Journal of Mathematics and Statistics

Research Article

On Bessel's Correction: Unbiased Sample Variance, the Bariance, and a Novel Runtime-Optimized Estimator

Felix Reichel

Department of Economics, Johannes Kepler University of Linz, Linz, Austria

Article history
Received: 07-04-2025
Revised: 23-05-2025
Accepted: 29-05-2025

Email: kontakt@felixreichel.com

Abstract: Bessel's correction adjusts the denominator in the sample variance formula from n to $n - 1$ to ensure an unbiased estimator of the population variance. This paper provides rigorous algebraic derivations to reinforce the necessity of this correction. It further introduces the concept of Bariance, an alternative dispersion measure based on average pairwise squared differences that avoids reliance on the arithmetic mean. Building on this, we address practical concerns raised in Rosenthal's article, which advocates for n-based estimates from a Mean Squared Error (MSE) perspective, particularly in pedagogical contexts and specific applied settings. Finally, the empirical component of this work, based on simulation studies, demonstrates that estimating the population variance via an algebraically optimized Bariance formula approach can yield a computational advantage. Specifically, the unbiased sample variance can be computed in linear time using the optimized Bariance estimator, resulting in shorter run-times while preserving statistical validity.

Keywords: Unbiased Sample Variance, Runtime-Optimized Linear Unbiased Sample Variance Estimators

Introduction

Variance Estimation and Multivariate Statistics

Variance estimation is a foundational task across statistics and econometrics, with the sample variance being the default estimator in most applications. The unbiased version, corrected by Bessel's factor (dividing by $n - 1$ rather than n), compensates for the loss of one degree of freedom due to pre-estimating the population mean. This correction is not just a simple algebraic trick—it admits deep geometric interpretations via orthogonal projections in $\mathbb{R}^n$ and can be derived rigorously from them.

Despite its theoretical appeal, the unbiased estimator is not always the most optimal in practice. In small samples especially, its higher variance may lead to suboptimal inference. This has led researchers to consider shrunken estimators that intentionally trade off a small amount of bias for a significant reduction in variance, thereby minimizing Mean Squared Error (MSE). For example, empirical Bayes methods shrink sample variances toward a global prior, stabilizing estimation across thousands of features in genomic studies (Smyth, 2004). Similar techniques based on James Stein shrinkage have been explored for variance estimation in high-dimensional settings (Efron and Morris, 1975).

Beyond the univariate case, shrinkage ideas are especially powerful in multivariate settings. In particular, shrinkage estimators for covariance matrices such as the Ledoit Wolf estimator (Ledoit and Wolf, 2004) have gained popularity in fields like econometrics and finance, particularly in the area of asset pricing. These estimators enhance the stability of sample covariance matrices by shrinking them toward structured targets (e.g., the identity matrix), significantly improving conditioning in high-dimensional models, which are known to perform poorly (Ledoit and Wolf, 2005). This has practical relevance in the construction of variance-covariance matrices for portfolio optimization, factor models, and robust standard error estimation in large-scale regression analysis for econometric applications.

In this broader context, this paper revisits classical variance estimation and introduces a novel perspective via an alternative measure of sample dispersion based on the average squared differences between all unordered pairs in a sample. We formally define this estimator as the Bariance, a term that reflects its construction from pairwise distances (between-variance) rather than deviations from a mean. It can be shown that for mean-centered data, the Bariance equals exactly twice the unbiased sample variance**.** Moreover, a linear-time optimized formulation of the Bariance can be derived using simple algebraic properties that avoids quadratic pairwise computation, making it both theoretically elegant and computationally very efficient.

SCIENCE Publications

Relation to Existing Statistics and Applications

Although the pairwise difference approach has roots in classical statistics such as U-statistics (Hoeffding, 1948), dissimilarity-based dispersion measures, and even the Gini coefficient (Cowell, 2011) the contribution here is a novel, unbiased estimator that is computationally optimized for runtime efficiency. In this respect, Bariance bridges theoretical variance estimation with algorithmic efficiency, a consideration critical in big-data contexts, real-time systems, and streaming analytics.

While computational efficiency is one of its key advantages, the Bariance measure may also prove valuable in applied scenarios where the concept of central tendency is unstable, ill-defined, or misleading. For example, in domains such as network analysis, genomics, ordinal survey research, or clustering, statistical dispersion is often better captured through relational or pairwise structures rather than deviations from a single global mean. In such contexts, the Bariance shares conceptual kinship with the Gini coefficient, which also operates on pairwise differences but in a distributional inequality framework. Unlike Gini, however, Bariance preserves unbiasedness for variance estimation under $i.i.d.$ sampling and scales naturally in high-dimensional or streaming environments. These features make it particularly attractive for modern applications in unsupervised learning, robust statistics, and high-throughput data pipelines—where traditional variance measures may either fail or become computationally prohibitive.

