COST WP6: Description of performance statistics

Dorleta García

May 7, 2009

1 Introduction

Let A denote the true value of the population parameter to be estimated. Given some sample data the estimation methods implemented in the previous Work Packages of the COST project will yield some estimate of the population parameter. The simulation Work Package consists in simulating n different samples from the population, estimating the parameter of interest E_j in each of the j samples according to the different estimation methods. Then, following the approach by [1], the performance of the estimation methods is studied in terms of some statistics that measure bias, precision and accuracy.

The term bias refers to the difference between the expected value of the estimator and the true value. Therefore, it leads to an over- or under-estimate of the true value. Alternatively, precision is a measure of the statistical variance of an estimator. An estimator is said to be precise if it has a small variance or standard error, and unlike bias, is independent of the true value of the population parameter. Both properties are desirable to occur together. The concept of accuracy combines both, bias and precision, and it is defined as the overall distance between estimated values and the true value of the parameter. Figure 1 illustrates these concepts.

This document describes the main statistics used to evaluate and compare the performance of the different estimation methods. Some of them are scaled by the true parameter value. This implies that they can be used to compare performance across populations that do not have the same true parameter value.

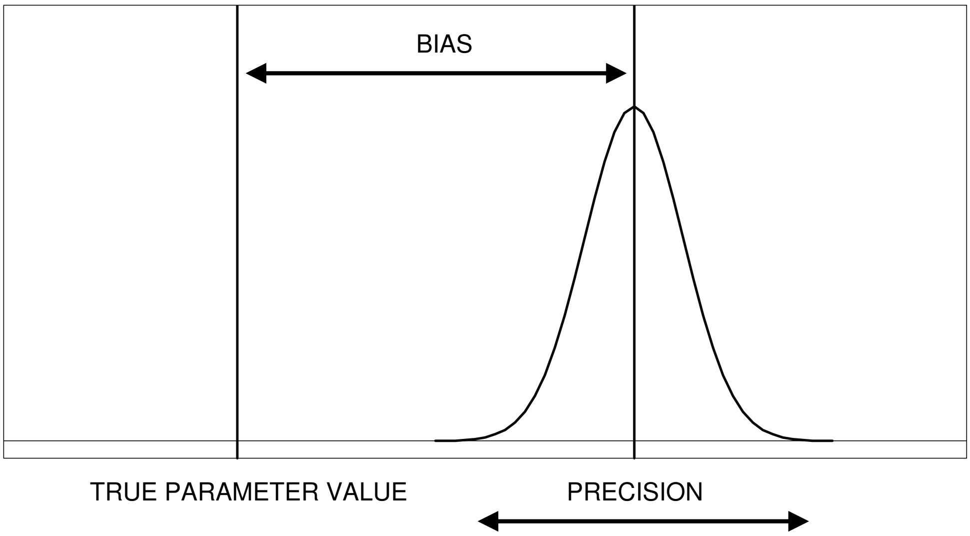

Figure 1: Illustration of bias and precision terms.

2 Performance statistics

2.1 Performance statistics of bias

Mean error

The mean error (ME) is the mean of all differences between the estimated values and the true value. It indicates whether the estimator consistently over- or underestimates the true parameter. It is calculated as follows:

$$\text{ME} = \frac{1}{n}\sum_{j=1}^{n}(E_j - A)\ . \tag{1}$$

Scaled mean error

The scaled mean error (SME) is the scaled counterpart of the mean error, that is, it is the mean error divided by the true population parameter:

$$\text{SME} = \frac{1}{An}\sum_{j=1}^{n}(E_j - A)\ . \tag{2}$$

Percent of actual parameter

The percent of actual parameter (PAR) is an scaled bias measure alternative to the SME. It is calculated as:

$$\mathrm{PAR} = \frac{1}{n}\sum_{j=1}^{n}(100E_j/A) \ , \tag{3}$$

and it is related with the SME according to:

$$\mathrm{PAR} = 100 + 100\mathrm{SME} \ . \tag{4}$$

Percentage of overestimates

The percentage of estimates that overestimates the true parameter (POE) is another unscaled bias statistic. It is calculated as:

$$\mathrm{POE} = \frac{1}{n}\sum_{j=1}^{n} I(E_j > A) \ , \tag{5}$$

where $I(\cdot)$ denotes an indicator function that takes the value 1 if the condition is fulfilled and 0 otherwise. If the estimator is unbiased, the percentage of overestimates should be close to 50%.

2.2 Performance measures of precision

Variance

The variance (Var) is the most common precision measure. In this case the unbiased estimate of variance is used:

$$\mathrm{Var} = \frac{1}{n-1}\sum_{j=1}^{n}(E_j - \bar{E})^2 \ . \tag{6}$$

Coefficient of variation

The coefficient of variation (CV) computed as the percentage of the standard deviation (square root of the variance) divided by the mean, is the most common scaled measure of precision:

$$\mathrm{CV} = 100\sqrt{\mathrm{Var}}/\bar{E} \ . \tag{7}$$

2.3 Performance measures of accuracy

Mean square error

The mean square error (MSE) is the mean of the squared differences between the sample estimates and the true value:

$$\mathrm{MSE} = \frac{1}{n} \sum_{j=1}^{n} (E_j - A)^2 \; . \tag{8}$$

In fact, this statistic is equal to the variance of the estimates plus the squared mean error.

Mean absolute error

The mean of the absolute value of the difference between the estimated and true population parameter is called the mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{n} \sum_{j=1}^{n} |E_j - A| \; . \tag{9}$$

Since replacing the squared difference in MSE by the absolute difference, this statistic avoids the potential problem of being dominated by outlying estimates far from the true value.

Scaled mean square error

The scaled mean square error (SMSE) is the scaled counterpart of the MSE and it is calculated by dividing the MSE by the squared true population parameter:

$$\mathrm{SMSE} = \frac{1}{A^2 n} \sum_{j=1}^{n} (E_j - A)^2 \; . \tag{10}$$

Scaled mean absolute error

The scaled mean absolute error (SMAE) is the scaled counterpart of the MAE and it is calculated by dividing the MAE by the true population parameter:

$$\mathrm{SMAE} = \frac{1}{An} \sum_{j=1}^{n} |E_j - A| \tag{11}$$

References

[1] Walther, B. and Moore, J. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography 28: 815-829.