Research Article

On Bessel's Correction: Unbiased Variance, Degrees of Freedom, and the Sum of Pairwise Differences

Felix Reichel[1]

1. Johannes Kepler University of Linz, Linz, Austria

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. We then introduce the concept of "bariance," an alternative pairwise interpretation of sample dispersion. Finally, we address practical concerns raised in Rosenthal's article advocating the use of n-based estimates from a MSE-based viewpoint for practical reasons and in certain contexts. Finally the empirical part using simulation reveals a shorter runtime for estimating population variance can be shortened using an optimized "bariance" approach using scalar sums.

1. Definitions and Setup

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

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

Define the sample mean and biased/unbiased variance estimates:

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

2. 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 sample variance with denominator n is defined as:

doi.org/10.32388/3GJGNA

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

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

Use the identity for variance in terms of raw moments.

We start by expanding the squared deviations:

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$$

Thus:

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

Take expectation of S^2.

We use linearity of expectation:

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

We now compute each part individually.

Compute $\mathbb{E}[X_i^2]$.

Using the identity:

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

So:

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

Compute $\mathbb{E}[\overline{X}^2]$.

Recall that:

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

Thus:

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

Final result.

 doi.org/10.32388/3GJGNA

Putting everything together:

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

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

This shows that the estimator S^2 is **biased**, underestimating the population variance σ^2.

Bessel's Correction.

To correct the bias, we define the **unbiased sample variance** as:

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

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

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 $\overline{X}$.

3. Geometric Interpretation of Variance

3.1. Orthogonal Projection

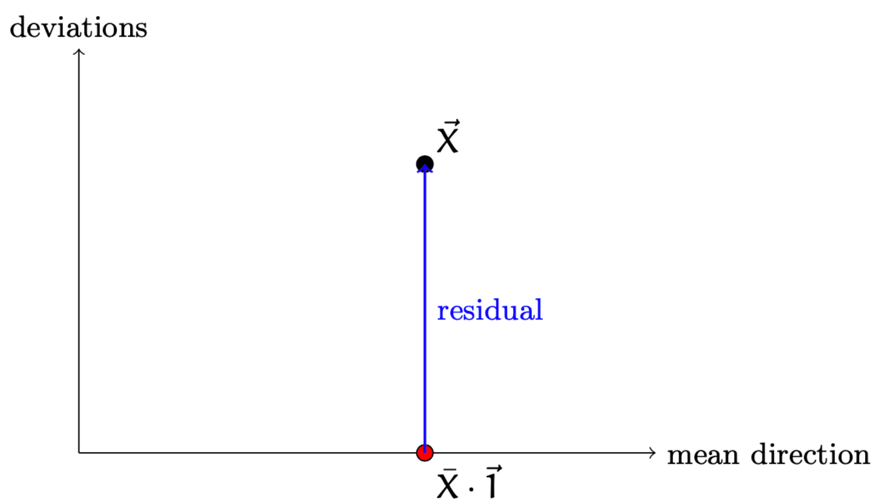

Figure 1. Projection of data vector onto mean direction.

doi.org/10.32388/3GJGNA

3.2. The Dimension Argument

- $\vec{X} \in \mathbb{R}^n$
- Projection onto $\vec{1}$: 1-dimensional mean component
- Residual lives in $\mathbb{R}^{n-1}$
- Hence, degrees of freedom = $n-1$

4. Introducing the "Bariance"

Definition and Expansion of 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.

Expand the square.

We expand 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:

doi.org/10.32388/3GJGNA

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

Now consider the double sum:

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

Putting it all together:

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

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

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

Substitute into the Bariance formula

Now divide by $n(n-1)$:

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

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

4.1. Special case — 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:

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

Therefore, following equality holds for the defined sum of pairwise differences:

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

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.

