

# Avoiding Zero and Infinity in Sample Based Algorithms

N.A.Thacker

Last updated  
24 / 06 / 2010

This document forms part of the **Statistics and Segmentation Series (2008-001)**  
available from [www.tina-vision.net](http://www.tina-vision.net).

- 2007-008 Tutorial: Defining Probability for Science.
- 2001-007 Performance Characterisation in Computer Vision:  
The Role of Statistics in Testing and Design.
- 2002-007 The Effects of an Arcsin Square Root Transform on a Binomial Distributed Quantity.
- 2001-010 The Effects of a Square Root Transform on a Poisson Distributed Quantity.
- 2004-004 Shannon Entropy, Renyi Entropy, and Information.
- 2002-002 Validating MRI Field Homogeneity Correction Using Image Information Measures.
- 2004-001 Empirical Validation of Covariance Estimates for Mutual Information Coregistration.
- 2004-005 The Equal Variance Domain: Issues Surrounding the Use of Probability Densities in  
Algorithm Design.
- 2009-008 Avoiding Zero and Infinity in Sample Based Algorithms.
- 2001-008 Derivation of the Renormalisation Formula for the Product of Uniform Probability  
Distributions and Extension to Non-Integer Dimensionality.
- 2001-005 Model Selection and Convergence of the EM Algorithm.
- 2003-007 Noise Filtering and Testing for MR Using a Multi-Dimensional Partial Volume Model.
- 2002-004 A Novel Method for Non-Parametric Image Subtraction:  
Identification of Enhancing Lesions in Multiple Sclerosis from MR Images.
- 2001-014 Bayesian and Non-Bayesian Probabilistic Models for Image Analysis.
- 1997-001 The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data.
- 1999-001 The Bhattacharyya Measure requires no Bias Correction.
- 1999-004 B-Fitting: An Estimation Technique With Automatic Parameter Selection.
- 2005-008 Tutorial: Beyond Likelihood.



Imaging Science and Biomedical Engineering Division,  
Medical School, University of Manchester,  
Stopford Building, Oxford Road,  
Manchester, M13 9PT.

# Avoiding Zero and Infinity in Sample Based Algorithms

N.A.Thacker  
Imaging Science and Biomedical Engineering Division  
Medical School, University of Manchester  
Manchester, M13 9PT, UK

## Abstract

This document derives estimators of frequency and probability for the elimination of numerical instability when processing small samples. The analysis uses the quantitative methodology motivated in previous documents [6, 7]. In particular we show that the expectation of the mean of a Poisson sample is  $n + 1/2$ , and for a Binomial sample is  $(n + 1/2)/(N + 1)$ . We argue that this is a better estimator of the underlying generator than conventional Likelihood in these cases. Some of the consequences of these results are illustrated on the ‘reference class problem’<sup>12</sup>.

## Introduction

If we believe that the basis for all quantitative analysis is probability theory, then we should find that a well constructed theory, which takes proper account of the uncertainty in our data, should result in algorithms free from catastrophic failure. This should mean that the algorithm avoids statistical instabilities. As the effects of noise are generally much larger than numerical problems we should also have few problems with numerical implementation. However, even a small amount of experience developing analysis software based upon probability expressions will soon show that infinities are often generated when sampling data. The problem of computing a division by zero, from a noisy measurement has been discussed elsewhere [5]. Here we restrict ourselves to the problems associated with counting, and the associated theory.

For a finite sample of  $n$  from  $N$  trials, if we use  $\mu = n/N$  as an estimate of probability, we generate infinities whenever we need to take the log (ie : for  $n = 0$  we have  $\log 0 = -\infty$ ). A common example where the need to compute logarithms of frequencies occurs is in attempts to estimate entropies of the form

$$E = \sum_i p_i \log(p_i)$$

using data samples to estimate  $p_i$ . The calculation is protected from generating infinities for  $p_i$  estimates of 0, as  $0 \log(0) = 0$ , but generalisations of this expression, such as a bootstrap Likelihood<sup>3</sup> for use in Expectation Maximisation, will generate data samples at probabilities of 0 during optimisation of the Maximisation step. The standard technique for preventing this from having a catastrophic effect on this and other algorithms is called the ‘principle of maximum ignorance’, and we set  $\mu = (n + 1)/N$ .

