ELSEVIER

Journal of Statistical Planning and Inference 59 (1997) 167–177

journal of statistical planning and inference

Stopping rules and the likelihood function

J.K. Lindsey *

Limburgs Universitair Centrum, 3590 Diepenbeek, Belgium

Abstract

Conditions are investigated whereby the likelihood function contains all of the relevant information from the data necessary for inference, with no knowledge of the sample design. Certain designs which result in the same *reported* likelihood for the final stopped experiment in fact have different underlying likelihood *functions*.

For a likelihood function to be valid, it must, at least, contain the minimum information necessary for the experiment to be performable; this is shown to be the minimal filtration of the experiment.

Keywords: Bayesian inference; Direct likelihood inference; Exponential family; Frequentist inference; Likelihood function; Performable experiment; Sample design; Sequential methods; Stopping rule; Sufficient statistic

1. Introduction

Many statisticians are in agreement that all of the information about a given model, obtainable from a data set, is contained in the likelihood function; see, for example, Barnard et al. (1962), Birnbaum (1962), or Berger and Wolpert (1988). Here, we shall investigate exactly what might be meant by such an affirmation. We shall investigate the conditions under which two reported likelihoods can be taken to be identical, so that they do indeed contain all of the *relevant* information from the experiment. To this end, we need to distinguish between the likelihood *function* which will result from some planned observations and the presently *reported* likelihood at the end of a stopped experiment. One may think of the former as having observations which can only be represented algebraically, say by $\mathbf{y}$, while the latter has actual numerical values.

To understand why such conditions might be important, we look at a first example.

Example 1. Consider the reported likelihood for n observations, $\mathbf{y}$, with sample mean, $\bar{y}_\bullet$, from a normal distribution with unknown mean parameter, μ, and unit variance

$$L(\mu) \propto e^{-(n/2)(\bar{y}_\bullet-\mu)^2}. \quad (1)$$

* Correspondence address: Department of Social Sciences, University of Liege, Sart Tilman B31, 4000 Liege, Belgium. E-mail: jlindsey@luc.ac.be.

PII S0378-3758(96)00096-1

With the information so far provided, it is impossible to specify the minimal sufficient statistic. We also require information as to which quantities were random when the observations were made, i.e. at least one aspect of the sample design. Thus, if the sample size, N, were fixed at n in advance, the minimal sufficient statistic would be $\bar{y}_\bullet$. But, if N were random, as, for example, by flipping a fair coin after each observation to decide whether or not to continue, it would be $(\bar{y}_\bullet, n)$. In the first case, we have a $(1,1)$ linear exponential family, and, in the second, a $(2,1)$ family.

We have two different likelihood functions resulting in the same reported likelihood: in the first case only, n has a numerical value before the experiment is performed. Even a concept as simple as the minimal sufficient statistic for a member of the linear exponential family cannot be determined simply by looking at the reported likelihood. When N is random, the marginal distribution of $\bar{y}_\bullet$ generally depends heavily on the stopping rule. The conditional distribution of $\bar{y}_\bullet$, given $N = n$, is only independent of the stopping rule if the latter does not depend on the sequence observed.

This has immediate implications for certain aspects of inference. In the second case above, with N random, asymptotic properties, even of a Bayesian posterior distribution of the mean, are difficult to derive. The reported likelihood does not contain all of the information involved in making inferences. However, one may argue that this has no relevance for inferences based on the *observed* data. The reported likelihood contains the necessary information, and is minimal. Such a conclusion is misleading for certain sampling schemes, as we shall see.

The statement that the likelihood function contains all the information in the data about the parameters means that it provides all the information for point and interval estimation, *given the family of models* under consideration. Our central thesis is that relevant information about the parameters, that for *goodness of fit*, can be available from a given experiment, but is absent from the likelihood usually reported. The appropriate likelihood function for an experiment must be constructed to take this type of information into account.

2. Stopping rules

One important area of disagreement between the Bayesian and frequentist approaches to inference lies in the relevance of the stopping rule (see, for example, Anscombe, 1953, 1963; Armitage, 1963; Cornfield, 1966; Wetherill and Glazebrook, 1986). The Bayesian stopping rule principle states that 'the reason for stopping experimentation (the *stopping rule*) should be irrelevant to evidentiary conclusions about θ'. (Berger and Wolpert, 1988, p. 74). In contrast, because the stopping rule determines the sample space, knowledge about it is essential for frequentist inference.

