Unbiasedness of the Sample Variance: A Linear Algebra Perspective

Your Friendly Math Teacher

April 5, 2025

1 Motivation: Why Estimate Variance?

In statistics, we often work with a sample $Y_1, Y_2, \ldots, Y_n$ drawn from a larger population. We assume these are independent draws from a distribution with some unknown mean μ and unknown variance σ^2. While the sample mean $\bar{Y} = \frac{1}{n}\sum Y_i$ is a natural estimator for μ, estimating the population's spread, σ^2, requires a bit more thought.

A key measure of spread in our sample is the *sample variance*, defined as:

Definition 1 (Sample Variance)**.** *The sample variance, denoted S_n^2, is given by*

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

You might wonder: why divide by $n-1$ instead of the seemingly more natural n? This factor is known as *Bessel's correction.* Our goal today is to prove that this correction makes S_n^2 an *unbiased estimator* for σ^2.

Definition 2 (Unbiased Estimator)**.** *An estimator $\hat{\theta}$ for a parameter θ is called unbiased if its expected value equals the true value of the parameter, i.e., $\mathbb{E}[\hat{\theta}] = \theta$.*

So, we want to rigorously show that $\mathbb{E}[S_n^2] = \sigma^2$. We'll use the power of linear algebra, assuming for this derivation that our observations Y_i are independent and identically distributed (i.i.d.) as $N(\mu, \sigma^2)$.

2 Setting the Stage: Vector Notation

Let's represent our sample as a vector in $\mathbb{R}^n$:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

Since $Y_i \sim N(\mu, \sigma^2)$ independently, the random vector $\boldsymbol{Y}$ follows a multivariate normal distribution:

$$\boldsymbol{Y} \sim N(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$$

where

- The mean vector is $\boldsymbol{\mu}_{\boldsymbol{Y}} = \mathbb{E}[\boldsymbol{Y}] = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} = \mu\mathbf{1}$, where $\mathbf{1}$ is the $n \times 1$ vector of ones.
- The covariance matrix is $\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \text{Cov}(\boldsymbol{Y}) = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $n \times n$ identity matrix. The off-diagonal zeros reflect the independence of the Y_i.

3 The Sum of Squares as a Quadratic Form

The core of S_n^2 is the sum of squared deviations: $\sum_{i=1}^n (Y_i - \bar{Y})^2$. Let's see how this looks in matrix form. This sum measures the squared length of the vector of deviations from the mean. We can express this using a special matrix operation.

Consider the *centering matrix* $\mathbf{M}$:

$$\mathbf{M} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$$

This matrix has the remarkable property that when it multiplies a vector $\boldsymbol{Y}$, it yields a vector whose components are the deviations from the mean, i.e., $Y_i - \bar{Y}$. (Although we won't explicitly show $\mathbf{M}\boldsymbol{Y}$ gives that exact vector, its quadratic form $\boldsymbol{Y}^T\mathbf{M}\boldsymbol{Y}$ achieves the desired sum.)

Let's verify the connection:

$$\begin{aligned} \boldsymbol{Y}^T\mathbf{M}\boldsymbol{Y} &= \boldsymbol{Y}^T \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \boldsymbol{Y} \\ &= \boldsymbol{Y}^T\mathbf{I}\boldsymbol{Y} - \boldsymbol{Y}^T \left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \boldsymbol{Y} \\ &= \boldsymbol{Y}^T\boldsymbol{Y} - \frac{1}{n}(\boldsymbol{Y}^T\mathbf{1})(\mathbf{1}^T\boldsymbol{Y}) \\ &= \sum_{i=1}^n Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^n Y_i\right)\left(\sum_{j=1}^n Y_j\right) \\ &= \sum Y_i^2 - \frac{1}{n}(n\bar{Y})(n\bar{Y}) \\ &= \sum Y_i^2 - n\bar{Y}^2 \end{aligned}$$

Recall the computational formula for the sum of squares: $\sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2$. So, we've shown:

Lemma 1. *The sum of squared deviations can be expressed as the quadratic form:*

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \boldsymbol{Y}^T\mathbf{M}\boldsymbol{Y}$$

where $\mathbf{M} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$.

