DATA ANALYSIS – A DIALOGUE WITH THE DATA

D. S. SIVIA
Rutherford Appleton Laboratory,
Chilton, OX11 0QX, England
E-mail: dss@isise.rl.ac.uk

A modern Bayesian physicist, Steve Gull from Cambridge, described data analysis as simply being 'a dialogue with the data'. This paper aims to illustrate this viewpoint with the aid of a simple example: *Peelle's pertinent puzzle*.

1. Introduction

The training in data analysis that most of us are given as undergraduates consists of being taught a collection of disjointed statistical recipes. This is generally unsatisfactory because the prescriptions appear *ad hoc* by lacking a unifying rationale. While the various tests might individually seem sensible at an intuitive level, the underlying assumptions and approximations are not obvious. It is far from clear, therefore, exactly what question is being addressed by their use.

Although attempts to give guidelines on 'best practice' are laudable, the shortcomings above will not be remedied without a programme of education on the fundamental principles of data analysis. To this end, scientists and engineers are increasingly finding that the Bayesian approach to probability theory advocated by mathematical physicists such as Laplace[1], Jeffreys[2] and Jaynes[3] provides the most suitable framework. This viewpoint is outlined in Section 2, and its use illustrated with an analysis and resolution of Peelle's pertinent puzzle[a] in Sections 3 and 4 respectively; we conclude with Section 5.

2. Bayesian Probability Theory

The origins of the Bayesian approach to probability theory dates back over three hundred years, to people such as the Bernoullis, Bayes and Laplace,

[a]Posed by Robert Peelle, from the Oak Ridge National Laboratory, Tennessee, in 1987.

and was developed as a tool for reasoning in situations where it is not possible to argue with certainty. This subject is relevant to all of us because it pertains to what we have to do everyday of our lives, both professionally and generally: namely, make inferences based on incomplete and/or unreliable data. In this context, a probability is seen as representing a *degree of belief*, or a *state of knowledge*, about something given the information available. For example, the probability of rain in the afternoon, given that there are dark clouds in the morning, is denoted by a number between zero and one, where the two extremes correspond to certainty about the outcome. Since the assessment of rain could easily be very different with additional access to the current weather maps, it means that probabilities are always *conditional* and that the associated information, assumptions and approximations must be stated clearly.

2.1. *Manipulating Probabilities*

In addition to the convention that probabilities should lie between 0 and 1, there are just two basic rules that they must satisfy:

$$\Pr(X\,|\,I) + \Pr\big(\overline{X}\,|\,I\big) = 1\,, \tag{1}$$

$$\Pr(X, Y\,|\,I) = \Pr(X|Y, I) \times \Pr(Y\,|\,I)\,. \tag{2}$$

Here X and Y are two specific propositions, $\overline{X}$ denotes that X is false, the vertical bar '$|$' means 'given' (so that all items to the right of this conditioning symbol are taken as being true) and the comma is read as the conjunction 'and'; I subsumes all the pertinent background information, assumptions and approximations. Equations (1) and (2), known as the *sum* and *product* rule respectively, are the same as those found in orthodox or conventional statistics; this later school of thought differs from the Bayesian one in its interpretation of probability, restricting it to apply only to frequencies, which limits its sphere of direct application.

Many other relationships can be derived from Eqs. (1) and (2). Among the most useful are:

$$\Pr(X\,|\,Y, I) = \frac{\Pr(Y\,|\,X, I) \times \Pr(X\,|\,I)}{\Pr(Y\,|\,I)}\,, \tag{3}$$

$$\Pr(X\,|\,I) = \Pr(X, Y\,|\,I) + \Pr\big(X, \overline{Y}\,|\,I\big)\,. \tag{4}$$

Equation (3) is called *Bayes' theorem*. Its power lies in the fact that it turns things around with respect to the conditioning symbol: it relates $\Pr(X\,|\,Y, I)$

to $\Pr(Y|X,I)$. Equation (4) is the simplest form of *marginalisation*. Its generalisations provide procedures for dealing with *nuisance* parameters and hypothesis uncertainties.

