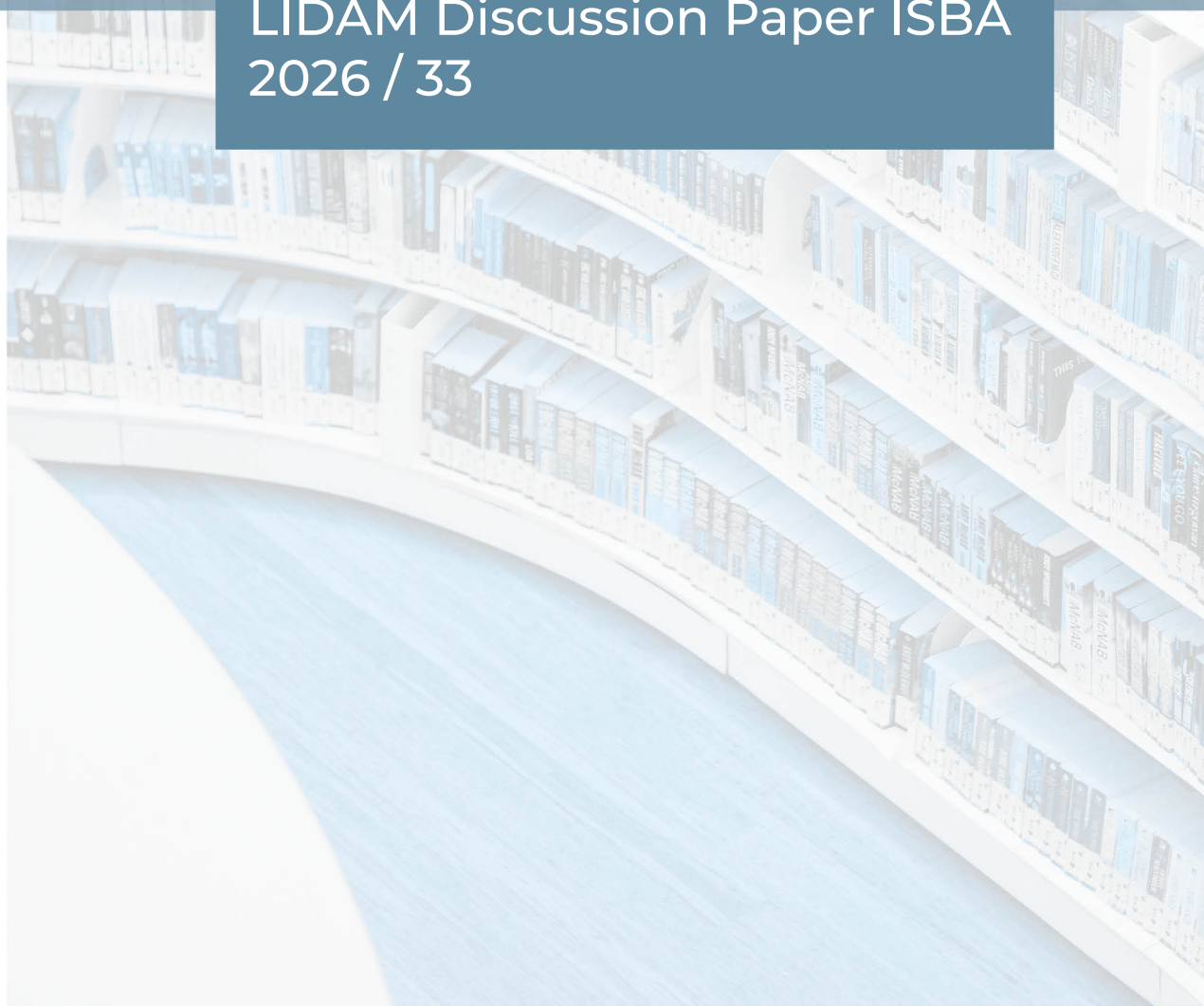
EXPLICIT GAMMA–STEIN BOUNDS FOR SATTERTHWAITE'S APPROXIMATION

Gabriel Bailly, Ferdinand Rapin, Yvik Swan, Rainer von Sachs

LIDAM Discussion Paper ISBA
2026 / 33



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

Explicit Gamma–Stein Bounds for Satterthwaite's Approximation

Gabriel Bailly[1], Ferdinand Rapin[2], Yvik Swan[2,3], and Rainer von Sachs[1]

[1]LIDAM, UCLouvain
[2]Département de mathématiques, Université libre de Bruxelles
[3]Vrije Universiteit Brussel

October 6, 2026

Abstract

We derive explicit Gamma–Stein bounds for the Satterthwaite approximation of weighted sums of independent chi-square random variables and, more generally, of independent Gamma random variables. The approximating Gamma distribution is chosen by matching the first two moments. Our approach relies on new kernel representations of the solution to the Gamma–Stein equation and of its first derivatives on the positive real line. These representations yield weighted Stein-factor bounds adapted to the moment-matching structure and show that the quality of the approximation is controlled directly by the discrepancies between the individual scale parameters and the matched scale parameter. Kolmogorov bounds are obtained by smoothing. The results provide quantitative guarantees for the widely used Satterthwaite approximation and are applied to pooled variances, sample variances of stationary Gaussian time series, Wishart traces and sums of Wishart traces, and high-dimensional repeated-measures statistics.

Keywords: Satterthwaite approximation; Gamma approximation; Stein's method; weighted chi-square distribution; Wishart distribution

Mathematics Subject Classification (2020): 60F05, 60E05, 62E17, 62H10

Introduction

Satterthwaite's approximation is a classical moment-matching procedure for approximating the distribution of a weighted sum of independent chi-square random variables by a scaled chi-square distribution. Its origins go back to Welch (1938), in the context of the two-sample t-test with unequal population variances. In that setting, Welch proposed to replace the exact distribution of the weighted sum appearing in the test statistic by a chi-square distribution whose degrees of freedom are chosen to reproduce the first two moments. This idea was subsequently generalized by Satterthwaite (1941, 1946) to arbitrary weighted sums of independent chi-square random variables. In modern terminology, the approximation consists in replacing such a weighted sum by a scaled chi-square random variable, with the scale and the degrees of freedom chosen so that the approximating distribution has the same mean and variance as the original statistic.

The appeal of this approximation is its simplicity. Exact formulas for the density of weighted chi-square sums are available, but they typically involve infinite series or integral representations and are therefore not always convenient for statistical practice; see, for instance, Moschopoulos (1985). Beyond the two-moment Satterthwaite approximation, several higher-order moment-matching procedures have

been proposed, including the Hall–Buckley–Eagleson, Wood F, and Lindsay–Pilla–Basak methods; see Bodenham and Adams (2015) for a comparative review. Nevertheless, the original Satterthwaite approximation remains widely used because it is explicit, interpretable, and easy to implement.

Satterthwaite-type approximations are especially useful when the exact null distribution is available in principle but too complicated to serve as a practical reference distribution. This occurs for many quadratic-form statistics, whose laws can be expressed as weighted sums of independent chi-square, or more generally Gamma, random variables. The examples developed below illustrate such situations: they include pooled variances, sample variances of stationary Gaussian time series, traces of Wishart random matrices, sums of independent Wishart traces, and high-dimensional repeated-measures designs. Together, these examples cover classical variance estimation, dependence, multivariate covariance analysis, and modern high-dimensional inference, illustrating why such approximations remain important in statistical practice.

Despite this broad use, quantitative theoretical guarantees for Satterthwaite's approximation remain relatively limited. Classical approaches often work directly with distribution functions or through asymptotic expansions. The question addressed here is slightly different: how close is the law of a weighted chi-square, or more generally a weighted Gamma sum, to its moment-matched Gamma approximation when both variables are kept on their original scale? This is the formulation that appears naturally in statistical applications, where the Satterthwaite approximation is used directly as a reference distribution.

Stein's method provides a natural framework for this problem. Gamma approximation by Stein's method goes back at least to Luk (1994), and was subsequently developed in several directions, including chi-square approximation via Stein's method (Pickett, 2004), Stein-factor estimates for Gamma approximation (Gaunt et al., 2017; Döbler and Peccati, 2018), and general density-based constructions of Stein operators (Ley et al., 2017a,b). The work of Döbler and Peccati (2018) also belongs to the Malliavin–Stein line of research on Gamma approximation, where Gamma or centered-Gamma Stein equations are used to prove non-central limit theorems for nonlinear functionals. These works provide powerful general tools, while the present paper focuses specifically on the moment-matching structure of Satterthwaite's approximation for positive weighted Gamma sums.

A related approach, closer in spirit to our setting, was proposed by Liu and Xia (2022), who use a characterization of the Gamma distribution through size biasing and zero biasing. Their results yield bounds for Gamma approximation of sums of independent nonnegative random variables, and in particular their Corollary 3.4 applies to convolutions of independent Gamma random variables. Our aim is complementary: we seek bounds stated directly in terms of the scale discrepancies between the summands and the moment-matched Gamma approximation.

We now introduce the notation used throughout the paper. We write $X \cong Y$ when X and Y have the same distribution, that is, $\mathscr{L}(X) = \mathscr{L}(Y)$, and we write $X \perp\!\!\!\perp Y$ for independence. The notation

$$(X_i)_{i=1}^n \overset{\text{iid}}{\sim} P$$

means that $X_i \sim P$ for every $i \in \{1, \dots, n\}$ and that $X_1, \dots, X_n$ are mutually independent.

For $a, b > 0$, we denote by $\mathcal{G}_{a,b}$ the Gamma distribution with shape a and scale b, namely the distribution with density

$$p_{a,b}(x) = \frac{1}{\Gamma(a) b^a} x^{a-1} e^{-x/b} \mathbf{1}_{\{x>0\}}.$$

We write $\mathcal{Z}_1$ for the class of continuously differentiable functions $h : (0, \infty) \to \mathbb{R}$ such that $\|h'\|_\infty \le 1$, where $\|\cdot\|_\infty$ denotes the usual supremum norm. For positive random variables X and Y with finite first moments, define

$$d_1(X, Y) := \sup_{h \in \mathcal{Z}_1} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|.$$

This is the usual Wasserstein distance of order one, written here in the smooth test-function formulation

that is convenient for Stein's method.

We also write $\mathcal{Z}_2$ for the class of twice continuously differentiable functions $h : (0,\infty) \to \mathbb{R}$ such that $\|h''\|_\infty \le 1$. For positive random variables X and Y with finite second moments, set

$$d_2(X,Y) := \sup_{h\in\mathcal{Z}_2} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|\,.$$

This is the Zolotarev ideal metric of order two (see Zolotarev (1979)), again expressed in the convenient smooth test-function formulation. It is finite in the natural setting of this paper, where the two variables have matching first two moments.

We shall also use the Kolmogorov distance

$$d_{\mathrm{Kol}}(X,Y) := \sup_{z\in\mathbb{R}} |\mathbb{P}[X \le z] - \mathbb{P}[Y \le z]|\,.$$

Using a smoothing procedure (e.g. Chen et al. (2011, p.48), see also Gaunt and Li (2023)) we shall use the following standard inequality to pass from Wasserstein to Kolmogorov bounds. If Y has a density p_Y satisfying

$$C_Y^1 := \sup_{x\in\mathbb{R}} p_Y(x) < \infty,$$

then

$$d_{\mathrm{Kol}}(X,Y) \le \sqrt{2C_Y^1\, d_1(X,Y)}. \tag{1}$$

These three distances play complementary roles. The Kolmogorov distance is closest to classical testing, since it controls distribution functions, critical values, and p-values. The Wasserstein distance d_1 controls expectations of Lipschitz transformations and provides, through smoothing, a route to Kolmogorov bounds. The distance d_2 is especially natural for Satterthwaite's approximation: since the approximating Gamma law is chosen by matching the first two moments, affine components are already accounted for, and the remaining discrepancy is governed by second-order smoothness.

The contributions of the paper are twofold. First, we develop a harmonised and fully explicit version of Stein's method for the Gamma distribution on $(0,\infty)$. The approach is based on kernel representations of the Stein solution and of its first two derivatives. It is self-contained, keeps the weights generated by the Gamma operator visible, and gives constants that can be tracked explicitly. This yields weighted Stein factors adapted to the distances d_1 and d_2.

Several uniform bounds for solutions of the Gamma–Stein equation and their derivatives are already available in the literature, starting with the generator approach developed in Luk (1994); Pickett (2004); Gaunt (2013) and given in detail in Gaunt et al. (2017); Döbler and Peccati (2018). The perspective taken here is different. Rather than deriving only global derivative bounds, we give pointwise kernel representations for the solution and for the derivatives entering the Stein identity. These representations make the dependence on the distribution parameters explicit and transparent, which is crucial for the asymptotic analysis carried out later in the paper. In particular, they allow us to obtain sharp large-shape asymptotics together with simple explicit bounds valid in the corresponding asymptotic regimes.

