DEPARTMENT OF ECONOMICS
JOHANNES KEPLER UNIVERSITY OF LINZ

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

by

Felix REICHEL

Working Paper No. 2506
April 2025

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

felix.reichel@jku.at

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

Felix Reichel[‡]

June 14, 2024; Last Revised April 16, 2025

Abstract

Bessel's correction adjusts the denominator in the sample variance formula from n to $n-1$ to produce an unbiased estimator for the population variance. This paper includes rigorous derivations, geometric interpretations, and visualizations. It then introduces the concept of "bariance," an alternative pairwise distances intuition of sample dispersion without an arithmetic mean. Finally, we address practical concerns raised in Rosenthal's article [1] advocating the use of n-based estimates from a more holistic MSE-based viewpoint for pedagogical reasons and in certain practical contexts. Finally, the empirical part using simulation reveals that the run-time of estimating population variance can be shortened when using an algebraically optimized "bariance" approach to estimate an unbiased variance.

JEL Codes: C10, C80
Keywords: Unbiased sample variance, Runtime-optimized linear unbiased sample variance estimators

[‡]Corresponding author. Email: mailto:felix.reichel@jku.at, mailto:kontakt@felixreichel.com. This paper is in an early draft stage; suggestions for improvement are welcome and greatly appreciated.

Contents

1 Introduction and Motivation

Variance estimation is a foundational task in 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 [2]. Similar techniques based on James-Stein shrinkage have been explored for variance estimation in high-dimensional settings [3].

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 [4]—have gained popularity in fields like econometrics and finance, particularly in the field 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 break down [4]. 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 [5].

In this broader context, this paper revisits classical variance estimation and introduces a novel perspective through the concept of an alternative measure of sample dispersion based on the average squared differences between all unordered pairs in a sample. 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 efficient.

Through a simulation study, I demonstrate that this **optimized unbiased sample variance estimator** remains unbiased and improves runtime. We then revisit the controversial idea—advocated by Rosenthal [1]—that dividing by n (rather than $n - 1$) may yield lower-MSE variance estimators in practice, especially when unbiasedness is not strictly required.

This paper tries to bridge classical econometric and statistical theory with modern considerations of efficiency, robustness, and computational scalability, while re-instating the often under-estimated choices in estimator design or usage.

2 Definitions and Setup

Let $X_1, X_2, \dots, X_n \in \mathbb{R}$ be i.i.d. random variables with:

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2$$

Define the sample mean and biased/unbiased variance estimates:

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

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:

$$\mathbb{E}[\hat{\theta}] = \theta$$

The normal n-based **sample variance with denominator** n is defined as:

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

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:

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 \equiv \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

Thus:

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

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

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

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

$$\mathbb{E}[X_i^2] \equiv \mathrm{Var}(X_i) + (\mathbb{E}[X_i])^2 = \sigma^2 + \mu^2$$

So, the n cancels out, eventually:

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

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

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \Rightarrow \quad \mathbb{E}[\bar{X}] = \mu, \quad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Thus:

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

Combining both terms now:

$$\mathbb{E}[S^2] = (\mu^2+\sigma^2) - \left(\mu^2 + \frac{\sigma^2}{n}\right) = \sigma^2 - \frac{\sigma^2}{n} = \left(\frac{n-1}{n}\right)\sigma^2$$

$$\boxed{\mathbb{E}[S^2] = \frac{n-1}{n}\sigma^2}$$

This shows that the estimator S^2 is **biased**, underestimating the population variance σ^2, because the denominator is larger then the numerator.

Bessel's Correction. To correct the bias, we define the **unbiased sample variance** as:

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

$$\boxed{\mathbb{E}[\hat{S}^2] = \sigma^2 \quad \text{(unbiased)}}$$

This is known as **Bessel's correction —** using $n-1$ instead of n in the denominator compensates for the loss of one degree of freedom from estimating the mean μ with $\bar{X}$.