2.2. *Assigning Probabilities*

While Eqs. (1) and (2), and their corollaries, specify how probabilities are to be manipulated, the rules for their assignment are less well defined. This is inevitable to some extent as 'states of knowledge' can take a myriad different forms, often rather vague. Nevertheless, there are some simple but powerful ideas on the issue based on arguments of self-consistency: if two people have the same information then they should assign the same probability. We refer the reader to some recent textbooks for a good discussion of this topic, and for examples of Bayesian analyses in general: Jaynes[3], Sivia[4], MacKay[5] and Gregory[6].

3. Peelle's Pertinent Puzzle

In 1987, Robert Peelle, from the Oak Ridge National Laboratory, posed the following simple problem as a way of highlighting an anomalous result from a standard *least-squares* analysis that is sometimes encountered by the nuclear data community[b]:

> "Suppose we are required to obtain the weighted average of two experimental results for the same physical quantity. The first result is 1.5 and the second result is 1.0 . The full covariance matrix of these data is believed to be the sum of three components. The first component is fully correlated with standard error of 20% of each respective value. The second and third components are independent of the first and of each other, and correspond to 10% random uncertainties in each experimental result.
>
> The weighted average obtained from the least-squares method is 0.88 ± 0.22, a value outside the range of the input values! Under what conditions is this the reasonable result that we sought to achieve by use of an advanced data reduction technique?"

[b] Oh and Seo[7] quote this from a secondary source, Chiba and Smith[8].

3.1. *The Least-Squares Approximation*

Let us begin by reviewing the Bayesian justification for least-squares. Recasting the problem in symbolic terms, we wish to infer the value of a quantity μ given two measurements $\mathbf{x} = \{x_1, x_2\}$ and related *covariance* information I_1:

$$\left\langle (x_1-\mu)^2 \right\rangle = \sigma_1^2\,, \quad \left\langle (x_2-\mu)^2 \right\rangle = \sigma_2^2 \quad \text{and} \quad \left\langle (x_1-\mu)\,(x_2-\mu) \right\rangle = \epsilon\,\sigma_1\sigma_2\,, \tag{5}$$

where the angled brackets denote *expectation* values and the coefficient of *correlation*, ϵ, is in the range $-1 \leq \epsilon \leq 1$. This means that we need to ascertain the conditional probability $\Pr(\mu|\mathbf{x}, I_1)$, since it encapsulates our state of knowledge about μ given the relevant data. Bayes' theorem allows us to relate this probability distribution function (pdf) to others that are easier to assign:

$$\Pr(\mu|\mathbf{x}, I_1) \;\propto\; \Pr(\mathbf{x}|\mu, I_1) \times \Pr(\mu|I_1)\;, \tag{6}$$

where the equality has been replaced by a proportionality due to the omission of $\Pr(\mathbf{x}|I_1)$ in the denominator, which simply acts as a *normalisation* constant here. Armed only with the covariance information in Eq. (5), the principle of *maximum entropy*[3] (MaxEnt) leads us to assign a *Gaussian likelihood* function:

$$\Pr(\mathbf{x}|\mu, I_1) = \frac{\mathrm{e}^{-Q_1/2}}{2\pi\sigma_1\sigma_2\sqrt{1-\epsilon^2}}\;, \tag{7}$$

$$\text{where} \quad Q_1 = \begin{pmatrix} x_1-\mu & x_2-\mu \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \epsilon\,\sigma_1\sigma_2 \\ \epsilon\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} x_1-\mu \\ x_2-\mu \end{pmatrix}. \tag{8}$$