Second, we apply this framework to Satterthwaite's approximation for sums of independent Gamma random variables. Let

$$\mathsf{G}_N = \sum_{i=1}^N G_i, \qquad G_i \sim \mathcal{G}_{\alpha_i,\beta_i}, \qquad i = 1,\dots,N,$$

where $G_1,\dots,G_N$ are independent. Let $U_N \sim \mathcal{G}_{\alpha,\beta}$ be chosen by matching the mean and variance of G_N, namely

$$\alpha\beta = \sum_{i=1}^N \alpha_i\beta_i, \qquad \alpha\beta^2 = \sum_{i=1}^N \alpha_i\beta_i^2.$$

Writing

$$A_N := \max_{1\le i\le N}\left\{\left|\frac{\beta_i}{\beta}-1\right|\beta_i\right\}, \qquad B_N := \max_{1\le i\le N}\left\{\left|\frac{\beta_i}{\beta}-1\right|\beta_i^2\right\},$$

we prove the non-asymptotic bounds

$$d_1(\mathsf{G}_N, U_N) \le 2A_N$$

and

$$d_2(\mathsf{G}_N, U_N) \le 2B_N + \alpha\beta e_\alpha A_N.$$

Thus the quality of the approximation is controlled directly by the discrepancies between the individual scale parameters β_i and the matched scale parameter β. Combining the d_1-estimate with (1) also gives a Kolmogorov bound whenever the matched shape parameter satisfies $\alpha > 1$. To the best of our knowledge, these are the first Gamma–Stein bounds stated directly for the Satterthwaite moment-matched approximation in terms of the parameters of the underlying weighted Gamma sum.

Overview of the paper. Section 1 develops the Gamma–Stein framework used throughout the paper. We derive the Stein equation, its kernel solution, and the weighted Stein-factor estimates needed for the d_1- and d_2-bounds. Section 2 applies these estimates to sums of independent Gamma random variables and proves the main non-asymptotic bounds. Section 3 specializes the abstract result to Satterthwaite approximations in statistical examples, including pooled variances, Gaussian time series, Wishart traces, sums of independent Wishart traces, and high-dimensional repeated-measures designs. Technical bounds for the Gamma kernels and the explicit constants are collected in Appendix A.

1 Kernel-based Gamma–Stein's method on the positive real line

In this section we develop the Gamma–Stein machinery used in the proof of the main approximation result. We start from the classical Gamma integration-by-parts identity, introduce the corresponding Stein equation, and then derive kernel representations for the solution and its first two derivatives.

Throughout, $a > 0$ and $b > 0$ are fixed. We write $L^1(\mathcal{G}_{a,b})$ for the Borel functions $h : (0,\infty) \to \mathbb{R}$ such that $\mathbb{E}[|h(X)|] < \infty$ for $X \sim \mathcal{G}_{a,b}$, and set

$$\mathcal{G}_{a,b}h := \int_0^\infty h(u)p_{a,b}(u)\,du.$$

For $x > 0$, write

$$P_{a,b}(x) := \int_0^x p_{a,b}(u)\,du, \qquad \overline{P}_{a,b}(x) := \int_x^\infty p_{a,b}(u)\,du.$$

The basic Gamma–Stein machinery follows from the density approach; see, for instance, Döbler (2015); Ley et al. (2017a,b). We first record the integration-by-parts identity and the corresponding characterization of the Gamma distribution.

Theorem 1.1. *Let $X \sim \mathcal{G}_{a,b}$. If $f : (0,\infty) \to \mathbb{R}$ is absolutely continuous and $x \mapsto xf'(x)$ belongs to $L^1(\mathcal{G}_{a,b})$, then $x \mapsto (x-ab)f(x)$ belongs to $L^1(\mathcal{G}_{a,b})$ and*

$$\mathbb{E}[(X-ab)f(X)] = b\,\mathbb{E}[Xf'(X)]. \tag{2}$$

Conversely, if a real-valued random variable X satisfies (2) for some $a, b > 0$ and every smooth compactly supported function $f : \mathbb{R} \to \mathbb{R}$, then $X \sim \mathcal{G}_{a,b}$.

The Stein equation associated with (2) is obtained by inverting the operator $f \mapsto bxf'(x) - (x-ab)f(x)$.

Definition 1.2. Let $h \in L^1(\mathcal{G}_{a,b})$. The Gamma–Stein equation associated with h is

$$bxf'(x) - (x - ab)f(x) = h(x) - \mathcal{G}_{a,b}h, \qquad x > 0. \tag{3}$$

A solution of (3) is a locally absolutely continuous function $f_{h;a,b} : (0,\infty) \to \mathbb{R}$ for which there exists a version of its a.e. derivative $f'_{h;a,b}$ such that (3) holds for every $x > 0$.

We then collect the kernels and kernel bounds needed in the sequel. For $y > 0$, define the weighted lower and upper Gamma–Mills ratios by

$$\underline{M}_{a,b}(y) := \frac{P_{a,b}(y)}{byp_{a,b}(y)}, \qquad \overline{M}_{a,b}(y) := \frac{\overline{P}_{a,b}(y)}{byp_{a,b}(y)}.$$

We begin with an elementary property which seems to be new; it follows from a simple integral representation given in Appendix A. Related convexity and reciprocal convexity properties of Gamma–Mills ratios can be found in Baricz (2012).

Lemma 1.3. *The function $y \mapsto \underline{M}_{a,b}(y)$ is absolutely monotone on $(0,\infty)$, while $y \mapsto \overline{M}_{a,b}(y)$ is completely monotone on $(0,\infty)$. Equivalently, for every $n \in \mathbb{N}$ and $y > 0$,*

$$\underline{M}^{(n)}_{a,b}(y) \geq 0, \qquad (-1)^n \overline{M}^{(n)}_{a,b}(y) \geq 0.$$

In particular, $\underline{M}_{a,b}$ is increasing and convex with increasing second derivative, whereas $\overline{M}_{a,b}$ is decreasing and convex with decreasing second derivative.

In particular the lower and upper Gamma–Mills ratios are therefore three times continuously differentiable on $(0,\infty)$. For $x, y > 0$, define

$$K^{0,1}_{a,b}(x,y) = -\overline{M}_{a,b}(y)\frac{P_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x\leq y\}} - \underline{M}_{a,b}(y)\frac{\overline{P}_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x>y\}}, \tag{4}$$

$$K^{1,1}_{a,b}(x,y) = -\overline{M}'_{a,b}(y)\frac{P_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x\leq y\}} - \underline{M}'_{a,b}(y)\frac{\overline{P}_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x>y\}}, \tag{5}$$

$$K^{2,1}_{a,b}(x,y) = -\overline{M}''_{a,b}(y)\frac{P_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x\leq y\}} - \underline{M}''_{a,b}(y)\frac{\overline{P}_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x>y\}}. \tag{6}$$

We also use the second-order kernels

$$K^{2,2}_{a,b}(x,y) = \overline{M}''_{a,b}(y)\frac{\underline{\mathrm{P}}^1_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x\leq y\}} - \underline{M}''_{a,b}(y)\frac{\overline{\mathrm{P}}^1_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x>y\}}, \tag{7}$$

$$K^{3,2}_{a,b}(x,y) = \overline{M}'''_{a,b}(y)\frac{\underline{\mathrm{P}}^1_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x\leq y\}} - \underline{M}'''_{a,b}(y)\frac{\overline{\mathrm{P}}^1_{a,b}(x)}{p_{a,b}(x)}\mathbf{1}_{\{x>y\}}, \tag{8}$$

where

$$\underline{\mathrm{P}}^1_{a,b}(x) := \int_0^x P_{a,b}(u)\,du, \qquad \overline{\mathrm{P}}^1_{a,b}(x) := \int_x^\infty \overline{P}_{a,b}(u)\,du.$$

Finally, for $y > 0$, set

$$C^1_{a,b}(y) := -\overline{M}'_{a,b}(y)\underline{\mathrm{P}}^1_{a,b}(y) + \underline{M}'_{a,b}(y)\overline{\mathrm{P}}^1_{a,b}(y), \tag{9}$$

$$C^2_{a,b}(y) := \overline{M}''_{a,b}(y)\int_0^y \underline{\mathrm{P}}^1_{a,b}(x)\,dx + \underline{M}''_{a,b}(y)\int_y^\infty \overline{\mathrm{P}}^1_{a,b}(x)\,dx. \tag{10}$$

By scaling,

$$C^1_{a,b}(y) = \frac{1}{b}C^1_{a,1}\left(\frac{y}{b}\right), \qquad C^2_{a,b}(y) = \frac{1}{b}C^2_{a,1}\left(\frac{y}{b}\right). \tag{11}$$

The basic kernel estimates are as follows.

Lemma 1.4. *Let $X \sim \mathcal{G}_{a,b}$. Then, for every $y > 0$,*

$$\mathbb{E}[|K_{a,b}^{0,1}(X,y)|] = 1, \qquad \mathbb{E}[|K_{a,b}^{2,1}(X,y)|] = \mathbb{E}[|K_{a,b}^{3,2}(X,y)|] = \frac{1}{by},$$

and

$$\mathbb{E}[|K_{a,b}^{1,1}(X,y)|] = C_{a,b}^{1}(y), \qquad \mathbb{E}[|K_{a,b}^{2,2}(X,y)|] = C_{a,b}^{2}(y).$$

We now connect the kernels with solutions of the Gamma–Stein equation (3). Let $X \sim \mathcal{G}_{a,b}$. For $x > 0$, define

$$f_{h;a,b}(x) = \mathbb{E}[h'(X)K_{a,b}^{0,1}(X,x)], \tag{12}$$

$$g_{h;a,b}(x) = \mathbb{E}[h'(X)K_{a,b}^{1,1}(X,x)], \tag{13}$$

$$\ell_{h;a,b}(x) = \mathbb{E}[h'(X)K_{a,b}^{2,1}(X,x)] + \frac{h'(x)}{bx}, \tag{14}$$

$$\widetilde{\ell}_{h;a,b}(x) = \mathbb{E}[h''(X)K_{a,b}^{2,2}(X,x)], \tag{15}$$

$$m_{h;a,b}(x) = \mathbb{E}[h''(X)K_{a,b}^{3,2}(X,x)] + \frac{h''(x)}{bx}. \tag{16}$$

These representations require the corresponding regularity of h: (12) and (13) require h to be a.e. differentiable; (14) requires differentiability on $(0,\infty)$; (15) requires h to be twice a.e. differentiable; and (16) requires twice differentiability on $(0,\infty)$.

Lemma 1.5. *Fix $a > 0$ and $b > 0$. If $h \in \mathcal{Z}_1$, the function $f_{h;a,b}$ defined in (12) solves (3). Moreover, $f_{h;a,b}$ and $g_{h;a,b}$ are differentiable on $(0,\infty)$, with $f'_{h;a,b} = g_{h;a,b}$ and $g'_{h;a,b} = \ell_{h;a,b}$. The functions $f_{h;a,b}$, $g_{h;a,b}$, $xg_{h;a,b}$, and $x\ell_{h;a,b}$ are bounded on $(0,\infty)$, hence integrable with respect to $\mathcal{G}_{u,v}$ for every $u, v > 0$.*

If $h \in \mathcal{Z}_2$, then, in addition, $\ell_{h;a,b} = \widetilde{\ell}_{h;a,b}$ on $(0,\infty)$. Moreover, $\ell_{h;a,b}$ is differentiable on $(0,\infty)$, with $\ell'_{h;a,b} = m_{h;a,b}$. The functions $\ell_{h;a,b}$ and $xm_{h;a,b}$ are bounded on $(0,\infty)$, hence integrable with respect to $\mathcal{G}_{u,v}$ for every $u, v > 0$.

Lemma 1.5 gives pointwise representatives of the Stein solution and of its derivatives. From now on, for convenience and under the regularity assumptions on h required by Lemma 1.5, we use the notation

$$f_h := f_{h;a,b}, \qquad f'_h := g_{h;a,b}, \qquad f''_h := \ell_{h;a,b}, \qquad f'''_h := m_{h;a,b},$$

where the derivatives are understood in the pointwise representative sense given by Lemma 1.5. In particular, these representatives satisfy the weighted boundedness properties needed to apply the Gamma integration-by-parts identity. The following bounds then follow directly from Lemma 1.4 after taking absolute values.

Lemma 1.6 (Non-uniform Stein factors)**.** *Let $a > 0$ and $b > 0$, and put $z = x/b$. If $h \in \mathcal{Z}_1$, then, for every $x > 0$,*

$$|f_h(x)| \le 1, \qquad |f'_h(x)| \le \frac{1}{b}C_{a,1}^{1}(z), \qquad |xf'_h(x)| \le zC_{a,1}^{1}(z), \qquad |xf''_h(x)| \le \frac{2}{b}.$$

If $h \in \mathcal{Z}_2$, then, for every $x > 0$,

$$|f''_h(x)| \le \frac{1}{b}C_{a,1}^{2}(z), \qquad |xf'''_h(x)| \le \frac{2}{b}.$$

Remark 1.7 (Sharpness of the non-uniform Stein factors). The non-uniform bounds in Lemma 1.6 cannot be improved as bounds uniform over $h \in \mathcal{Z}_1$ or $h \in \mathcal{Z}_2$. Indeed, each estimate is obtained from a kernel representation by taking the absolute value inside the expectation. This loss is unavoidable.

