

Sharp Risk Bounds and a Shrinking-Variance Transition for Normal and Gamma Fits

Alper Ferudun*

16 October 2026

Abstract

Suppose lognormal data are fitted by an ordinary normal model and a shape-scale gamma model. If $s > 0$ is the log variance, the difference $\Delta(s)$ between their optimized forward Kullback–Leibler risks satisfies $2s/3 < \Delta(s) < s$ and $2/3 < \Delta'(s) < 1$. Both constants are optimal. We also establish the boundary behavior of the maximized gamma-to-normal log-likelihood ratio R_n when $s = s_n \rightarrow 0$. If $ns_n \rightarrow c \in (0, \infty)$, then $R_n \Rightarrow N(2c/3, 2c/3)$. If $ns_n \rightarrow \infty$, gamma is selected with probability tending to one; if $ns_n \rightarrow 0$, that probability tends to one half. The proof uses an exact profile identity and a uniform sample-moment expansion, not an extrapolation of a fixed-variance central limit theorem. We include fixed-variance likelihood and compact-prior marginal asymptotics, while distinguishing these from boundary Bayes factors. The KL projection formulas and general misspecification theory are established prior ingredients.

1 The comparison problem

Let $Z \sim N(0, 1)$ and

$$X = e^{\mu + \sqrt{s}Z}, \quad s > 0, \quad M = \mathbb{E}X = e^{\mu + s/2}, \quad V = M^2(e^s - 1). \quad (1)$$

We compare the ordinary normal density $p_N(x; m, v)$, $m \in \mathbb{R}$, $v > 0$, and the gamma density

$$p_G(x; k, \theta) = \frac{x^{k-1}e^{-x/\theta}}{\Gamma(k)\theta^k}, \quad x, k, \theta > 0.$$

All logarithms are natural. With q denoting the density in (1), define

$$K_j(s) = \inf_{\vartheta} \mathbb{E}_q \log \frac{q(X)}{p_j(X; \vartheta)}, \quad \Delta(s) = K_N(s) - K_G(s).$$

The normal is not truncated, and the divergence is forward KL.

Problem 1.3 of the AIM *Singular learning theory* list asks about asymptotics for precisely this pair of fitted families under lognormal sampling [1]. Its dataset identifier is AIM-ALGEBRAIC_GEOMETRY-0271 in the frozen UnsolvedMath v1.6.0 dataset [5]. That record’s metadata says 11.3; the archived primary page labels the question 1.3. The prompt does not specify priors or a particular asymptotic functional. Our results give an explicit frequentist comparison and one sufficient fixed-prior Bayesian formulation, not a resolution of every interpretation.

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: [AlperTheKing](https://github.com/AlperTheKing).

The optimal gamma approximation is treated by Dang, Ye, Beach and Chafii [2]; normal approximation of lognormal variability is studied by Ellassaïss-Schaap and Duisters [3]. The frozen dataset's machine-generated research aid already records the exact projections, a positive increasing risk gap, endpoint expansions and fixed-variance fluctuations. We do not claim those ingredients as new. The emphasis here is the sharp global linear bound and the direct shrinking-variance likelihood theorem. Classical nonnested misspecified likelihood theory, including the distinction between observationally equal and distinct projections, is due to Vuong [6]. Our boundary argument is derived explicitly below.

2 A sharp global risk bound

Write $h(k) = \log k - \psi(k)$, where $\psi = \Gamma'/\Gamma$.

Theorem 1. *The unique normal KL projection is $N(M, V)$. The unique gamma projection has shape $k = k(s)$ and scale M/k , where*

$$h(k) = s/2. \quad (2)$$

The optimized risks are

$$K_N(s) = \frac{1}{2} \left[s + \log \frac{e^s - 1}{s} \right], \quad (3)$$

$$K_G(s) = \log \Gamma(k) - k \log k + k + \frac{ks}{2} - \frac{1}{2} \log(2\pi es). \quad (4)$$

For every $s > 0$,

$$\frac{2}{3} < \Delta'(s) < 1, \quad \frac{2}{3}s < \Delta(s) < s. \quad (5)$$

Both linear constants are optimal: $\Delta(s)/s \rightarrow 2/3$ as $s \downarrow 0$ and $\Delta(s)/s \rightarrow 1$ as $s \rightarrow \infty$. Moreover $0 < K_G(s) < s/12$ and $3s/4 < K_N(s) < s$.