Therefore, we can write the quantity related to the sample variance as:

$$(n-1)S_n^2 = \boldsymbol{Y}^T\mathbf{M}\boldsymbol{Y}$$

4 The Main Tool: Expectation of a Quadratic Form

To find $\mathbb{E}[(n-1)S_n^2]$, we need the expectation of $\boldsymbol{Y}^T\mathbf{M}\boldsymbol{Y}$. There's a beautiful general theorem for this:

Theorem 1 (Expectation of a Quadratic Form)**.** *Let* $\mathbf{X}$ *be a random vector with mean vector* $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$ *and covariance matrix* $\mathrm{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$*. For any constant matrix* $\mathbf{A}$ *of appropriate dimensions, the expected value of the quadratic form* $\mathbf{X}^T\mathbf{A}\mathbf{X}$ *is given by:*

$$\mathbb{E}[\mathbf{X}^T\mathbf{A}\mathbf{X}] = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$$

where $\mathrm{tr}(\cdot)$ *denotes the trace of a matrix (the sum of its diagonal elements).*

Remark 1. *This theorem is a cornerstone result in multivariate statistics. It arises from applying the linearity of expectation and properties of the trace operator.*

5 Applying the Theorem

Let's apply this theorem to our situation. We have:

- Random vector: $\mathbf{X} = \boldsymbol{Y}$
- Constant matrix: $\mathbf{A} = \mathbf{M}$
- Mean vector: $\boldsymbol{\mu} = \boldsymbol{\mu}_{\boldsymbol{Y}} = \mu\mathbf{1}$
- Covariance matrix: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{Y}} = \sigma^2\mathbf{I}$

Plugging these into the formula:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{Y}^T\mathbf{M}\boldsymbol{Y}] &= \mathrm{tr}(\mathbf{M}(\sigma^2\mathbf{I})) + (\mu\mathbf{1})^T\mathbf{M}(\mu\mathbf{1}) \\
&= \mathrm{tr}(\sigma^2\mathbf{M}\mathbf{I}) + \mu^2(\mathbf{1}^T\mathbf{M}\mathbf{1}) \\
&= \sigma^2\,\mathrm{tr}(\mathbf{M}) + \mu^2(\mathbf{1}^T\mathbf{M}\mathbf{1}) \quad \text{(since } \mathbf{M}\mathbf{I} = \mathbf{M} \text{ and trace is linear)}
\end{aligned}$$

Our task now boils down to calculating two key quantities: $\mathrm{tr}(\mathbf{M})$ and $\mathbf{1}^T\mathbf{M}\mathbf{1}$.

6 Calculating the Components

6.1 The Trace of the Centering Matrix: $\mathrm{tr}(\mathbf{M})$

We need the trace of $\mathbf{M} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$.

$$\begin{aligned}
\mathrm{tr}(\mathbf{M}) &= \mathrm{tr}\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \\
&= \mathrm{tr}(\mathbf{I}) - \mathrm{tr}\left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \quad \text{(Linearity of trace)} \\
&= n - \frac{1}{n}\,\mathrm{tr}(\mathbf{1}\mathbf{1}^T)
\end{aligned}$$

What is $\mathrm{tr}(\mathbf{1}\mathbf{1}^T)$? The matrix $\mathbf{1}\mathbf{1}^T$ is an $n \times n$ matrix where every entry is 1.

$$\mathbf{1}\mathbf{1}^T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

The trace is the sum of the diagonal elements, which are all 1s. There are n diagonal elements, so $\mathrm{tr}(\mathbf{1}\mathbf{1}^T) = n$. Alternatively, note that $\mathbf{P_1} = \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is the projection matrix onto the one-dimensional subspace spanned by $\mathbf{1}$. For any projection matrix, the trace equals its rank. The rank of $\mathbf{P_1}$ is 1. Thus $\mathrm{tr}(\mathbf{P_1}) = 1$, which means $\mathrm{tr}(\frac{1}{n}\mathbf{1}\mathbf{1}^T) = 1$.

Substituting back:

$$\mathrm{tr}(\mathbf{M}) = n - \frac{1}{n}(n) = n - 1$$

So, the trace of the centering matrix is $n-1$. This number might look familiar – it's exactly the divisor in our sample variance formula! It represents the degrees of freedom associated with the variance estimation after accounting for estimating the mean.