More precisely, the sign properties of the Gamma–Mills' ratios from Lemma 1.3 determine the signs of the relevant kernels: $K_{a,b}^{0,1}$, $K_{a,b}^{2,1}$, and $K_{a,b}^{3,2}$ have a fixed sign, whereas $K_{a,b}^{1,1}$ and $K_{a,b}^{2,2}$ change sign only

(a) $a = 5$ (b) $a = 50$ (c) $a = 500$

Figure 1: *Normalized non-uniform Stein-factor envelopes for increasing values of the Gamma shape parameter. From left to right, the plots correspond to $a = 5$, $a = 50$, and $a = 500$. In each panel, the solid blue curve is $C^1_{a,1}(z)/(\sqrt{2/(\pi a)})$, the dashed orange curve is $zC^1_{a,1}(z)/(\sqrt{2a/\pi})$, and the dotted green curve is $C^2_{a,1}(z)/\sqrt{\pi/(8a)}$. The grey dot-dashed horizontal line marks the reference level 1. The stabilization of the curves illustrates the large-shape asymptotics in (17).*

at the evaluation point. Therefore the usual L^∞–L^1 duality identity

$$\sup_{\|\varphi\|_\infty \le 1} |\mathbb{E}[\varphi(X)K(X,x)]| = \mathbb{E}[|K(X,x)|]$$

shows that the envelopes appearing in Lemma 1.6 are the best possible ones. The supremum is attained over bounded measurable functions by taking φ to be the sign of the relevant kernel; within the original smooth test classes, the same conclusion follows by smooth approximation.

The pointwise optimal bounds $C^1_{a,1}(z)$, $zC^1_{a,1}(z)$, and $C^2_{a,1}(z)$ are given explicitly in terms of the Gamma density and distribution function. Their analytic properties can therefore be studied directly. In Proposition A.1, we obtain their behaviour at zero and at infinity, and show in particular that they are bounded on $(0,\infty)$. Thus

$$c_a := \sup_{z>0} C^1_{a,1}(z), \qquad d_a := \sup_{z>0} zC^1_{a,1}(z), \qquad e_a := \sup_{z>0} C^2_{a,1}(z)$$

are finite. Taking the supremum over the evaluation point z in Lemma 1.6 yields the following bounds, now uniform in z, for the corresponding Stein factors.

Lemma 1.8 (Stein factors)**.** *Let $a > 0$ and $b > 0$. If $h \in \mathcal{Z}_1$, then*

$$\|f_h\|_\infty \le 1, \qquad \|f_h'\|_\infty \le \frac{c_a}{b}, \qquad \|xf_h'\|_\infty \le d_a, \qquad \|xf_h''\|_\infty \le \frac{2}{b}.$$

If $h \in \mathcal{Z}_2$, then

$$\|f_h''\|_\infty \le \frac{e_a}{b}, \qquad \|xf_h'''\|_\infty \le \frac{2}{b}.$$

Finally, Proposition A.1 gives, as $a \to \infty$,

$$\sqrt{a}\, c_a \to \sqrt{\frac{2}{\pi}}, \qquad \frac{d_a}{\sqrt{a}} \to \sqrt{\frac{2}{\pi}}, \qquad \sqrt{a}\, e_a \to \sqrt{\frac{\pi}{8}}. \tag{17}$$

These asymptotics, illustrated in Figure 1, make explicit the dependence of the Stein factors on the shape parameter in the large-a regime, and will be used below to compare our sharp estimates with existing bounds in the literature.

Remark 1.9. For completeness, we also mention that one can derive simple elementary upper bounds for the constants appearing above. For instance, as shown in Appendix A, for every $a \ge 16$,

$$e_a \le \frac{5}{2\sqrt{a}}.$$

The constant $5/2$ is not sharp and is chosen only for convenience. This bound is not needed in its explicit form below, where we keep the notation e_a, but it confirms that e_a has the correct order $a^{-1/2}$, in agreement with (17).

Remark 1.10. The closest comparable bounds we are aware of are those of Gaunt et al. (2017); Gaunt

(2013). Translated into our notation, they imply that, for every $m \geq 1$,

$$\|f^{(m)}_{h;a,b}\|_\infty \leq \frac{1}{b}\left\{\sqrt{\frac{2\pi}{a+m}} + \frac{\log(a+m+1)}{e(a+m)} + \frac{2}{a+m} + \frac{1}{e(a+m+1)}\right\}\|h^{(m)}\|_\infty.$$

For $m = 1$ and $m = 2$, these bounds have the same order as the estimates obtained above, namely $a^{-1/2}/b$, but with different leading constants: our leading constants are smaller by asymptotic factors π for the first derivative and 4 for the second derivative. Since we do not pursue higher-order derivatives here, we make no formal comparison for $m \geq 3$.

A similar comparison can be made for weighted bounds. Lemma 2.1 in Gaunt et al. (2017) translates, in our notation, into

$$\|x f^{(m+1)}_{h;a,b}\|_\infty \leq \frac{4}{b}\|h^{(m)}\|_\infty, \qquad m \geq 1.$$

For the derivatives considered here, our corresponding weighted estimates have constant $2/b$, and hence improve this bound by a factor 2. This constant is the sharp one for the kernel representations used above, in the sense explained in Remark 1.7.

Finally, the comparison with Döbler and Peccati (2018) is less direct, since there the Gamma Stein equation is studied on the whole real line rather than only on the support $(0, \infty)$ of the target distribution. The kernel representations used here are intrinsically tied to the positive half-line, so it is not immediate whether they admit a natural analogue in that setting.

We can now wrap up this section with the final ingredient of the Gamma–Stein method: the transfer principle, which rewrites the d_j-discrepancy from a Gamma target as an expected Stein operator.

Corollary 1.11 (Stein transfer principle)**.** *Let Y be a positive random variable. Fix $a > 0$, $b > 0$, and let $G_{a,b} \sim \mathcal{G}_{a,b}$. Assume that Y has a finite first moment if $j = 1$, and a finite second moment if $j = 2$. Then*

$$d_j(Y, G_{a,b}) = \sup_{h \in \mathcal{Z}_j} \left|\mathbb{E}\left[bY f'_{h;a,b}(Y)\right] - \mathbb{E}\left[(Y - ab) f_{h;a,b}(Y)\right]\right|, \tag{18}$$

where $f_{h;a,b}$ is the Stein solution defined in (12), and $f'_{h;a,b}$ denotes the pointwise representative $g_{h;a,b}$ from (13).

The regularity and boundedness guaranteed by Lemma 1.5, together with the Stein factors from Lemma 1.8, make (18) tractable in Gamma comparison problems.

2 Gamma approximation for sums of independent Gamma variates

In this section we turn the Gamma–Stein machinery of Section 1 into quantitative bounds for sums of independent Gamma random variables. Let $N \geq 1$, let $G_i \sim \mathcal{G}_{\alpha_i,\beta_i}$, $i = 1, \ldots, N$, be independent, and set

$$\mathsf{G}_N := \sum_{i=1}^N G_i. \tag{19}$$

Let $U_N \sim \mathcal{G}_{\alpha,\beta}$, where α and β are chosen by matching the mean and variance of G_N:

$$\alpha\beta = \sum_{i=1}^N \alpha_i\beta_i, \qquad \alpha\beta^2 = \sum_{i=1}^N \alpha_i\beta_i^2. \tag{20}$$

Equivalently,

$$\beta = \frac{\sum_{i=1}^N \alpha_i\beta_i^2}{\sum_{i=1}^N \alpha_i\beta_i}, \qquad \alpha = \frac{\left(\sum_{i=1}^N \alpha_i\beta_i\right)^2}{\sum_{i=1}^N \alpha_i\beta_i^2}.$$

With this choice, G_N and U_N have the same mean and variance. The approximation error is controlled by the discrepancies between the individual scales β_i and the matched scale β.

Theorem 2.1. *Assume that $G_i \sim \mathcal{G}_{\alpha_i,\beta_i}$, $i = 1, \dots, N$, are independent, and let $U_N \sim \mathcal{G}_{\alpha,\beta}$, where α, β are chosen by (20). Set*

$$A_N := \max_{1\le i\le N} \left\{ \left| \frac{\beta_i}{\beta} - 1 \right| \beta_i \right\}, \qquad B_N := \max_{1\le i\le N} \left\{ \left| \frac{\beta_i}{\beta} - 1 \right| \beta_i^2 \right\}.$$

Then

$$d_1(\mathsf{G}_N, U_N) \le 2A_N, \qquad d_2(\mathsf{G}_N, U_N) \le 2B_N + \alpha\beta e_\alpha A_N.$$

Moreover, if $\alpha > 1$, then

$$d_{\mathrm{Kol}}(\mathsf{G}_N, U_N) \le 2 \left(\frac{A_N}{\beta\sqrt{2\pi(\alpha-1)}\, r_\alpha} \right)^{1/2}, \qquad r_\alpha = \exp\left(\frac{1}{12\alpha - 11} \right). \tag{21}$$

Proof. Let $h \in \mathcal{Z}_1$. Write

$$f_h = f_{h;\alpha,\beta}, \qquad g_h = g_{h;\alpha,\beta}, \qquad \ell_h = \ell_{h;\alpha,\beta}.$$

By Lemma 1.5, $f_h' = g_h$ and $g_h' = \ell_h$ on $(0,\infty)$. The Stein transfer principle gives

$$\mathbb{E}[h(\mathsf{G}_N)] - \mathbb{E}[h(U_N)] = \mathbb{E}[\beta \mathsf{G}_N g_h(\mathsf{G}_N)] - \mathbb{E}[(\mathsf{G}_N - \alpha\beta) f_h(\mathsf{G}_N)]. \tag{22}$$

For each i, set $\mathsf{G}_N^{(i)} := \mathsf{G}_N - G_i$. Since $\sum_i \alpha_i\beta_i = \alpha\beta$,

$$\mathbb{E}[(\mathsf{G}_N - \alpha\beta) f_h(\mathsf{G}_N)] = \sum_{i=1}^N \mathbb{E}[(G_i - \alpha_i\beta_i) f_h(G_i + \mathsf{G}_N^{(i)})].$$

For fixed i, put $\psi_i(x) := \mathbb{E}[f_h(x + \mathsf{G}_N^{(i)})]$. Since G_i and $\mathsf{G}_N^{(i)}$ are independent, and since $\psi_i'(x) = \mathbb{E}[g_h(x + \mathsf{G}_N^{(i)})]$, the Gamma integration-by-parts identity gives

$$\mathbb{E}[(\mathsf{G}_N - \alpha\beta) f_h(\mathsf{G}_N)] = \sum_{i=1}^N \beta_i \mathbb{E}[G_i g_h(\mathsf{G}_N)].$$

The required integrability follows from Lemma 1.5, since

$$\mathbb{E}[G_i |\psi_i'(G_i)|] \le \mathbb{E}\left[\frac{G_i}{\mathsf{G}_N} \mathsf{G}_N |g_h(\mathsf{G}_N)| \right] < \infty.$$

Substituting this identity into (22) yields

$$\mathbb{E}[h(\mathsf{G}_N)] - \mathbb{E}[h(U_N)] = -\beta \sum_{i=1}^N \left(\frac{\beta_i}{\beta} - 1 \right) \mathbb{E}[G_i g_h(\mathsf{G}_N)].$$

By variance matching,

$$\sum_{i=1}^N \left(\frac{\beta_i}{\beta} - 1 \right) \alpha_i\beta_i = \frac{1}{\beta} \sum_{i=1}^N \alpha_i\beta_i^2 - \sum_{i=1}^N \alpha_i\beta_i = 0.$$

Hence

$$\mathbb{E}[h(\mathsf{G}_N)] - \mathbb{E}[h(U_N)] = -\beta \sum_{i=1}^N \left(\frac{\beta_i}{\beta} - 1 \right) \mathbb{E}[(G_i - \alpha_i\beta_i) g_h(\mathsf{G}_N)].$$

Applying integration by parts once more, now to $x \mapsto \mathbb{E}[g_h(x + \mathsf{G}_N^{(i)})]$, gives

$$\mathbb{E}[h(\mathsf{G}_N)] - \mathbb{E}[h(U_N)] = -\beta \sum_{i=1}^{N} \left(\frac{\beta_i}{\beta} - 1 \right) \beta_i \mathbb{E}[G_i \ell_h(\mathsf{G}_N)]. \tag{23}$$

Using $\|x\ell_h\|_\infty \le 2/\beta$, we obtain

$$|\mathbb{E}[h(\mathsf{G}_N)] - \mathbb{E}[h(U_N)]| \le 2A_N \sum_{i=1}^{N} \mathbb{E}\left[\frac{G_i}{\mathsf{G}_N} \right] = 2A_N.$$

Taking the supremum over $h \in \mathcal{Z}_1$ proves the d_1-bound.

