sampleVarianceProof

Alex Hahn

September 6, 2022

$$
\begin{aligned}
E[s^2_{biased}] &= E\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2\right] = E\left[\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\frac{1}{N}\sum_{j=1}^{N}x_j\right)^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[x_i^2-\frac{2}{N}x_i\sum_{j=1}^{N}x_j+\frac{1}{N^2}\sum_{j=1}^{N}x_j\sum_{k=1}^{N}x_k\right] \\
&= \frac{1}{N}\sum_{i=i}^{N}\left[\frac{N-2}{N}E[x_i^2]-\frac{2}{N}\sum_{j\neq i}E[x_ix_j]+\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k\neq j}E[x_jx_k]+\frac{1}{N^2}\sum_{j=1}^{N}E[x_j^2]\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{N-2}{N}(\sigma^2+\mu^2)=\frac{2}{N}(N-1)\mu^2+\frac{1}{N^2}N(N-1)\mu^2+\frac{1}{N}(sigma^2+\mu^2)\right] \\
&= \frac{N-1}{N}\sigma^2
\end{aligned}
\tag{1}
$$

By using the $N-1$ in the denominator instead of N, we eliminate the extra multiplicative term $\frac{N-1}{N}$

$$
E[s^2] = E\left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2\right] = \sigma^2 \tag{2}
$$

We can reason that the $N-1$ term in the demoninator servers to boost the numerical result enough to compensate for the defree of freedom (one degree used for the mean, rest for var)

Note that although this type of reasoning works for simple examples where the definition of the estimate of the sample mean and variance appear this is not the case with even slightly more complicated examples. Eg considet the unbiased estimator of the population standard deviation, it is not simply the square root of the unbiased variance estimator:

$$s \neq \frac{\sqrt{\frac{\sum_i^N (X_i - \mu)^2}{N-1}}}{c_4} \tag{3}$$

The formula for the correction factor c_4 is given by

$$c_4 = \sqrt{\frac{2}{N-1}} \frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N-1}{2})} \tag{4}$$

Where the gamma function $\Gamma(t)$ is given by

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx \tag{5}$$

where for positive integers N the gamma function becomes

$$\Gamma(N) = (N-1)! \tag{6}$$

This c_4 correction factor is a direct consequence of Jensen's Inequality (the secant line of a convex function lies above the fraph of the convex function) therefore allowing us to easily prove that the expectation of a convex function is greater than the function of the expectation: $E(f(x)) \geq f(E(x))$