4 Geometric Interpretation of Estimated Variance and $n-1$ Degrees of Freedom

4.1 Orthogonal Projection

Let $\vec{X} \in \mathbb{R}^n$ be a data vector. Define the mean:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the mean vector:

$$\vec{\mu} = \bar{X} \cdot \vec{1}$$

We decompose:

$$\vec{X} = \vec{\mu} + \vec{r}, \quad \vec{r} = \vec{X} - \vec{\mu}, \quad \vec{r} \in \vec{1}^{\perp}$$

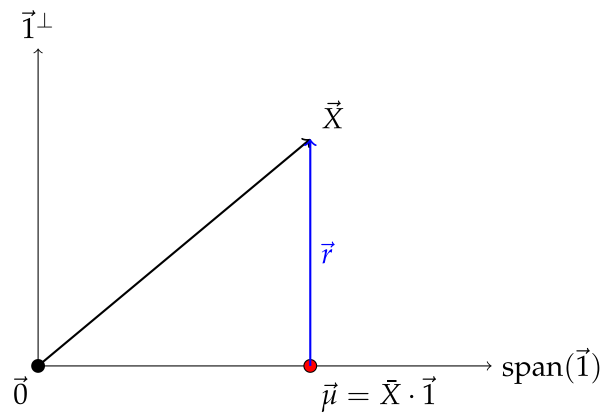

Figure 1: Projection of $\vec{X}$ onto mean direction span($\vec{1}$) and residual in $\vec{1}^{\perp}$

4.2 Dimension Argument

$$\begin{aligned}
&\vec{X} \in \mathbb{R}^n \\
&\vec{\mu} = \bar{X} \cdot \vec{1} \in \text{span}(\vec{1}), \quad \dim = 1 \\
&\vec{r} = \vec{X} - \vec{\mu} \in \vec{1}^{\perp}, \quad \dim = n-1
\end{aligned}$$

$$\Rightarrow \text{Degrees of Freedom} = n-1$$

4.3 Sample Variance

$$s^2 = \frac{1}{n-1}\|\vec{r}\|^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

The normalization by $n-1$ accounts for the loss of one degree of freedom due to estimation of the sample mean.

5 Introducing the "Bariance"

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

$$\text{Bariance} := \frac{1}{n(n-1)} \sum_{i \neq j} (X_i - X_j)^2$$

This can be interpreted as the average squared length of all edges in the complete graph on the sample points.
We begin by expanding the inner squared difference:

$$(X_i - X_j)^2 = X_i^2 - 2X_iX_j + X_j^2$$

Summing over all distinct $i \neq j$:

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

We split this into three terms:

$$= \sum_{i \neq j} X_i^2 + \sum_{i \neq j} X_j^2 - 2\sum_{i \neq j} X_iX_j$$

Note the following observations: - For fixed i, there are $n-1$ values of $j \neq i$, so:

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

Similarly, $\sum_{i \neq j} X_j^2 = (n-1)\sum_{j=1}^{n} X_j^2$
So the first two terms become:

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

Now consider the double sum:

$$\sum_{i \neq j} X_iX_j = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_iX_j\right) - \sum_{i=1}^{n} X_i^2 = \left(\sum_{i=1}^{n} X_i\right)^2 - \sum_{i=1}^{n} X_i^2$$

Combine:

$$\sum_{i \neq j} (X_i - X_j)^2 = 2(n-1)\sum X_i^2 - 2\left(\left(\sum X_i\right)^2 - \sum X_i^2\right)$$

$$= 2(n-1)\sum X_i^2 - 2\left(\left(\sum X_i\right)^2 - \sum X_i^2\right) = 2(n-1)\sum X_i^2 - 2\left(\sum X_i\right)^2 + 2\sum X_i^2$$

$$= 2n\sum X_i^2 - 2\left(\sum X_i\right)^2$$

Substitute back into the Bariance formula. Now divide by $n(n-1)$:

$$\text{Bariance} = \frac{1}{n(n-1)}\sum_{i\neq j}(X_i - X_j)^2 \equiv \frac{2n}{n(n-1)}\sum X_i^2 - \frac{2}{n(n-1)}\left(\sum X_i\right)^2$$

See empirical verification in appendix A.

$$\boxed{\text{Bariance} = \frac{2}{n-1}\left(\sum X_i^2\right) - \frac{2}{n(n-1)}\left(\sum X_i\right)^2}$$

5.1 In the Case Of Mean-centered data

.