We now let $h \in \mathcal{Z}_2$. Starting from (23), write

$$\begin{aligned} \mathbb{E}[h(\mathsf{G}_N)] - \mathbb{E}[h(U_N)] = &-\beta \sum_{i=1}^{N} \left(\frac{\beta_i}{\beta} - 1 \right) \beta_i \mathbb{E}[(G_i - \alpha_i\beta_i)\ell_h(\mathsf{G}_N)] \\ &- \beta \sum_{i=1}^{N} \left(\frac{\beta_i}{\beta} - 1 \right) \alpha_i \beta_i^2 \mathbb{E}[\ell_h(\mathsf{G}_N)]. \end{aligned}$$

Applying integration by parts to $x \mapsto \mathbb{E}[\ell_h(x + \mathsf{G}_N^{(i)})]$, and using $\ell_h' = m_h$, gives

$$\mathbb{E}[h(\mathsf{G}_N)] - \mathbb{E}[h(U_N)] = -\beta \sum_{i=1}^{N} \left(\frac{\beta_i}{\beta} - 1 \right) \beta_i^2 \mathbb{E}[G_i m_h(\mathsf{G}_N)] - \beta \sum_{i=1}^{N} \left(\frac{\beta_i}{\beta} - 1 \right) \alpha_i \beta_i^2 \mathbb{E}[\ell_h(\mathsf{G}_N)].$$

The Stein-factor bounds $\|xm_h\|_\infty \le 2/\beta$ and $\|\ell_h\|_\infty \le e_\alpha/\beta$ imply

$$|\mathbb{E}[h(\mathsf{G}_N)] - \mathbb{E}[h(U_N)]| \le 2 \sum_{i=1}^{N} \left| \frac{\beta_i}{\beta} - 1 \right| \beta_i^2 \mathbb{E}\left[\frac{G_i}{\mathsf{G}_N} \right] + e_\alpha \sum_{i=1}^{N} \left| \frac{\beta_i}{\beta} - 1 \right| \alpha_i \beta_i^2.$$

The first term is at most $2B_N$, because $\sum_i \mathbb{E}[G_i/\mathsf{G}_N] = 1$. For the second term,

$$\sum_{i=1}^{N} \left| \frac{\beta_i}{\beta} - 1 \right| \alpha_i \beta_i^2 \le A_N \sum_{i=1}^{N} \alpha_i \beta_i = A_N \alpha \beta.$$

Thus

$$|\mathbb{E}[h(\mathsf{G}_N)] - \mathbb{E}[h(U_N)]| \le 2B_N + \alpha\beta e_\alpha A_N.$$

Taking the supremum over $h \in \mathcal{Z}_2$ proves the d_2-bound.

It remains to prove the Kolmogorov bound. By the smoothing inequality (1),

$$d_{\mathrm{Kol}}(\mathsf{G}_N, U_N) \le \sqrt{2C_{\alpha,\beta}\, d_1(\mathsf{G}_N, U_N)}, \qquad C_{\alpha,\beta} := \sup_{x>0} p_{\alpha,\beta}(x).$$

If $\alpha > 1$, the density of $\mathcal{G}_{\alpha,\beta}$ is maximized at $x = \beta(\alpha - 1)$. Robbins' refinement of Stirling's formula (Robbins, 1955) gives

$$C_{\alpha,\beta} \le \frac{1}{\beta\sqrt{2\pi(\alpha-1)}\, r_\alpha}, \qquad r_\alpha = \exp\left(\frac{1}{12\alpha - 11} \right).$$

Combining this estimate with the d_1-bound gives (21). □

Theorem 2.1 reduces the approximation problem to A_N and B_N. Under (20), the matched scale β is

a weighted mean of the individual scales:

$$\beta = \sum_{i=1}^{N} w_i \beta_i, \qquad w_i := \frac{\alpha_i \beta_i}{\sum_{j=1}^{N} \alpha_j \beta_j}. \tag{24}$$

Here $w_i > 0$ and $\sum_i w_i = 1$. Thus, whenever the individual scales are all of the same order, the matched scale is of that order as well.

Lemma 2.2. *Let $N \geq 1$, and let $\{\alpha_i\}_{i=1}^N$ and $\{\beta_i\}_{i=1}^N$ be positive real numbers. Let $\alpha, \beta > 0$ be defined by (20). Assume that there exists a positive quantity τ, possibly depending on the parameters, such that $\beta_i \asymp \tau$ uniformly in $i = 1, \ldots, N$. Then $\beta \asymp \tau$,*

$$A_N = O(\beta), \qquad B_N = O(\beta^2),$$

and consequently

$$d_1(\mathsf{G}_N, U_N) = O(\beta), \qquad d_2(\mathsf{G}_N, U_N) = O(\alpha\beta^2 e_\alpha + \beta^2).$$

If, in addition, $\alpha > 1$, then

$$d_{\mathrm{Kol}}(\mathsf{G}_N, U_N) = O\big((\alpha - 1)^{-1/4}\big).$$

Proof. Since (24) expresses β as a weighted mean of $\beta_1, \ldots, \beta_N$,

$$\min_i \beta_i \leq \beta \leq \max_i \beta_i.$$

The assumption $\beta_i \asymp \tau$ uniformly in i therefore implies $\beta \asymp \tau$. Hence there exist constants $0 < c < C < \infty$, independent of the regime, such that $c\beta \leq \beta_i \leq C\beta$ for all i. It follows immediately that

$$A_N = O(\beta), \qquad B_N = O(\beta^2).$$

Combining these estimates with Theorem 2.1 gives

$$d_1(\mathsf{G}_N, U_N) = O(\beta), \qquad d_2(\mathsf{G}_N, U_N) = O(\beta^2 + \alpha\beta^2 e_\alpha).$$

For the Kolmogorov bound, use (21), the estimate $A_N = O(\beta)$, and $r_\alpha \geq 1$ for $\alpha > 1$. This gives

$$d_{\mathrm{Kol}}(\mathsf{G}_N, U_N) = O\left(\frac{A_N}{\beta\sqrt{\alpha - 1}}\right)^{1/2} = O\big((\alpha - 1)^{-1/4}\big).$$

□

3 Satterthwaite approximations

We now return to the statistical motivation of the paper. The examples below come from different settings—heteroscedastic variance estimation, dependent Gaussian observations, Wishart traces, and high-dimensional repeated-measures testing—but they share the same algebraic core. In each case, the statistic can be diagonalized as a finite sum

$$W = \sum_{i=1}^{N} Y_i$$

of independent Gamma random variables with possibly unequal scale parameters. The Satterthwaite approximation replaces this sum by a single Gamma distribution with matching mean and variance.

Throughout this section, we write

$$Y_i \sim \mathcal{G}_{\alpha_i, 2\tau_i}, \qquad i = 1, \dots, N,$$

with $Y_1, \dots, Y_N$ independent. Thus the scale parameters in Theorem 2.1 are $\beta_i = 2\tau_i$. If the moment-matched Gamma approximation is $U \sim \mathcal{G}_{\nu/2, 2\tau}$, with $\mu = \mathbb{E}[W]$ and $\tau = \mu/\nu$, define

$$\Delta^{(r)} := \max_i \left\{ \left| \frac{\tau_i}{\tau} - 1 \right| \tau_i^r \right\}, \qquad r = 1, 2.$$

Then Theorem 2.1 gives directly

$$d_1(W, U) \le 4\Delta^{(1)}, \qquad d_2(W, U) \le 8\Delta^{(2)} + 2\mu e_{\nu/2} \Delta^{(1)},$$

and, if $\nu > 2$,

$$d_{\mathrm{Kol}}(W, U) \le 2 \left(\frac{\nu \Delta^{(1)}}{\mu \sqrt{2\pi(\nu/2 - 1)}\, r_{\nu/2}} \right)^{1/2}.$$

Indeed, $A_N = 2\Delta^{(1)}$, $B_N = 4\Delta^{(2)}$, and $\alpha\beta = \mu$.

Related Gamma approximation bounds for sums of independent nonnegative random variables were obtained by Liu and Xia (2022), in particular through their Corollary 3.4 for convolutions of independent Gamma variables. The bounds below are different in spirit: they are tailored to the Satterthwaite construction and express the error directly through the discrepancy between the individual half-scales τ_i and the matched half-scale $\tau = \mu/\nu$.

3.1 Unequal-variance pooled variances

We begin with the classical setting behind Welch–Satterthwaite type approximations. Let K be fixed and suppose that, for $k = 1, \dots, K$, $X_{k,1}, \dots, X_{k,n_k} \overset{\text{iid}}{\sim} N(0, \sigma_k^2)$, with the samples mutually independent across groups. Consider

$$W_n := \sum_{k=1}^{K} \frac{S_k^2}{n_k}, \qquad S_k^2 := \frac{1}{n_k - 1} \sum_{i=1}^{n_k} \left(X_{k,i} - \overline{X}_k \right)^2. \tag{25}$$

Since

$$\frac{S_k^2}{n_k} \sim \frac{\sigma_k^2}{n_k(n_k - 1)} \chi^2_{n_k - 1} \sim \mathcal{G}_{(n_k - 1)/2,\, 2\sigma_k^2/(n_k(n_k - 1))},$$

the half-scales are $\tau_{k,n} := \sigma_k^2/(n_k(n_k - 1))$. The Satterthwaite approximation is $U_n \sim \mathcal{G}_{\nu_n/2, 2\mu_n/\nu_n}$, where

$$\mu_n := \sum_{k=1}^{K} \frac{\sigma_k^2}{n_k}, \qquad \nu_n := \frac{\left(\sum_{k=1}^{K} \sigma_k^2/n_k \right)^2}{\sum_{k=1}^{K} \sigma_k^4/(n_k^2(n_k - 1))}. \tag{26}$$

The matched half-scale is $\tau_n = \mu_n/\nu_n$.

Corollary 3.1. *Let W_n be defined by (25), and let $U_n \sim \mathcal{G}_{\nu_n/2, 2\mu_n/\nu_n}$, with μ_n, ν_n defined by (26). For $r = 1, 2$, set*

$$\Delta_n^{(r)} := \max_{1 \le k \le K} \left\{ \left| \frac{\tau_{k,n}}{\mu_n/\nu_n} - 1 \right| \tau_{k,n}^r \right\}.$$

Then

$$d_1(W_n, U_n) \le 4\Delta_n^{(1)}, \qquad d_2(W_n, U_n) \le 8\Delta_n^{(2)} + 2\mu_n e_{\nu_n/2} \Delta_n^{(1)}.$$

If $\nu_n > 2$, then

$$d_{\mathrm{Kol}}(W_n, U_n) \le 2\left(\frac{\nu_n \Delta_n^{(1)}}{\mu_n \sqrt{2\pi(\nu_n/2-1)}\, r_{\nu_n/2}}\right)^{1/2}.$$

Remark 3.2. Assume that K is fixed and that $n_k = c_{k,n} n$, where $n = \sum_{k=1}^K n_k$, $\sum_{k=1}^K c_{k,n} = 1$, and the proportions $c_{k,n}$ are bounded away from 0. This balanced regime ensures that all group sizes grow at the same order; otherwise a small group may dominate the largest scale discrepancy. If the variances σ_k^2 are fixed, then $\mu_n \asymp n^{-1}$, $\nu_n \asymp n$, $\tau_{k,n} \asymp n^{-2}$, and $\tau_n \asymp n^{-2}$. Consequently $\Delta_n^{(1)} = O(n^{-2})$ and $\Delta_n^{(2)} = O(n^{-4})$. Since $e_{\nu_n/2} = O(n^{-1/2})$, Corollary 3.1 gives

$$d_1(W_n, U_n) = O(n^{-2}), \qquad d_2(W_n, U_n) = O(n^{-7/2}), \qquad d_{\mathrm{Kol}}(W_n, U_n) = O(n^{-1/4}).$$

For $K = 2$, this is the usual Welch–Satterthwaite approximation.

3.2 Sample variance of a stationary Gaussian time series

The second example shows how serial dependence enters the approximation through the spectrum of the covariance matrix. Let $(X_t)_{t\in\mathbb{Z}}$ be a zero-mean stationary Gaussian process with $\mathrm{Cov}[X_t, X_{t+h}] = \gamma_h = \gamma_{-h}$, and assume $\sum_{h\in\mathbb{Z}} |\gamma_h| < \infty$. Its spectral density is

$$\mathcal{F}(\omega) := \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \gamma_h e^{\imath\omega h}, \qquad \omega \in [-\pi, \pi].$$

Write $\sigma^2 := \gamma_0 = \mathrm{Var}[X_t]$. For $\boldsymbol{X}_n := (X_1, \ldots, X_n)'$, we have $\boldsymbol{X}_n \sim N_n(0, \Gamma_n)$, where $\Gamma_n = (\gamma_{a-b})_{1\le a,b\le n}$. Write the normalized spectral decomposition as

$$\Gamma_n = \sigma^2 Q_n \Lambda_n Q_n', \qquad \Lambda_n = \mathrm{diag}(\lambda_{1,n}, \ldots, \lambda_{n,n}), \qquad \sum_{j=1}^n \lambda_{j,n} = n.$$

By Szegö's theorem and its refinements, for every continuous g,

$$\frac{1}{n}\sum_{j=1}^n g(\lambda_{j,n}) \to \frac{1}{2\pi}\int_{-\pi}^{\pi} g\left(\frac{2\pi\mathcal{F}(\omega)}{\sigma^2}\right) d\omega. \tag{27}$$

Consider the normalized sample variance

$$W_n := \frac{1}{n\sigma^2}\sum_{t=1}^n X_t^2. \tag{28}$$

If $Z \sim N_n(0, I_n)$, then

$$W_n \cong \frac{1}{n} Z' Q_n \Lambda_n Q_n' Z \cong \sum_{j=1}^n \frac{\lambda_{j,n}}{n} \xi_j, \qquad \xi_1, \ldots, \xi_n \overset{\text{iid}}{\sim} \chi_1^2.$$

Thus $\tau_{j,n} := \lambda_{j,n}/n$. Since $\sum_j \lambda_{j,n} = n$, we have $\mu_n = 1$. Put

$$\omega_n := \frac{1}{n}\sum_{j=1}^n \lambda_{j,n}^2, \qquad \nu_n := \frac{n}{\omega_n}, \qquad \tau_n := \frac{\omega_n}{n}. \tag{29}$$

The Satterthwaite approximation is $U_n \sim \mathcal{G}_{\nu_n/2, 2, 2/\nu_n}$.