Given that such modifications seem common place in software when implementing what look like reasonable probability estimates one might begin to feel that there is something wrong with the idea that we can legitimately take a frequentist view of probability. The reference class problem has been cited as a reason to abandon quantitative use of probability altogether and simply accept the subjective nature of all analysis. In [4] we find the following: *In the end, even a strict frequentist position involves subjective analysis... The reference class problem illustrates the intrusion of subjectivity. Suppose that a frequentist doctor wants to know the chances that a patient has a particular disease. The doctor wants to consider other patients who are similar in important ways - age, symptoms, perhaps sex - and see what proportion of them had the disease. But if the doctor considered everything that is known about the patient - weight to the nearest gram, hair colour, mother’s maiden name, etc. - the result would be that there are no other patients who are exactly the same and thus no reference class from which to collect experimental data. This has been a vexing problem in the philosophy of science.*

Such statements are enough to convince us that the presentation of probability in many standard texts may have missed something of fundamental importance. This is the issue of how to develop a self-consistent methodology

<sup>1</sup>I have no intention to publish this at this time as I believe someone must have figured this out a century ago. I would be interested to hear from anyone who thinks otherwise, or can provide a relevant reference.

<sup>2</sup>Thanks to Jamie Gilmore for pointing out that this is related to Laplace’s rule of succession, see below

<sup>3</sup>By which I mean using the data sample itself to obtain a non-parametric estimate the probability density from simple frequency ratios, rather than fitting a parametric form via bootstrap resampling.

for the analysis of finite samples of data in the real world. This document aims to explain the origins of these problems. We show that for each there is a fully justifiable approach, often similar to the empirical approaches adopted, which is consistent with quantitative use of probability. Like previous documents in this series, we will see the key is to remember that probability notation is strictly defined for boolean events, its use with continuous variables therefore requires care. In particular we need to make a distinction between probabilities  $P()$  and densities  $p()$ .

## Estimating Mean Frequency from Poisson Samples

When we observe 0 samples of a particular event, we know that this does not mean that the event will never occur. Intuitively, we appreciate that we have only put a limit on this frequency which is set by how often we looked for it. Von Mises goes further, and states that because a frequentist probability is defined in the limit of a large number of samples, a probability of 0 does not imply that something can never happen <sup>4</sup>.

For a Poisson distribution, the probability of observing  $n$  samples when the generator of the distribution has mean  $\mu$  is

$$P(n) = \frac{\exp(-\mu)\mu^n}{n!}$$

We therefore can see that there are more ways of generating  $n = 0$  than only  $\mu = 0$ . What we are seeking is a way of turning this observation into a quantitative expression of probability.

What we would really like to know is, ‘having observed  $n$ , what was the most likely value of  $\mu$ ?’. One obvious approach is to use  $P(n)$  to define a density distribution over  $\mu$  and use this to make a Likelihood estimate. If we try this we find we get no further forward, as the most likely generator for 0 in this basis is still only  $\mu = 0$ .

I believe the thing that may have gone wrong here is an example of the theoretical analysis presented in [7]. Maximum likelihood is only useful as an estimate in as much as it summarises the likelihood distribution. It works best when we can combine it with a parameter covariance as a summary of the full distribution. For the current task this distribution is skewed, not symmetrical. We therefore need a different summary variable, and given the definition of the problem the appropriate choice is the expectation of the mean.

$$\langle \mu \rangle = \frac{\int_0^\infty \mu p(\mu|n) d\mu}{\int_0^\infty p(\mu|n) d\mu}$$

Here we take the view that invariance under parameter redefinition, combined with the use of continuous variables allows us to make the association that Likelihood is proportional to the probability of the parameters lying within their error propagated uncertainty

$$\int_{\mu_e - \kappa\sigma_{\mu_e}}^{\mu_e + \kappa\sigma_{\mu_e}} p(\mu|n) d\mu$$

So that as  $\kappa \rightarrow 0$  we can write our Likelihood as

$$L(\mu_e, n) \propto P(n|\mu_e) \propto \sigma_{\mu_e} p(\mu_e|n)$$

where  $p(\mu|n)$  is the density and  $\sigma_\mu$  is the accuracy of the estimate, used to define the integral limits which relate probability to probability density, (ie: Fisher Information).