Proof. The normal cross-entropy is $\frac{1}{2} \log(2\pi v) + [V + (M - m)^2]/(2v)$, uniquely minimized at (M, V) . The gamma cross-entropy is

$$-(k-1)\mu + M/\theta + \log \Gamma(k) + k \log \theta.$$

For fixed k , its unique minimum occurs at $\theta = M/k$. The profile is $\mu + ks/2 + k + \log \Gamma(k) - k \log k$, with derivative $s/2 - h(k)$. The classical digamma integral [4, 5.9.13] reads

$$h(k) = \int_0^\infty e^{-kt} A(t) dt, \quad A(t) = \frac{1}{2} + \frac{1}{2} \left(\coth(t/2) - \frac{2}{t} \right). \quad (6)$$

For $x > 0$, one has $0 < \coth x - 1/x < x/3$. Indeed, $x \cosh x - \sinh x$ vanishes at zero and has positive derivative $x \sinh x$. For the upper inequality set $F(x) = (x^2 + 3) \sinh x - 3x \cosh x$; then $F(0) = 0$ and $F'(x) = x(x \cosh x - \sinh x) > 0$. Consequently

$$\frac{1}{2} < A(t) < \frac{1}{2} + \frac{t}{12}, \quad \frac{1}{2k} < h(k) < \frac{1}{2k} + \frac{1}{12k^2}. \quad (7)$$

The integral gives $h'(k) < 0$, and these bounds give limits infinity and zero at the respective ends of $(0, \infty)$. Thus (2) has a unique solution, where the profile derivative changes from negative to positive. Subtracting the lognormal entropy $\mu + \frac{1}{2} \log(2\pi es)$ gives (3)–(4).

At the root, (7) yields

$$\frac{1}{s} < k(s) < \frac{1 + \sqrt{1 + 2s/3}}{2s} < \frac{1}{s} + \frac{1}{6}. \quad (8)$$

The last step uses $\sqrt{1+x} < 1 + x/2$. The first-order condition cancels the $k'(s)$ term in the optimized gamma derivative, giving

$$K'_G(s) = \frac{1}{2} \left(k(s) - \frac{1}{s} \right), \quad \Delta'(s) = \frac{1}{2} \left(2 + \frac{1}{e^s - 1} - k(s) \right). \quad (9)$$

The hyperbolic lower inequality and $e^s > 1 + s$ imply

$$\frac{1}{s} - \frac{1}{2} < \frac{1}{e^s - 1} < \frac{1}{s}.$$

Together with (8) this proves $2/3 < \Delta' < 1$, $0 < K'_G < 1/12$, and $3/4 < K'_N < 1$.

To fix the integration constants, use the classical positive-real expansions [4, 5.11.1–2]:

$$\begin{aligned} h(k) &= \frac{1}{2k} + \frac{1}{12k^2} + O(k^{-4}), \\ \log \Gamma(k) &= (k - \tfrac{1}{2}) \log k - k + \tfrac{1}{2} \log(2\pi) + \frac{1}{12k} + O(k^{-3}). \end{aligned}$$

They give $k(s) = s^{-1} + 1/6 + O(s)$, $K_G(s) = s/12 + O(s^2)$ and $K_N(s) = 3s/4 + O(s^2)$. Both risks therefore vanish at zero. Integrating the strict derivative bounds proves all global inequalities and the lower sharpness limit. As $s \rightarrow \infty$, $k(s) \rightarrow 0$, so (9) gives $\Delta'(s) \rightarrow 1$. Integration and division by s prove the upper sharpness limit. $\square$

3 An exact profile identity

For a positive nonconstant sample $x_1, \dots, x_n$, $n \geq 2$, put

$$b = \bar{x}, \quad v = \frac{1}{n} \sum_i (x_i - b)^2, \quad a = \log b - \overline{\log x}, \quad c = v/b^2. \quad (10)$$

Strict Jensen gives $a > 0$. The normal MLE is (b, v) ; the gamma MLE is $(\hat{k}, b/\hat{k})$ with $h(\hat{k}) = a$. Let $R_n = \max \ell_G - \max \ell_N$, with no factor of two.