Through an empirical simulation study, I demonstrate that this optimized unbiased sample variance estimator remains unbiased and improves runtime. The simulated empirical runtimes section includes hardware specifications and multiple replications, thereby addressing robustness, reproducibility, and statistical reliability. We then revisit the controversial idea advocated by Rosenthal (2015) that dividing by n (rather than $n-1$) may yield lower-MSE variance estimators in practice, especially when unbiasedness is not strictly required.

To sum up, the Bariance framework bridges computational efficiency with applied relevance, offering a theoretically grounded yet practically flexible alternative to traditional variance estimators. This paper thus aims to bridge classical econometric and statistical theory with modern considerations of efficiency, robustness, and computational scalability, while highlighting the often-underestimated choices in estimator design or usage.

Setup

Let $X_i, \dots X_n \in \mathbb{R}$ be $i.i.d.$ random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}(X_i) = \sigma^2$. Define the sample mean and biased/unbiased sample variance:

$$\bar{X} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i}^{n} X_i \quad (1)$$

$$S^2 \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i}^{n} (X_i - \bar{X})^2 \quad (2)$$

$$\hat{S}^2 \stackrel{\text{def}}{=} \frac{1}{n-1}\sum_{i}^{n} (X_i - \bar{X})^2 \quad (3)$$

Derivation of Bias and Bessel's Correction

An estimator $\hat{\theta}$ for a parameter θ is called unbiased if its expected value equals the true value:

$$\mathbb{E}[\hat{\theta}] = \theta \quad (4)$$

The normal n-based sample variance with denominator n is defined as $S^2 = \frac{1}{n}\sum_i^n (X_i - \bar{X})^2$ as in Eq. 2.

We aim to compute $\mathbb{E}[S^2]$, the expected value of this estimator, to show that it is biased.

We start by expanding the squared deviations:

$$\sum_{i}^{n} (X_i - \bar{X})^2 \equiv \sum_{i}^{n} X_i^2 - n\bar{X}^2 \quad (5)$$

Thus:

$$S^2 = \frac{1}{n}\left(\sum_{i}^{n} X_i^2 - n\,\bar{X}^2\right) = \frac{1}{n}\sum_{i}^{n} X_i^2 - \bar{X}^2 \quad (6)$$

Then, take expectation of S^2. By linearity of expectation:

$$\mathbb{E}[S^2] = \frac{1}{n}\sum_{i}^{n} \mathbb{E}[X_i^2] - \mathbb{E}[\bar{X}^2] \quad (7)$$

Compute $\mathbb{E}\,[X_i^2]$. Using the identity:

$$\mathbb{E}[X_i^2] \equiv \mathbb{V}(X_i) + (\mathbb{E}[X_i])^2 = \sigma^2 + \mu^2 \quad (8)$$

Thus:

$$\frac{1}{n}\sum_{i}^{n} \mathbb{E}[X_i^2] = \mu^2 + \sigma^2 \quad (9)$$

Because the n cancels out.

Compute $\mathbb{E}[\bar{X}^2]$. Recall first:

$$\bar{X} = \frac{1}{n}\sum_{i}^{n} X_i \Rightarrow \mathbb{E}[\bar{X}] = \mu, \quad (10)$$

$$\mathbb{V}(\bar{X}) = \frac{\sigma^2}{n} \quad (11)$$

DOI: 10.3844/jmssp.2025.44.49

Thus:

$$\mathbb{E}[\bar{X}^2] = \mathbb{V}(\bar{X}) + \mathbb{E}[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2 \quad (12)$$

Combining both terms now yields:

$$\mathbb{E}[S^2] = (\mu^2 + \sigma^2) - \left(\mu^2 + \frac{\sigma^2}{n}\right) = \sigma^2 - \frac{\sigma^2}{n} \quad (13)$$

$$\mathbb{E}[S^2] = \frac{n-1}{n}\,\sigma^2, \quad (14)$$

This shows S^2 is biased. Q.E.D.

Bessel's Correction defines the unbiased sample variance as in Eq. 3:

$$\hat{S}^2 \coloneqq \frac{1}{n-1}\sum_{i}^{n}(X_i - \bar{X})^2 \Rightarrow \mathbb{E}[\hat{S}^2] = \sigma^2\ (unbiased) \quad (15)$$

This is Bessel's Correction: dividing by $n-1$ compensates for the loss of one degree of freedom (df).