For the case of a Poisson distribution the appropriate interval is proportional to  $\sqrt{\mu}$  so that

$$p(\mu|n) \propto \exp(-\mu)\mu^n / \sqrt{\mu}$$

Notice, that we could (if we wished) relate this additional term to a Bayesian ‘prior’<sup>5</sup>. Instead, we say that the prior here is uniform,  $\sqrt{\mu}$  is the term needed so that we apply probability notation consistently when considering continuous variables, an assertion which will be discussed in the end notes.

Using the standard integral

$$\int_0^\infty e^{-x} x^{n-1/2} dx = 1.3.5\dots(2n-1)\sqrt{\pi}/2^{n+1}$$

<sup>4</sup>I would disagree with this and say instead that any honest estimate of probability made using a finite set of samples cannot imply 0 probability. It seems perfectly reasonable to me to reserve probabilities of both 0 and 1 for those circumstances when something will **never** or **always** happen. Otherwise we cannot use probability to encompass the normal rules of logic. I agree (as we will see below) otherwise with the implication that a frequentist definition of probability precludes definitive statements.

<sup>5</sup>Mathematically it is related to a Jeffreys prior. The relationship between these priors and equal variance transforms has been noted before [3], though the consequences do not seem to have been generally appreciated.

we have

$$\langle \mu \rangle = (2(n + 1) - 1)/2 = n + 1/2$$

which is very similar to the “principle of maximum ignorance”.

## Estimating Mean Ratio from Binomial Samples

For many algorithms, we are not interested just in the number, but in the proportion such as  $n/N$ . Here we cannot simply use  $\langle \mu \rangle / N$ , as now we have a restriction on the behaviour of the estimate around  $\mu = 1$ . For example, as with the previous case, we know that a sample of  $N$  from  $N$  only puts an upper limit on the most likely ratio, it does not guarantee that we will continue to see data with 100 % probability. Most statisticians generally say that an observation of  $N$  can only be interpreted as a ratio of approximately  $(N - 1)/N$ .

In this case we must estimate the most likely proportion using the Binomial distribution rather than a Poisson as the data generator. ie:

$$P(n|\mu, N) = \frac{N!}{n!(N - n)!} \mu^n (1 - \mu)^{N - n}$$

and the corresponding interval term is  $\sigma \propto (\mu - \mu^2)^{1/2} / \sqrt{N}$ .

The expectation of  $\mu$  is given by

$$\langle \mu \rangle = \frac{\int_0^1 \mu^{n+1/2} (1 - \mu)^{N - n - 1/2} dx}{\int_0^1 \mu^{n-1/2} (1 - \mu)^{N - n - 1/2} dx}$$

Making the substitution  $\mu = \sin^2(\theta)$

$$\int_0^1 \mu^{n+1/2} (1 - \mu)^{N - n - 1/2} dx = 2 \int_0^{\pi/2} \sin^{2n+2} \cos^{2N-2n} d\theta$$

Which is in the form of the standard integral

$$\int_0^{\pi/2} \sin^p \theta \cos^q \theta d\theta = \frac{p-1}{p+q} \int_0^{\pi/2} \sin^{p-2} \theta \cos^q \theta d\theta$$

Therefore,

$$\langle \mu \rangle = \frac{(2n + 2) - 1}{(2n + 2) + (2N - 2n)} = \frac{n + 1/2}{N + 1}$$

which differs from the conventional Likelihood result of  $n/N$  [1, 9]. Our result will again have approximately the same effect on algorithm behaviour as the “principle of maximum ignorance”, and now also restricts the maximum ratio to  $(N + 1/2)/(N + 1)$ .