Lemma 2. *For every such sample,*

$$\frac{R_n}{n} = a + \frac{1}{2} \log \frac{c}{2a} - K_G(2a) = \Delta(2a) + \frac{1}{2} \log \frac{c}{e^{2a} - 1}. \quad (11)$$

In particular, R_n is invariant under multiplying the observations by a common positive constant.

Proof. The gamma maximized log likelihood per observation is

$$-\log b + a - \hat{k}a + \hat{k} \log \hat{k} - \log \Gamma(\hat{k}) - \hat{k},$$

while the normal value is $-\log b - \frac{1}{2} \log(2\pi ec)$. Subtract and apply (4) at $2a$. The second identity follows from (3). Both a and c are scale-invariant. $\square$

Population risk ordering does not imply $R_n > 0$ for every sample. The first fluctuating term in the next proof can have either sign.

4 The shrinking-variance transition

Theorem 3. For each n , let $X_{n,i} = \exp(\mu_n + \sigma_n Z_{n,i})$, $1 \leq i \leq n$, where the $Z_{n,i}$ are independent standard normals within each row. Let $\mu_n \in \mathbb{R}$ be arbitrary and $\sigma_n > 0$ tend to zero. Set $s_n = \sigma_n^2$. For the unrestricted maximized log-likelihood ratio R_n :

1. If $ns_n \rightarrow \infty$, then $R_n/(ns_n) \rightarrow 2/3$ in probability and $\mathbb{P}(R_n > 0) \rightarrow 1$.
2. If $ns_n \rightarrow c \in (0, \infty)$, then $R_n \Rightarrow N(2c/3, 2c/3)$ and $\mathbb{P}(R_n > 0) \rightarrow \Phi(\sqrt{2c/3})$.
3. If $ns_n \rightarrow 0$, then $R_n/\sqrt{ns_n} \Rightarrow N(0, 2/3)$, $R_n \rightarrow 0$ in probability, and $\mathbb{P}(R_n > 0) \rightarrow 1/2$.

Here Φ is the standard normal distribution function. Ordinary AIC and BIC have the same selection probabilities, since both models have two parameters.

Proof. Suppress row subscripts. By scale invariance replace X_i by $e^{\sigma z_i}$, where $z_i = Z_i - \bar{Z}$. Define $m_j = n^{-1} \sum_i z_i^j$ and $A_r = n^{-1} \sum_i e^{r\sigma z_i}$. Then $m_1 = 0$, $a = \log A_1$ and $c = A_2/A_1^2 - 1$. Taylor expansion gives

$$a = \frac{m_2}{2}\sigma^2 + \frac{m_3}{6}\sigma^3 + \frac{m_4 - 3m_2^2}{24}\sigma^4 + O_p(\sigma^5), \quad (12)$$

$$c = m_2\sigma^2 + m_3\sigma^3 + \left(\frac{7}{12}m_4 - \frac{5}{4}m_2^2\right)\sigma^4 + O_p(\sigma^5). \quad (13)$$

These remainders hold along every deterministic sequence $\sigma_n \rightarrow 0$. To see this, fix $\sigma_0 > 0$. For $r = 1, 2$ the degree-four exponential remainder is bounded by a constant times $\sigma^5 |z_i|^5 e^{2\sigma_0 |z_i|}$. Its empirical coefficient is $O_p(1)$, by $|z_i| \leq |Z_i| + |\bar{Z}|$, $\bar{Z} = O_p(n^{-1/2})$, and finite Gaussian expectations of polynomials times $e^{2\sigma_0 |Z|}$. All empirical moments in use are $O_p(1)$. Also $A_1 = 1 + O_p(\sigma^2)$ and $m_2 \rightarrow 1$ in probability. Thus the logarithmic and rational expansions have denominators bounded away from zero with probability tending to one. After removing the leading σ^2 factors, their remainders are $O_p(\sigma^3)$. No condition on $\sigma_n \max_i |Z_{n,i}|$ is needed.

