The problem with Sturges' rule for constructing histograms

Rob J Hyndman[1]

5 July 1995

Abstract Most statistical packages use Sturges' rule (or an extension of it) for selecting the number of classes when constructing a histogram. Sturges' rule is also widely recommended in introductory statistics textbooks. It is known that Sturges' rule leads to oversmoothed histograms, but Sturges' derivation of his rule has never been questioned. In this note, I point out that the argument leading to Sturges' rule is wrong.

Key words: Histogram, bin-width selection, statistical computer packages.

Herbert Sturges (1926) considered an idealised frequency histogram with k bins where the ith bin count is the binomial coefficient $\binom{k-1}{i}$, $i = 0, 1, \ldots, k-1$. As k increases, this ideal frequency histogram approaches the shape of a normal density. The total sample size is

$$n = \sum_{i=0}^{k-1} \binom{k-1}{i} = (1+1)^{k-1} = 2^{k-1}$$

by the binomial expansion. So the number of classes to choose when constructing a histogram from normal data is

$$k = 1 + \log_2 n.$$

This is Sturges' rule. If the data are not normal, additional classes may be required; Doane (1976) proposed a modification to Sturges' rule to allow for skewness. It is known that Sturges' rule and Doane's rule lead to oversmoothed histograms, especially for large samples (see, for example, Scott, 1992). But the derivation of Sturges' rule does not seem to have been questioned, and has been reproduced by several authors including both Doane (1976) and Scott (1992).

The problem with the above argument is that any multiple of the binomial coefficients could have been used and the frequency histogram would still approach the normal density. So, any number of classes could be obtained, depending on the multiple used.

In the derivation above, Sturges is implicitly using a binomial distribution to approximate an underlying normal distribution. That is, if the underlying density is normal and is appropriately scaled to have mean $(k-1)/2$ and variance $(k-1)/4$, then it can be approximated by a binomial distribution $B(k-1, 0.5)$ for large k. Then the histogram classes correspond to the discrete values of the binomial distribution. The expected class frequencies are $n\binom{k-1}{i}(0.5)^{k-1}$. So if $n = 2^{k-1}$ we obtain Sturges' idealised histogram. But any other value of n is equally valid and does not lead to Sturges' rule.

---

[1]Rob Hyndman is Senior Lecturer, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria, Australia, 3168.

Alternative rules for constructing histograms include Scott's (1979) rule for the class width:

$$h = 3.5sn^{-1/3}$$

and Freedman and Diaconis's (1981) rule for the class width:

$$h = 2(\text{IQ})n^{-1/3}$$

where s is the sample standard deviation and IQ is the sample interquartile range. Either of these are just as simple to use as Sturges' rule, but are well-founded in statistical theory. Wand (1995) has recently extended Scott's rule to give consistent estimators of the underlying density.

Sturges' rule has probably survived as long as it has because, for moderate n (less than 200), it gives similar results to the alternative rules above (see Scott, 1992, p.56), and so produces reasonable histograms. However, it does not work for large n.

The problem with Sturges' rule is that its derivation is wrong. It is a rule which no longer deserves a place in statistics textbooks or as a default in statistical computer packages.

References

DOANE, D.P. (1976) Aesthetic frequency classification. *American Statistician*, **30**, 181–183.

FREEDMAN, D. and DIACONIS, P. (1981) On this histogram as a density estimator: L_2 theory. *Zeit. Wahr. ver. Geb.*, **57**, 453–476.

SCOTT, D.W. (1979) On optimal and data-based histograms. *Biometrika*, **66**, 605–610.

SCOTT, D.W. (1992) *Multivariate density estimation: theory, practice, and visualization*, John Wiley & Sons: New York.

STURGES, H. (1926) The choice of a class-interval. *J. Amer. Statist. Assoc.*, **21**, 65–66.

WAND, M.P. (1995) Data-based choice of histogram bin-width. Technical report, Australian Graduate School of Management, University of NSW.