6.2 The Mean Term: $\mathbf{1}^T\mathbf{M}\mathbf{1}$

Now let's figure out the second part, $\mathbf{1}^T\mathbf{M}\mathbf{1}$. The key is to first calculate $\mathbf{M}\mathbf{1}$:

$$\begin{aligned}
\mathbf{M}\mathbf{1} &= \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{1} \\
&= \mathbf{I}\mathbf{1} - \frac{1}{n}\mathbf{1}(\mathbf{1}^T\mathbf{1}) \quad \text{(Matrix multiplication associativity)} \\
&= \mathbf{1} - \frac{1}{n}\mathbf{1}(n) \quad \text{(since } \mathbf{I}\mathbf{1} = \mathbf{1} \text{ and } \mathbf{1}^T\mathbf{1} = \textstyle\sum 1^2 = n) \\
&= \mathbf{1} - \mathbf{1} \\
&= \mathbf{0} \quad \text{(the } n \times 1 \text{ zero vector)}
\end{aligned}$$

This is a crucial property: the centering matrix $\mathbf{M}$ annihilates the vector $\mathbf{1}$ (and any vector proportional to it). This makes intuitive sense: if all data points were the same ($Y_i = c$), their mean would be c, and all deviations $(Y_i - \bar{Y})$ would be zero. $\mathbf{M}$ effectively removes the 'average level' component represented by $\mathbf{1}$.

Now, we can easily compute the quadratic form involving the mean:

$$\mathbf{1}^T\mathbf{M}\mathbf{1} = \mathbf{1}^T(\mathbf{M}\mathbf{1}) = \mathbf{1}^T\mathbf{0} = 0$$

7 Putting It All Together: The Expected Sum of Squares

Let's substitute our findings for $\mathrm{tr}(\mathbf{M})$ and $\mathbf{1}^T\mathbf{M}\mathbf{1}$ back into the expectation formula:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{Y}^T\mathbf{M}\boldsymbol{Y}] &= \sigma^2\,\mathrm{tr}(\mathbf{M}) + \mu^2(\mathbf{1}^T\mathbf{M}\mathbf{1}) \\
&= \sigma^2(n-1) + \mu^2(0) \\
&= \sigma^2(n-1)
\end{aligned}$$

So, we have found the expected value of the sum of squared deviations:

$$\mathbb{E}\left[\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] = \sigma^2(n-1)$$

8 Conclusion: S_n^2 is Unbiased!

We are just one step away. Recall that $(n-1)S_n^2 = \sum(Y_i - \bar{Y})^2 = \boldsymbol{Y}^T\mathbf{M}\boldsymbol{Y}$. Taking the expectation:

$$\mathbb{E}[(n-1)S_n^2] = \mathbb{E}[\boldsymbol{Y}^T\mathbf{M}\boldsymbol{Y}]$$

Using our result from the previous section:

$$\mathbb{E}[(n-1)S_n^2] = \sigma^2(n-1)$$

By the linearity property of expectation, we can pull the constant $(n-1)$ out:

$$(n-1)\mathbb{E}[S_n^2] = \sigma^2(n-1)$$

Assuming $n > 1$ (we need at least two data points to estimate variance), we can divide both sides by $(n-1)$:

$$\boxed{\mathbb{E}[S_n^2] = \sigma^2}$$

Corollary 1. *The sample variance* $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ *is an unbiased estimator for the population variance* σ^2*.*

This elegant result confirms that the inclusion of Bessel's correction $(n-1)$ in the denominator is precisely what's needed to ensure that, on average, our sample variance correctly estimates the true population variance. Without it, dividing by n would lead to an estimator that systematically underestimates σ^2.

Remark 2. *While we used the normality assumption ($\boldsymbol{Y} \sim N(\mu\mathbf{1}, \sigma^2\mathbf{I})$) to cleanly apply the standard theorem for the expectation of a quadratic form of a multivariate normal vector, the result that S_n^2 is an unbiased estimator for σ^2 holds more generally. It only requires that the Y_i are i.i.d. with finite mean μ and finite variance σ^2. The proof in that general case typically uses algebraic manipulation of the sums directly, rather than matrix forms.*