Dividing (13) by twice (12) and taking a logarithm yields

$$\frac{1}{2} \log \frac{c}{2a} = \frac{m_3}{3m_2}\sigma + \left(\frac{m_4}{4m_2} - \frac{m_2}{2} - \frac{2m_3^2}{9m_2^2}\right)\sigma^2 + O_p(\sigma^3).$$

Since $K_G(2a) = a/6 + O(a^2)$ by Theorem 1, Lemma 2 gives

$$\frac{R_n}{n} = \frac{m_3}{3m_2}\sigma + \sigma^2 C_n + O_p(\sigma^3), \quad C_n = \frac{m_4}{4m_2} - \frac{m_2}{2} - \frac{2m_3^2}{9m_2^2}. \quad (14)$$

The Gaussian moment laws give $m_2 \rightarrow 1$, $m_4 \rightarrow 3$, $m_3 \rightarrow 0$ and $C_n \rightarrow 2/3$ in probability. Furthermore

$$\sqrt{n} m_3 = \frac{1}{\sqrt{n}} \sum_i (Z_i^3 - 3Z_i) + o_p(1) \Rightarrow N(0, 6). \quad (15)$$

Indeed $m_3 = \bar{Z}^3 - 3\bar{Z}\bar{Z}^2 + 2\bar{Z}^3$; $\bar{Z}^2 - 1 = O_p(n^{-1/2})$ controls the centering remainder. The variance is $\mathbb{E}(Z^3 - 3Z)^2 = 15 - 18 + 9 = 6$. The rowwise CLT applies because every row has the same standard-normal distribution.

If $ns \rightarrow \infty$, divide (14) by s . The first term is $O_p((ns)^{-1/2})$, the remainder is $O_p(\sigma)$ and $C_n \rightarrow 2/3$. If $ns \rightarrow c > 0$, multiply by n . The first term is $\sqrt{ns}(\sqrt{nm_3})/(3m_2)$, the second tends to $2c/3$, and the remainder $O_p(ns\sigma)$ vanishes. Apply (15). Finally, if $ns \rightarrow 0$, divide n times (14) by $\sqrt{ns}$. The drift $\sqrt{ns}C_n$ and remainder $O_p(\sqrt{ns}\sigma)$ vanish, leaving $N(0, 2/3)$. Each selection probability follows from continuity of the limiting normal law at zero, or from convergence to the positive constant $2/3$. $\square$

5 Fixed variance and marginal likelihood

The ordinary fixed-variance comparison follows from a finite-dimensional delta method. We record explicit sufficient Bayesian hypotheses to separate it from Theorem 3.

Theorem 4. *Fix $\mu \in \mathbb{R}$ and $s > 0$, and let X_i be iid as in (1). Then $R_n/n \rightarrow \Delta(s)$ almost surely and*

$$\frac{R_n - n\Delta(s)}{\sqrt{n}} \Rightarrow N(0, \tau^2(s)), \quad \tau^2(s) = \text{Var}_q \log \frac{p_G(X; k, M/k)}{p_N(X; M, V)} > 0. \quad (16)$$

Suppose each model has a proper bounded prior density on a compact subset of its open parameter space, with its pseudo-true point in the interior and a continuous positive density near that point. Then (16) also holds with the gamma-to-normal log Bayes factor $\log B_{G,N}$ in place of R_n , and $n^{-1} \log B_{G,N} \rightarrow \Delta(s)$ almost surely. In each model,

$$\log m_j(X^n) = \ell_j(\hat{\vartheta}_j) - \log n + \log \pi_j(\vartheta_j^*) + \log(2\pi) - \frac{1}{2} \log \det J_j + o_p(1), \quad (17)$$

where J_j is the expected negative Hessian in the prior's coordinates.

Proof. The SLLN for $(X, X^2, \log X)$ and continuity of h^{-1} give consistency of the explicit MLEs and the almost-sure profile rate. The joint CLT for that vector holds because all polynomial moments of X and $\log X$ are finite. The delta method gives $\sqrt{n}$ -consistent MLEs; here $h'(k) < 0$. Taylor expansion at each pseudo-true point, with integrable local derivative envelopes, gives $\ell_j(\hat{\vartheta}_j) = \ell_j(\vartheta_j^*) + O_p(1)$. The iid contrast CLT proves (16).