4.2. A Short Numerical Example with Five Numbers

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

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

Metric	**Value**
Sample Mean $\overline{X}$	6
Variance (biased)	8
Variance (unbiased)	10
Bariance	20

4.3. Graph-Theoretic View of Bariance

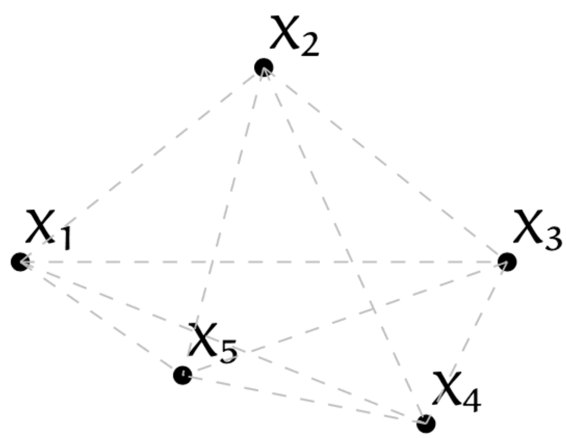

Figure 2. Complete graph of sample values — each edge contributes to the "bariance".

doi.org/10.32388/3GJGNA

4.4. Deviation from Mean vs. Pairwise Differences

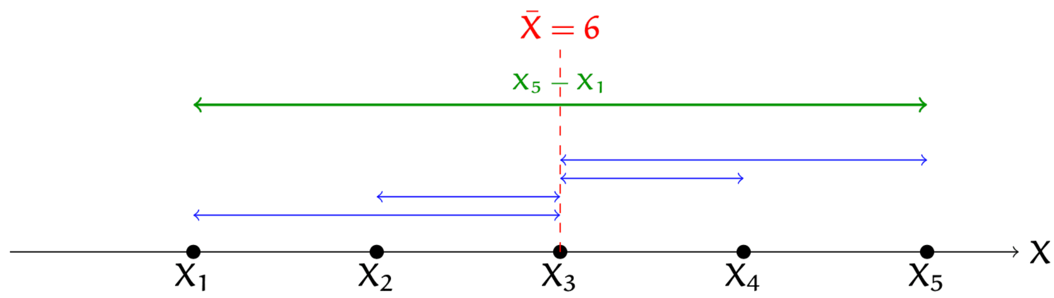

Figure 3. Blue: variance (mean-deviation). Green: pairwise distance = a bariance component.

4.5. 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. Grid of $(X_i - X_j)^2$ for $X = \{2, 4, 6, 8, 10\}$

5. Discussion: Should Just We 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:

 doi.org/10.32388/3GJGNA

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

Now compute Mean Squared Error (MSE):

$$\text{MSE}(S^2) = \underbrace{\left(\mathbb{E}[S^2] - \sigma^2\right)^2}_{\text{Bias}^2} + \underbrace{Var(S^2)}_{\text{Variance}}$$
$$\text{MSE}(\hat{S}^2) = \underbrace{0^2}_{\text{Unbiased}} + Var(\hat{S}^2)$$

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

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

Simulation Setup: Estimator Behavior for $n = 5$ and $a \in [3.5, 8.5]$

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

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

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

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

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

doi.org/10.32388/3GJGNA

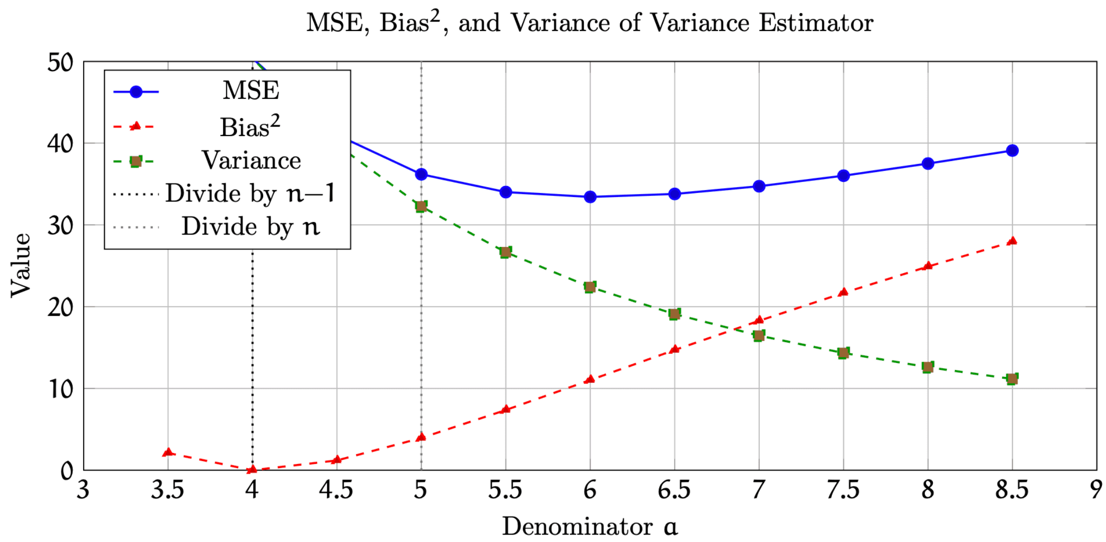

Figure 5. Empirical MSE, Bias^2, and Variance of the sample variance estimator for $a \in [3.5, 8.5]$ and $n = 5$. Minimum MSE occurs between $a = 5.5$ and $a = 6.5$.

