Submitted to the Annals of Statistics

A COUNTEREXAMPLE TO GAUSSIAN APPROXIMATION FOR MAXIMA OF SELF-NORMALIZED SUMS

BY WEI BIAO WU[1,a]

[1]*Department of Statistics, University of Chicago,* [a]*wbwu@uchicago.edu*

Marginal moment bounds do not suffice for Gaussian approximation of maxima of coordinatewise self-normalized sums under unrestricted coordinate dependence. We give an elementary example with polynomial dimension, unit variances and uniformly bounded marginal moments of any prescribed order greater than two. Its self-normalized maximum converges to $\max\{G, 1\}$, while the Gaussian maximum with the same population covariance converges to $G \sim N(0,1)$. A second construction has uniformly bounded marginal moments of every fixed order, dimension satisfying $\log p = O((\log n)^3)$, and Kolmogorov approximation error tending to one. The latter failure can also be realized by coordinate projections of a single fixed infinite-dimensional population. Both constructions apply to absolute maxima and Studentized means. They disprove the unrestricted, dimension-uniform extension proposed in Remark 2.4 of Qiu, Chen and Shao (2025), while leaving its proved theorems intact.

1. Introduction. Let $X_{n1}, \ldots, X_{nn}$ be iid p_n-dimensional random vectors with mean zero, unit coordinate variances and covariance matrix R_n. Write

$$S_{nj} = \sum_{i=1}^{n} X_{nij}, \qquad V_{nj}^2 = \sum_{i=1}^{n} X_{nij}^2, \qquad U_{nj} = S_{nj}/V_{nj}, \tag{1.1}$$

and let $Z_n \sim N(0, R_n)$. We study the approximation of

$$M_n = \max_{j \le p_n} U_{nj}, \qquad M_n^{\mathrm{abs}} = \max_{j \le p_n} |U_{nj}| \tag{1.2}$$

by $H_n = \max_j Z_{nj}$ and $H_n^{\mathrm{abs}} = \max_j |Z_{nj}|$, respectively. For real random variables A, B, set

$$d_{\mathrm{K}}(A,B) = \sup_{x \in \mathbb{R}} |\mathbb{P}(A \le x) - \mathbb{P}(B \le x)|.$$

The classical self-normalized moderate deviation theory permits polynomial moment assumptions; see Shao (1999) and Jing, Shao and Wang (2003). In high dimension, Qiu, Chen and Shao (2025) prove a relative Gaussian approximation under uncorrelatedness and extend it to specified correlation classes. Their Remark 2.4, equations (2.16)–(2.17), proposes that the marginal bound

$$\sup_n \max_{j \le p_n} \mathbb{E}|X_{nij}|^{2+\tau} < \infty, \qquad 0 < \tau \le 1,$$

suffices for arbitrary correlation matrices:

$$\frac{\mathbb{P}(M_n > x)}{\mathbb{P}(H_n > x)} \longrightarrow 1, \qquad \frac{\mathbb{P}(M_n^{\mathrm{abs}} > x)}{\mathbb{P}(H_n^{\mathrm{abs}} > x)} \longrightarrow 1, \qquad 0 \le x = o\left(n^{\tau/(4+2\tau)}\right), \tag{1.3}$$

MSC2020 subject classifications. Primary 60F05; secondary 62E20, 60F10.

Keywords and phrases. Gaussian approximation, self-normalization, Studentization, high dimension, rare observations.

uniformly in dimension. The counterexamples below address this proposed unrestricted extension.

The mechanism combines a common Gaussian variable with coordinate-specific rare observations. The rare component contributes negligibly to each coordinate's variance, so the matching Gaussian vector is almost constant across coordinates. Nevertheless, maximizing over sufficiently many coordinates selects a column containing several rare observations. A column with k such observations has self-normalized sum approximately $\sqrt{k}$.

Section 2 gives the construction and two elementary lemmas. Section 3 provides a polynomial-dimensional example and a stronger example with all marginal moments bounded and approximation error tending to one. Section 4 realizes the latter failure within a fixed population. All limits are taken as $n \to \infty$, unless stated otherwise, and all logarithms are natural.