If we also assign a uniform *prior* for μ over a suitably large range, to naïvely represent gross initial ignorance,

$$\Pr(\mu|I_1) = \begin{cases} (\mu_{\max}-\mu_{\min})^{-1} & \text{for } \mu_{\min} \leq \mu \leq \mu_{\max} \\ 0 & \text{otherwise} \end{cases}\;, \tag{9}$$

then the logarithm of the *posterior* pdf being sought, $\mathcal{L}_1$, becomes

$$\mathcal{L}_1 = \ln\big[\Pr(\mu|\mathbf{x}, I_1)\big] = \text{constant} - \frac{Q_1}{2} \tag{10}$$

for $\mu_{\min} \leq \mu \leq \mu_{\max}$, and $-\infty$ otherwise. Since Eq. (10) tells us that $\mathcal{L}_1$ is largest when Q_1 is smallest, our 'best' estimate μ_0 is given by that value of μ which minimises the quadratic scalar mismatch of Eq. (8) — this is the least-squares solution.

A *Taylor* series expansion of $\mathcal{L}_1$ shows that $\Pr(\mu|\mathbf{x}, I_1)$ is Gaussian, with the mean and variance parameters, μ_0 and σ^2, being defined by

$$\left.\frac{\mathrm{d}Q_1}{\mathrm{d}\mu}\right|_{\mu_0} = 0 \qquad \text{and} \qquad \frac{\mathrm{d}^2Q_1}{\mathrm{d}\mu^2} = \frac{2}{\sigma^2} \; . \tag{11}$$

Hence, our inference about μ can be summarised by a best estimate and an associated error-bar, $\mu_0 \pm \sigma$, in the standard way:

$$\mu \; = \; \frac{\alpha\, x_1 + \beta\, x_2}{\alpha + \beta} \; \pm \; \sigma_1\sigma_2 \sqrt{\frac{1-\epsilon^2}{\alpha+\beta}} \; , \tag{12}$$

$$\text{where} \quad \alpha = \sigma_2\,(\sigma_2 - \epsilon\,\sigma_1) \quad \text{and} \quad \beta = \sigma_1\,(\sigma_1 - \epsilon\,\sigma_2) \; . \tag{13}$$

For Peelle's pertinent puzzle, $x_1 = 1.5$, $x_2 = 1.0$, $\sigma_1 = 0.3354$, $\sigma_2 = 0.2236$ and $\epsilon = 0.8$; this leads to the conclusion that $\mu = 0.88 \pm 0.22$.

3.2. *Understanding the Puzzle*

The result of the above analysis is anomalous, because it's at odds with our expectation that the best estimate should be bounded by the two measurements. Although seemingly weird, is it unacceptable?

From Eq. (12), it's not too difficult to see that μ_0 will lie between x_1 and x_2 as long as both α and β are positive. Equation (13) translates this into the requirement that

$$\epsilon \; \leq \; \min\left\{\frac{\sigma_1}{\sigma_2}\,, \frac{\sigma_2}{\sigma_1}\right\} \; . \tag{14}$$

This will always be satisfied for negative correlations, but will fail as $\epsilon \to 1$ when $\sigma_1 \neq \sigma_2$. Upon reflection, this is not surprising.

Independence, or $\epsilon = 0$, indicates that x_2 is no more likely to be higher or lower than the true value of μ no matter what the corresponding deviation of x_1; and *vice versa*. It seems reasonable that μ_0 should then lie between x_1 and x_2, since this minimises the (sum of squared *residuals*) mismatch with the measurements. There is an additional reason for this outcome when $\epsilon < 0$, as there is also the expectation that $x_1 - \mu$ and $x_2 - \mu$ have opposite signs. By the same token, there is increasing pressure for μ_0 to be outside the range spanned by the data as $\epsilon \to 1$ so as to satisfy the growing expectation from the positive correlation that both measurements deviate from the true value in the same sense. With $\epsilon = 0.8$ in Peelle's case, a best estimate of 0.88 for μ doesn't now seem quite so ludicrous.

