Benford's Law and background field errors in data assimilation

R S Seaman
Bureau of Meteorology Research Centre
Melbourne, Australia

Benford's Law is a logarithmic distribution of first significant digits that is observed in many naturally occurring non-meteorological data sets, such as catchment areas of rivers, stock market prices and tax returns. In such data sets, the frequency of occurrence (f) of first digit (d) is given by

$$f = \log (1 + 1/d), d = 1,2 .. 9$$

where the logarithm is base 10. That is, the first significant digit 1 occurs with a frequency of about 30%, while 9 occurs with a frequency of less than 5%. Benford's Law has been used as a tool for the detection of accounting and taxation fraud.

The present application is motivated by a recent proof by Hill (1996) that Benford's Law can be considered as the asymptotic result of combining random data from random distributions. The goodness of fit to Benford's Law may therefore indicate the extent of mixing in a data set.

Background field errors in most data assimilation systems are assumed to be normally distributed. Ensembles of background field errors from the Australian Bureau of Meteorology's data assimilation system have been used to calculate (i) the goodness of fit of the errors to a Gaussian distribution, and (ii) the goodness of fit of the first significant digits of the errors to Benford's Law. A quasi-inverse relation was found between these two goodness of fit measures. This tends to support the suggestion (Ingleby and Lorenc 1992, Devenyi and Schlatter 1994) that background field errors in general may be considered as mixtures of Gaussian distributions. Confirmation is provided by statistical simulation, and by studies of individual ensembles.

References

Devenyi D and Schlatter T W. 1994. Statistical properties of three hour prediction "errors" derived from the mesoscale prediction system. Mon. Weath. Rev., 122, 1263-80.

Hill T P. 1996. A statistical derivation of the significant digit law. Statistical Science, 10, 354-63.

Ingleby N B and Lorenc A C. 1992. Forecast errors and buddy checks. Preprints. 12th conference in probability and statistics in the atmospheric sciences. Toronto, Amer. Met. Soc., 207-13.

Seaman R S. 2002. The relevance of Benford's Law to background field errors in data assimilation. Aust. Met. Mag., 51, 25-33.