Corollary 3.3. *Let W_n be defined by (28), and let $U_n \sim \mathcal{G}_{\nu_n/2,2,2/\nu_n}$, with ν_n defined in (29). For $r = 1, 2$,*

set

$$\Delta_n^{(r)} := \max_{1\le j\le n} \left\{ \left| \frac{\lambda_{j,n}}{\omega_n} - 1 \right| \left(\frac{\lambda_{j,n}}{n} \right)^r \right\}.$$

Then

$$d_1(W_n, U_n) \le 4\Delta_n^{(1)}, \qquad d_2(W_n, U_n) \le 8\Delta_n^{(2)} + 2e_{\nu_n/2}\Delta_n^{(1)}.$$

If $\nu_n > 2$, *then*

$$d_{\mathrm{Kol}}(W_n, U_n) \le 2 \left(\frac{n\Delta_n^{(1)}}{\omega_n \sqrt{2\pi(\nu_n/2 - 1)}\, r_{\nu_n/2}} \right)^{1/2}.$$

Remark 3.4. The quantity ω_n measures the spectral dispersion of the normalized covariance matrix. By (27) with $g(x) = x^2$, one has $\omega_n = O(1)$, while Cauchy–Schwarz gives $\omega_n \ge 1$. Thus $\nu_n = n/\omega_n \asymp n$. Moreover,

$$\max_{1\le j\le n} \lambda_{j,n} \le \frac{1}{\sigma^2} \sum_{h\in\mathbb{Z}} |\gamma_h|.$$

It follows that $\Delta_n^{(1)} = O(n^{-1})$ and $\Delta_n^{(2)} = O(n^{-2})$. Corollary 3.3 therefore gives

$$d_1(W_n, U_n) = O(n^{-1}), \qquad d_2(W_n, U_n) = O(n^{-3/2}), \qquad d_{\mathrm{Kol}}(W_n, U_n) = O(n^{-1/4}).$$

An illustration is given in Figure 2. We consider stationary Gaussian AR(1) processes

$$X_t = \rho X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, 1-\rho^2), \qquad |\rho| < 1,$$

so that $\mathrm{Var}[X_t] = 1$. Hence W_n is the normalized sample variance in Corollary 3.3. For $n = 10, 20, \ldots, 200$, we approximate $d_1(W_n, U_n)$ by Monte Carlo simulation with $M = 1{,}000{,}000$ replications. We also compare the Satterthwaite approximation with the normal approximation $Z_n \sim N(1, 2/\nu_n)$.

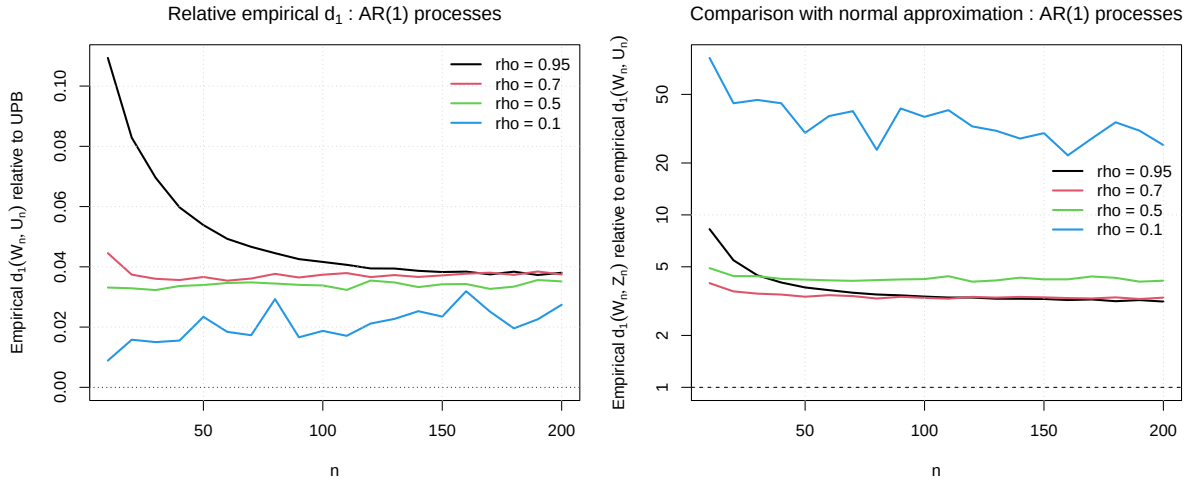


Figure 2: *Left: empirical* $d_1(W_n, U_n)$ *relative to the upper bound from Corollary 3.3. Right: empirical* $d_1(W_n, Z_n)$ *relative to empirical* $d_1(W_n, U_n)$.

The left panel shows that the theoretical upper bound has the correct scale in this example: the empirical Wasserstein distance remains below the bound and the ratio is fairly stable over the displayed range of sample sizes. The effect of dependence is visible. Larger values of $|\rho|$ produce a more heterogeneous covariance spectrum and therefore a larger scale-discrepancy term $\Delta_n^{(1)}$.

The right panel compares the Satterthwaite approximation with the normal approximation having the same first two moments. The ratios are systematically larger than one, indicating that the Gamma approximation is closer in Wasserstein distance throughout the simulation. This is natural here: W_n is a

positive quadratic form, and the Gamma approximation preserves both positivity and the skewed finite-sample shape, whereas the normal approximation is symmetric. When $\rho = 0$, the covariance spectrum is flat and the Satterthwaite approximation is exact.

3.3 Traces of Wishart random matrices

We next consider trace functionals of Wishart matrices. This example is useful because the same Satterthwaite mechanism covers both the standard one-sample Wishart model and sums of independent Wishart matrices with different covariance structures.

Let $\mathfrak{W}_n \sim \mathcal{W}_d(n, n^{-1}\Sigma)$, where Σ is symmetric positive definite, and define $W_n := \mathrm{tr}(\mathfrak{W}_n)$. If $\Sigma = \sigma^2 I_d$, then $W_n \sim (\sigma^2/n)\chi^2_{nd}$. For general Σ, writing $\lambda_1, \dots, \lambda_d$ for its eigenvalues, the Laplace transform gives (Khatri and Srivastava, 1971)

$$W_n \cong \sum_{j=1}^{d} \frac{\lambda_j}{n} \xi_j, \qquad \xi_1, \dots, \xi_d \overset{\text{iid}}{\sim} \chi^2_n.$$

Thus $\tau_{j,n} := \lambda_j/n$. Here

$$\mu = \mathrm{tr}(\Sigma), \qquad \nu_n = n\frac{\mathrm{tr}(\Sigma)^2}{\mathrm{tr}(\Sigma^2)}, \qquad \tau_n = \frac{\mathrm{tr}(\Sigma^2)}{n\mathrm{tr}(\Sigma)}. \tag{30}$$

The Satterthwaite approximation is $U_n \sim \mathcal{G}_{\nu_n/2, 2\mathrm{tr}(\Sigma)/\nu_n}$.

Corollary 3.5. *Let $W_n = \mathrm{tr}(\mathfrak{W}_n)$, with $\mathfrak{W}_n \sim \mathcal{W}_d(n, n^{-1}\Sigma)$, and let $U_n \sim \mathcal{G}_{\nu_n/2, 2\mathrm{tr}(\Sigma)/\nu_n}$, with ν_n defined in (30). For $r = 1, 2$, set*

$$\Delta_n^{(r)} := \max_{1 \le j \le d} \left\{ \left| \frac{\lambda_j \mathrm{tr}(\Sigma)}{\mathrm{tr}(\Sigma^2)} - 1 \right| \left(\frac{\lambda_j}{n} \right)^r \right\}.$$

Then

$$d_1(W_n, U_n) \le 4\Delta_n^{(1)}, \qquad d_2(W_n, U_n) \le 8\Delta_n^{(2)} + 2\mathrm{tr}(\Sigma) e_{\nu_n/2} \Delta_n^{(1)}.$$

If $\nu_n > 2$, then

$$d_{\mathrm{Kol}}(W_n, U_n) \le 2 \left(\frac{\mathrm{tr}(\Sigma)}{\mathrm{tr}(\Sigma^2)} \frac{n\Delta_n^{(1)}}{\sqrt{2\pi(\nu_n/2 - 1)}\, r_{\nu_n/2}} \right)^{1/2}.$$

Remark 3.6. Assume that $0 < c \le \lambda_j \le C < \infty$, $j = 1, \dots, d$, with constants independent of d. Then $\mathrm{tr}(\Sigma) \asymp d$, $\mathrm{tr}(\Sigma^2) \asymp d$, $\nu_n \asymp nd$, $\Delta_n^{(1)} = O(n^{-1})$, and $\Delta_n^{(2)} = O(n^{-2})$. Since $e_{\nu_n/2} = O((nd)^{-1/2})$, Corollary 3.5 gives

$$d_1(W_n, U_n) = O(n^{-1}), \qquad d_2(W_n, U_n) = O(n^{-3/2}d^{1/2}) + O(n^{-2}), \qquad d_{\mathrm{Kol}}(W_n, U_n) = O((nd)^{-1/4}).$$

Thus increasing the dimension improves the Kolmogorov bound through the effective degrees of freedom ν_n, while the d_2-bound also carries the natural size factor $\mathrm{tr}(\Sigma) \asymp d$. If $d \asymp n$, the Kolmogorov rate is of Berry–Esseen order in n.

The same calculation extends directly to sums of independent Wishart matrices. Let $K \in \mathbb{N}$ be fixed, let $n_1, \dots, n_K \in \mathbb{N}$, set $n := \sum_{k=1}^K n_k$, and consider

$$\mathsf{W}_n = \sum_{k=1}^{K} \mathfrak{W}_{k,n}, \qquad \mathfrak{W}_{k,n} \sim \mathcal{W}_d(n_k, n_k^{-1}\Sigma_k), \tag{31}$$

with the summands independent. A Satterthwaite approximation for such sums was proposed by Tan and Gupta (1983), based on matching mean and variance through the generalized variance, that is, the

determinant of the $d(d+1)/2 \times d(d+1)/2$ covariance matrix of W_n. Here we consider only the trace. If $\lambda_{k1}, \dots, \lambda_{kd}$ are the eigenvalues of Σ_k, then

$$W_n := \operatorname{tr}(\mathsf{W}_n) \cong \sum_{k=1}^{K} \sum_{l=1}^{d} \frac{\lambda_{kl}}{n_k} \xi_{kl}, \qquad \xi_{kl} \sim \chi^2_{n_k}$$

independently. Hence $\tau_{kl} := \lambda_{kl}/n_k$. Write

$$\mu = \sum_{k=1}^{K} \operatorname{tr}(\Sigma_k), \qquad \nu = \frac{\left(\sum_{k=1}^{K} \operatorname{tr}(\Sigma_k)\right)^2}{\sum_{k=1}^{K} n_k^{-1} \operatorname{tr}(\Sigma_k^2)}, \qquad \tau = \frac{\mu}{\nu}. \tag{32}$$

The Satterthwaite approximation is $U_n \sim \mathcal{G}_{\nu/2, 2\mu/\nu}$.

Corollary 3.7. *Let $W_n = \operatorname{tr}(\mathsf{W}_n)$, with W_n defined in (31), and let $U_n \sim \mathcal{G}_{\nu/2,2\mu/\nu}$, with μ, ν defined in (32). For $r = 1, 2$, set*

$$\Delta_n^{(r)} := \max_{\substack{1 \le k \le K \\ 1 \le l \le d}} \left\{ \left| \frac{\lambda_{kl}/n_k}{\mu/\nu} - 1 \right| \left(\frac{\lambda_{kl}}{n_k} \right)^r \right\}.$$

Then

$$d_1(W_n, U_n) \le 4\Delta_n^{(1)}, \qquad d_2(W_n, U_n) \le 8\Delta_n^{(2)} + 2\mu e_{\nu/2} \Delta_n^{(1)}.$$

If $\nu > 2$, then

$$d_{\mathrm{Kol}}(W_n, U_n) \le 2 \left(\frac{\nu \Delta_n^{(1)}}{\mu \sqrt{2\pi(\nu/2 - 1)}\, r_{\nu/2}} \right)^{1/2}.$$

Remark 3.8. Assume that K is fixed, $n_k = c_{k,n} n$ with $\sum_{k=1}^{K} c_{k,n} = 1$, and the $c_{k,n}$ are bounded away from 0. Assume also that $0 < c \le \lambda_{kl} \le C < \infty$ uniformly in k, l, n, d. Then $\mu \asymp d$, $\nu \asymp nd$, $\tau_{kl} \asymp n^{-1}$, $\tau \asymp n^{-1}$, $\Delta_n^{(1)} = O(n^{-1})$, and $\Delta_n^{(2)} = O(n^{-2})$. Since $e_{\nu/2} = O((nd)^{-1/2})$, Corollary 3.7 gives

$$d_1(W_n, U_n) = O(n^{-1}), \qquad d_2(W_n, U_n) = O(n^{-3/2} d^{1/2}) + O(n^{-2}), \qquad d_{\mathrm{Kol}}(W_n, U_n) = O((nd)^{-1/4}).$$

If d is fixed, the Kolmogorov rate reduces to $O(n^{-1/4})$; if d grows with n, the effective degrees of freedom increase accordingly.