The “rule of succession” by Laplace used the expectation of the Binomial distribution (without converting first to the density) to obtain a value of  $\frac{n+1}{N+2}$ , though it has been pointed out by several mathematicians that this is based on inconsistent assumptions. In particular it is hard to justify the use of a uniform prior and the use of the Likelihood function as though it is a density. The Likelihood is invariant to parameter re-transformation and a density should not be. It is my opinion that the theory here solves these problems.

We can now say that the sample entropy of a data can be written as

$$E = \sum_i \frac{n_i}{N} \log\left(\frac{n_i + 1/2}{N + 1}\right)$$

which happens to define the term for  $n_i = 0$  as 0, as in the conventional approaches (other  $n_i \neq 0$  terms differ slightly), and also prevents exact zero’s even for bootstrapped Likelihoods.

## Generalisation to Noisy Measurements

The above analysis assumes that we can obtain unambiguous integer values from a counting process. However, many counting processes already have a problem which prevents this. If we are comparing an integer value  $d$  to a threshold  $t$ , what do we conclude for  $t = d$ ? In the real world the problem is further complicated by the presence of noise on real valued measurements.

We can only conclude that for  $t = d$  our total count must be incremented by  $1/2$ . For noisy data the logical approach which is consistent with this idea is to integrate the noise distribution  $p(x|d)$  below the threshold. In this way the soft rank of  $d_j$  can be defined as

$$\sum_i^N \int_{-\infty}^{t=d_j} p(x|d_i) dx$$

However, a bootstrap Likelihood for arbitrary values of  $t$  computed on this basis would allow probabilities with values outside the  $(1/2)/N$  and  $(N - 1/2)/N$  limits derived above. We solve this problem by observing that the hypothesis we are testing can be defined such that the value  $t$  is itself an observation assumed to be from the required class, ie:

$$s_t = \sum_i^N \int_{-\infty}^t p(x|d_i) dx + \int_{-\infty}^t p(x|t) dx$$

where the second term is by definition  $1/2$ .<sup>6</sup> We can see that not only is this in accord with counting for a Poisson process, but also normalisation of  $s_t$  to the total density distribution regenerates the minimum and maximum values defined for the Binomial process. The values obtained for some notable values of  $t$  are given in Table 1.

data	Bootstrap Likelihood
$t \ll d_{min}$	$(1/2)/(N + 1)$
$t = d_{min}$	$1/(N + 1)$
$t = d_{max}$	$N/(N + 1)$
$t \gg d_{max}$	$(N + 1/2)/(N + 1)$

Table 1: Characteristic values for Likelihoods estimated from  $N$  noisy data samples with maximal and minimal values of  $d_{min}$  and  $d_{max}$ , using a soft rank.

This process will generate constant limits at the extremes of the distribution over  $t$ , similar to the approach taken to Likelihood estimation in robust statistics, so eliminating statistical and numerical problems. An example of use can be found in [8].

Those interested in computing confidence limits for histograms of noisy real variables can do so by applying error propagation to transform the effects of measurement error into uncertainty in histogram frequency. This will generate terms proportional to the derivative of the histogram as suggested in [9]. This approach will be quantitatively valid for homogenous measurement errors on individual samples (and across the histogram). Further, by adopting the method used here, for quantitative use of probability via interval construction, use of an ad-hoc Bayesian prior is unnecessary.

## The Reference Class Problem

For a binary set of symptom labels estimating the probability for a particular outcome is again an example of a Binomial sampling process. The probability that the test subject is from the selected class is then  $(n + 1/2)/(N + 1)$  in accordance with the previous section, where  $N$  is the total number of cases with the diagnosis  $n$  is the number with the true diagnosis. Note that for  $n = 0$ , as in the reference class problem, the estimated probability based upon this data is 0.5, ie: there is no information, (as expected). Obtaining the best group of symptoms from the data-base will not generate an empty set provided we choose them partly on the basis of this quantity. Though this solves the specific problem identified, there is more to practical solution of this task than getting this right.