Introducing the Bariance and an Optimized Linear Runtime Estimator

We define the Bariance of a random sample $\{X_1, X_2, \ldots, X_n\}$ as the average squared difference over all unordered pairs:

$$Bar(X) \coloneqq \frac{1}{n(n-1)} \cdot \sum_{i\neq j}^{n}(X_i - X_j)^2 \quad (16)$$

The selected term "Bariance" emphasizes the estimator's foundation on pairwise between-sample variance rather than deviations from a mean.

We begin by expanding the inner squared difference:

$$(X_i - X_j)^2 = X_i^2 - 2X_iX_j + X_j^2 \quad (17)$$

Summing over all distinct $i \neq j$:

$$\sum_{i\neq j}^{n}(X_i - X_j)^2 = \sum_{i\neq j}^{n}(X_i^2 - 2X_iX_j + X_j^2) \quad (18)$$

Split this into three terms:

$$= \sum_{i\neq j}^{n}(X_i^2) + \sum_{i\neq j}^{n}(X_j^2) - \sum_{i\neq j}^{n}(2X_iX_j) \quad (19)$$

For fixed i, there are $n-1$ values of $j \neq i$:

$$\sum_{i\neq j}^{n}(X_i^2) = (n-1)\sum_{i}^{n}X_i^2 \quad (20)$$

$$\sum_{i\neq j}^{n}(X_j^2) = (n-1)\sum_{j}^{n}X_j^2 \quad (21)$$

So the first two terms become:

$$2(n-1)\sum_{i}^{n}X_i^2 \quad (22)$$

Now consider:

$$\sum_{i\neq j}^{n}X_i\,X_j = \left(\sum_{i}^{n}X_i\right)^2 - \sum_{i}^{n}X_i^2 \quad (23)$$

Combine:

$$\sum_{i\neq j}^{n}X_i\,X_j = 2(n-1)\sum_{i}^{n}X_i^2 - 2\left[\left(\sum_{i}^{n}X_i\right)^2 - \sum_{i}^{n}X_i^2\right] \quad (24)$$

$$= 2n\sum_{i}^{n}X_i^2 - 2\left(\sum_{i}^{n}X_i\right)^2 \quad (25)$$

Substitute back:

$$Bariance = \left(\frac{2n}{n(n-1)}\right)\sum_i^n X_i^2 - \left(\frac{2}{n(n-1)}\right)\left(\sum_i^n X_i\right)^2 \quad (26)$$

The Obtained Optimized Bariance Expression

$$Bariance_{opt} \coloneqq \frac{2}{n-1}\sum_{i}^{n}X_i^2 - \frac{2}{n(n-1)}\left(\sum_{i}^{n}X_i\right)^2 \quad (27)$$

In the Case of Mean Centered Data

Iff $\sum_i^n X_i = 0$ we can simplify the expressiom as:

$$Bariance \equiv \frac{2}{n-1}\sum_{i}^{n}X_i^2 \quad (28)$$

Relate this to the unbiased sample variance:

$$S^2 = \frac{1}{n-1}\sum_{i}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i}^{n}(X_i)^2 \quad (29)$$

Thus:

$$Bariance = 2 \cdot S^2 \quad (30)$$

Properties of the Bariance

$$Let\ \theta := \sigma^2 = 1. \tag{31}$$

Then:

$$E[S^2] = \theta \tag{32}$$

$$E[Bariance] = 2\theta \tag{33}$$

$$Bias(S^2) = 0 \tag{34}$$

$$Bias(Bariance) = \theta \tag{35}$$

$$Var(S^2) = 2\theta^2/(n-1) \tag{36}$$

$$Var(Bariance) = 4\,Var(S^2) \tag{37}$$

$$MSE(S^2) = Var(S^2) \tag{38}$$

$$MSE(Bariance) = 4\,Var(S^2) + \theta^2 \tag{39}$$

Materials and Methods

Data-Generating Processes (DGPs)

All numerical experimentation in this study relies on synthetic data generated under known statistical.

Distributions, enabling exact comparisons between theoretical results and empirical outcomes. Unless otherwise.