3.4 High-dimensional repeated-measures designs

We finally consider an ANOVA-type statistic for high-dimensional repeated measurements. The distinctive feature of this example is that, in the oracle case where Σ is known, the quality of the Gamma approximation is governed by the rank q of the hypothesis projection rather than directly by the number of subjects.

Let $\mathbf{H} \in \mathbb{R}^{q \times d}$ be a full-rank hypothesis matrix, and let $\boldsymbol{X}_i \sim N_d(\boldsymbol{\mu}, \Sigma)$, $i = 1, \dots, n$, be independent repeated-measurement vectors. The null hypothesis is $H_0 : \mathbf{H}\boldsymbol{\mu} = 0_q$, or equivalently $H_0 : \mathbf{T}\boldsymbol{\mu} = 0_d$, where

$$\mathbf{T} = \mathbf{H}'(\mathbf{H}\mathbf{H}')^{+}\mathbf{H}.$$

Here $(\cdot)^+$ denotes the Moore–Penrose inverse. The matrix $\mathbf{T}$ is the orthogonal projection onto the hypothesis space generated by $\mathbf{H}$, and $\operatorname{rk}(\mathbf{T}) = q$.

When the number of repeated measurements exceeds the number of subjects, classical Wald-type tests cannot be used because they require inversion of the covariance matrix. In this setting, Rauf Ahmad et al. (2008) proposed an ANOVA-type statistic whose null distribution is approximated by a Satterthwaite

approximation. In the oracle setting where Σ is known, the statistic is

$$W_n = \frac{n}{\mathrm{tr}(\mathbf{T}\Sigma)} \overline{\boldsymbol{X}}' \mathbf{T} \overline{\boldsymbol{X}}. \tag{33}$$

Under H_0, $\overline{\boldsymbol{X}} \sim N_d(0, n^{-1}\Sigma)$, and a spectral decomposition gives

$$W_n \cong \sum_{j=1}^{q} \frac{\lambda_j}{\mathrm{tr}(\mathbf{T}\Sigma)} \xi_j, \qquad \xi_1, \ldots, \xi_q \overset{\text{iid}}{\sim} \chi^2_1,$$

where $\lambda_1, \ldots, \lambda_q$ are the non-zero eigenvalues of $\Sigma^{1/2}\mathbf{T}\Sigma^{1/2}$. Since $\sum_{j=1}^q \lambda_j = \mathrm{tr}(\mathbf{T}\Sigma)$ and $\sum_{j=1}^q \lambda_j^2 = \mathrm{tr}((\mathbf{T}\Sigma)^2)$, the half-scales are $\tau_j := \lambda_j / \mathrm{tr}(\mathbf{T}\Sigma)$. Thus $\mu = 1$ and

$$\nu = \frac{\mathrm{tr}(\mathbf{T}\Sigma)^2}{\mathrm{tr}((\mathbf{T}\Sigma)^2)}, \qquad \tau = \frac{1}{\nu}. \tag{34}$$

The Satterthwaite approximation is $U_q \sim \mathcal{G}_{\nu/2, 2/\nu}$.

Corollary 3.9. *Let W_n be defined in (33), and let $U_q \sim \mathcal{G}_{\nu/2,2/\nu}$, with ν defined in (34). For $r = 1, 2$, set*

$$\Delta_q^{(r)} := \max_{1 \le j \le q} \left\{ \left| \frac{\lambda_j \mathrm{tr}(\mathbf{T}\Sigma)}{\mathrm{tr}((\mathbf{T}\Sigma)^2)} - 1 \right| \left(\frac{\lambda_j}{\mathrm{tr}(\mathbf{T}\Sigma)} \right)^r \right\}.$$

Then

$$d_1(W_n, U_q) \le 4\Delta_q^{(1)}, \qquad d_2(W_n, U_q) \le 8\Delta_q^{(2)} + 2e_{\nu/2}\Delta_q^{(1)}.$$

If $\nu > 2$, then

$$d_{\mathrm{Kol}}(W_n, U_q) \le 2 \left(\frac{\nu \Delta_q^{(1)}}{\sqrt{2\pi(\nu/2 - 1)}\, r_{\nu/2}} \right)^{1/2}.$$

Remark 3.10. Assume that the non-zero eigenvalues of $\Sigma^{1/2}\mathbf{T}\Sigma^{1/2}$ satisfy $0 < c \le \lambda_j \le C < \infty$, $j = 1, \ldots, q$. This condition is implied, for instance, by corresponding uniform eigenvalue bounds on Σ, since $\mathbf{T}$ is an orthogonal projection of rank q. Then $\mathrm{tr}(\mathbf{T}\Sigma) \asymp q$, $\mathrm{tr}((\mathbf{T}\Sigma)^2) \asymp q$, $\nu \asymp q$, $\Delta_q^{(1)} = O(q^{-1})$, and $\Delta_q^{(2)} = O(q^{-2})$. Since $e_{\nu/2} = O(q^{-1/2})$, Corollary 3.9 gives

$$d_1(W_n, U_q) = O(q^{-1}), \qquad d_2(W_n, U_q) = O(q^{-3/2}), \qquad d_{\mathrm{Kol}}(W_n, U_q) = O(q^{-1/4}).$$

If $q = q_n$ grows with n, these rates are read along the chosen high-dimensional regime. For instance, $q_n \asymp n$ yields

$$d_1(W_n, U_q) = O(n^{-1}), \qquad d_2(W_n, U_q) = O(n^{-3/2}), \qquad d_{\mathrm{Kol}}(W_n, U_q) = O(n^{-1/4}).$$

In many high-dimensional repeated-measures regimes, the sample size, the number of repeated measurements, and the rank of the hypothesis projection are assumed to grow at comparable rates.

4 Outlook

The results of this paper show that the Satterthwaite approximation can be studied directly on its natural Gamma scale, without passing through distribution-function expansions or asymptotic normalizations. The kernel representations developed above make the relevant Stein factors explicit and lead to bounds whose dependence on the individual scale parameters is transparent. Several extensions seem worth pursuing.

First, the non-uniform Stein factors obtained here could be used to sharpen other Gamma approximation results. In normal approximation, non-uniform bounds often capture local features that are invisible

in global estimates; see (Chen et al., 2011, Chapter 8). The Gamma kernels derived in this paper suggest that an analogous refinement is possible for positive approximations, especially in regimes where the target shape parameter is large.

Second, it would be natural to develop higher-order versions of the present kernel calculus. The bounds of Gaunt (2013); Gaunt et al. (2017) indicate that Gamma–Stein estimates should persist at arbitrary derivative order, but the representations used here are currently tailored to the first two derivatives needed for moment matching. Extending the construction to higher orders could provide a Stein-based treatment of higher-order moment-matching approximations, including Wood's F-approximation Wood (1989).

Third, the present kernels are intrinsically adapted to the positive half-line. As discussed in Remark 1.10, this makes the approach different from whole-line Gamma Stein equations such as those studied in Döbler and Peccati (2018). A natural next step is to understand whether analogous one-sided kernel representations can be built for signed targets or for approximations involving both positive and negative components.

Finally, the method should apply beyond Gamma summands. The proof of Theorem 2.1 separates two ingredients: a Gamma–Stein transfer identity for the target law, and integration-by-parts identities for the summands. Replacing the latter by suitable Stein identities for other positive variables would lead to Satterthwaite-type bounds for linear combinations of independent squared variables that are not necessarily chi-square distributed.

Declaration on AI-assisted language editing.

The authors used ChatGPT (OpenAI, GPT-5.5 Thinking, accessed in May 2026) for language editing and clarity improvements during manuscript preparation. The authors reviewed and edited all AI-assisted output and take full responsibility for the final content of the manuscript.

References

Baricz, Á. (2012). Mills' ratio: reciprocal convexity and functional inequalities. *Acta Universitatis Sapientiae, Mathematica* **4**, 26–35.

Bodenham, D. A. and Adams, N. M. (2015). A comparison of efficient approximations for a weighted sum of chi-squared random variables. *Statistics and Computing* **26**, 917–928.

Chen, L. H. Y., Goldstein, L., and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method*. Probability and Its Applications. Springer, Berlin.

Döbler, C. (2015). Stein's method of exchangeable pairs for the beta distribution and generalizations. *Electronic Journal of Probability* **20**, 1–34.

Döbler, C. and Peccati, G. (2018). The gamma Stein equation and noncentral de Jong theorems. *Bernoulli* **24**.

Gaunt, R. (2013). *Rates of Convergence of Variance-Gamma Approximations via Stein's Method*. PhD thesis, University of Oxford.

Gaunt, R. E. and Li, S. (2023). Bounding Kolmogorov distances through Wasserstein and related integral probability metrics. *Journal of Mathematical Analysis and Applications* **522**, 126985.

Gaunt, R. E., Pickett, A. M., and Reinert, G. (2017). Chi-square approximation by Stein's method with application to Pearson's statistic. *The Annals of Applied Probability* **27**.

Khatri, C. G. and Srivastava, M. S. (1971). On exact non-null distributions of likelihood ratio criteria for sphericity test and equality of two covariance matrices. *Sankhyā: The Indian Journal of Statistics, Series A* **33**, 201–206.

Ley, C., Reinert, G., and Swan, Y. (2017a). Distances between nested densities and a measure of the impact of the prior in Bayesian statistics. *The Annals of Applied Probability* **27**, 216–241.

Ley, C., Reinert, G., and Swan, Y. (2017b). Stein's method for comparison of univariate distributions. *Probability Surveys* **14**, 1–52.

Liu, Q. and Xia, A. (2022). Geometric sums, size biasing and zero biasing. *Electronic Communications in Probability* **27**, 1–13.

Luk, H. M. (1994). *Stein's Method for the Gamma Distribution and Related Statistical Applications*. PhD thesis, University of Southern California.

Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics* **37**, 541–544.

Pickett, A. (2004). *Rates of convergence of chi-square approximations via Stein's method*. PhD thesis, University of Oxford.

Rauf Ahmad, M., Werner, C., and Brunner, E. (2008). Analysis of high-dimensional repeated measures designs: the one sample case. *Computational Statistics & Data Analysis* **53**, 416–427.

Robbins, H. (1955). A remark on Stirling's formula. *The American Mathematical Monthly* **62**, 26–29.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika* **6**, 309–316.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2**, 110–114.

Tan, W. Y. and Gupta, R. P. (1983). On approximating a linear combination of central Wishart matrices with positive coefficients. *Communications in Statistics – Theory and Methods* **12**, 2589–2600.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika* **29**, 350–362.

Wood, A. T. A. (1989). An F approximation to the distribution of a linear combination of chi-squared variables. *Communications in Statistics – Simulation and Computation* **18**, 1439–1456.

Zolotarev, V. M. (1979). Ideal metrics in the problems of probability theory and mathematical statistics. *Australian Journal of Statistics* **21**, 193–208.

A Proofs from Section 1

Proof of Lemma 1.3. By scaling, it is enough to prove the result for $b = 1$. For $y > 0$, we get

$$\underline{M}_{a,1}(y) = \int_0^1 t^{a-1} e^{y(1-t)}\, dt.$$

Since the integration range is bounded, differentiation under the integral sign is immediate. Hence, for every $n \in \mathbb{N}$ and every $y > 0$,

$$\underline{M}^{(n)}_{a,1}(y) = \int_0^1 (1-t)^n t^{a-1} e^{y(1-t)}\, dt \geq 0.$$

Thus $y \mapsto \underline{M}_{a,1}(y)$ is absolutely monotone on $(0, \infty)$.

