On Götze's multivariate central limit therem: doubts, clarification and improvements

Martin Raič

University of Ljubljana
and
University of Primorska
Slovenia

Symposium in Memory of Charles Stein
NUS, IMS, Singapore

June 2019

STATEMENT

- $X_1, X_2, \dots, X_n$ ... independent $\mathbb{R}^d$-valued random vectors with sum W
- $\mathbb{E}\, X_i = \mathbf{0}, \quad \mathrm{Var}(W) = \mathbf{I}_d$
- $\mathscr{C}_d$ ... the family of all measurable convex subsets of $\mathbb{R}^d$
- $\|\cdot\|$ ... Euclidean norm
- $\beta_3 := \sum_{i=1}^{n} \mathbb{E}\, \|X_i\|^3$

Theorem (Götze, 1991)

For each d, there exists K_d, such that

$$\left|\mathbb{P}(W \in C) - \mathrm{N}(0, \mathbf{I}_d)\{C\}\right| \leq K_d \beta_3$$

for all W and all $C \in \mathscr{C}_d$.
For $d \geq 6$, we have $K_d \leq 157.85\, d + 10$.

DOUBT

Theorem (Götze, 1991)

For each d, there exists K_d, such that

$$\left|\mathbb{P}(W \in C) - \mathrm{N}(0, \mathbf{I}_d)\{C\}\right| \leq K_d \beta_3$$

for all W and all $C \in \mathscr{C}_d$.
For $d \geq 6$, we have $K_d \leq 157.85\, d + 10$.

Theorem (Bhattacharya and Holmes, 2010)

$$\left|\mathbb{P}(W \in C) - \mathrm{N}(0, \mathbf{I}_d)\{C\}\right| \leq K\, d^{5/2} \beta_3$$

STEIN'S METHOD FOR THE MULTIVARIATE NORMAL DISTRIBUTION

We have

$$\mathbb{E}\big[f(W)\big] - \mathrm{N}(0, \mathbf{I}_d)\{f\} = \mathbb{E}\big[\mathscr{S} g(W)\big]\,,$$

where

$$\mathscr{S} g(W) = \Delta g(w) - \langle \nabla g(w)\,,\, w\rangle$$

and where g solves $\mathscr{S} g(w) = f(w) - \mathrm{N}(0, \mathbf{I}_d)\{f\}$.

$$g(w) = \int_0^\infty \int_{\mathbb{R}^d} \Big(\mathrm{N}(\mathbf{0}, \mathbf{I}_d)\{f\} - f\big(e^{-t}w + z\sqrt{1 - e^{-2t}}\big)\Big) \phi_d(z)\, dz\, dt$$

$$= \int_0^{\pi/2} \int_{\mathbb{R}^d} \Big(\mathrm{N}(\mathbf{0}, \mathbf{I}_d)\{f\} - f\big(w \cos\alpha + z \sin\alpha\big)\Big) \phi_d(z)\, dz \tan\alpha\, d\alpha$$

where $\phi_d(z) = (2\pi)^{-d/2} e^{-\|z\|^2/2}$ is the standard normal density.

Let $w, x_1, \ldots, x_r$ in $\mathbb{R}^d$ with $w_i = (w_1, \ldots, w_d)$ and $x_i = (x_{i1}, \ldots, x_{id})$. Define

$$\langle \nabla^r g(w)\,,\, x_1 \otimes x_2 \otimes \cdots \otimes x_r \rangle := \sum_{j_1,\ldots,j_r=1}^{d} D_{j_1 j_2 \cdots j_r} g(w)\, x_{1j_1} x_{2j_2} \cdots x_{dj_d}\,.$$

$\nabla^r g(w)$ can be regarded as a symmetric r-fold tensor.

$$\mathbb{E}\big[f(W)\big] - \mathrm{N}(0, \mathbf{I}_d)\{f\} = -\int_0^{\pi/2} \mathbb{E}\big[\mathscr{S}\,\mathscr{U}_\alpha f(W)\big] \tan\alpha \, d\alpha \,,$$

where

$$\mathscr{U}_\alpha f(w) := \int_{\mathbb{R}^d} f\big(w\cos\alpha + z\sin\alpha\big)\, \phi_d(z)\, dz \,.$$