If the data is centered, i.e., $\sum X_i = 0$, then:

$$\text{Bariance} = \frac{2n}{n(n-1)}\sum X_i^2 = \frac{2}{n-1}\sum X_i^2$$

We now relate this to the unbiased sample variance:

$$\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum X_i^2 \quad (\text{since } \bar{X} = 0)$$

Therefore, following equality holds for the defined “Bariance“:

$$\boxed{\text{Bariance} = 2 \cdot \hat{S}^2}$$

This result shows that bariance represents twice the unbiased sample variance when the sample is mean-centered. It provides an elegant **pairwise perspective** on variance: instead of summing squared deviations from a central value, we sum squared differences between all pairs and average, despite the one we are currently looking from.

5.2 A Short Numerical Example with Five Numbers

Let $X = \{2, 4, 6, 8, 10\} \Rightarrow \bar{X} = 6$

$$\sum(X_i - \bar{X})^2 = 40, \quad \hat{S}^2 = \frac{40}{4} = 10, \quad \text{Bariance} = \frac{2 \cdot 200}{20} = 20$$

Metric	Value
Sample Mean $\bar{X}$	6
Variance (biased)	8
Variance (unbiased)	10
Bariance	20
Bariance / 2	10

5.3 Graph-Theoretic View of Bariance

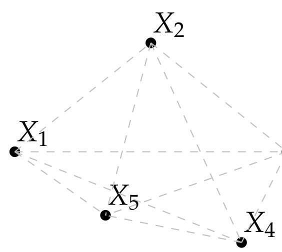

Figure 2: Complete graph of sample values — each edge contributes as a component to the "bariance".

5.4 Deviation from Mean (Variance) vs. Pairwise Differences (Bariance)

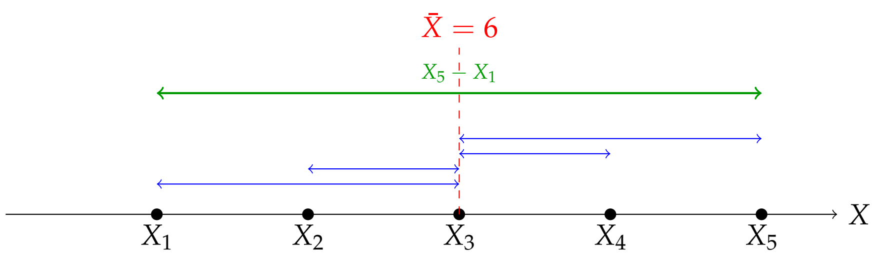

Figure 3: Blue: variance (mean-deviation, n-1 degrees of freedom adjustment). Green: pairwise distance = a bariance component.

5.5 The Pairwise Difference Grid

	2	4	6	8	10
2	0.0	4.0	16.0	36.0	64.0
4	4.0	0.0	4.0	16.0	36.0
6	16.0	4.0	0.0	4.0	16.0
8	36.0	16.0	4.0	0.0	4.0
10	64.0	36.0	16.0	4.0	0.0

Figure 4: Symmetric Grid of $(X_i - X_j)^2$, 0.0 for $i = j$, for $X = \{2, 4, 6, 8, 10\}$

6 Discussion: Should We Just Divide by n?

In Rosenthal [1] 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 might come at the cost of higher variance. In some cases, a biased but lower-MSE estimator using n is preferable:

> "...a smaller, shrunken, biased estimator actually reduces the MSE..." — [1]

This introduces another viewpoint: unbiasedness isn't always the ultimate goal **— minimizing error in practice often is.**

Numerical Example: Bias vs. MSE

Suppose we have $n = 5$ observations drawn from a population with true variance $\sigma^2 = 10$. Then:

- The biased estimator divides by $n = 5$: $\quad S^2 \approx \frac{4}{5} \cdot \sigma^2 = 8$
- The unbiased estimator divides by $n - 1 = 4$: $\quad \hat{S}^2 = 10$

Now compute Mean Squared Error (MSE):

$$
\begin{aligned}
\mathrm{MSE}(S^2) &= \underbrace{\left(\mathbb{E}[S^2] - \sigma^2\right)^2}_{\text{Bias}^2} + \underbrace{\mathrm{Var}(S^2)}_{\text{Variance}} \\
\mathrm{MSE}(\hat{S}^2) &= \underbrace{0^2}_{\text{= Unbiased (Bessel's correction)}} + \mathrm{Var}(\hat{S}^2)
\end{aligned}
$$

It turns out (and Rosenthal notes this explicitly) that: - $\mathrm{Var}(S^2) < \mathrm{Var}(\hat{S}^2)$ - So in some cases, even though S^2 is biased, its total MSE is still smaller!