Similarly, with the change of variables $u = yt$, we obtain

$$\overline{M}_{a,1}(y) = \int_1^\infty t^{a-1} e^{-y(t-1)}\, dt.$$

For every compact interval $I \subset (0, \infty)$ and every $n \in \mathbb{N}$, the function

$$t \mapsto (t-1)^n t^{a-1} e^{-\eta(t-1)}, \qquad \eta := \inf I > 0,$$

is integrable on $(1,\infty)$. Hence differentiation under the integral sign is justified locally uniformly in $y \in I$. It follows that, for every $n \in \mathbb{N}$ and every $y > 0$,

$$\overline{M}_{a,1}^{(n)}(y) = (-1)^n \int_1^\infty (t-1)^n t^{a-1} e^{-y(t-1)}\, dt.$$

Consequently, $y \mapsto \overline{M}_{a,1}(y)$ is completely monotone on $(0,\infty)$. □

Proof of Lemma 1.4. Recall that

$$\underline{\mathrm{P}}_{a,b}^1(y) = \int_0^y P_{a,b}(u)\, du, \qquad \overline{\mathrm{P}}_{a,b}^1(y) = \int_y^\infty \overline{P}_{a,b}(u)\, du.$$

We shall use

$$\underline{\mathrm{P}}_{a,b}^1(y) = (y-ab)P_{a,b}(y) + by\, p_{a,b}(y), \qquad \overline{\mathrm{P}}_{a,b}^1(y) = (ab-y)\overline{P}_{a,b}(y) + by\, p_{a,b}(y), \tag{35}$$

which follows by integration by parts from

$$(bxp_{a,b}(x))' = (ab-x)p_{a,b}(x).$$

Since the two pieces in $K_{a,b}^{0,1}$ have disjoint supports and $\underline{M}_{a,b}, \overline{M}_{a,b} \geq 0$,

$$\mathbb{E}[|K_{a,b}^{0,1}(X,y)|] = \overline{M}_{a,b}(y)\underline{\mathrm{P}}_{a,b}^1(y) + \underline{M}_{a,b}(y)\overline{\mathrm{P}}_{a,b}^1(y).$$

Using (35), we obtain

$$\mathbb{E}[|K_{a,b}^{0,1}(X,y)|] = \overline{P}_{a,b}(y) + P_{a,b}(y) = 1.$$

Then, using the signs from Lemma 1.3,

$$\mathbb{E}[|K_{a,b}^{1,1}(X,y)|] = -\overline{M}'_{a,b}(y)\underline{\mathrm{P}}_{a,b}^1(y) + \underline{M}'_{a,b}(y)\overline{\mathrm{P}}_{a,b}^1(y) = C_{a,b}^1(y).$$

Introduce the second iterated tails

$$I_{a,b}(y) := \int_0^y \underline{\mathrm{P}}_{a,b}^1(x)\, dx, \qquad J_{a,b}(y) := \int_y^\infty \overline{\mathrm{P}}_{a,b}^1(x)\, dx. \tag{36}$$

A second integration by parts yields

$$I_{a,b}(y) = \frac{1}{2}\left\{(y-ab-b)\underline{\mathrm{P}}_{a,b}^1(y) + byP_{a,b}(y)\right\}, \tag{37}$$

and

$$J_{a,b}(y) = \frac{1}{2}\left\{(ab+b-y)\overline{\mathrm{P}}_{a,b}^1(y) + by\overline{P}_{a,b}(y)\right\}. \tag{38}$$

Now, direct differentiation gives

$$\underline{M}'_{a,b}(y) = \frac{\underline{\mathrm{P}}_{a,b}^1(y)}{b^2y^2p_{a,b}(y)}, \qquad \overline{M}'_{a,b}(y) = -\frac{\overline{\mathrm{P}}_{a,b}^1(y)}{b^2y^2p_{a,b}(y)}. \tag{39}$$

Differentiating one more time, and using (37)–(38), gives

$$\underline{M}''_{a,b}(y) = \frac{2I_{a,b}(y)}{b^3y^3p_{a,b}(y)}, \qquad \overline{M}''_{a,b}(y) = \frac{2J_{a,b}(y)}{b^3y^3p_{a,b}(y)}. \tag{40}$$

Hence by (37), (38) and (40),

$$\mathbb{E}[|K^{2,1}_{a,b}(X,y)|] = \overline{M}''_{a,b}(y)\underline{\mathrm{P}}^1_{a,b}(y) + \underline{M}''_{a,b}(y)\overline{\mathrm{P}}^1_{a,b}(y) = \frac{1}{by}$$

For $K^{2,2}_{a,b}$, since $\underline{M}''_{a,b}(y), \overline{M}''_{a,b}(y) \geq 0$,

$$\mathbb{E}[|K^{2,2}_{a,b}(X,y)|] = \overline{M}''_{a,b}(y)\int_0^y \underline{\mathrm{P}}^1_{a,b}(x)\,dx + \underline{M}''_{a,b}(y)\int_y^\infty \overline{\mathrm{P}}^1_{a,b}(x)\,dx = C^2_{a,b}(y).$$

Finally, differentiating (40) gives

$$\underline{M}'''_{a,b}(y) = \frac{2}{b^3y^3p_{a,b}(y)}\left\{\underline{\mathrm{P}}^1_{a,b}(y) - \left(\frac{a+2}{y} - \frac{1}{b}\right)I_{a,b}(y)\right\},$$

and

$$\overline{M}'''_{a,b}(y) = \frac{2}{b^3y^3p_{a,b}(y)}\left\{-\overline{\mathrm{P}}^1_{a,b}(y) - \left(\frac{a+2}{y} - \frac{1}{b}\right)J_{a,b}(y)\right\}.$$

Using again the signs from Lemma 1.3,

$$\mathbb{E}[|K^{3,2}_{a,b}(X,y)|] = -\overline{M}'''_{a,b}(y)I_{a,b}(y) + \underline{M}'''_{a,b}(y)J_{a,b}(y) = \frac{1}{by},$$

by the identity already proved above. This completes the proof. □

Proof of Lemma 1.5. We first prove the assertions for smooth compactly supported h; the general case follows by approximation, using the bounds obtained below.

Expanding the kernel in (12), we get, for every $x > 0$,

$$f_{h;a,b}(x) = -\frac{\overline{P}_{a,b}(x)}{bxp_{a,b}(x)}\int_0^x P_{a,b}(t)h'(t)\,dt - \frac{P_{a,b}(x)}{bxp_{a,b}(x)}\int_x^\infty \overline{P}_{a,b}(t)h'(t)\,dt. \tag{41}$$

For brevity, set

$$\underline{\Phi}_h(x) := \int_0^x P_{a,b}(t)h'(t)\,dt, \qquad \overline{\Phi}_h(x) := \int_x^\infty \overline{P}_{a,b}(t)h'(t)\,dt.$$

Then

$$f_{h;a,b}(x) = -\overline{M}_{a,b}(x)\underline{\Phi}_h(x) - \underline{M}_{a,b}(x)\overline{\Phi}_h(x).$$

Since

$$\underline{\Phi}'_h(x) = P_{a,b}(x)h'(x), \qquad \overline{\Phi}'_h(x) = -\overline{P}_{a,b}(x)h'(x),$$

differentiating gives

$$\begin{aligned} f'_{h;a,b}(x) = &-\overline{M}'_{a,b}(x)\underline{\Phi}_h(x) - \underline{M}'_{a,b}(x)\overline{\Phi}_h(x) \\ &- \overline{M}_{a,b}(x)P_{a,b}(x)h'(x) + \underline{M}_{a,b}(x)\overline{P}_{a,b}(x)h'(x). \end{aligned}$$

The last two terms cancel, because

$$\overline{M}_{a,b}(x)P_{a,b}(x) = \underline{M}_{a,b}(x)\overline{P}_{a,b}(x) = \frac{P_{a,b}(x)\overline{P}_{a,b}(x)}{bxp_{a,b}(x)}.$$

Therefore

$$f'_{h;a,b}(x) = -\overline{M}'_{a,b}(x)\underline{\Phi}_h(x) - \underline{M}'_{a,b}(x)\overline{\Phi}_h(x) = \mathbb{E}\left[h'(X)K^{1,1}_{a,b}(X,x)\right] = g_{h;a,b}(x).$$

We now check the Stein equation. By integration by parts,

$$\underline{\Phi}_h(x) = P_{a,b}(x)h(x) - \int_0^x h(t)p_{a,b}(t)\,dt,$$

and, since h is compactly supported,

$$\overline{\Phi}_h(x) = -\overline{P}_{a,b}(x)h(x) + \int_x^\infty h(t)p_{a,b}(t)\,dt.$$

It follows that

$$\overline{P}_{a,b}(x)\underline{\Phi}_h(x) + P_{a,b}(x)\overline{\Phi}_h(x) = P_{a,b}(x)\mathcal{G}_{a,b}h - \int_0^x h(t)p_{a,b}(t)\,dt.$$

Hence (41) is equivalently

$$f_{h;a,b}(x) = \frac{1}{bxp_{a,b}(x)}\int_0^x \{h(t) - \mathcal{G}_{a,b}h\}\,p_{a,b}(t)\,dt.$$

Using

$$(bxp_{a,b}(x))' = (ab - x)p_{a,b}(x),$$

we obtain

$$bxf'_{h;a,b}(x) - (x - ab)f_{h;a,b}(x) = h(x) - \mathcal{G}_{a,b}h.$$

Thus $f_{h;a,b}$ solves the Gamma–Stein equation.

Next, differentiating the expression

$$g_{h;a,b}(x) = -\overline{M}'_{a,b}(x)\underline{\Phi}_h(x) - \underline{M}'_{a,b}(x)\overline{\Phi}_h(x)$$

gives

$$\begin{aligned} g'_{h;a,b}(x) = &-\overline{M}''_{a,b}(x)\underline{\Phi}_h(x) - \underline{M}''_{a,b}(x)\overline{\Phi}_h(x) \\ &- \overline{M}'_{a,b}(x)P_{a,b}(x)h'(x) + \underline{M}'_{a,b}(x)\overline{P}_{a,b}(x)h'(x). \end{aligned}$$

The first line is

$$\mathbb{E}\left[h'(X)K^{2,1}_{a,b}(X,x)\right].$$

For the diagonal contribution, use

$$\underline{M}'_{a,b}(x) = \frac{\underline{\mathrm{P}}^1_{a,b}(x)}{b^2x^2p_{a,b}(x)}, \qquad \overline{M}'_{a,b}(x) = -\frac{\overline{\mathrm{P}}^1_{a,b}(x)}{b^2x^2p_{a,b}(x)}.$$

Then

$$\begin{aligned} &- \overline{M}'_{a,b}(x)P_{a,b}(x) + \underline{M}'_{a,b}(x)\overline{P}_{a,b}(x) \\ &= \frac{P_{a,b}(x)\overline{\mathrm{P}}^1_{a,b}(x) + \overline{P}_{a,b}(x)\underline{\mathrm{P}}^1_{a,b}(x)}{b^2x^2p_{a,b}(x)}. \end{aligned}$$

By (35),

$$P_{a,b}(x)\overline{\mathrm{P}}^1_{a,b}(x) + \overline{P}_{a,b}(x)\underline{\mathrm{P}}^1_{a,b}(x) = bxp_{a,b}(x).$$

Therefore

$$-\overline{M}'_{a,b}(x)P_{a,b}(x) + \underline{M}'_{a,b}(x)\overline{P}_{a,b}(x) = \frac{1}{bx}.$$

Consequently,

$$g'_{h;a,b}(x) = \mathbb{E}\left[h'(X)K^{2,1}_{a,b}(X,x)\right] + \frac{h'(x)}{bx} = \ell_{h;a,b}(x).$$

Assume now that $h \in \mathcal{Z}_2$. From the preceding identity,

$$\ell_{h;a,b}(x) = -\overline{M}''_{a,b}(x)\underline{\Phi}_h(x) - \underline{M}''_{a,b}(x)\overline{\Phi}_h(x) + \frac{h'(x)}{bx}.$$

We integrate by parts in the representation of $\widetilde{\ell}_{h;a,b}$. Since h is compactly supported,

$$\int_0^x \underline{\mathrm{P}}^1_{a,b}(t)h''(t)\,dt = \underline{\mathrm{P}}^1_{a,b}(x)h'(x) - \int_0^x P_{a,b}(t)h'(t)\,dt$$

and

$$\int_x^\infty \overline{\mathrm{P}}^1_{a,b}(t)h''(t)\,dt = -\overline{\mathrm{P}}^1_{a,b}(x)h'(x) + \int_x^\infty \overline{P}_{a,b}(t)h'(t)\,dt.$$

Therefore, using (7),

$$\begin{aligned}\widetilde{\ell}_{h;a,b}(x) &= \mathbb{E}\left[h''(X)K^{2,2}_{a,b}(X,x)\right] \\ &= -\overline{M}''_{a,b}(x)\underline{\Phi}_h(x) - \underline{M}''_{a,b}(x)\overline{\Phi}_h(x) \\ &\quad + h'(x)\left\{\overline{M}''_{a,b}(x)\underline{\mathrm{P}}^1_{a,b}(x) + \underline{M}''_{a,b}(x)\overline{\mathrm{P}}^1_{a,b}(x)\right\}.\end{aligned}$$

By Lemma 1.4, or directly from (40) and (35),

$$\overline{M}''_{a,b}(x)\underline{\mathrm{P}}^1_{a,b}(x) + \underline{M}''_{a,b}(x)\overline{\mathrm{P}}^1_{a,b}(x) = \frac{1}{bx}.$$

Hence

$$\widetilde{\ell}_{h;a,b}(x) = \ell_{h;a,b}(x).$$

It remains to differentiate this second-order representation. Write

$$\underline{\Phi}^1_h(x) := \int_0^x \underline{\mathrm{P}}^1_{a,b}(t)h''(t)\,dt, \qquad \overline{\Phi}^1_h(x) := \int_x^\infty \overline{\mathrm{P}}^1_{a,b}(t)h''(t)\,dt.$$

Then

$$\ell_{h;a,b}(x) = \overline{M}''_{a,b}(x)\underline{\Phi}^1_h(x) - \underline{M}''_{a,b}(x)\overline{\Phi}^1_h(x).$$

Since

$$\underline{\Phi}^{1\,\prime}_h(x) = \underline{\mathrm{P}}^1_{a,b}(x)h''(x), \qquad \overline{\Phi}^{1\,\prime}_h(x) = -\overline{\mathrm{P}}^1_{a,b}(x)h''(x),$$

we get

$$\begin{aligned}\ell'_{h;a,b}(x) &= \overline{M}'''_{a,b}(x)\underline{\Phi}^1_h(x) - \underline{M}'''_{a,b}(x)\overline{\Phi}^1_h(x) \\ &\quad + h''(x)\left\{\overline{M}''_{a,b}(x)\underline{\mathrm{P}}^1_{a,b}(x) + \underline{M}''_{a,b}(x)\overline{\mathrm{P}}^1_{a,b}(x)\right\}.\end{aligned}$$

The first line is

$$\mathbb{E}\left[h''(X)K^{3,2}_{a,b}(X,x)\right],$$

and the factor in braces is again $1/(bx)$. Thus

$$\ell'_{h;a,b}(x) = \mathbb{E}\left[h''(X)K^{3,2}_{a,b}(X,x)\right] + \frac{h''(x)}{bx} = m_{h;a,b}(x).$$

We have proved the differentiability statements and the identities for the pointwise representatives. It remains to prove the boundedness assertions. If $h \in \mathcal{Z}_1$, then $\|h'\|_\infty \le 1$, and Lemma 1.4 gives

$$|f_{h;a,b}(x)| \le \mathbb{E}[|K^{0,1}_{a,b}(X,x)|] = 1,$$

$$|g_{h;a,b}(x)| \le \mathbb{E}[|K^{1,1}_{a,b}(X,x)|] = C^1_{a,b}(x),$$

and

$$|\ell_{h;a,b}(x)| \le \mathbb{E}[|K^{2,1}_{a,b}(X,x)|] + \frac{1}{bx} = \frac{2}{bx}.$$

Using the scaling relation (11) and Proposition A.1, it follows that $f_{h;a,b}$, $g_{h;a,b}$, $xg_{h;a,b}$, and $x\ell_{h;a,b}$ are bounded.