Taylor's expansion yields

$$\begin{aligned}
&\mathbb{E}\big[\mathscr{S} g(W)\big] \\
&\quad = \sum_{i=1}^{n} \mathbb{E}\Big[\Big\langle \nabla^3 g(W_i + \theta X_i)\,,\, X_i \otimes \tilde{X}_i^{\otimes 2} - (1-\theta) X_i^{\otimes 3} \Big\rangle\Big] \,,
\end{aligned}$$

where $W_i = W - X_i$, $\tilde{X}_i$ is an independent copy of X_i, θ is uniformly distributed over $[0, 1]$, and $\tilde{X}_i$ and θ are independent of each other and all other variates.

Bhattacharya and Holmes estimate

$$\left|\langle \nabla^3 g(w)\,,\, x\otimes y\otimes z\rangle\right| \le \sum_{j,k,l=1}^{d} \left|D_{jkl}(w)\right| |x_i|\,|y_j|\,|z_l|\,.$$

$$\mathscr{U}_\alpha f(w) = \int_{\mathbb{R}^d} f\big(w\cos\alpha + z\sin\alpha\big)\,\phi_d(z)\,dz\,,$$

$$D_{jkl}\mathscr{U}_\alpha f(w) = -\cot^3\alpha \int_{\mathbb{R}^d} f\big(w\cos\alpha + z\sin\alpha\big)\,D_{jkl}\phi_d(z)\,dz\,,$$

$$\Big|\Big\langle \nabla^3 \mathscr{U}_\alpha f(w)\,,\, x\otimes y\otimes z\Big\rangle\Big| \le c_3\|f\|_\infty \sum_{j,k,l=1}^{d} |x_j|\,|y_k|\,|z_l|\cot^3\alpha\,,$$

where

$$c_3 = \max_{j,k,l}\int_{\mathbb{R}^d}\left|D_{jkl}\phi_d(z)\right|dz = \int_{\mathbb{R}^d}\left|D_{111}\phi_d(z)\right|dz = \int_{\mathbb{R}}\left|\phi_1'''(z)\right|dz$$

$$\left|\langle \nabla^3 \mathscr{U}_\alpha f(w)\,,\, x \otimes y \otimes z\rangle\right| \leq c_3 \|f\|_\infty \sum_{j,k,l=1}^{d} |x_j|\,|y_k|\,|z_l| \cot^3 \alpha$$
$$\leq c_3 \|f\|_\infty \cdot d^{3/2} \|x\|\,\|y\|\,\|z\| \cot^3 \alpha\,.$$

Can we do better?

Yes, thanks to the fact that $\nabla^3 g(w)$ is a symmetric 3-tensor.

Theorem (Banach, 1938)

If T is a symmetric r-tensor, then

$$\|T\|_\vee := \sup_{\|x_1\|,\ldots,\|x_r\|\leq 1} \left|\langle T\,,\, x_1 \otimes \cdots \otimes x_r\rangle\right| = \sup_{\|x\|\leq 1} \left|\langle T\,,\, x^{\otimes r}\rangle\right|.$$

Therefore,

$$\left|\langle \nabla^3 \mathscr{U}_\alpha f(w)\,,\, x \otimes y \otimes z\rangle\right| \leq \left\|\nabla^3 \mathscr{U}_\alpha f(w)\right\|_\vee \|x\|\,\|y\|\,\|z\|$$
$$\leq c_3 \|f\|_\infty \|x\|\,\|y\|\,\|z\| \cot^3 \alpha\,.$$

For $f = f_C^{\downarrow} := f_C - \frac{1}{2}$, where $f_C(w) = \mathbf{1}(w \in C)$, we obtain

$$\Big|\mathbb{E}\big[\mathscr{S}\mathscr{U}_\alpha f_C(W)\big]\Big| \tan\alpha = \Big|\mathbb{E}\big[\mathscr{S}\mathscr{U}_\alpha f_C^{\downarrow}(W)\big]\Big| \tan\alpha \le \tfrac{3}{4} c_3 \beta_3 \cot^2\alpha\,,$$

but this is not integrable with respect to α.

