Why Sample Variance has $n-1$ on denominator?

Recall the population variance formula:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$$

Sample variance:

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

Rewrite $\mathbb{E}[\sum_{i=1}^{n}(X_i - \bar{X})^2]$ as:

$$\mathbb{E}[\sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2]$$

Expand the squared term:

$$\mathbb{E}[\sum_{i=1}^{n}((X_i - \mu)^2 + 2(X_i - \mu)(\mu - \bar{X}) + (\mu - \bar{X})^2)]$$

Simplify:

$$\mathbb{E}[\sum_{i=1}^{n}(X_i - \mu)^2] + \mathbb{E}[\sum_{i=1}^{n}2(X_i - \mu)(\mu - \bar{X})] + \mathbb{E}[\sum_{i=1}^{n}(\mu - \bar{X})^2]$$

The middle term is

$$\begin{aligned}\mathbb{E}[\sum_{i=1}^{n}2(X_i - \mu)(\mu - \bar{X})] &= 2\mathbb{E}[\sum_{i=1}^{n}(X_i - \mu)(\mu - \bar{X})]\\ &= 2\mathbb{E}[(n\bar{X} - n\mu)(\mu - \bar{X})]\\ &= 2n\mathbb{E}[(\mu - \bar{X})^2]\end{aligned}$$

So we have

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

We know that $\mathbb{E}[\sum_{i=1}^{n}(X_i-\mu)^2]=n\sigma^2$

$$\mathbb{E}[(\mu-\bar{X})^2]=\mathbb{E}[\mu^2-2\mu\bar{X}+\bar{X}^2]=\mathbb{E}[\bar{X}^2]-\mu^2$$

And

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[X^2-2X\mathbb{E}[X]+\mathbb{E}[X]^2\right] \\
&= \mathbb{E}[X^2]-2\mathbb{E}[X]\mathbb{E}[X]+\mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2]-2\mathbb{E}[X]^2+\mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2]-\mathbb{E}[X]^2
\end{aligned}$$

So

$$\mathbb{E}[\bar{X}^2]=\mathrm{Var}[\bar{X}]+\mathbb{E}[\bar{X}]^2$$

So

$$\begin{aligned}
\mathbb{E}[(\mu-\bar{X})^2] &= \mathbb{E}[\bar{X}^2]-\mu^2 \\
&= \mathrm{Var}[\bar{X}] \\
&= \mathrm{Var}[\frac{1}{n}\sum_{i=1}^{n}X_i] \\
&= \frac{1}{n^2}\mathrm{Var}[\sum_{i=1}^{n}X_i] \\
&= \frac{1}{n^2}\times n\times \mathrm{Var}[X_i] \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

So we have

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

If we use n in the denominator:

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

This is biased because it's less than the true population variance σ^2.

But if we use $(n-1)$ in the denominator:

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

This gives us an unbiased estimator of the population variance.