4. Peelle's Pertinent Ambiguity

Although we can understand the reason for the anomalous result in Peelle's pertinent puzzle, there is a curious feature in the statement of the problem: the covariance elements are given in relative, rather than absolute, terms. Is this significant?

4.1. *Least-Squares for Magnitude Data*

The least-squares analysis of the preceding section relied, for the most part, on the Gaussian likelihood of Eqs. (7) and (8). This assignment was based on the constraints of Eq. (5), and motivated by the principle of MaxEnt. A more literal interpretation of the covariance information in Peelle's statement, however, would be

$$\left\langle \left(\frac{\delta x_1}{x_1}\right)^2 \right\rangle = s_1^2 \,, \quad \left\langle \left(\frac{\delta x_2}{x_2}\right)^2 \right\rangle = s_2^2 \quad \text{and} \quad \left\langle \left(\frac{\delta x_1}{x_1}\right)\left(\frac{\delta x_2}{x_2}\right) \right\rangle = \epsilon\, s_1 s_2 \;, \quad (15)$$

where δx_1 and δx_2 are the deviations of the measurements from the true value of μ; the fractional error-bars are equal, with $s_1 = s_2 = 0.2236$, but the correlation coefficient remains unchanged ($\epsilon = 0.8$). These constraints can be turned into the simpler form of Eq. (5) through the substitution of $y_1 = \ln x_1$ and $y_2 = \ln x_2$, so that

$$\left\langle \delta y_1^2 \right\rangle = s_1^2 \,, \quad \left\langle \delta y_2^2 \right\rangle = s_2^2 \quad \text{and} \quad \left\langle \delta y_1 \delta y_2 \right\rangle = \epsilon\, s_1 s_2 \;. \quad (16)$$

The MaxEnt principle would then lead us to assign a Gaussian likelihood for the logarithm of the data,

$$\Pr(\ln \mathbf{x} | \ln \mu, I_2) = \frac{\mathrm{e}^{-Q_2/2}}{2\pi s_1 s_2 \sqrt{1-\epsilon^2}} \;, \quad (17)$$

$$\text{where} \quad Q_2 = \begin{pmatrix} \ln[x_1/\mu] & \ln[x_2/\mu] \end{pmatrix} \begin{pmatrix} s_1^2 & \epsilon\, s_1 s_2 \\ \epsilon\, s_1 s_2 & s_2^2 \end{pmatrix}^{-1} \begin{pmatrix} \ln[x_1/\mu] \\ \ln[x_2/\mu] \end{pmatrix} \;, \quad (18)$$

where I_2 denotes the covariance information in Eq. (15).

The above discussion suggests that μ is a *scale* parameter, or something that is positive and pertains to a magnitude. As such, the prior that expresses gross initial ignorance[3] is

$$\Pr(\ln \mu | I_2) = \begin{cases} \left(\ln[\mu_{\max}/\mu_{\min}]\right)^{-1} & \text{for } \mu_{\min} \le \mu \le \mu_{\max} \\ 0 & \text{otherwise} \end{cases} \;. \quad (19)$$