However,

$$\begin{aligned}
&\mathscr{U}_\varepsilon f_C(w) - \mathrm{N}(0, \mathbf{I}_d)\{f_C\} \\
&\qquad = \int_\varepsilon^{\pi/2} \big(\mathrm{N}(\mathbf{0}, \mathbf{I}_d)\{f\} - \mathscr{U}_\alpha f_C(w)\big) \tan\alpha \, d\alpha\,, \\
&\Big|\mathbb{E}\big[\mathscr{U}_\varepsilon f_C(W)\big] - \mathrm{N}(0, \mathbf{I}_d)\{f_C\}\Big| \le \tfrac{3}{4} c_3 \beta_3 \int_\varepsilon^{\pi/2} \cot^2\alpha \, d\alpha \\
&\qquad \le \tfrac{3}{4} c_3 \beta_3 \cot\varepsilon\,.
\end{aligned}$$

SMOOTHING (2)

$$\left|\mathbb{E}\left[\mathscr{U}_\varepsilon f_C(w)\right] - \mathrm{N}(0,\mathbf{I}_d)\{f_C\}\right| \leq \tfrac{3}{4} c_3 \beta_3 \cot\varepsilon$$

Smoothing cost:

$$\sup_{C\in\mathscr{C}_d}\left|\mathbb{E}\left[f_C(W)\right] - \mathrm{N}(0,\mathbf{I}_d)\{f_C\}\right|$$
$$\leq \tfrac{4}{3}\sup_{C\in\mathscr{C}_d}\left|\mathbb{E}\left[\mathscr{U}_\varepsilon f_C(W)\right] - \mathrm{N}(0,\mathbf{I}_d)\{f_C\}\right| + \tfrac{5}{2}\chi_{7/8;d}\,\gamma_d \tan\varepsilon\,,$$

where $\chi_{p;d}$ is the p-th quantile of the $\chi(d)$-distribution and where $\gamma_d := \sup_{C\in\mathscr{C}_d}\int_{\partial C}\phi_d\, d\,\mathrm{Vol}_{d-1}$.
Optimization in ε gives

$$\sup_{C\in\mathscr{C}_d}\left|\mathbb{P}(W\in C) - \mathrm{N}(0,\mathbf{I}_d)\{C\}\right| \leq \sqrt{10\,\chi_{7/8;d}\,\gamma_d\beta_3}\,.$$

For $d \geq 6$, $\chi_{7/8;d} \leq 1.27\sqrt{d}$.
In 1991, it was only known that $\gamma_d \leq 2\sqrt{d}$.

INDUCTION (1)

In order to improve the rate of convergence, we can condition the expression

$$\mathbb{E}\big[\mathscr{S} g(W)\big] = \sum_{i=1}^{n} \mathbb{E}\Big[\Big\langle \nabla^3 g(W_i + \theta X_i)\,,\, X_i \otimes \tilde{X}_i^{\otimes 2} - (1-\theta) X_i^{\otimes 3} \Big\rangle\Big]\,,$$

on $X_i, \tilde{X}_i$, apply independence and use the approximate normality of W_i as an induction hypothesis. Approximate normality helps because

$$\big\| \mathrm{N}(\mu, \boldsymbol{\Sigma})\big\{\nabla^3 \mathscr{U}_\alpha f\big\}\big\|_\vee \le c_3 \|f\|_\infty \frac{\cos^3\alpha}{\sigma^3}\,,$$

where σ^2 is the smallest eigenvalue of $\boldsymbol{\Sigma}$.