Stated, samples $X_1, X_2, \dots, X_n$ were drawn independently and identically distributed (i.i.d.). Two sampling distributions were used:

1. Normal distribution: $X_i \sim N(0,1)$
2. Gamma distribution: $X_i \sim \Gamma(k = 2, \theta = 2)$, with $E[X_i] = 4$ and $Var\,[X_i] = 8$

Sample sizes and replications: n = 100, τ = 1000.

Estimators Evaluated

The following variance estimators were evaluated.

Unbiased sample variance:

$$\hat{S}^2 := \frac{1}{n-1}\sum_{i}^{n}(X_i - \bar{X})^2 \tag{40}$$

Id as in Eq. 3. bariance estimator (pairwise form):

$$Bar(X) := \frac{1}{n(n-1)} \cdot \sum_{i \neq j}^{n}\left(X_i - X_j\right)^2 \tag{41}$$

As in Eq. 16. optimized Bariance (linear form):

$$Bar(X) := \frac{2}{n-1}\sum_{i}^{n} X_i^2 - \frac{2}{n(n-1)}\left(\sum_{i}^{n} X_i\right)^2 \tag{42}$$

As in Eq. 27.

Computational Environment

Simulations were executed in Python and R on a Linux x86_64 machine with vectorized numerical libraries. The method time.perf_counter in Python was used for precision timing. Random seeds were fixed for reproducibility.

Monte-Carlo Protocol

For each replication (τ = 1000):

1. Generate n = 100 i.i.d. observations
2. Compute $\hat{S}^2$ and Bariance
3. Record point estimates, bias, variance, and MSE
4. Aggregate across replications. (as SI unit seconds)

Runtime Benchmarking

Runtime evaluation was performed by repeatedly computing each estimator and measuring wall-clock time. Bootstrapped confidence intervals were computed from τ = 20 repeated measurements. Both naive pairwise and optimized Bariance implementations were benchmarked.

Results

This section summarizes empirical results for unbiased variance and Bariance estimators under normal and gamma-distributed data, as well as runtime comparisons.

Table 1: Empirical results for X ~ N (0, 1) (n = 100, τ = 1000)

Estimator	Mean	Bias	Variance	MSE
Unbiased Sample Variance	1.00091	0.00091	0.02156	0.02151
Bariance	2.00181	1.00181	0.08625	1.08968

Table 2: Empirical results for X ~ Γ (2, 2) (n = 100, τ= 1000)

Estimator	Mean	Bias	Variance	MSE
Unbiased Sample Variance	8.00087	0.00087	2.92574	2.92281
Bariance	16.00174	8.00174	11.70295	75.71907

DOI: 10.3844/jmssp.2025.44.49

Table 3: Runtime for Normal-Distributed Data (in seconds)

n	Biased Var	Unbiased Var	Naïve Bariance	Optimized Bariance
10.0	0.0131	0.0142	0.0601	**0.0119**
20.0	0.0208	0.0143	0.2191	**0.0092**
30.0	0.0115	0.0115	0.4872	**0.0091**
40.0	0.0121	0.0123	0.8767	**0.0104**
50.0	0.0134	0.0132	1.5155	**0.0092**
60.0	0.0124	0.0122	2.105	**0.009**
70.0	0.0186	0.0176	2.7712	**0.0087**
80.0	**0.0126**	0.0205	3.6592	0.0155
90.0	0.0139	0.0135	5.0322	**0.0095**
100.0	0.0127	0.0125	5.6617	**0.0098**

Execution Environment

All simulations were executed in Python 3.11 within a virtualized Linux environment (kernel 4.4.0), using 1 GB RAM and a 32-core $x86\_64$ processor. Timing was measured via time. Perf counter and repeated $\tau = 20$ times for each estimator and sample size.

Runtime Results

Tables 1, 2 and 3 display the empirical runtime results from the manuscript, reproduced exactly, including highlighting of the fastest estimator in each row.

Discussion

Rosenthal (2015) argues that using n instead of $n - 1$ may lead to a smaller Mean Squared Error (MSE), especially when teaching or in practical settings.

He shows that while dividing by n − 1 yields an unbiased estimator, this may come at the cost of increased variance. In some cases, a biased but lower-MSE estimator using n is preferable: "A smaller, shrunken, biased estimator actually reduces the MSE." (Rosenthal, 2015).

This introduces another viewpoint: Unbiasedness is not always the ultimate goal, minimizing error in practice often is.

From a theoretical perspective, unbiasedness ensures that the expected value of the estimator exactly matches the true population variance. However, unbiasedness alone does not guarantee minimal estimation error in finite samples.

Allowing a small bias can reduce this variance enough to yield a lower overall MSE (Casella and Berger, 2002; Shao, 2003).