7 A Simulation Study: Bias^2, Variance, and MSE Across Denominator Values

We consider the family of estimators for the population variance σ^2:

$$\hat{\sigma}_a^2 := \frac{1}{a}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{for varying } a > 0$$

The simulation is carried out with the following parameters:

- Sample size: $n = 5$
- True variance: $\sigma^2 = 10$
- Distribution: $X_i \sim \mathcal{N}(0, \sigma^2)$
- Number of simulations: 100,000

For each value of $a \in [3.5, 8.5]$ (in increments of 0.5), we compute the following empirically:

$$\begin{aligned}
\text{Bias}(\hat{\sigma}_a^2) &= \mathbb{E}[\hat{\sigma}_a^2] - \sigma^2 \\
\text{Bias}^2 &= \left(\mathbb{E}[\hat{\sigma}_a^2] - \sigma^2\right)^2 \\
\text{Variance} &= \text{Var}[\hat{\sigma}_a^2] \\
\text{MSE} &= \text{Bias}^2 + \text{Variance}
\end{aligned}$$

Empirical Results

Table 1: Empirical results averaged from 100,000 simulations for variance estimator using $n = 5$ and $a \in [3.5, 8.5]$ in 0.5 increments. The bold rows highlight $a \in n-1$, n and $n+1$, respectively.

a (denominator)	Bias^2	Variance	MSE
3.5	2.1044	65.8077	67.9122
4.0	**0.0004**	50.3840	50.3844
4.5	1.1967	39.8096	41.0063
5.0	3.9384	32.2458	**36.1842**
5.5	7.3615	26.6494	34.0109
6.0	11.0254	22.3929	**33.4183**
6.5	14.7015	19.0803	33.7819
7.0	18.2728	16.4519	34.7247
7.5	21.6817	14.3315	36.0131
8.0	24.9034	12.5960	37.4994
8.5	27.9314	11.1577	39.0891

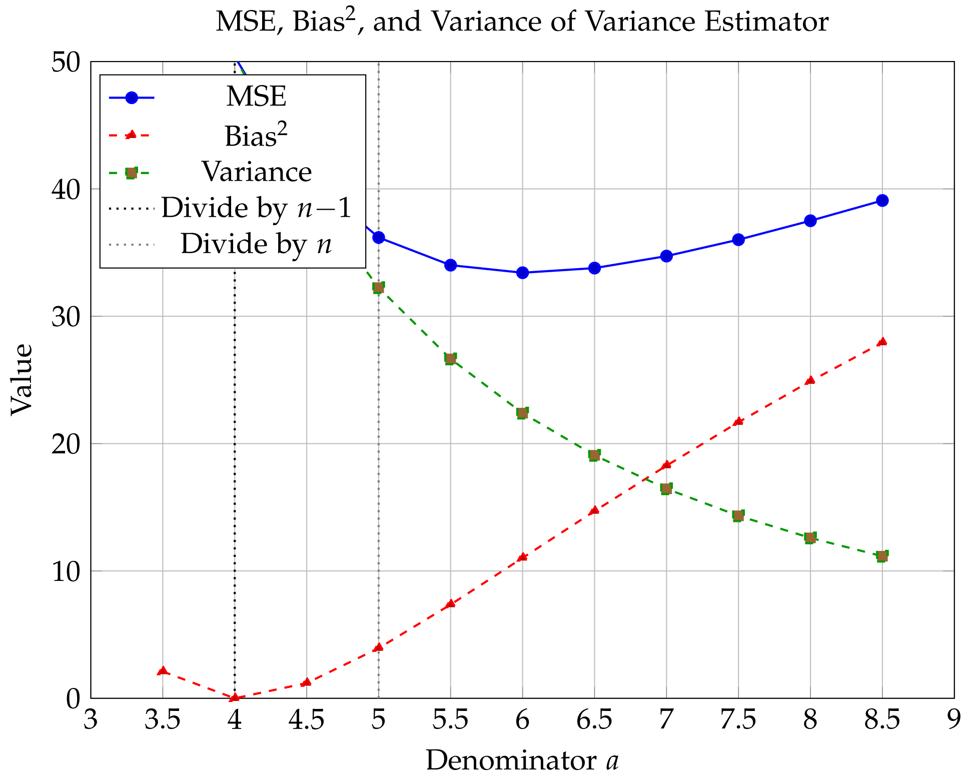

Figure 5: Empirical MSE, Bias^2, and Variance of the sample variance estimator for $a \in [3.5, 8.5]$ and $n = 5$. Minimum MSE occurs between $a = 5.5$ and $a = 6.5$. ($6 = n + 1$ in the case of $\mathcal{N}$)