2. A common-factor construction. Let $p = p_n \ge 2$ and $q = q_n \in (0,1)$. Let $Y_1, Y_2, \ldots$ be iid standard normals, and let B_{nij} be independent Bernoulli(q_n) variables, independent of the Y_i. Define

$$v_n = n^2 q_n(1-q_n), \qquad X_{nij} = \frac{Y_i + n(B_{nij} - q_n)}{\sqrt{1+v_n}}. \tag{2.1}$$

For each n, the rows are iid and

$$\mathbb{E}X_{nij} = 0, \quad \mathbb{E}X_{nij}^2 = 1, \qquad R_n = \frac{\mathbf{1}_p\mathbf{1}_p' + v_n I_p}{1+v_n}. \tag{2.2}$$

In particular, R_n is positive definite. For every $r \ge 1$,

$$\mathbb{E}|X_{nij}|^r \le 2^{r-1}\left\{\mathbb{E}|Y_1|^r + n^r(q_n + q_n^r)\right\}. \tag{2.3}$$

Both denominators in the self-normalized and Studentized statistics below are positive almost surely for $n \ge 2$: conditional on a column of B, its observations have a nondegenerate Gaussian density on $\mathbb{R}^n$.

LEMMA 2.1 (Column occupancies). *Let $K_{nj} = \sum_{i=1}^n B_{nij}$ and $\lambda_n = nq_n$. Suppose $\lambda_n \to 0$, $k_n \ge 1$ is integer, $k_n^2/n \to 0$, and*

$$\theta_n := \frac{p_n \lambda_n^{k_n}}{k_n!} \longrightarrow \infty, \qquad \frac{\theta_n \lambda_n}{k_n+1} \longrightarrow 0. \tag{2.4}$$

Then

$$\mathbb{P}\left(\min_{j\le p_n} K_{nj} = 0,\ \max_{j \le p_n} K_{nj} = k_n\right) \longrightarrow 1. \tag{2.5}$$

PROOF. For $k = k_n$,

$$\mathbb{P}(K_{n1} = k) = \frac{\lambda_n^k}{k!}\prod_{s=0}^{k-1}\left(1 - \frac{s}{n}\right)(1-q_n)^{n-k} = (1+o(1))\frac{\lambda_n^k}{k!}.$$

The product tends to one because $k_n^2/n \to 0$ and $nq_n \to 0$. Independence among columns and the first condition in (2.4) give

$$\mathbb{P}(K_{nj} \ne k_n \text{ for all } j) \le \exp\{-p_n \mathbb{P}(K_{n1} = k_n)\} \longrightarrow 0.$$

The factorial-moment bound and the second condition give

$$\mathbb{P}(\max_j K_{nj} \ge k_n + 1) \le \frac{p_n (nq_n)^{k_n+1}}{(k_n+1)!} = \frac{\theta_n \lambda_n}{k_n+1} \longrightarrow 0.$$

Finally, if every column is nonempty, then $\sum_j K_{nj} \ge p_n$. Markov's inequality gives

$$\mathbb{P}(\min_j K_{nj} \ge 1) \le nq_n = \lambda_n \to 0.$$

$\square$

LEMMA 2.2 (Self-normalized and Gaussian maxima). *For (2.1), suppose $n^{3/2}q_n \to 0$, the positive integers k_n satisfy $k_n = O(\log n)$, and (2.5) holds. Set*

$$D_{ni} = Y_i - nq_n, \qquad A_n = \sum_i D_{ni}, \qquad Q_n = \sum_i D_{ni}^2, \qquad W_n = A_n/\sqrt{Q_n}. \tag{2.6}$$

Then $W_n \Rightarrow N(0,1)$ and

$$M_n = \max\{W_n, \sqrt{k_n}\} + o_{\mathbb{P}}(1), \qquad M_n^{\text{abs}} = \max\{|W_n|, \sqrt{k_n}\} + o_{\mathbb{P}}(1). \tag{2.7}$$

If, in addition, $v_n \log p_n \to 0$, then

$$H_n \Rightarrow G, \qquad H_n^{\text{abs}} \Rightarrow |G|, \qquad G \sim N(0,1). \tag{2.8}$$