The reference class problem, as stated in the introduction, implies that if we could define the probability of observing a particular group of symptoms in a patient, then this would automatically be the best thing to use as an aid to diagnosis. This is however, not the case. In fact the problem is an example of discriminant analysis. Over the set of all possible symptom permutations there are corresponding samples of various sizes with more or less discriminant capabilities. Also, depending upon the data, we may need to combine the information in more sophisticated ways than simply taking the ‘and’ of a subset of features as a reference class. Logically the most general form would be to take ‘or’ combinations of ‘and’ s, but even the choice of a simple reference class requires

<sup>6</sup>This constraint also specifies the way that we must parameterise non-symmetric noise distributions, ie: the position as defined by the centre of mass.

us to take account of specificity and sensitivity. That is, we must decide on a way of trading off the consequences of missing someone who has the disease in comparison to mis-diagnosis of someone who hasn't. Whatever we pick, (eg: resulting deaths or financial cost), we can quantitatively predict the likely consequence of any decision. On the basis of an additive assumption of total risk (which seems a valid assumption in these cases), the process can be visualised as a series of parallel risk lines on an ROC plot. For data close to the extreme corners of these plots the methods derived for counting here avoid over-interpretation of data (figure 1).

We can see that the reference class is only a part of the larger problem of optimal use of data for prediction. In accordance with the notation of probability we could construct any boolean expression of variables for use in diagnosis. We are not restricted solely to the identification of a subset of factors.

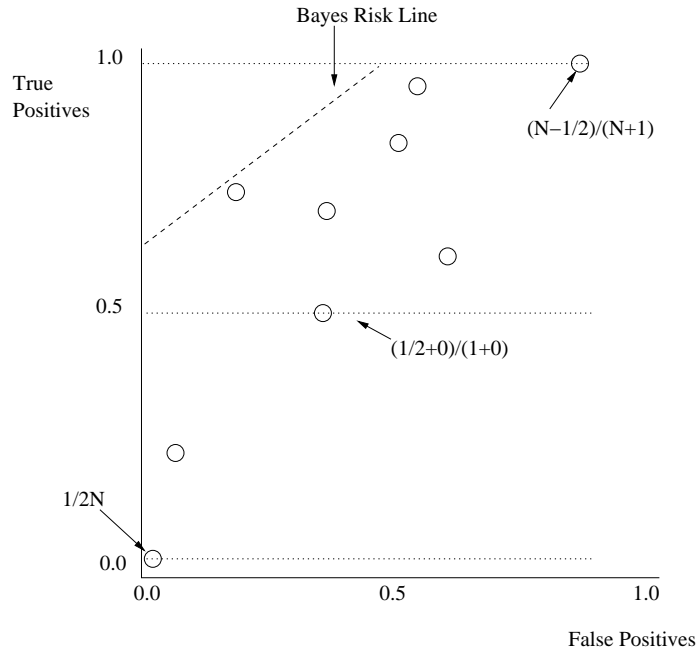


Figure 1: Circles represent candidate reference classes. The line shows an iso-contour of "Bayes Risk". For a linear risk calculation all such contours are parallel and we seek the reference class which minimises the distance to the top left-hand corner of the plot, in the direction perpendicular to these lines.

When treated as an exercise in discriminant analysis the reference class problem is not arbitrary at all, but objective as we can make clear statements regarding the utility of any choice. This assessment can be made prior to the use of the data for decision support. The subjectivity is introduced only at the point that we select the outcome criteria (ie: the Bayes risk). Arguing from a position of subjective use back to a motivation for general use of subjective probability is therefore flawed logic. This approach would prevent us from predicting the likely consequences of decisions. Compare the statements; "Using this reference class might be a rather bad idea for some patients.", and "Using this reference class will result in typically 12 deaths per 100 cases."