For its variance write $Y = X/M$, $b = e^s$ and

$$A = \frac{1}{2(b-1)}, \quad B = -k - \frac{1}{b-1}, \quad C = k-1.$$

The random part of the contrast is $AY^2 + BY + C \log Y$. It is not constant on $(0, \infty)$ because $A > 0$, so its variance is strictly positive. Using $\mathbb{E}Y^r = b^{r(r-1)/2}$ gives explicitly

$$\begin{aligned} \tau^2 &= A^2(b^6 - b^2) + B^2(b-1) + C^2s \\ &\quad + 2AB(b^3 - b) + 4ACsb + 2BCs. \end{aligned} \quad (18)$$

On the fixed compact prior support, a constant times $1 + X^2 + |\log X|$ controls each log density and its local derivatives. Uniform SLLN and the unique expected optimum separate every fixed outer neighborhood by a positive log-likelihood gap. The local Hessians are positive definite. In normal coordinates (m, v) , $J_N = \text{diag}(V^{-1}, (2V^2)^{-1})$; in gamma natural coordinates $(k-1, -1/\theta)$,

$$J_G = \begin{pmatrix} \psi_1(k) & \theta \\ \theta & k\theta^2 \end{pmatrix}, \quad \det J_G = \theta^2(k\psi_1(k) - 1) > 0,$$

the last inequality following from $h'(k) < 0$. Uniformly positive empirical Hessians on a small neighborhood control the intermediate integration tails by a Gaussian. Taylor expansion on $n^{-1/2}$ neighborhoods, the positive continuous prior density, and the exponentially small contribution of the remaining compact support then yield the two-dimensional Laplace formula (17). Taking its difference cancels $-\log n$ and proves the Bayes-factor CLT.

For the almost-sure marginal rate separately, bound each integral above by the maximum likelihood and below by its restriction to a fixed small ball about the optimum, which has positive prior mass. Uniform SLLN and then shrinking the ball give the matching rate. An almost-sure claim is not inferred from an $O_p(1)$ remainder. $\square$

Remark 5. The compact supports in Theorem 4 contain fixed interior optima. As $s_n \rightarrow 0$, the normal variance tends to zero and the gamma shape tends to infinity. Those hypotheses are lost. The boundary likelihood theorem is therefore not a Bayes-factor theorem for arbitrary priors. The equality of the AIC and BIC decisions here is algebraic, not a claim that BIC uniformly approximates a boundary marginal likelihood. Growing log variance, sharp error large deviations, truncated normals and other losses are not treated.

The derivation and exposition were developed with substantial AI assistance and checked by the originating researcher through a written audit and reproducible symbolic and numerical calculations. The source package supplies 42 exact symbolic checks, 31 high-precision regression points and sample-profile tests; these supplement the analytic proofs. The work is unrefereed, with no claim of independent review, formal verification or absolute priority.

References

- [1] American Institute of Mathematics, *Singular learning theory*, Section 1, Problem 1.3, <http://aimpl.org/singularlearning/1/>. Archived primary page, 19 December 2024: [Internet Archive snapshot](#).
- [2] S. Dang, J. Ye, M. A. Beach and M. Chaffi, On optimality of gamma approximation for lognormal shadowing models, *IEEE Antennas and Wireless Propagation Letters* **22** (2023), 1099–1103. [doi:10.1109/LAWP.2022.3233522](https://doi.org/10.1109/LAWP.2022.3233522).
- [3] J. Ellassaïss-Schaap and K. Duisters, Variability in the log domain and limitations to its approximation by the normal distribution, *CPT: Pharmacometrics & Systems Pharmacology* **9** (2020), 245–257. [doi:10.1002/psp4.12507](https://doi.org/10.1002/psp4.12507).
- [4] NIST Digital Library of Mathematical Functions, Sections 5.9 and 5.11, equations 5.9.13, 5.11.1 and 5.11.2, <https://dlmf.nist.gov/5.9> and <https://dlmf.nist.gov/5.11>.
- [5] ulam.ai, *UnsolvedMath*, version 1.6.0, record AIM-ALGEBRAIC_GEOMETRY-0271 and accompanying research aid, frozen commit c6e7f41b4eb6ce876feb83e7eb626be1b243fd7e, <https://huggingface.co/datasets/ulamai/UnsolvedMath>. The research aid is machine-generated and not peer-reviewed.
- [6] Q. H. Vuong, Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica* **57** (1989), 307–333. [doi:10.2307/1912557](https://doi.org/10.2307/1912557).