7. Computational Complexity of Variance and Bariance Estimators and Optimization using Scalar Sums

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

- Biased sample variance:

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

- Unbiased sample variance (Bessel corrected):

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

- Bariance (pairwise variance):

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

doi.org/10.32388/3GJGNA

Estimator	Operations	Complexity
Biased Variance		
$S^2 = \frac{1}{n}\sum (X_i - \overline{X})^2$	• 1 pass to compute mean $\overline{X}$ • 1 pass to compute squared deviations • Total: 2 linear scans • For $n = 5$: 5 additions, 5 subtractions, 5 squarings	$\mathcal{O}(n)$
Unbiased Variance		
$\hat{S}^2 = \frac{1}{n-1}\sum (X_i - \overline{X})^2$	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)		
$\frac{2n}{n(n-1)}\sum X_i^2 - \frac{2}{n(n-1)}(\sum X_i)^2$	• Uses 2 scalar sums: $\sum X_i, \sum X_i^2$ • Each computed in 1 pass • For $n = 5$: 5 additions, 5 squarings	$\mathcal{O}(n)$

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

7.1. 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:

doi.org/10.32388/3GJGNA

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

Unbiased Variance:

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

Naïve Bariance:

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

Optimized Bariance:

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

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

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.

7.2. Empirical Runtime Simulation

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 above defined estimators.

Parameters of the Simulation Study

- 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 reflect real computational effort and make the comparison fair between estimator types.

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

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

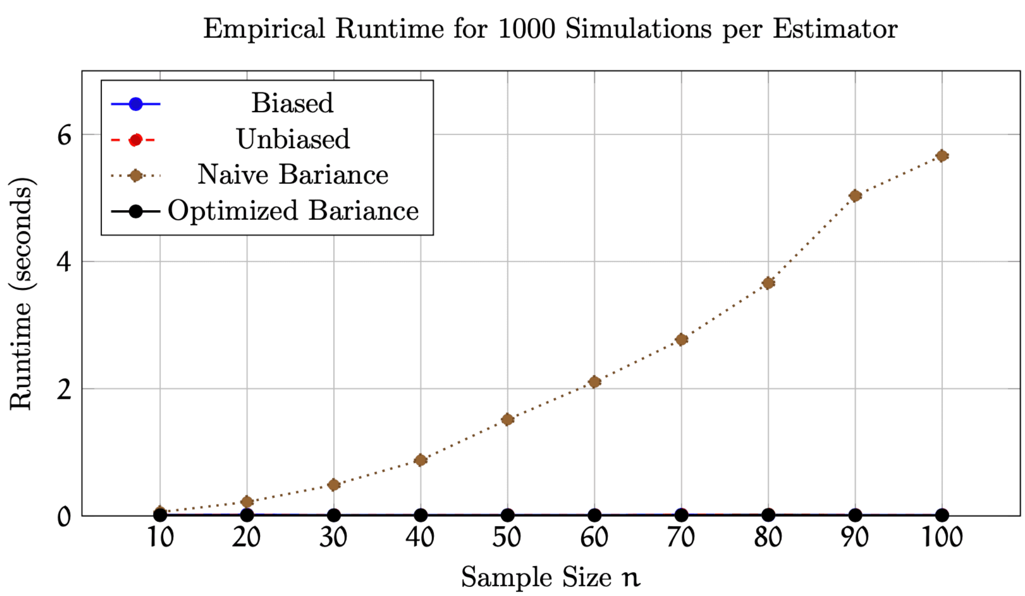

Figure 6. Empirical runtime comparison of variance and bariance estimators over 1000 simulations per sample size.

8. 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 (sum of pairwise differences), building both intuition and understanding. Additionally, we considered a pedagocial and practical perspective, such as Rosenthal's MSE-based view.

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 practice. Furthermore, the empirical results reveal a faster runtime in our simulationon example using the average pairwise differences definition as variance estimator when using the algebraic optimized definition with scalar sums.

References

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

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.