[3] N. H. McCoy, The Theory of Rings. Macmillan: New York, 1964.

D. MacHale,
Department of Mathematics,
University College,
Cork.

# A RE-ANALYSIS OF BESSEL'S ERROR DATA

A. Kinsella

## Introduction

The Gaussian (Normal) probability model

$$f(x;\mu,\sigma) = \frac{\exp(-(x-\mu)^2/2\sigma^2)}{\sigma(2\pi)^{1/2}}$$

is, arguably, the most widely used probability model because of

1. the fact that it is found as a limiting form of other common probability models;

2. the operation of the Central Limit Theorem which gives rise to the Gaussian form;

3. the intuitive appeal of the model as a description of measurement errors in that it postulates that, in the long run, measurements will zone in on the "true but unknown" quantity of interest, $\mu$, and will be close to this value, lying between $(\mu-\sigma)$ and $(\mu+\sigma)$ some 68% of the time;

4. the mathematical tractability of linear and quadratic functions of Gaussian random variables which are used in Student's $t$ and $F$ ratio tests;

5. the ability of the model to readily change location and shape because of the independence of $\mu$, the location parameter, and $\sigma$, the shape parameter.

A simple transformation of the random variable, namely,

$$y = |x|$$

gives rise to the Folded Normal probability model

$$g(y;\mu,\sigma) = \frac{\exp(-(y-\mu)^2/2\sigma^2) + \exp(+(y-\mu)^2/2\sigma^2)}{\sigma(2\pi)^{1/2}}$$

which can arise when empirical observational data are recorded without regard to the sign. If the signed data are postulated to conform to the Gaussian probability model, the unsigned data will conform to the Folded Normal model. An example of the possible use of this model is provided by the data set given by Topping, [1], in his discussion of "Normal error distributions". The data, which are displayed in the first two columns of Table 1, are the "much discussed example of observational data satisfying the normal error law (which) was given by Bessel". This data set tabulates "the errors involved in measuring the right ascension of stars.", the magnitude of the error of observation, in seconds, being shown in the first column of Table 1 with the corresponding frequency in the second column. In his subsequent analysis of this data set Topping assumes that since "positive and negative errors are grouped together, ... so we can only assume that they are equally divided". This assumption is the basis of the subsequent analysis in terms of a Gaussian model. This note analyses the data set on the assumption that the Folded Normal model is appropriate.

**Table 1: Right Ascension Error Data Set**

| Limits of Error | Observed Frequency ($n$) | Predicted Frequency ($e$) | Pearson Residual ($r$) |
|---|---|---|---|
| 0.0 – 0.1 | 114 | 102.1 | +1.174 |
| 0.1 – 0.2 | 84 | 84.4 | −0.041 |
| 0.2 – 0.3 | 53 | 57.6 | −0.005 |
| 0.3 – 0.4 | 24 | 32.5 | −1.487 |
| 0.4 – 0.5 | 14 | 15.1 | −0.290 |
| 0.5 – 0.6 | 6 | 5.8 | +0.074 |
| 0.6 – 0.7 | 3 | 1.9 | +0.845 |
| 0.7 – 0.8 | 1 | 0.5 | +0.737 |
| 0.8 – 0.9 | 1 | 0.1 | +2.756 |



### Parameter Estimation

The Maximum Likelihood method of parameter estimation was used to extract numerical estimates of the unknown parameters, $\mu$ and $\sigma$, from the data set. In general, if the data set is a random sample of size $n$, denoted by $(x_1, x_2, \ldots, x_n)$, the Likelihood Function is

$$L(\theta) = f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta),$$

where $f(x;\theta)$ denotes the probability model of interest and $\theta$ denotes the set of parameters of the model. Since the data are, or will be, known, the Likelihood Function is a function of the elements of the set of parameters, namely, $\theta$, which are continuous *nonrandom* variables. The Maximum Likelihood Estimator(s) of the parameter(s) are the value(s) which maximize the Likelihood Function, namely, the "most likely" value(s) which can be found using the data set. In simpler cases exact functions of the data are found to be the Maximum Likelihood Estimators, these being the solutions(s) to the equation(s)

$$\frac{\delta L(\theta)}{\delta\theta} = 0 \text{ or } \frac{\delta \mathrm{Log}(L(\theta))}{\delta\theta} = 0.$$

A complication arises in the case of the data set in Table 1 because the observational data, the error, is censored in that the number of occasions on which an error lies within an interval of length 0.1 seconds is recorded rather than the actual value of the error. This means that the Likelihood Function has to be rewritten as

$$L(\mu,\sigma) = \frac{N!}{n_1!n_2!\ldots n_9!}(F_1)^{n_1}(F_2)^{n_2}\ldots(F_9)^{n_9}$$

which is a multinomial probability model. In this Likelihood Function, $N$ denotes the total number of observations,

$$N = \sum_{i=1}^{9} n_i = 300,$$

$$n_1 = 114, n_2 = 84, \ldots, n_9 = 1,$$

and $F_i$ denotes the integral of the Folded Normal probability model over the $i^{th}$ interval,

$$F_1 = \int_{0.0}^{0.1} g(y; \mu, \sigma) dy, \ldots, F_9 = \int_{0.8}^{0.9} g(y; \mu, \sigma) dy.$$

These integrals give the probability that a randomly chosen observation will fall in any given interval.

