Question) Is the estimation of the variance unbiased?

In statistics, we evaluate the 'goodness' of the estimation by checking if the estimation is 'unbiased'. By saying 'unbiased', it means the expectation of the estimator equals to the true value, e.g. if $E|\bar{x}| = \mu$ then the mean estimator is unbiased. Now we will show that the equation actually holds for mean estimator.

$$E|x| = E\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] = \frac{1}{N}\sum_{i=1}^{N} E|x|$$
$$= \frac{1}{N}.N.E|x|$$
$$= E|x| =$$

The first line makes use of the assumption that the samples are drawn from the true distribution, thus $E|x_i|$ is actually $E|x|$. From the proof above, it is shown that the mean estimator is unbiased.

Now we move to the variance estimator. At the first glance, the variance estimator $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$ should follow because mean estimator x is unbiased. However, it is not the case:

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

We know $\sum_{i=1}^{N} x_i = N.x$ and $\sum_{i=1}^{N} x^2 = N.x^2$. Plug these into derivation:

$$E\left|\sigma^2\right| = \frac{1}{N}E\left[\sum_{i=1}^{N} x_i^2 - 2.N.\bar{x}^2 + N.\bar{x}^2\right]$$
$$= \frac{1}{N}E\left[\sum_{i=1}^{N} x_i^2 - N.\bar{x}^2\right]$$
$$= \frac{1}{N}E\left[\sum_{i=1}^{N} x_i^2\right] - E\left[\bar{x}^2\right]$$
$$= E\left|x_i^2\right| - E|\bar{x}^2|$$

According to the alternative definition of variance, $\sigma_x^2 = E\left[x^2\right] - E[x]^2$, and similarly, $\sigma_{\bar{x}}^2 = E\left[\bar{x}^2\right] - E[\bar{x}]^2$, where the random variable is $\bar{x}$. Note that $E[x]=E[\bar{x}]=\mu$. Plug the 2 equations to the derivation:

$$E\left[\sigma^2\right] = \left(\sigma_x^2 + \mu^2\right) - \left(\sigma_{\bar{x}}^2 + \mu^2\right)$$
$$= \sigma_x^2 - \sigma_{\bar{x}}^2$$

$$\sigma_{\bar{x}}^{2} = VAR[\bar{x}] = VAR\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] = \frac{1}{N^2}VAR[]\sum_{i=1}^{N} x_i$$

Since, the samples are drawn,

$$VAR\left[\sum_{i=1}^{N} x_i\right] = \sum_{i=1}^{N} VAR[x] = N.VAR[x]$$

Thus,

$$\sigma_{\bar{x}}^{2} = \frac{1}{N}VAR[x] = \frac{1}{N}\sigma_x^{2}$$

Plug back to the $E[\sigma^2]$ derivation,

$$E\left[\sigma^2\right] = \frac{N-1}{N}\sigma_x^{2}$$

Therefore, $E[\sigma^2] \neq \sigma_x^{2}$ and it is shown that we tend to underestimate the variance. In order to overcome this biased problem, the maximum likelihood estimator of the variance can be slightly modified to take this into account.

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})^2$$

It is easy to show that this modified variance estimator is unbiased.