The relations in (2.7) also hold with U_{nj} replaced by the Student statistic $t_{nj} = \sqrt{n}\,\bar{X}_{nj}/\hat{\sigma}_{nj}$, where

$$\bar{X}_{nj} = n^{-1}S_{nj}, \qquad \hat{\sigma}_{nj}^2 = (n-1)^{-1}\sum_i (X_{nij} - \bar{X}_{nj})^2. \tag{2.9}$$

PROOF. We have

$$A_n/\sqrt{n} \Rightarrow N(0,1), \qquad Q_n/n \to_{\mathbb{P}} 1, \qquad \max_{i\le n} |D_{ni}| = O_{\mathbb{P}}(\sqrt{\log n}). \tag{2.10}$$

For the first assertion, use $A_n/\sqrt{n} = n^{-1/2}\sum_i Y_i - n^{3/2}q_n$; the other assertions follow from the Gaussian law of large numbers and a union bound.

The factor $(1+v_n)^{-1/2}$ cancels in U_{nj}. If $K_{nj} = 0$, then $U_{nj} = W_n$. If $K = K_{nj} \ge 1$, direct expansion gives

$$U_{nj} = \frac{A_n + nK}{\left(Q_n + 2n\sum_{i:B_{nij}=1} D_{ni} + n^2K\right)^{1/2}} = \frac{\sqrt{K} + A_n/(n\sqrt{K})}{\sqrt{1+r_{nj}}}, \tag{2.11}$$

where

$$|r_{nj}| \le \eta_n := \frac{Q_n}{n^2} + \frac{2}{n}\max_{i\le n}|D_{ni}| = O_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n}\right).$$

On $\{\eta_n \le 1/2,\ \max_j K_{nj} \le k_n\}$, the mean value theorem implies

$$\max_{j:K_{nj}\ge 1} |U_{nj} - \sqrt{K_{nj}}| \le C\left(\frac{|A_n|}{n} + \sqrt{k_n}\,\eta_n\right) = O_{\mathbb{P}}\left(n^{-1/2} + \frac{\sqrt{k_n \log n}}{n}\right) = o_{\mathbb{P}}(1). \tag{2.12}$$

The existence of both an empty column and a column with k_n observations now proves (2.7). The assertion for absolute maxima follows also from $\big||U_{nj}| - \sqrt{K_{nj}}\big| \le |U_{nj} - \sqrt{K_{nj}}|$.

For the Gaussian comparison, take independent standard normals $G_0, G_1, \ldots, G_p$ and write

$$Z_{nj} = \frac{G_0 + \sqrt{v_n}\,G_j}{\sqrt{1+v_n}}. \tag{2.13}$$

Since $\max_{j\le p}|G_j| = O_{\mathbb{P}}(\sqrt{\log p})$, the additional assumption gives $\max_j |Z_{nj} - G_0| \to_{\mathbb{P}} 0$, proving (2.8).

Finally, the exact transformation between the two normalizations is

$$t_{nj} = f_n(U_{nj}), \qquad f_n(u) = \frac{\sqrt{n-1}\,u}{\sqrt{n-u^2}}, \qquad |u| < \sqrt{n}. \tag{2.14}$$

The map is odd and strictly increasing. Thus the Studentized maximum is $f_n(M_n)$, and its absolute maximum is $f_n(M_n^{\mathrm{abs}})$. By (2.7), both arguments are $O_{\mathbb{P}}(\sqrt{\log n})$. The elementary estimate

$$|f_n(u) - u| \le C\frac{|u| + |u|^3}{n}, \qquad u^2 \le n/2,$$

shows that both changes are $o_{\mathbb{P}}(1)$. □

3. Two counterexamples.

3.1. *Polynomial dimension.*

THEOREM 3.1. *Fix $a > 2$. In (2.1), take $p_n = \lceil n^a \rceil$ and $q_n = 1/p_n$. Then*

$$\sup_{n\ge 2}\max_{j\le p_n} \mathbb{E}|X_{nij}|^a \le 2^{a-1}\{\mathbb{E}|G|^a + 2\} < \infty, \tag{3.1}$$

and R_n is positive definite with unit diagonal. Moreover,

$$M_n \Rightarrow \max\{G, 1\}, \quad M_n^{\mathrm{abs}} \Rightarrow \max\{|G|, 1\}, \quad H_n \Rightarrow G, \quad H_n^{\mathrm{abs}} \Rightarrow |G|. \tag{3.2}$$