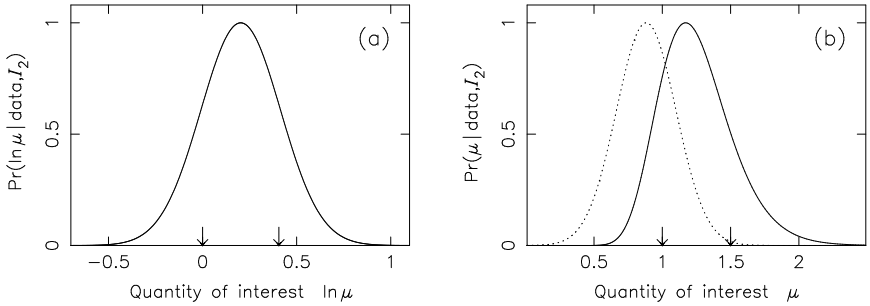

Figure 1. The posterior pdfs of Eqs. (10) and (21), with Peelle's measurements marked by arrows, which have been scaled vertically to have a maximum value of unity to aid comparison. (a) The posterior pdf of Eq. (21), $\Pr(\ln\mu|\ln\mathbf{x}, I_2)$, which is a Gaussian. (b) The same pdf transformed to linear μ, $\Pr(\mu|\ln\mathbf{x}, I_2)$, where it is non-Gaussian; the posterior pdf of Eq. (10), $\Pr(\mu|\mathbf{x}, I_1)$, is plotted as a dotted line.

Using Bayes' theorem,

$$\Pr(\ln\mu|\ln\mathbf{x}, I_2) \propto \Pr(\ln\mathbf{x}|\ln\mu, I_2) \times \Pr(\ln\mu|I_2) \;, \tag{20}$$

where we have again omitted the denominator $\Pr(\ln\mathbf{x}|I_2)$, we find that the logarithm of the posterior pdf for $\ln\mu$ is

$$\mathcal{L}_2 = \ln\big[\Pr(\ln\mu|\ln\mathbf{x}, I_2)\big] = \text{constant} - \frac{Q_2}{2} \tag{21}$$

for $\mu_{\min} \le \mu \le \mu_{\max}$, and $-\infty$ otherwise. Thus $\Pr(\ln\mu|\ln\mathbf{x}, I_2)$ is also a Gaussian pdf which, for the case of equal relative error-bars $s_1 = s_2$, can be succinctly summarised by

$$\ln\mu = \ln\sqrt{x_1 x_2} \pm s_1\sqrt{\frac{1+\epsilon}{2}} \;. \tag{22}$$

The substitution of Peelle's data yields $\ln\mu = 0.20 \pm 0.21$ or, through a standard (linearised) propagation of errors[4], $\mu \approx 1.22 \pm 0.26$. The posterior pdfs of Eqs. (10) and (21) are shown graphically in Fig. 1.

4.2. *Looking at the Evidence*

The above two analyses of Peelle's data give noticeably different optimal estimates of μ, although there is a substantial degree of overlap between them. This should not be too surprising as each is predicated on a different set of assumptions, I_1 and I_2, corresponding to alternative interpretations of the information provided. Hanson *et al.*[9] correctly point out that the real solution to this problem rests with the experimentalists giving more

details on the nature of the uncertainties in the measurements. Whatever the response, probability theory provides a formal mechanism for dealing with such ambiguities; it is based on marginalisation.

If I represents the statement of Peelle's puzzle, and any other information pertinent to it, then our inference about the value of μ is encapsulated by $\Pr(\mu|\mathbf{x}, I)$. This can be related to analyses based on alternative interpretations of the data, I_1, I_2, ..., I_M, by

$$\Pr(\mu|\mathbf{x}, I) = \sum_{j=1}^{M} \Pr(\mu, I_j|\mathbf{x}, I) \ , \tag{23}$$

which is a generalisation of Eq. (4). Using the product rule of probability and Bayes' theorem, each term in the summation becomes

$$\Pr(\mu, I_j|\mathbf{x}, I) = \Pr(\mu|\mathbf{x}, I_j) \times \frac{\Pr(\mathbf{x}|I_j) \times \Pr(I_j|I)}{\Pr(\mathbf{x}|I)} \ , \tag{24}$$

where the conditioning on I has been dropped, as being unnecessary, when I_j is given. Since $\Pr(\mathbf{x}|I)$ does not depend on μ or j, it can be treated as a normalisation constant. Without a prior indication of the 'correct' interpretation of the data, when all the $\Pr(I_j|I)$ can be set equal, Eq. (23) simplifies to

$$\Pr(\mu|\mathbf{x}, I) \propto \sum_{j=1}^{M} \Pr(\mu|\mathbf{x}, I_j) \times \Pr(\mathbf{x}|I_j) \ . \tag{25}$$