## Conclusions

In conclusions we can say that when observing finite quantities of data and trying to construct estimation or inference systems, quantitative application of probability theory requires that we take appropriate account of the stability of estimated parameters. In this document we have used the associated theory to show that a Poisson sample of  $n$  has an expected generator with mean  $n + 1/2$ , and that a frequency ratio of  $n/N$  is better estimated as  $\langle \mu \rangle = (n + 1/2)/(N + 1)$ . As  $\langle \mu \rangle$  is based upon a quantitative approach to the use of probability, this is a frequentist estimate<sup>7</sup>.  $n/N$  is the *maximum likelihood* result, but for situations where we do not believe that ML summaries the relevant distributions we can also estimate an *expectation value*. As an expectation computed from a density is no longer invariant under non-linear parameter redefinition, we can only do this if we have a uniquely definable parameter. In the case of mean values and ratios we have such a definition.

These solutions are consistent with empirically driven methods, and can be used to exclude definitive statements for a frequentist definition of probability (as expected). This result suggests that algorithms based upon maximum

<sup>7</sup>Contradicting the common assertion that the  $n/N$  solution represents a failure of frequentist probability.

likelihood derived without such correction might result in bias.

We have applied this result to the counting of noisy data, providing a probabilistic form for such calculations. The approach seems to suggest a corresponding hypothesis where the test value itself is treated as a sample from the test distribution. This is an interesting conclusion which would not necessarily be obvious as the way to approach this problem in the absence of the preceding analysis.

We suggest that the reference class problem (which has vexed scientists for decades), can be accordingly solved as a discriminant analysis in the appropriate metric space. We would urge people to be wary of arguments for subjectivity based upon an inability to apply quantitative probability.

## Notes

*There is no consensus on these kind of issues in the statistics community. I think it is appropriate here to expend some effort explaining the key philosophical ideas concerning the approach taken. The equal variance result presented can be seen to be related to Bayesian concepts, whereby additional terms are used to multiply initial probability densities. As will be explained below, these terms are not priors because we are still computing  $P(n|\mu)$  at this point and not  $P(\mu|n)$ . The prior itself, if you feel you need one, is constant, so that  $P(\mu|n) \propto P(n|\mu)$ . Regardless of our definition of probability, this is logically the correct choice when we have no other information (see below).*

*The similarity to Bayesian approaches is obvious, and I could doubtless have used this as a justification for the methods without raising too many objections. As the overall approach is mathematically identical to using a “Jeffreys Prior” the resulting estimation process has all of the properties (eg:uniqueness) which Jeffreys intended. But I believe that this would have failed to make some important points. There are conceptual differences here which are worthy of further consideration.*

*Our approach of using an integral to compute the required probability mass function from a density is based in frequentist principles.*

$$P(n|\mu) = \int_{\mu-\delta}^{\mu+\delta} p(n|\mu)d\mu \approx 2\delta p(n|\mu)$$

*This appears unnecessary and unjustifiable if the link to quantitation is abandoned as part of a subjective methodology. We must then appeal to Bayes theory in an attempt to explain the origins of the additional terms and if we replace  $\delta \rightarrow p(\mu)$  we have instead;*

$$p(\mu|n) \propto p(n|\mu) p(\mu)$$

*In this case, by definition any uninformative priors must logically be uniform. Uniform priors have the potential to extend to infinity. Some mathematicians may not like this property as it directly contradicts the basic axioms, and such probabilities (which have no definite interval for integration) are called “improper”<sup>8</sup>. Though this implies the presence of an unknown constant in a Likelihood formulation, it does not make the approach arbitrary (subjective).*

*If we believe that “Jeffreys’ priors” are the uninformative choice then as they are not uniform we have a logical contradiction. If we can see no flaw in the argument that the **only** self-consistent way to perform estimation is to use these terms then this raises a further issue. As these priors are computed using the data they are a function not only of the parameter but also the data sample, and are therefore not consistent with the use of probability notation (i.e. the concept of a ‘prior’, being conditional on nothing).*