The Studentized maxima have the same limits as their self-normalized counterparts. In particular,

$$\liminf_n d_{\mathrm{K}}(M_n, H_n) \ge \Phi(1), \qquad \liminf_n d_{\mathrm{K}}(M_n^{\mathrm{abs}}, H_n^{\mathrm{abs}}) \ge 2\Phi(1) - 1. \tag{3.3}$$

PROOF. Since $n^a(q_n + q_n^a) \le 2$, (3.1) follows from (2.3). In Lemma 2.1, choose $k_n = 1$. Then

$$\lambda_n = n/p_n \to 0, \qquad \theta_n = n \to \infty, \qquad \theta_n\lambda_n/2 = n^2/(2p_n) \to 0.$$

Furthermore, $n^{3/2}q_n \to 0$ and $v_n \log p_n = O(n^{2-a}\log n) \to 0$. Lemma 2.2 proves (3.2) and its Studentized version.

For every fixed $0 < x < 1$, the limits are evaluated at continuity points and yield

$$\frac{\mathbb{P}(M_n > x)}{\mathbb{P}(H_n > x)} \longrightarrow \frac{1}{\overline{\Phi}(x)}, \qquad \frac{\mathbb{P}(M_n^{\mathrm{abs}} > x)}{\mathbb{P}(H_n^{\mathrm{abs}} > x)} \longrightarrow \frac{1}{2\overline{\Phi}(x)}, \qquad \overline{\Phi}(x) = 1 - \Phi(x). \tag{3.4}$$

The corresponding absolute tail differences converge to $\Phi(x)$ and $2\Phi(x) - 1$. Taking the supremum over $0 < x < 1$ proves (3.3); no convergence assertion at the atom $x = 1$ is needed. □

Here $\log p_n = O(\log n) = o(n^c)$ for every $c > 0$. Thus the failure cannot be removed by decreasing a positive exponent in a condition of the form $\log p_n = o(n^c)$.

3.2. *All marginal moments and maximal approximation error.* The next choice permits the largest column occupancy to diverge while the Gaussian maximum remains tight.

THEOREM 3.2. *For $n \geq 3$, set*

$$q_n = \exp\{-(\log n)^2\}, \quad \lambda_n = nq_n, \quad k_n = \lfloor \log n \rfloor, \quad p_n = \left\lceil k_n!\, \lambda_n^{-(k_n+1/2)} \right\rceil, \tag{3.5}$$

and use (2.1). Then

$$\sup_{n\geq 3} \max_{j\leq p_n} \mathbb{E}|X_{nij}|^r < \infty \quad \textit{for every fixed } r > 0, \qquad \log p_n \sim (\log n)^3. \tag{3.6}$$

In fact, $\mathbb{E}|X_{n11}|^r \to \mathbb{E}|G|^r$ for every fixed $r > 0$. Nevertheless,

$$M_n - \sqrt{k_n} \to_{\mathbb{P}} 0, \qquad M_n^{\mathrm{abs}} - \sqrt{k_n} \to_{\mathbb{P}} 0, \qquad H_n \Rightarrow G, \qquad H_n^{\mathrm{abs}} \Rightarrow |G|, \tag{3.7}$$

so that

$$\boxed{d_{\mathrm{K}}(M_n, H_n) \longrightarrow 1, \qquad d_{\mathrm{K}}(M_n^{\mathrm{abs}}, H_n^{\mathrm{abs}}) \longrightarrow 1.} \tag{3.8}$$

All assertions about the maxima and approximation errors remain valid for Studentized means. At $x_n = \sqrt{k_n}/2$, both tail ratios in (1.3) diverge to infinity, although $x_n = o(n^{\tau/(4+2\tau)})$ for every fixed $\tau > 0$.

PROOF. For every fixed $r \geq 1$,

$$n^r(q_n + q_n^r) \leq 2\exp\{r\log n - (\log n)^2\} \longrightarrow 0.$$

Equation (2.3) proves uniform boundedness of all such moments. Also $v_n \to 0$, so (2.1) gives $X_{n11} - Y_1 \to 0$ in L^r for every fixed $r \geq 1$. Moment convergence and the assertions for $0 < r < 1$ follow as well. The coordinates have identical marginal distributions.

