bias-variance-risk

May 22, 2021

1 The Bias-Variance Tradeoff in Estimating Risk

Disclaimer: I'm going to be a bit hand-wavey and maybe even wrong with notation and calculations in this one, because it's a work-in-progress.

Suppose you have a series of observations $(r_1, r_2, \ldots, r_T)$ where each observation is i.i.d. Normal with ground-truth mean μ and ground-truth variance σ_2.

1.1 Estimators

Consider three estimators for σ_2. We're implicitly going to consider Mean Squared Error (MSE) as our loss function when evaluating them, but MSE isn't necessarily the best one. Which loss function is most appropriate can depend on the setting and application. For example, maybe in your particular use case, underestimating σ^2 is more dangerous than overestimating.

1.1.1 Standard Bessel-Corrected Sample Variance Estimator

Define

$$s_A^2 := \frac{1}{T-1} \sum (r_t - \bar{r})^2.$$

This will be distributed as

$$\frac{1}{T-1} \sigma^2 \chi_{T-1}^2.$$

Its bias is 0, so its squared bias is also 0.

Its squared standard error is $\frac{1}{(T-1)^2} \sigma^4 2(T-1) = 2 \frac{1}{T-1} \sigma^4$.

The sum of its squared bias plus squared standard error is

$$2 \frac{1}{T-1} \sigma^4.$$

This has one undesirable property in the case I mentioned before: If underestimating σ^2 is more dangerous than overestimating. This estimator will grossly underestimate, for instance, if all the r's just randomly happen to come out to the same number.

1.1.2 Overriden Zero-Mean Sample Variance Estimator

Define
$$s_B^2 := \frac{1}{T} \sum r_t^2.$$

This will be distributed as
$$\mu^2 + \frac{1}{T} \sigma^2 \chi_T^2.$$

Its bias is μ^2, so its squared bias is μ^4.

Its squared standard error is $\frac{1}{T^2} \sigma^4 2T = 2\frac{1}{T} \sigma^4$.

The sum of its squared bias plus squared standard error is
$$\mu^4 + 2\frac{1}{T} \sigma^4.$$

1.1.3 Minimum-MSE Sample Variance Estimator

Define
$$s_C^2 := \frac{1}{T+1} \sum (r_t - \bar{r})^2.$$

This is the best you can do in terms of MSE [cf], but I'm not sure what its distribution is.

1.2 Upshot

Let's compare s_B^2 vs s_A^2. When will the overriden estimator's sum of squared bias plus squared standard error be better (i.e. smaller) than the standard's?

Well, when
$$\mu^4 + 2\frac{1}{T} \sigma^4 < 2\frac{1}{T-1} \sigma^4$$
$$\mu^4 < 2\left(\frac{1}{T-1} - \frac{1}{T}\right) \sigma^4$$
$$\mu^4 < 2\frac{1}{(T-1)T} \sigma^4$$
$$\mu < \sqrt[4]{2\frac{1}{(T-1)T}} \sigma$$
$$\frac{\mu}{\sigma} < \sqrt[4]{2\frac{1}{(T-1)T}}.$$

So for example if $T = 25$, then s_B^2 will have a lower sum-of-squared-bias-plus-squared-standard-error (TODO: is this equivalent to saying "lower MSE"?) than s_A^2 as long as the ratio of μ to σ is less than ≈ 0.24. Keep in mind that both μ and σ here are for a single observation.