To illustrate, consider the generalized family of estimators:

$$\sigma^2{}_a = \frac{1}{a}\sum_{i}^{n}(X_i - \bar{X})^2$$

As derived in (Rosenthal, 2015), the MSE-minimizing denominator is shown graphically to be as:

$$a^* = n + 1$$

In sum, relaxing unbiasedness for variance estimation is a principled and context-dependent choice.

Conclusion

Bessel's correction is a foundational concept that ensures unbiased estimates of variance. We explored its necessity through algebraic, and pairwise differences reasoning (now formalized as the Bariance construct), building both intuition and understanding. Additionally, we considered a pedagogical and practical perspective, such as Rosenthal's MSE-based view for estimating variance (Rosenthal, 2015) in finite sample.

Although the unbiased estimator is mathematically correct in expectation, the biased version can sometimes be more intuitive and, in certain contexts, statistically preferable across various sampling distributions. This aligns with insights from modern treatments of mathematical statistics, which often emphasize the trade-off between bias and variance in estimator performance (Casella and Berger, 2002; Shao, 2003). Furthermore, empirical results revealed a faster runtime in our simulation example using the average pairwise differences definition as an unbiased variance estimator, referred to as the Bariance estimator particularly when employing the algebraically optimized formula using scalar sums allowing for vectorised implementations in high-level programming langauges.

To sum up, the main finding the run-time optimized estimator for the Bariance formula was a coincidental yet significant observation, that the unbiased estimator can be computed in linear time and statistically outperforms the conventional unbiased sample variance estimator in all tested empirical runtime performance scenarios. Naturally, many other estimators exist for sample variance, including those designed to trade off bias for computational gains. Someone trained in complexity theory, theoretical computer scientist or mathematician could derive theoretical bounds on the time complexity of such estimators.

Beyond its computational efficiency, however, the Bariance measure may also offer benefits in applied settings where deviation from a central mean is either unstable, undefined, or conceptually inappropriate. In fields such as genomics, network analysis, robust statistics, or ordinal survey research, dispersion may be more meaningfully characterized by average pairwise differences rather than deviations from a global average. Moreover, distance-based methods like clustering, energy statistics, and nonparametric ANOVA (e.g. the case of permutation ANOVA) can benefit from the geometric and symmetry-preserving properties of Bariance, particularly in high-dimensional or irregularly structured data where the mean offers little interpretive value. These contexts highlight how the pairwise construction of Bariance is not only computationally attractive but also methodologically appropriate.

Thus, the optimized Bariance formula alone stands as a viable alternative with promising practical implications for real-time multivariate big data applications, including forecasting (especially with shrunken variance–covariance estimators), computational biology, chemistry, finance, and big-data streaming applications (such as online learning) where unbiased and scalable variance estimation is essential or outperforming a competitor by the means of runtime is important.

Acknowledgment

The author thanks four open-access commentators on Qeios (under *qeios.com/read/3GJGNA.2*) for helpful suggestions.

Funding Information

No external funding was received for this research.

Ethics

This research uses simulated data only and requires no ethical approval.

References

Casella, G., & Berger, R. L. (2002). *Statistical Inference*. (2nd ed.). Duxbury.

Cowell, F. A. (2011). *Measuring Inequality*. (3rd ed.). London School of Economics Perspectives in Economic Analysis. Oxford University Press.

Efron, B., & Morris, C. (1975). Data Analysis Using Stein's Estimator and its Generalizations. *Journal of the American Statistical Association*, *70*(350), 311–319. https://doi.org/10.1080/01621459.1975.10479864

Hoeffding, W. (1948). *A Class of Statistics with Asymptotically Normal Distribution*. The Annals of Mathematical Statistics. https://doi.org/10.1214/aoms/1177730196

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. https://doi.org/10.1016/s0047-259x(03)00096-4

Ledoit, O., & Wolf, M. N. (2005). Honey, I Shrunk the Sample Covariance Matrix. *SSRN Electronic Journal*, *691*. https://doi.org/10.2139/ssrn.433840

Rosenthal, J. S. (2015). The kids are alright: Divide by n when estimating variance. *Institute of Mathematical Statistics*. https://imstat.org/2015/11/17/ the-kids-are-alright-divide-by-n-when-estimating-variance/

Shao, J. (2003). *Mathematical Statistics*. 19(3), 308–334. https://doi.org/10.1214/aoms/1177730196

Smyth, G. K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology*, *3*(1), 1–25. https://doi.org/10.2202/1544-6115.1027