8 Computational Complexity of Variance Bariance Estimators and Optimization

Let $X = \{X_1, X_2, \ldots, X_n\} \subset \mathbb{R}$ be a sample of size n. Define:

- **Biased sample variance:**
$$S^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- **Unbiased sample variance (Bessel corrected):**
$$\hat{S}^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- **Bariance (pairwise variance):**
$$\text{Bariance}(X) := \frac{1}{n(n-1)}\sum_{i \neq j}(X_i - X_j)^2$$

Table 2: Computational complexity of variance and bariance estimators with explanation

Estimator	Operations	Complexity
Biased Variance $S^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$	• 1 pass to compute mean $\bar{X}$ • 1 pass to compute squared deviations • Total: 2 linear scans • For $n = 5$: 5 additions, 5 subtractions, 5 squarings	$\mathcal{O}(n)$
Unbiased Variance $\hat{S}^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$	Same steps as biased estimator; only the divisor differs. No added computation.	$\mathcal{O}(n)$
Bariance (Naïve) $\frac{1}{n(n-1)}\sum_{i\neq j}(X_i - X_j)^2$	• All $n(n-1)$ ordered pairs evaluated • Each requires subtraction + squaring • For $n = 5$: $5 \times 4 = 20$ pairs • Cost grows quadratically with sample size	$\mathcal{O}(n^2)$
Bariance (Optimized) $\frac{2n}{n(n-1)}\sum X_i^2 - \frac{2}{n(n-1)}\left(\sum X_i\right)^2$	• Uses 2 scalar sums: $\sum X_i$, $\sum X_i^2$ • Each computed in 1 pass • For $n = 5$: 5 additions, 5 squarings	$\mathcal{O}(n)$

8.1 Computational Complexity Comparison with Numerical Illustration

We compare the computational cost of the biased variance, unbiased variance, and bariance estimators using both theoretical analysis and a numerical example for $n = 5$.

Example: $X = \{2, 4, 6, 8, 10\}$

Biased Variance:

$$\bar{X} = 6, \quad S^2 = \frac{1}{5}\sum(X_i - 6)^2 = \frac{40}{5} = 8$$

Unbiased Variance:

$$\hat{S}^2 = \frac{1}{4}\sum(X_i - 6)^2 = \frac{40}{4} = 10$$

Naïve Bariance:

$$\sum_{i<j}(X_i - X_j)^2 = 200, \quad \text{(10 unordered pairs)} \Rightarrow \text{Bariance} = \frac{2 \cdot 200}{20} = 20$$

Optimized Bariance:

$$\sum X_i = 30, \quad \sum X_i^2 = 220$$

$$\text{Bariance} = \frac{2 \cdot 5}{5 \cdot 4} \cdot 220 - \frac{2}{5 \cdot 4} \cdot 900 = 110 - 90 = 20$$

Thus, all estimators yield consistent results, but the number of operations differs significantly with growing n.

While the pairwise form of the bariance appears quadratic, algebraic reduction allows it to be computed in linear time, just like classical variance. This makes it a viable alternative even in large-scale statistical computations.

8.2 Empirical Runtime

To evaluate the practical performance of variance and bariance estimators, we conducted an empirical benchmark based on simulated data. The goal was to measure actual computation time across increasing sample sizes for the four as above defined estimators.

8.2.1 Parameters of the Normal-Based Simulation

- **Number of simulations per sample size:** 1000
- **Sample sizes tested:** $n \in \{10, 20, \ldots, 100\}$
- **Distribution:** $X_i \sim \mathcal{N}(0, 1)$
- **Timing measurement:** Wall-clock time per estimator (summed over 1000 replications)
- **Hardware environment:** CPU timing measured in Python on a standard workstation

All implementations were naïvely vectorized using broadcasting or looped to mimic real computational effort and make the comparison fair between estimator types.