This is an average of the alternative analyses weighted by the *evidence* of the data, $\Pr(\mathbf{x}|I_j)$. The latter, which is also known as the *global* or *marginal likelihood*, or the *prior predictive*, is simply the denominator term that is usually omitted in applications of Bayes' theorem as an uninteresting normalisation constant:

$$\Pr(\mu|\mathbf{x}, I_j) = \frac{\Pr(\mathbf{x}|\mu, I_j) \times \Pr(\mu|I_j)}{\Pr(\mathbf{x}|I_j)} \ . \tag{26}$$

Using the assignments of Eqs. (7) and (9), the evidence for I_1 is given by

$$\Pr(\mathbf{x}|I_1) = \int \Pr(\mathbf{x}, \mu|I_1) \, \mathrm{d}\mu = \frac{1}{\mu_{\max} - \mu_{\min}} \int_{\mu_{\min}}^{\mu_{\max}} \frac{\mathrm{e}^{-Q_1/2} \, \mathrm{d}\mu}{2\pi\sigma_1\sigma_2\sqrt{1-\epsilon^2}} \ . \tag{27}$$

The dependence of the analysis on $\mu_{\min}$ and $\mu_{\max}$ might seem surprising, but that is because their exact values tend to be irrelevant for the more familiar problem of parameter estimation: the posterior pdf $\Pr(\mu|\mathbf{x}, I_j)$ is independent of the bounds as long as they cover a sufficiently large μ-range

to encompass all the significant region of the likelihood function $\Pr(\mathbf{x}|\mu, I_j)$. For the assignments of Eqs. (17) and (19), the corresponding evidence is best evaluated in log-space:

$$\begin{aligned}\Pr(\mathbf{x}|I_2) &= \frac{\Pr(\ln \mathbf{x}|I_2)}{x_1 x_2} \\ &= \frac{1}{x_1 x_2 \ln[\mu_{\max}/\mu_{\min}]} \int_{\ln \mu_{\min}}^{\ln \mu_{\max}} \frac{\mathrm{e}^{-Q_2/2}\, \mathrm{d}\ln\mu}{2\pi s_1 s_2 \sqrt{1-\epsilon^2}} \,, \end{aligned} \tag{28}$$

where the $x_1 x_2$ in the denominator is the *Jacobian* for the transformation from $\Pr(\ln \mathbf{x}|I_2)$ to $\Pr(\mathbf{x}|I_2)$. It should be noted that $\mu_{\min}$ and $\mu_{\max}$ do not have to have the same values in Eqs. (27) and (28): these bounds must be positive in Eq. (28), in keeping with the scale parameter view of μ implied by I_2, whereas they are free from this restriction in Eq. (27).

Carrying out the evidence-weighted averaging of Eq. (25) for $M = 2$, with $\mu_{\min}$ and $\mu_{\max}$ set somewhat arbitrarily to 0.1 and 3.0 in both Eqs. (27) and (28), we obtain the marginal posterior pdf for Peelle's problem shown in Fig. 2; it has a mean of 0.96, a standard deviation of 0.27, a maximum at 0.91 and is asymmetric with a tail towards higher μ. Although the precise result necessarily depends on the μ-bounds chosen, it does so fairly weakly. The essential conclusion is that a value of μ between 1.5 and 2.0, which is on the upper-side of the larger measurement, cannot be excluded with such high certainty if the possibility of I_2 is admitted (in addition to I_1).