INDUCTION (2)

Suppose that we have proved

$$\sup_{C\in\mathscr{C}_d}\left|\mathbb{P}(W\in C)-\mathrm{N}(0,\mathbf{I}_d)\{C\}\right|\leq K_d\beta_3$$

for all suitable sums W of $n-1$ random vectors. If W is a suitable sum of n random vectors, then

$$\left|\mathbb{E}\left[\mathscr{S}\mathscr{U}_\alpha f_C(W)\right]\right|\tan\alpha\leq\frac{\frac{3}{4}c_3\beta_3\cos^2\alpha\sin\alpha+\frac{3}{2}c_3\beta_3^2\cot^2\alpha}{\sigma^3},$$

where σ^2 is the smallest one of all eigenvalues of $\mathrm{Var}(W_i)$. Therefore,

$$\sup_{C\in\mathscr{C}_d}\left|\mathbb{P}(W\in C)-\mathrm{N}(0,\mathbf{I}_d)\{C\}\right|$$
$$\leq\frac{\frac{1}{3}c_3\beta_3+2c_3K_d\beta_3^2\cot\varepsilon}{\sigma^3}+\frac{5}{2}\chi_{7/8;d}\,\gamma_d\tan\varepsilon\,.$$

INDUCTION (3)

$$\sup_{C \in \mathscr{C}_d} \left| \mathbb{P}(W \in C) - \mathrm{N}(0, \mathbf{I}_d)\{C\} \right|$$
$$\leq \frac{\frac{1}{3} c_3 \beta_3 + 2 c_3 K_d \beta_3^2 \cot \varepsilon}{\sigma^3} + \tfrac{5}{2} \chi_{7/8;d}\, \gamma_d \tan \varepsilon \,.$$

Choosing $\tan \varepsilon = \dfrac{4 c_3 \beta_3}{\sigma^3}$, the right hand side reduces to

$$\left(\frac{\frac{1}{3} c_3 + 10 c_3\, \chi_{7/8;d}\, \gamma_d}{\sigma^3} + \frac{K_d}{2} \right) \beta_3 \,.$$

Using $\sigma^2 \leq 1 - \beta_3^{2/3}$ and assuming w.l.o.g. that β_3 is sufficiently small, we can complete the induction.

For $d \geq 6$, we have $\chi_{7/8;d}\, \gamma_d \leq 2.54\, d$.

- In 1991, it was known that $\gamma_d \leq 2\sqrt{d}$.
- Ball (1994): $\gamma_d \leq 4\, d^{1/4}$
- Nazarov (2003): $0.28 < \liminf_{d \to \infty} \frac{\gamma_d}{d^{1/4}} \leq \limsup_{d \to \infty} \frac{\gamma_d}{d^{1/4}} < 0.76$
- R. (2018): $\gamma_d \leq 0.59\, d^{1/4} + 0.21$

This reduces $K_d = O(d)$ to $K_d = O(d^{3/4})$.

MORE EFFICIENT SMOOTHING (1)

Senatov (1980), Bentkus (1986, 2003):

- $C^{\varepsilon} = \{x\,;\,\mathrm{dist}(x, C) \le \varepsilon\}$; $\quad C^{-\varepsilon} = \{x\,;\,\mathrm{dist}(x, C^c) \ge \varepsilon\}$;
- $f_{C^{-\varepsilon}} \le f_C^{-\varepsilon} \le f_C \le f_C^{\varepsilon} \le f_{C^{\varepsilon}}$;
- $\big|\langle \nabla f(w) - \nabla f(z)\,,\, x\rangle\big| \le \dfrac{8\|w - z\|\|x\|}{\varepsilon^2}$ for $f \in \{f_C^{-\varepsilon}, f_C^{\varepsilon}\}$.

This allows us to bound $\|\nabla^3 \mathscr{U}_\alpha f\|_\vee \le \dfrac{4c_1 \cot\alpha}{\varepsilon^2}$
and we already know that $\|\nabla^3 \mathscr{U}_\alpha f\|_\vee \le \frac{1}{2} c_3 \cot^3 \alpha$.
Combining both estimates, the induction step changes to

$$\sup_{C \in \mathscr{C}_d} \big|\mathbb{P}(W \in C) - \mathrm{N}(0, \mathbf{I}_d)\{C\}\big|$$

$$\le \beta_3 \int_0^{\pi/2} \min\biggl\{ \frac{12 c_1 \cos\alpha}{\varepsilon} \left(\frac{\gamma_d}{\sigma} + \frac{2 K_d \beta_3}{\varepsilon \sigma^3} \right),$$

$$+ 3c_3 \left(\frac{\cos^2\alpha \sin\alpha}{4\sigma^3} + \frac{K_d \beta_3 \cot^2\alpha}{2\sigma^3} \right) \biggr\} d\alpha + \gamma_d \varepsilon\,.$$