Table 3: Empirical runtime (in seconds) for 1000 simulations per estimator at different sample sizes

n	Biased Variance	Unbiased Variance	Bariance (Naïve)	Bariance (Optimized)
10	0.0131	0.0142	0.0601	**0.0119**
20	0.0208	0.0143	0.2191	**0.0092**
30	0.0115	0.0115	0.4872	**0.0091**
40	0.0121	0.0123	0.8767	**0.0104**
50	0.0134	0.0132	1.5155	**0.0092**
60	0.0124	0.0122	2.1050	**0.0090**
70	0.0186	0.0176	2.7712	**0.0087**
80	**0.0126**	0.0205	3.6592	0.0155
90	0.0139	0.0135	5.0322	**0.0095**
100	0.0127	0.0125	5.6617	**0.0098**

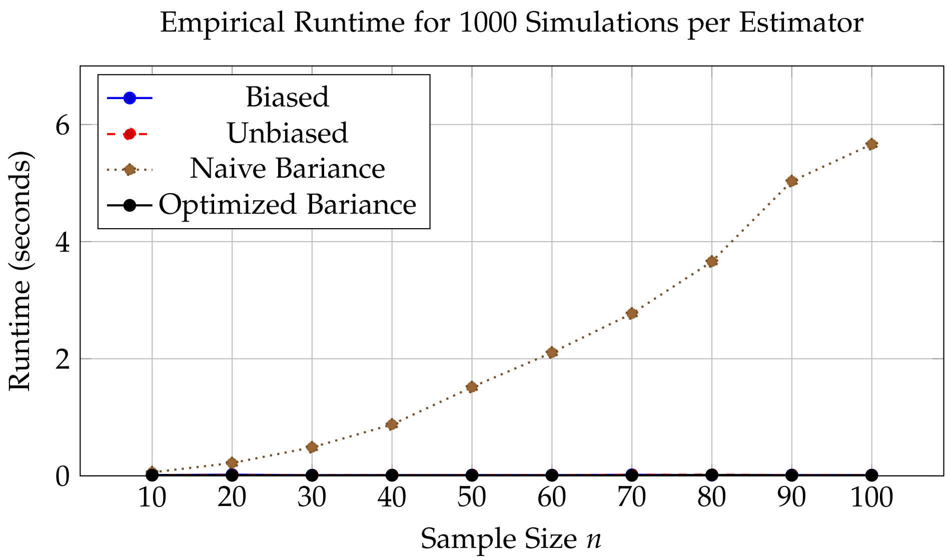

Figure 6: Empirical runtime comparison of variance and bariance estimators over 1000 simulations per sample size.

8.2.2 Gamma-Distributed Data

To examine runtime behavior under non-Gaussian conditions, we conducted a second simulation study using data generated from a Gamma distribution. The Γ-distribution is positively skewed, making it a useful alternative to test estimator performance beyond the symmetric $\mathcal{N}$ case.

Parameters of the Gamma-Based Simulation

- **Number of simulations per sample size:** 500
- **Sample sizes tested:** $n \in \{100, 200, 300, 400, 500\}$
- **Distribution:** $X_i \sim \Gamma(2, 2)$
- **Timing measurement:** Wall-clock time per estimator (summed over 500 replications)
- **Hardware environment:** Standard workstation with vectorized Python implementation

Table 4: Empirical runtime (in seconds) for 500 simulations per estimator using Gamma-distributed data

n	Biased Variance	Unbiased Variance	Bariance (Naïve)	Bariance (Optimized)
100	0.0073	0.0105	0.0149	**0.0065**
200	**0.0083**	0.0101	0.0430	0.0084
300	0.0080	0.0102	0.1075	**0.0073**
400	0.0077	0.0101	0.1937	**0.0074**
500	0.0128	0.0164	0.3266	**0.0095**

8.2.3 Highly Dispersed Gamma-Distributed Data

To further assess runtime robustness under high skew and dispersion, we generated data from a Γ distribution with increased variance. This setup simulates conditions with greater variability, which are common in skewed real-world datasets.