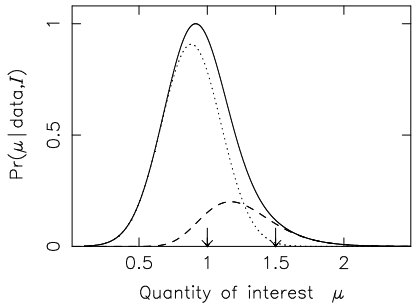

Figure 2. The marginal posterior pdf of Eq. (25), $\Pr(\mu|\mathbf{x}, I)$, for $M = 2$. The evidence-weighted contributions from the two alternative interpretations of the data considered, $\Pr(\mu|\mathbf{x}, I_1)$ and $\Pr(\mu|\mathbf{x}, I_2)$, are shown with a dotted and dashed lines; $\mu_{\min}$ and $\mu_{\max}$ were taken to be 0.1 and 3.0 in both cases. Peelle's measurements are marked by arrows and, to aid comparison with Fig. 1, all the pdfs have been scaled vertically so that $\Pr(\mu|\mathbf{x}, I)$ has a maximum value of unity.

5. Conclusions

We have used Peelle's pertinent puzzle as a simple example to illustrate how the analysis of data is a dynamic process akin to holding a conversation. When the initial least-squares analysis of Section 3.1 led to results that seemed 'wrong', we reacted by looking more carefully at the validity of the assumptions that underlie that procedure. This prompted us to formulate a different question, addressed in Section 4.1, defined by an alternative interpretation of the information provided. In the absence of experimental details regarding the nature of the uncertainties associated with the given measurements, we again turned to probability theory to ask, in Section 4.2, what we could conclude in face of the ambiguity.

To avoid any confusion, let us clarify further a few points regarding what we have done in this analysis of Peelle's pertinent puzzle and about our Bayesian viewpoint in general.

We have not said that the least-squares analysis was wrong. Indeed, in Section 3.2, we have explained why the counter-intuitive result could actually be quite reasonable. We simply asked a series of questions, defined by alternative assumptions, and addressed them through probability theory — it was just a dialogue with the data.

The Bayesian viewpoint expounded here follows the approach of mathematical physicists such as Laplace[1], Jeffreys[2] and Jaynes[3], and is still not widely taught to science and engineering undergraduates today. It differs markedly in its accessibility for scientists from the works of many statisticians engaged in the Bayesian field; the latter carry over much of the vocabulary and mind-set of their classical frequentist training, which we believe to be neither necessary nor helpful. We refer the reader to some recent textbooks, such as Jaynes[3], Sivia[4], MacKay[5] and Gregory[6], for a good introduction to our viewpoint.

To conclude, a black-box approach to the subject of data analysis, even with useful guidelines, is best avoided because it can be both limiting and misleading. All analyses are conditional on assumptions and approximations, and it's important to understand and state them clearly. While the evaluation of an arithmetic mean might seem objective and incontrovertible, for example, its status as a crucial number requires some qualified justification. We believe that an understanding of the principles underlying data analysis, along the lines outlined here, is at least as important as formal prescriptions of best practice.

Acknowledgments

I am grateful to Soo-youl Oh for bringing this fun little problem to my attention, and to Stephen Gull and David Waymont for giving me useful feedback on my analysis of it.

References

1. P. S. de Laplace, *Théorie analytique des probabilités*, Courcier Imprimeur, Paris (1812).
2. H. Jeffreys, *Theory of probability*, Clarendon Press, Oxford (1939).
3. E. T. Jaynes, *Probability theory: the logic of science*, edited by G. L. Bretthorst, Cambridge University Press, Cambridge (2003).
4. D. S. Sivia, *Data analysis – a Bayesian tutorial*, Oxford University Press, Oxford (1996).
5. D. J. C. MacKay *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge (2003).
6. P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press, Cambridge (2005).
7. S.-Y. Oh and C.-G. Seo, *PHYSOR 2004*, American Nuclear Society, Lagrange Park, Illinois (2004).
8. S. Chiba and D. L. Smith, *ANL/NDM-121*, Argonne National Laboratory, Chicago (1991).
9. K. Hanson, T. Kawano and P. Talou, *AIP Conf. Proc.* **769**, 304-307.