

3. Why Unbiasedness?

GULP Winter 2015

January 27, 2015

1 Why we can not search in everything?

Remember we have defined a proper measure of the estimator performance which is risk function $R_\theta(\delta, g) = \mathbb{E}_\theta L(g(\theta), \delta(\mathbf{X}))$. Then naturally, we want an estimator which can perform really good under different θ . Or mathematically, a **universal minimizer of risk function**. That is, there exists δ^* such that

$$\forall \delta, R_\theta(\delta, g) \geq R_\theta(\delta^*, g), \forall \theta.$$

But is that possible? Let's look at a simple case.

Example(Silly guesser) Suppose we have a set of samples from normal distribution $N(\theta, 1)$ and we want to estimate θ . One wise guy uses the sample mean to estimate θ , and his competitor, a person with no statistics knowledge, claims $\theta = 5$, always.

Under the square loss, with simple calculation we will know that the risk for the first person is always $1/n$. While the risk for the second person is $(\theta - 5)^2$. This is fairly interesting. If the true parameter is 0, or 8, the first person is clearly better. But what if, in a parallel world, the θ is just 5? Then the guesser beat the wise statistician in the case.

This example shows a cruel but important fact. If our estimator δ yields $R_\theta(\delta, g) > 0$ for some $\theta = \theta_0$, a guesser betting on θ_0 will beat δ . Therefore such δ is not a universal minimizer of risk function. Therefore, the only possible universal risk minimizer will make the risk function zero, always.

This simply means we know $g(\theta)$, which is not possible.

2 Unbiasedness

So this is the reason we restrict the set of estimators. If we apply some regulations, and we find the universal risk minimizer in that subset of estimators, then we may feel good about it. Unbiased estimators are such a subset that is widely used.

The definition of unbiasedness is as follows. If $\delta(\mathbf{X})$ is unbiased, then

$$\mathbb{E}_\theta(\delta(\mathbf{X})) = g(\theta), \forall \theta.$$

Pay attention that the equation holds for all possible values of θ . For example, the estimator $\sum_i X_i / (n + 1)$ for the mean is unbiased when mean is zero, but biased otherwise. So this is not an unbiased estimator. Here are some examples of unbiased estimators.

For normal distribution $N(\theta, \sigma^2)$, the following estimator

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

is unbiased for (λ, σ^2) . It is a direct task to check this.

For Poisson distribution λ , both $\bar{X}$ and S^2 , defined as before, are unbiased estimator for λ . (Why?)

For an geometric distribution with only one observation, the unbiased estimator for the success rate p , is $\mathbb{1}(X = 0)$.

3 Bias-Variance Trade-off

Under squared loss, we have an important result:

$$R(\delta, g) = \mathbb{E}(\delta(X) - g(\theta))^2 = \mathbb{E}(\delta(X) - \mathbb{E} \delta(X))^2 + \mathbb{E} \delta(X) - g(\theta))^2 = \mathbb{E}(\delta(X) - \mathbb{E} \delta(X))^2 + (\mathbb{E} \delta(X) - g(\theta))^2$$

The first part is the variance of $\delta(X)$, and the second part is the square of the bias. This generally means that if you reduce the estimator's variance, you probably introduce more bias into the model.

For example, for normal model, the unbiased estimator is

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

and we have the MLE

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

Here MLE has smaller variance, but MLE is biased, so MLE trades the bias with smaller variance.