Because of the necessity of numerically integrating the Folded Normal probability model, it is necessary to use a search method to find the values of $\mu$ and $\sigma$ which maximize the Likelihood Function, the Maximum Likelihood estimates. A simple "trial and error" search of the two dimensional parameter space was used in this case. The values of the negative of the natural logarithm of the Likelihood Function, excluding the constant factor involving $N!$ and $n_i!$, are shown for a wide grid of values in Table 2.

**Table 2: Logarithm of Likelihood Function**

| $\mu/\sigma$ | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 |
|---|---|---|---|---|---|---|---|
| −0.2 | 566.2 | 492.4 | 487.7 | 499.6 | 517.3 | 537.5 | 558.8 |
| −0.1 | 615.4 | 489.4 | 466.3 | 473.5 | 492.1 | 515.1 | 539.3 |
| 0.0 | 805.3 | 529.9 | 470.2 | 468.0 | 484.7 | 507.9 | 533.0 |
| 0.1 | 615.4 | 489.4 | 466.3 | 473.5 | 492.1 | 515.1 | 539.3 |
| 0.2 | 566.2 | 492.4 | 487.7 | 499.6 | 517.3 | 537.5 | 558.8 |

This function is chosen because the Maximum Likelihood estimate will minimize its value so that, in general, a function minimization algorithm can be used to obtain the required values. The use of a more refined grid in the region

$$-0.1 < \mu < +0.1, \quad 0.20 < \sigma < 0.30$$

indicated that the logarithm of the Likelihood Function was more sensitive to changes in $\sigma$ than in $\mu$. The final values which were chosen were $\mu = 0.0$ and $\sigma = 0.227$. The changes in the value of the logarithm of the Likelihood Function were of the order of 0.001 for corresponding changes in the magnitudes of $\mu$ and $\sigma$.



### Model Evaluation

The "goodness of fit" of the Folded Normal probability model to the data set was judged by Pearson residual, [2, pp.37-39], which is defined as

$$r_i = (n_i - e_i)/(e_i)^{1/2},$$

where $e_i$ is the expected frequency in the $i$-th censoring interval, being equal to

$$N \int g(y; \hat{\mu}, \hat{\sigma})\, dy,$$

the integral being over the appropriate interval. Here $\hat{\mu}$ and $\hat{\sigma}$ denote the Maximum Likelihood estimates. The square of the Pearson residual is the the $i$-th component of the familiar Chi-squared test statistic

$$X^2 = \sum_{i=1}^{9} (n_i - e_i)^2 / e_i$$

and is useful in indicating which components of the overall test are making the largest contributions. On the basis that the Pearson residual has, approximately, a Standard Gaussian distribution ($\mu = 0$, $\sigma = 1$) one value of $r_i$, namely, +2.76, is sufficiently large to warrant some attention. This arises because the expected frequency is approximately one tenth of the observed frequency but since this apparent problem arises in a low frequency tail it is of no practical significance.

On the basis of this analysis the claim that the error has a Gaussian distribution would appear to be vindicated in view of the connection of that probability model with the Folded Normal Probability model.

### References

[1] J. Topping, Errors of Observation and their Treatment. Chapman and Hall: London, 1971.

[2] P. McCullagh and J. A. Nelder, Generalized Linear Models, 2nd ed. Chapman and Hall: London, 1990.

A. Kinsella,
Department of Mathematics, Statistics and Computer Science
Dublin Institute of Technology,
Kevin Street,
Dublin 8.

# SOME MATHEMATICAL ASPECTS OF INFORMATION TECHNOLOGY: FIXED POINTS AND THE FORMAL SEMANTICS OF PROGRAMMING LANGUAGES

Anthony Karel Seda

## 1. Introduction

> "It is reasonable to hope that the relationship between computation and mathematical logic will be as fruitful in the next century as that between analysis and physics in the last. The development of this relationship demands a concern for both applications and for mathematical elegance." John McCarthy[1], 1967.

In describing Information Technology, the Web page of the recently formed Information Technology Centre at University College, Galway says this: "During the past decade Information Technology (IT) has transformed business life, from the boardroom to the shopfloor. As we generally understand it, Information Technology is an outgrowth from the computer, microelectronics, and telecommunications industries, and now comprises: computer processors and data storage devices, telecommunications, software, microprocessors, automation technologies and user interface media."

Generally speaking, users of IT need not be expert in, nor even familiar with, the technologies which support it. If this is true, it is even more true that these same users need have no knowledge of the *theory* which supports the technologies which support IT. Nevertheless, the issue of the theories underlying IT and, in particular, which areas of mathematics are important in

[1]Inventor of the programming language Lisp and pioneer of AI.