*Jeffreys asks the question; “How can we make a subjective Bayesian approach to data analysis self consistent?”. Our answer to which seems to be; “Use a prior which results in an equivalent expression to that obtained using frequentist probability.”. Further, if (under that very rare situation) a quantitatively valid prior were known (constructed once again via integration from its own density estimate using an appropriate interval), it would have to be multiplied by our  $P(n|\mu)$  in order to construct the final probability distribution over  $\mu$ . This step seems to be missing in conventional use, where the Jeffreys (uninformative) prior would simply be replaced by the new (informative) one.*

*This is enough for me to form a definite opinion on these issues. While I accept the need, and mathematical form for the Jeffreys Prior, I prefer to call this an interval term. It is equivalent to mapping the estimated parameter in to a homoscedastic space so that (in Bayesian terms) a uniform prior can be assumed. However, we must note that this analysis is at odds with popular interpretations which describes anything multiplying a density as a prior and criticises Jeffreys Priors for being improper [3]. History seems to be telling us that rationalisation of these*

---

<sup>8</sup>There are two ways to deal with this. The first is to only consider such priors, and the effect they have on any estimation process, in the large interval limit. In this limit the estimation process can be shown trivially to converge on a fixed value. Secondly, in the real world no physically meaningful interval can ever be infinite. For theories of the real world it is therefore legitimate to make a distinction between “infinity”, and “large but unknown”. From a frequentist point of view, use of an improper uninformative prior is simply quantitative use of probability and is the logical way to perform an estimate.

*ideas cannot be made by a community operating with Kolmogorov's axioms, a subjective definition of probability, and general tolerance of arbitrariness. Purely mathematical considerations are too weak to permit the reason for needing Jeffreys priors to be universally accepted and his motivations are seen as a 'choice' rather than fundamental for a meaningful (scientific) analysis. This allows anyone seeing the conceptual problems with the idea to simply disregard it rather than address the deeper issues*<sup>9</sup>.

*As practical problems necessarily involve the real world we would also expect that any 'engineering' approach should also take account of this result. Under the general banner of 'Bayesian' or MAP estimation, careful selection of terms (fudge factors) to compensate for poorly motivated likelihoods (or other methodological problems) can always demonstrate some level of merit. Though persisting in referring to these fudge factors as "priors" when they are not consistent with such a description under any definition of probability will only undermine any theoretical basis, hinder interpretation and continue to cause general confusion among algorithm developers.*

## References

- [1] J. Berkson, Minimum Chi-Square, Not Maximum Likelihood!, *Annals of Statistics*, Vol 8, 3, 457-487, 1980.
- [2] P.A. Bromiley and N.A. Thacker, The Effects of an Arcsin Square Root Transform on a Binomial Distributed Quantity., Tina Memo 2002-007.
- [3] K. Ord and S. Arnold. *Kendall's Advanced Theory of Statistics: Classical Inference*. Arnold, 1998.
- [4] S. Russell and P. Norvig, *Artificial Intelligence, A Modern Approach* (second edition), Prentice Hall, 2003.
- [5] N.A.Thacker and A.J.Reader, Modal Division and its Application to Medical Image Analysis. *Proc. MIUA*, pp 7-10, London. 10th-11th July, 2000 (also Tina memo 2000-011).
- [6] N.A.Thacker, P.Bromiley, The Equal Variance Domain: Issues Surrounding the use of Probability Densities for Algorithm Construction. Tina memo, 2004-005, 2004.
- [7] N.A.Thacker, Tutorial: Statistical Design of Quantitative Vision Systems; Beyond Likelihood, Tina memo 2005-008, 2005.
- [8] N.A.Thacker, Useful Image Processing Methods, Tina Memo 2008-010, 2008.
- [9] P. Vermeesh, Statistical Uncertainty Associated with Histograms in the Earth Sciences. *Jou. Geophys. Res.* Vol 110, B02211, 1-15, 2005.

---

<sup>9</sup>If this were to happen it would be analogous to a physicist rejecting the 'equivalence principle'.