tilde{W}_\lambda^{(1)}), X \rangle \qquad & \tilde{W}_\lambda^{(1)} \stackrel{\mathrm{d}}{=} \tilde{\mathbf{Q}}_\lambda^{(1)} \tilde{W}_{\tilde{\lambda}_1} + \tilde{w}_\lambda^{(1)}\,, \\ \langle \mathbb{E}\, \nabla^3 g(\tilde{W}_\lambda^{(3)}), XXX \rangle \qquad & \tilde{W}_\lambda^{(3)} \stackrel{\mathrm{d}}{=} \tilde{\mathbf{Q}}_\lambda^{(3)} \tilde{W}_{\tilde{\lambda}_3} + \tilde{w}_\lambda^{(3)}\,. \end{aligned}$$

Let D stand either for:

$$\mathbb{E}\Big[\min\Big\{1, \operatorname{tr}\big|\mathbf{I} - \tilde{\mathbf{Q}}_\lambda^{(1)} \tilde{\mathbf{Q}}_\lambda^{(1)\mathrm{T}}\big| + \big(\beta(\tilde{\lambda}_1) - \beta(\lambda)\big)_+\Big\}\Big]$$

or for:

$$\mathbb{E}\Big[\min\Big\{1, \big\|\mathbf{I} - \tilde{\mathbf{Q}}_\lambda^{(3)} \tilde{\mathbf{Q}}_\lambda^{(3)\mathrm{T}}\big\| + \big(\beta(\tilde{\lambda}_3) - \beta(\lambda)\big)_+\Big\}\Big]\,.$$

MAIN RESULT

Theorem

If there exists a constant K_1, such that for all $\lambda \in \Lambda$,

$$\iiint |XXX| + \int |X|D + \int |X| \left(\iint |XX|D \right)^{2/3} \leq K_1 \, \beta(\lambda) \, ,$$

then there exists another constant K_2, such that for all $\lambda \in \Lambda$ and all $C \in \mathscr{C}$,

$$\left| \mathbb{P}(W_\lambda \in C) - \mathbb{P}(Z_d \in C) \right| \leq K_2 \, d^{1/4} \, \beta(\lambda) \, .$$

The combinatorial CLT and the result for the local dependence follow.

THANK YOU FOR YOUR ATTENTION!