Writing $L_n = \log n$, we have

$$\log(1/\lambda_n) = L_n^2 - L_n, \qquad k_n = L_n + O(1), \qquad \log(k_n!) = O(L_n \log L_n).$$

The ceiling in (3.5) has negligible relative effect. Consequently,

$$\log p_n = (k_n + 1/2)(L_n^2 - L_n) + \log(k_n!) + o(1) \sim L_n^3.$$

The same observation yields

$$\theta_n \sim \lambda_n^{-1/2} \longrightarrow \infty, \qquad \frac{\theta_n \lambda_n}{k_n + 1} \sim \frac{\lambda_n^{1/2}}{k_n + 1} \longrightarrow 0. \tag{3.9}$$

Lemma 2.1 applies because $\lambda_n \to 0$ and $k_n^2/n \to 0$. Moreover,

$$n^{3/2} q_n \to 0, \qquad v_n \log p_n = O\Big(e^{-(\log n)^2 + 2\log n}(\log n)^3\Big) \to 0.$$

Lemma 2.2, $k_n \to \infty$, and tightness of W_n prove (3.7), including Studentization.

At $x_n = \sqrt{k_n}/2$, both self-normalized upper tails tend to one, whereas the two Gaussian upper tails tend to zero. This proves (3.8), since a Kolmogorov distance is at most one, and also proves divergence of the two tail ratios. The same argument applies to Studentized maxima. □

Theorem 3.2 satisfies the marginal moment condition in (1.3) for every $0 < \tau \leq 1$. Its divergent thresholds lie inside every associated moderate deviation range. Thus restricting attention to thresholds tending to infinity does not rescue the unrestricted assertion.

4. A single fixed population. The previous constructions specify a family of row distributions. The following result shows that changing the underlying population with the sample size is unnecessary. Only the number of retained coordinates changes.

THEOREM 4.1. *There exist iid random sequences $\mathcal{X}_1, \mathcal{X}_2, \ldots \in \mathbb{R}^{\mathbb{N}}$, with a law independent of n, and integers P_n satisfying*

$$\mathbb{E}\mathcal{X}_{ij} = 0, \quad \mathbb{E}\mathcal{X}_{ij}^2 = 1, \quad \sup_{j\geq 1}\mathbb{E}|\mathcal{X}_{ij}|^r < \infty \ \ (r>0), \qquad \log P_n = O((\log n)^3), \tag{4.1}$$

such that every finite covariance matrix is positive definite and the following holds. Using n observations of the first P_n coordinates, the Kolmogorov distance between the self-normalized maximum and the Gaussian maximum with the same population covariance tends to one. The same assertion holds for absolute maxima and for both Studentized versions.

PROOF. Index the coordinates in successive finite blocks $\ell = 3, 4, \ldots$. For block ℓ, take $q_\ell, \lambda_\ell, k_\ell, p_\ell, v_\ell$ from (3.5) and (2.1) with $n = \ell$. On a single probability space, take iid standard normals Y_i and mutually independent Bernoulli(q_ℓ) variables $B_{i,\ell j}$, independent of all Y_i. Define

$$\mathcal{X}_{i,\ell j} = \frac{Y_i + \ell(B_{i,\ell j} - q_\ell)}{\sqrt{1+v_\ell}}, \qquad \ell \geq 3, \quad 1 \leq j \leq p_\ell, \qquad P_n = \sum_{\ell=3}^{n} p_\ell. \tag{4.2}$$

This specifies one population on the countable product space. The moment bound in Theorem 3.2, applied block by block, proves the first three properties in (4.1). Also,

$$\log P_n \leq \log n + \max_{3\leq \ell\leq n} \log p_\ell = O((\log n)^3).$$

For any finite selection of coordinates, its covariance is the sum of a rank-one positive semidefinite matrix and a diagonal matrix with entries $v_\ell/(1+v_\ell) > 0$. It is therefore positive definite.