We assume that any experiment under consideration is, at least theoretically, *performable* given prior knowledge, that the stopping rule is proper, i.e. that $\Pr(N < \infty) = 1$ for all values of the unknown parameter. This criterion is rather arbitrary and unrealistic, because it may still allow $E[N] = \infty$. As well, such conditions cannot guarantee that any actual experiment will stop; they are based on prior models which are almost surely wrong.

The stopping rule principle is held to be derivable from the likelihood principle: 'All information about θ obtainable from an experiment is contained in the likelihood function for θ given $\boldsymbol{x}$. Two likelihood functions for θ (from the same or different experiments) contain the same information about θ if they are proportional to one another' (Berger and Wolpert, 1988, p. 19). An argument for the pertinence of the stopping rule principle is that inferences should not depend 'upon the intentions of the experimenter concerning stopping the experiment' (Berger and Wolpert, 1988, p. 74.1). On the other hand, such intentions are necessary for a valid prior probability: 'The uncertainty will be up front in the prior where it belongs, however, with the data speaking for itself through the likelihood function' (Berger and Wolpert, 1988, p. 79). 'If the experimenter forgot to record the stopping rule and then died, it is unappealing to have to guess his stopping rule in order to conduct the analysis' (Berger and Wolpert, 1988, p. 78). But if the experimenter's prior was not recorded either, can a Bayesian legitimately conduct an analysis? Our claim here is that any such intentions which reflect prior knowledge of the phenomenon under study must be included in the likelihood function.

The most commonly used example of the irrelevance of the stopping rule for the information in the reported likelihood is the comparison between binomial and negative binomial sampling.

Example 2. Consider independent Bernoulli trials with constant unknown probability of success, π, where the likelihood function for y_1 successes and y_2 failures is

$$L(\pi) \propto \pi^{y_1}(1-\pi)^{y_2}. \tag{2}$$

We must have $\pi \in (0,1]$, because $\Pr(N < \infty; \pi = 0) = 0$ when a negative binomial stopping rule is applied. This stopping rule comparison is very special because even the problem of specifying the sufficient statistic in Example 1 does not occur. The minimal sufficient statistic is y_2, without further knowledge. It does not matter if $n = y_1 + y_2$ or y_1 was fixed at some numerical value in advance; in both cases, we have a $(1,1)$ linear exponential family.

However, let us look at a slight modification of this.

Example 3. Suppose that the reported likelihood for independent Bernoulli trials is

$$L(\pi) \propto \pi^1(1-\pi)^2 \tag{3}$$

with no further information available. According to the stopping rule principle, we are in the situation of Example 2. Thus, many statisticians would immediately be prepared to draw inferences based only on this information. However, suppose that the stopping rule, used by the experimenter but unknown to the statistician, had been: take the first three observations if the fourth is a failure or all four if the fourth is a success. Few statisticians who learned this would still be prepared to say that Eq. (3) is the same likelihood as if $N = 3$ had been fixed in advance. On the other hand, the apparently closely related rule, if the third observation is a failure stop, otherwise, make one more trial and stop, is acceptable.

Can we ignore the intentions of the experimenter in these cases? The reason for distinguishing between them is that the first stopping rule is not a martingale stopping time (Dawid, 1979); the decision to stop at some time point requires knowledge of events after that stopping point. Information has been discarded in constructing the reported likelihood. Such practice is not excluded by the stopping rule principle, as usually stated.

A second common example of the irrelevance of the stopping rule is based on Wald's sequential probability ratio test.

Example 4. Suppose that the following procedure had been used to obtain the reported likelihood of Eq. (1): stop if

$$\bar{y}_\bullet \leqslant B - \frac{k}{\sqrt{n}} \quad \text{or} \quad \bar{y}_\bullet \geqslant A + \frac{k}{\sqrt{n}}, \quad B \leqslant A \tag{4}$$

for $n < n^*$ or at some very large fixed value, n^*. Most authors using this example take $A = B = 0$, as, for example, Basu (1975), Berger and Wolpert (1988, p. 76), or Cox and Hinkley (1974, pp. 50–51), in which case the limit on n is not necessary ($n^* = \infty$) in order for the experiment to be performable. