

An Aid to Intuitively Understanding Bias and Consistency

Neil H. Spencer
University of Hertfordshire

Summary

Students often find the concepts of bias and consistency hard to understand intuitively. This paper presents means of demonstrating these concepts using a computer so that the students, once they have seen them in action, can more readily grasp their meanings and implications.

Keywords: Bias; Consistency.

Introduction and Scenario

The concepts of bias and consistency are ones that students often find hard to grasp in intuitive terms. They may well be able to show mathematically whether or not an estimator is biased and/or consistent, but the substantive meanings and implications of these properties may not be clear.

This paper presents a way of helping students to understand the concepts of bias and consistency using the random number generation and graphics capabilities available in a package such as Minitab.

Before showing how we demonstrate bias and consistency, we introduce the scenario that we use in this paper together with two estimators.

A well-respected publishing company has a policy that for aesthetic reasons, the final page of a series of books they publish must contain no more than a certain amount of blank space beneath the end of the text. The books in the series are all the same size. If a book is written which has a final page with too much blank space, a rewrite is undertaken to rectify the problem. To satisfy quality concerns, the rewrite has to meet certain standards with the result that the amount of blank space left after the rewrite is independent of the amount of blank space previously left.

The amount of blank space left (excluding a constant lower margin at the bottom of the page) thus follows a uniform distribution with minimum of zero and maximum set by the company's policy. As the company uses different sized fonts for text in the series of books, the measurement of the amount of blank space is made in millimetres and the uniform distribution is therefore continuous.

A rival company wishes to copy this policy of the well-respected company and thus needs to obtain an estimate of the maximum amount of blank space allowable. They aim to do this by taking a sample of n books from the series and measuring the amount of blank space on their final pages (they can ascertain the size of the lower margin from elsewhere in the book and exclude it). They consider two estimators of the maximum of the continuous uniform distribution with minimum zero, using X_{\max} to represent the largest amount of blank space measured.

$$\text{Estimator 1} = X_{\max} \quad \text{Estimator 2} = \left(\frac{n+1}{n}\right)X_{\max}$$

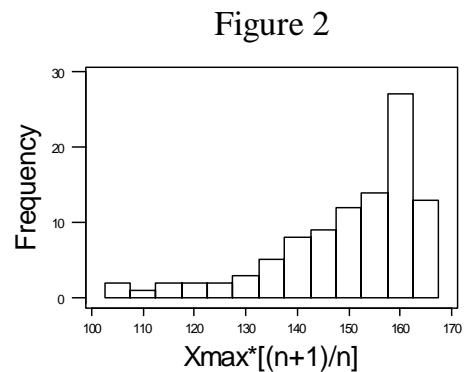
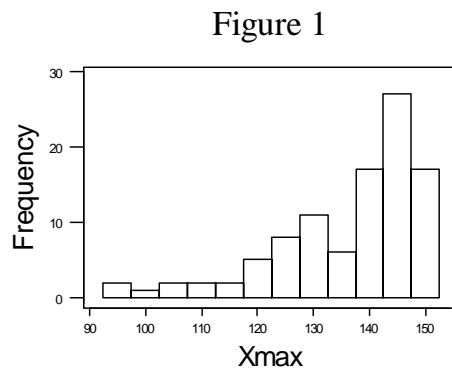
Other scenarios and estimators can be used in conjunction with the methods used to demonstrate bias and consistency illustrated here, but the advantages of using estimators of the maximum value of a continuous uniform distribution as an example are threefold. The first advantage is that the students are unlikely to have any strong preconceived ideas about the suggested estimators although there is likely to be a preference for the more natural looking estimator, X_{\max} . The second advantage is that this more natural looking estimator is actually the one that is biased and it is the more complicated estimator that is unbiased. Finally the third advantage is that the biased estimator is, in fact, consistent and can be considered adequate as an estimator of the maximum if the sample size is large enough.

Demonstration of Bias

To demonstrate the properties of the two suggested estimators in relation to bias, we firstly use Minitab to simulate 100 datasets, each of 10 values (representing 10 books), from the uniform distribution with minimum 0 and maximum 150 (thus defining our unknown maximum to be 150), asking for 100 rows of data to be stored in ten columns of the worksheet (menu commands Calc, Random Data, Uniform... in Minitab).

We then calculate the values of the estimator X_{\max} for each of our 100 rows and store the results in another column of the worksheet (menu commands Calc, Row Statistics..., statistic “maximum” in Minitab). We calculate the values of the other estimator from the values of X_{\max} that have just been obtained (Calc, Calculator... in Minitab).

Histograms of the resulting columns of estimators (e.g. figures 1 and 2) may be used to investigate the biased nature of the estimator X_{\max} and the unbiased nature of the other estimator. The calculation of the mean for each estimator reinforces what may be observable from the histograms (here the mean for the estimator X_{\max} is 136.04 and for the other estimator is 149.65, obtained using Stat, Basic Statistics, Descriptive Statistics... in Minitab).



Demonstration of Consistency

For the demonstration of consistency, the students generate individual random samples of varying sizes from a continuous uniform distribution with minimum 1 and maximum 150. For each sample they then calculate the value of each estimator of the maximum and its bias. The results are tabulated (in such a manner as in table 1) and it may be observed that although the estimator X_{\max} is biased, as the sample size increases, its bias decreases until it is virtually indistinguishable from the other estimator. It is inevitable that on occasions, the simulated data produced for this demonstration will not yield the nice patterns of bias displayed in table 1. It may thus be advisable to aggregate the calculated biases from all the students in the class and produce an overall summary table.

Table 1

| Sample Size | X_{\max} | | $\left(\frac{n+1}{n}\right) X_{\max}$ | |
|-------------|------------|----------|---------------------------------------|---------|
| | Value | Bias | Value | Bias |
| n = 10 | 134.825 | - 15.175 | 148.308 | - 1.692 |
| n = 50 | 148.097 | - 1.903 | 151.059 | + 1.059 |
| n = 100 | 149.614 | - 0.386 | 151.110 | + 1.110 |
| n = 500 | 149.601 | - 0.399 | 149.900 | - 0.100 |

Conclusions and Experiences of Use

The demonstrations described above provide a straightforward means to aid students in their understanding of the concepts of bias and consistency. They have been designed for use with the Minitab package, exploiting its easily comprehensible menu structures and dialog boxes, and work with the most commonly available versions of the package. The simplicity of the menus and dialog boxes needed to generate the simulated datasets, perform the necessary calculations and create histograms in Minitab means that the students do not lose sight of the concepts of bias and consistency amid the necessary computing. The demonstrations can also be carried out without major difficulty in SPSS for Windows and the Microsoft Excel spreadsheet package although the simulation of the datasets is slightly more complicated than in Minitab, and obtaining the histograms causes additional difficulty in Excel.

The techniques used here to demonstrate bias and consistency in action have been used in two different courses, one with Applied Statistics students and one with students studying for Business related degrees. For both groups, the main beneficiaries of the demonstrations were not the most able students (who could understand the concepts of bias and consistency well without the demonstrations) but were the rest of the student body whose understanding of the concepts

was enhanced. The demonstrations also provided a welcome diversion from the mathematical problems of ascertaining whether or not estimators are unbiased and/or consistent.

Acknowledgement

The author would like to thank an anonymous referee and the editor for helpful comments on this paper.