Parameters of the Highly Dispersed Gamma-Based Simulation

- **Number of simulations per sample size:** 1000
- **Sample sizes tested:** $n \in \{50, 100, 150, 200, 250\}$
- **Distribution:** $X_i \sim \Gamma(1.5, 4.0)$
- **Timing measurement:** Wall-clock time per estimator (summed over 1000 replications)
- **Hardware environment:** Standard workstation with vectorized Python implementation

Table 5: Empirical runtime (in seconds) for 1000 simulations using a highly dispersed Gamma distribution

n	Biased Variance	Unbiased Variance	Bariance (Naïve)	Bariance (Optimized)
50	**0.0134**	0.0171	0.0173	0.0141
100	0.0132	0.0171	0.0284	**0.0121**
150	0.0139	0.0184	0.0507	**0.0128**
200	0.0156	0.0183	0.0831	**0.0127**
250	0.0161	0.0179	0.1284	**0.0129**

9 Conclusion

Bessel's correction is a foundational concept that ensures unbiased estimates of variance. We explored its necessity through algebraic, geometric, and pairwise differences reasoning ("Bariance"), building both intuition and understanding. Additionally, we considered a pedagocial and practical perspective, such as Rosenthal's *MSE*-based view for estimating variance.

Although the unbiased estimator is mathematically correct in expectation, the biased version can sometimes be more intuitive, and, in certain contexts, **is statistically preferable in most cases – for many sampling distributions**. Furthermore, the empirical results reveal a faster runtime in our simulation example using the average pairwise differences definition as unbiased variance estimator when using the algebraically optimized definition with scalar sums. Further research could focus on experiments using unbiased or shrunken runtime-optimized covariance estimators with the application of speeding up construction of variance-covariance matrices.

References

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

[2] 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), Article 3.

[3] Efron, B., & Morris, C. (1975). *Data analysis using Stein's estimator and its generalizations*. *Journal of the American Statistical Association*, 70(350), 311–319.

[4] Ledoit, O., & Wolf, M. (2004). *A well-conditioned estimator for large-dimensional covariance matrices*. *Journal of Multivariate Analysis*, 88(2), 365–411.

[5] Ledoit, O., & Wolf, M. (2003). *Honey, I Shrunk the Sample Covariance Matrix*, *UPF Economics and Business Working Paper Series*. No. 691, Available at https://ssrn.com/abstract=433840 & http://dx.doi.org/10.2139/ssrn.433840.

A Proof of Equivalence: Naïve vs Optimized Bariance Estimators

To verify the theoretical equivalence between the naïve and optimized formulations of the bariance estimator, we conducted a simulation study using the exact formulas defined in Table 2. The data were drawn from a highly dispersed Γ- distribution.

Estimator Formulas

- **Naïve Bariance:**

$$\text{Bariance}_{\text{naïve}} = \frac{1}{n(n-1)} \sum_{i \neq j} (X_i - X_j)^2$$

- **Optimized Bariance:**

$$\text{Bariance}_{\text{opt}} = \frac{2n}{n(n-1)} \sum X_i^2 - \frac{2}{n(n-1)} \left(\sum X_i\right)^2$$

Simulation Parameters

- **Distribution:** Γ(1.5, 4.0)
- **Sample sizes:** $n \in \{50, 100, 150, 200, 250\}$
- **Number of simulations per n:** 1000
- **Language:** Python (NumPy)
- **Precision check:** `numpy.allclose` with `rtol` $= 10^{-9}$, `atol` $= 10^{-9}$

Results

Table 6: Bariance estimator comparison using formula-based definitions

n	Mean Naïve Bariance	Mean Optimized Bariance	Max Absolute Difference
50	47.0330	47.0330	8.53×10^{-14}
100	48.4181	48.4181	7.82×10^{-14}
150	47.9282	47.9282	7.11×10^{-14}
200	47.8339	47.8339	8.53×10^{-14}
250	47.6121	47.6121	4.97×10^{-14}

Conclusion

Across all sample sizes tested, the values of the bariance computed using both the naïve and optimized formulas were numerically equivalent within machine precision. This empirically confirms the algebraic identity:

$$\frac{1}{n(n-1)}\sum_{i\neq j}(X_i - X_j)^2 \equiv \frac{2n}{n(n-1)}\sum X_i^2 - \frac{2}{n(n-1)}\left(\sum X_i\right)^2$$