If $h \in \mathcal{Z}_2$, then $\|h''\|_\infty \le 1$, and the representation $\ell_{h;a,b} = \widetilde{\ell}_{h;a,b}$ gives

$$|\ell_{h;a,b}(x)| \le \mathbb{E}[|K^{2,2}_{a,b}(X,x)|] = C^2_{a,b}(x).$$

Moreover,

$$|m_{h;a,b}(x)| \le \mathbb{E}[|K^{3,2}_{a,b}(X,x)|] + \frac{1}{bx} = \frac{2}{bx}.$$

Again by (11) and Proposition A.1, $\ell_{h;a,b}$ and $xm_{h;a,b}$ are bounded.

Finally, the extension from smooth compactly supported h to $\mathcal{Z}_1$ and $\mathcal{Z}_2$ follows by mollification and truncation, using the same bounds and dominated convergence. □

Proposition A.1. *For every $a > 0$, the functions*

$$z \mapsto C^1_{a,1}(z), \qquad z \mapsto zC^1_{a,1}(z), \qquad z \mapsto C^2_{a,1}(z)$$

are bounded on $(0,\infty)$. Consequently,

$$c_a := \sup_{z>0} C^1_{a,1}(z), \qquad d_a := \sup_{z>0} zC^1_{a,1}(z), \qquad e_a := \sup_{z>0} C^2_{a,1}(z) \tag{42}$$

are finite. Moreover, as $a \to \infty$,

$$\sqrt{a}\, c_a \to \sqrt{\frac{2}{\pi}}, \qquad \frac{d_a}{\sqrt{a}} \to \sqrt{\frac{2}{\pi}}, \qquad \sqrt{a}\, e_a \to \sqrt{\frac{\pi}{8}}.$$

Proof. Throughout the proof we take $b = 1$. Recall the notation (36). From (39) and (40),

$$C^1_{a,1}(x) = 2\frac{\underline{\mathrm{P}}^1_{a,1}(x)\overline{\mathrm{P}}^1_{a,1}(x)}{x^2 p_{a,1}(x)}, \qquad C^2_{a,1}(x) = 4\frac{I_{a,1}(x)J_{a,1}(x)}{x^3 p_{a,1}(x)}.$$

If $X \sim \mathcal{G}_{a,1}$, then, by Fubini,

$$\underline{\mathrm{P}}^1_{a,1}(x) = \mathbb{E}[(x-X)_+], \quad \overline{\mathrm{P}}^1_{a,1}(x) = \mathbb{E}[(X-x)_+], \quad I_{a,1}(x) = \frac{1}{2}\mathbb{E}[(x-X)^2_+], \quad J_{a,1}(x) = \frac{1}{2}\mathbb{E}[(X-x)^2_+].$$

As $x \downarrow 0$, using $p_{a,1}(x) \sim x^{a-1}/\Gamma(a)$ and $P_{a,1}(x) \sim x^a/\Gamma(a+1)$, we get

$$\underline{\mathrm{P}}^1_{a,1}(x) \sim \frac{x^{a+1}}{\Gamma(a+2)}, \quad \overline{\mathrm{P}}^1_{a,1}(x) \to a, \quad I_{a,1}(x) \sim \frac{x^{a+2}}{(a+2)\Gamma(a+2)}, \quad J_{a,1}(x) \to \frac{a(a+1)}{2}.$$

Thus

$$C^1_{a,1}(x) \to \frac{2}{a+1}, \qquad xC^1_{a,1}(x) \to 0, \qquad C^2_{a,1}(x) \to \frac{2}{a+2}.$$

As $x \to \infty$, $\underline{\mathrm{P}}^1_{a,1}(x) - \overline{\mathrm{P}}^1_{a,1}(x) = x - a$, while the standard incomplete-gamma expansion gives $\overline{P}_{a,1}(x) \sim p_{a,1}(x)$ and $\overline{\mathrm{P}}^1_{a,1}(x) \sim p_{a,1}(x)$. Hence

$$C^1_{a,1}(x) \sim \frac{2}{x}, \qquad xC^1_{a,1}(x) \to 2.$$

Moreover $I_{a,1}(x) \sim x^2/2$ and $J_{a,1}(x) \sim p_{a,1}(x)$, so $C^2_{a,1}(x) \sim 2/x \to 0$. By continuity and the finite one-sided limits just computed, the three functions $C^1_{a,1}$, $xC^1_{a,1}$, and $C^2_{a,1}$ are bounded on $(0,\infty)$, and

therefore $c_a, d_a, e_a < \infty$.

We turn to the large-a behaviour. Put $x = a + t\sqrt{a}$ and $Y_a = (X - a)/\sqrt{a}$. The local central limit theorem gives, locally uniformly in t,

$$\sqrt{a}\, p_{a,1}(a + t\sqrt{a}) \to \varphi(t).$$

Since $\sup_{a\geq 1} \mathbb{E}[Y_a^4] < \infty$, the family $(Y_a^2)_{a\geq 1}$ is uniformly integrable, and hence, also locally uniformly in t,

$$\frac{\underline{\mathrm{P}}^1_{a,1}(a + t\sqrt{a})}{\sqrt{a}} \to \mathbb{E}[(t - Z)_+] = t\Phi(t) + \varphi(t), \qquad \frac{\overline{\mathrm{P}}^1_{a,1}(a + t\sqrt{a})}{\sqrt{a}} \to \mathbb{E}[(Z - t)_+] = \varphi(t) - t\overline{\Phi}(t),$$

where $Z \sim N(0, 1)$. Consequently,

$$\sqrt{a}\, C^1_{a,1}(a + t\sqrt{a}) \to K_1(t), \qquad \frac{(a + t\sqrt{a})C^1_{a,1}(a + t\sqrt{a})}{\sqrt{a}} \to K_1(t),$$

locally uniformly, with

$$K_1(t) = 2\frac{\mathbb{E}[(t - Z)_+]\mathbb{E}[(Z - t)_+]}{\varphi(t)}.$$

Similarly,

$$\frac{I_{a,1}(a + t\sqrt{a})}{a} \to \frac{1}{2}\mathbb{E}[(t - Z)^2_+], \qquad \frac{J_{a,1}(a + t\sqrt{a})}{a} \to \frac{1}{2}\mathbb{E}[(Z - t)^2_+],$$

locally uniformly, and therefore

$$\sqrt{a}\, C^2_{a,1}(a + t\sqrt{a}) \to K_2(t), \qquad K_2(t) = \frac{\mathbb{E}[(t - Z)^2_+]\mathbb{E}[(Z - t)^2_+]}{\varphi(t)}.$$

It remains only to pass from local convergence to convergence of the suprema. We use the standard Gamma Mills estimates, obtained by integration by parts at the endpoint. They imply that, for some universal constant C,

$$\limsup_{a\to\infty} \sup_{|x-a|\geq R\sqrt{a}} \max\left\{\sqrt{a}\, C^1_{a,1}(x), \frac{xC^1_{a,1}(x)}{\sqrt{a}}, \sqrt{a}\, C^2_{a,1}(x)\right\} \leq C(R^{-1} + R^{-3}).$$

Indeed, the smaller one-sided first and second truncated moments are $O(p_{a,1}(x)x^2/|x-a|^2)$ and $O(p_{a,1}(x)x^3/|x-a|^3)$, respectively. Thus no maximizer can escape the window $x = a+O(\sqrt{a})$. Since also $K_1(t), K_2(t) \to 0$ as $|t| \to \infty$, the locally uniform convergence above yields

$$\sqrt{a}\, c_a \to \sup_{t\in\mathbb{R}} K_1(t), \qquad \frac{d_a}{\sqrt{a}} \to \sup_{t\in\mathbb{R}} K_1(t), \qquad \sqrt{a}\, e_a \to \sup_{t\in\mathbb{R}} K_2(t).$$

A direct differentiation, using $\mathbb{E}[(t-Z)_+] = t\Phi(t)+\varphi(t)$ and $\mathbb{E}[(Z-t)_+] = \varphi(t)-t\overline{\Phi}(t)$, shows that K_1 is even and strictly decreasing on $(0, \infty)$. The analogous computation with the second truncated moments gives the same conclusion for K_2. Hence both are maximized at 0. Since

$$K_1(0) = \sqrt{\frac{2}{\pi}}, \qquad K_2(0) = \sqrt{\frac{\pi}{8}},$$

the announced limits follow. □

Proposition A.2. *For every $a \geq 16$,*

$$e_a \leq \frac{5}{2\sqrt{a}}.$$

Proof. We take $b = 1$ and write $I = I_{a,1}$, $J = J_{a,1}$, $p = p_{a,1}$. Let $X \sim \mathcal{G}_{a,1}$. By Fubini,

$$I(x) = \frac{1}{2}\mathbb{E}[(x - X)_+^2], \qquad J(x) = \frac{1}{2}\mathbb{E}[(X - x)_+^2],$$

so

$$I(x) + J(x) = \frac{1}{2}\{a + (x - a)^2\}. \tag{43}$$

Put $n = a - 1$. If $0 < x \le n$, set $\tau = (n - x)/\sqrt{n}$. With $s = x - y$,

$$I(x) = \frac{1}{2}\int_0^x s^2 p(x - s)\, ds.$$

For $0 < s < x$,

$$\frac{p(x - s)}{p(x)} = \left(1 - \frac{s}{x}\right)^n e^s.$$

Taking $s = xz/\sqrt{n}$ and using $x = n - \tau\sqrt{n}$, the inequality $\log(1 - u) \le -u - u^2/2$ gives

$$\frac{p(x - s)}{p(x)} \le e^{-\tau z - z^2/2}.$$

Thus

$$I(x) \le \frac{x^3 p(x)}{2n^{3/2}} L(\tau), \qquad L(\tau) := \int_0^\infty z^2 e^{-\tau z - z^2/2}\, dz.$$

Since $J(x) \le I(x) + J(x)$, (43) yields

$$\sqrt{a}\, C_{a,1}^2(x) \le L_n(\tau) := \frac{\sqrt{n+1}}{n^{3/2}}\{n + 1 + (\tau\sqrt{n} + 1)^2\} L(\tau).$$

Moreover L_n is decreasing: differentiating and using $L'(\tau) = -\int_0^\infty z^3 e^{-\tau z - z^2/2}\, dz$, one obtains $L_n'(\tau) \le 0$ after one integration by parts. Hence

$$L_n(\tau) \le L_n(0) = \frac{\sqrt{n+1}(n+2)}{n^{3/2}}\sqrt{\frac{\pi}{2}} \le \frac{5}{2}$$

for $a \ge 4$. Therefore

$$\sqrt{a}\, C_{a,1}^2(x) \le \frac{5}{2}, \qquad 0 < x \le a - 1.$$

If $x \ge n$, set $\tau = (x - n)/\sqrt{n}$. With $s = y - x$,

$$J(x) = \frac{1}{2}\int_0^\infty s^2 p(x + s)\, ds.$$

Taking $s = xz/\sqrt{n}$ and using $x = n + \tau\sqrt{n}$,

$$J(x) = \frac{x^3 p(x)}{2n^{3/2}} R_n(\tau),$$

where

$$R_n(\tau) := \int_0^\infty z^2 \exp\left\{n \log\left(1 + \frac{z}{\sqrt{n}}\right) - (\sqrt{n} + \tau)z\right\} dz.$$

Since $I(x) \le I(x) + J(x)$, (43) gives

$$\sqrt{a}\, C_{a,1}^2(x) \le \widetilde{R}_n(\tau) := \frac{\sqrt{n+1}}{n^{3/2}}\{n + 1 + (\tau\sqrt{n} - 1)^2\} R_n(\tau).$$

Again $\widetilde{R}_n$ is decreasing: differentiating under the integral sign and integrating by parts in z gives $\widetilde{R}_n'(\tau) \le$

0. Hence

$$\widetilde{R}_n(\tau) \le \widetilde{R}_n(0) = \frac{e^n(n+2)\mathrm{ExpIntegralE}(-n-2,n)}{\sqrt{n+1}} \le \frac{5}{2}$$

for $a \ge 16$, where

$$\mathrm{ExpIntegralE}(\nu, z) := \int_1^\infty e^{-zt} t^{-\nu}\, dt.$$

Thus

$$\sqrt{a}\, C_{a,1}^2(x) \le \frac{5}{2}, \qquad x \ge a - 1.$$

The two ranges cover $(0, \infty)$, and the result follows. □