MORE EFFICIENT SMOOTHING (2)

$$\sup_{C \in \mathscr{C}_d} \left| \mathbb{P}(W \in C) - \mathrm{N}(0, \mathbf{I}_d)\{C\} \right|$$

$$\leq \beta_3 \int_0^{\pi/2} \min\left\{ \frac{12 c_1 \cos\alpha}{\varepsilon} \left(\frac{\gamma_d}{\sigma} + \frac{2 K_d \beta_3}{\varepsilon \sigma^3} \right), \right.$$

$$\left. + 3 c_3 \left(\frac{\cos^2\alpha \sin\alpha}{4\sigma^3} + \frac{K_d \beta_3 \cot^2\alpha}{2\sigma^3} \right) \right\} d\alpha + \gamma_d \varepsilon \, .$$

The last term no longer contains a factor like $\chi_{7/8,d}$.

Important:

$$\int_0^{\pi/2} \min\{A, \; B \cot^2\alpha\} \, d\alpha \leq 2\sqrt{AB} \, .$$

This allows us to derive

$$\sup_{C \in \mathscr{C}_d} \left| \mathbb{P}(W \in C) - \mathrm{N}(0, \mathbf{I}_d)\{C\} \right| \leq (71\gamma_d + 1)\beta_3$$

$$\leq (42 \, d^{1/4} + 16)\beta_3 \, .$$

EXTENSION TO DEPENDENCE?

- For dependent random vectors, we would need induction over conditional distributions.
- Conditional β_3 can be larger than unconditional.
- It is challenging to bound the error in terms of third moments.

  In particular, if $(X_i)_{i\in\mathscr{I}}$ fit a dependence graph with maximum degree of D, is it true that

$$\sup_{C\in\mathscr{C}_d}\left|\mathbb{P}(W\in C)-\mathrm{N}(0,\mathbf{I}_d)\{C\}\right|\le Kd^{1/4}D^2\beta_3$$

  for some universal constant K?

POSSIBLE STRUCTURE OF RANDOM VECTORS

- W_λ, $\lambda \in \Lambda$, defined on possibly different probability spaces;
- $\mathbb{E}\, W_\lambda = \mathbf{0}, \quad \mathrm{Var}(W_\lambda) = \mathbf{I}_d$;
- $V_{\lambda\xi}$ satisfying

$$\mathbb{E}\big[f(W_\lambda)W_\lambda\big] = \int_{\Xi_\lambda} \mathbb{E}\big[f(V_{\lambda\xi})\big]\,\mu_\lambda(d\xi)\,;$$

  similar to the Stein coupling

$$\mathbb{E}\big[f(W_\lambda)W_\lambda\big] = \mathbb{E}\Big[G_\lambda\big(f(W'_\lambda) - f(W_\lambda)\big)\Big]$$

  (μ_λ are vector-valued measures);
- $\mathbb{E}\big[f(W_\lambda)\big] - \mathbb{E}\big[f(V_{\lambda\xi})\big] = \int_{\Xi_\lambda} \big\langle \mathbb{E}\big[\nabla f(V_{\lambda\eta})\big]\,,\, \nu_{\lambda\xi}(d\eta)\big\rangle$
  (the total variation of the vector-valued measure $\nu_{\lambda\xi}$ measures the proximity of $V_{\lambda\xi}$ to W_λ);
- $V_{\lambda\xi} \overset{\mathrm{d}}{=} Q_{\lambda\xi} W_{[\lambda,\xi]} + v_{\lambda\xi}$.

THANK YOU FOR YOUR ATTENTION!