To control all Gaussian comparisons simultaneously, take independent standard normals G_0 and $G_{\ell j}$ and set

$$\mathcal{Z}_{\ell j} = \frac{G_0 + \sqrt{v_\ell}\, G_{\ell j}}{\sqrt{1+v_\ell}}. \tag{4.3}$$

Every finite subvector has exactly the corresponding population covariance in (4.2). The Gaussian union bound gives

$$\mathbb{P}\left(\sqrt{v_\ell}\max_{j\leq p_\ell}|G_{\ell j}| > 1\right) \leq 2p_\ell \exp\{-1/(2v_\ell)\}. \tag{4.4}$$

The right-hand side is summable in ℓ: $\log p_\ell = O((\log \ell)^3)$, whereas

$$\frac{1}{v_\ell} \geq \exp\{(\log \ell)^2 - 2\log \ell\}$$

eventually dominates every power of ℓ. Borel–Cantelli and finiteness of every block imply

$$\sup_{\ell\geq 3}\max_{j\leq p_\ell}|\mathcal{Z}_{\ell j}| \leq |G_0| + \sup_{\ell\geq 3}\sqrt{v_\ell}\max_{j\leq p_\ell}|G_{\ell j}| < \infty \qquad \text{almost surely.} \tag{4.5}$$

Thus all Gaussian maxima over the first P_n coordinates are uniformly tight.

For sample size n, the newest block $\ell = n$ has exactly the distribution used in Theorem 3.2. Its self-normalized and Studentized maxima equal $\sqrt{k_n} + o_{\mathbb{P}}(1)$. The maxima over all P_n coordinates are at least as large. At $x_n = \sqrt{k_n}/2 \to \infty$, all four data-based upper tails therefore tend to one. The Gaussian upper tails tend to zero by (4.5). Their absolute differences tend to one, proving the theorem. □

Every fixed finite-dimensional subvector of this population satisfies the ordinary multivariate central limit theorem, and coordinatewise Studentization is valid by the law of large numbers. The obstruction appears only when the maximum is taken over a growing number of coordinates.

5. Implications and scope. The examples distinguish marginal moment control from control of the coordinate envelope. For the array in Theorem 3.2, $p_n q_n \to \infty$, so a given row contains a rare observation with probability tending to one. On the additional event $|Y_i| \le n/4$, whose probability also tends to one, (2.1) gives

$$\max_{j \le p_n} |X_{nij}| \ge n/2$$

for all sufficiently large n. Conversely, that maximum is at most $|Y_i| + n$. Consequently, for each fixed $r \ge 1$,

$$\mathbb{E} \max_{j \le p_n} |X_{nij}|^r \asymp n^r, \qquad \max_{j \le p_n} \mathbb{E}|X_{nij}|^r \longrightarrow \mathbb{E}|G|^r. \tag{5.1}$$

This explains why bounds involving joint tails or envelope moments, such as those in Chang et al. (2025), do not yield a vanishing error for these examples.

In (2.2), every off-diagonal correlation is $(1 + v_n)^{-1} \to 1$. The constructions therefore fall outside the correlation assumptions of the proved dependent-coordinate results of Qiu, Chen and Shao (2025); in particular, they violate its condition (2.4), which bounds absolute pairwise correlations away from one. The counterexamples concern its unrestricted conjecture, with the Gaussian target always defined by the population covariance.

The fixed-population result shows that neither strengthening a marginal polynomial moment assumption to moments of all orders nor requiring a common underlying population resolves the problem. Valid extensions under coordinate dependence require assumptions that exclude the concentration of rare observations across the coordinate maximum.

REFERENCES

CHANG, W., TAKATSU, K., URBAN, K. AND KUCHIBHOTLA, A. K. (2025). The Berry–Esseen bound for high-dimensional self-normalized sums. *arXiv preprint* arXiv:2501.08979. doi:10.48550/arXiv.2501.08979.

JING, B.-Y., SHAO, Q.-M. AND WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31**(4), 2167–2215. doi:10.1214/aop/1068646382.

QIU, J., CHEN, S. X. AND SHAO, Q.-M. (2025). Self-normalized Cramér type moderate deviation theorem for Gaussian approximation. *Ann. Statist.* **53**(3), 1319–1346. doi:10.1214/25-AOS2507.

SHAO, Q.-M. (1999). A Cramér type large deviation result for Student's t-statistic. *J. Theoret. Probab.* **12**(2), 385–398. doi:10.1023/A:1021626127372.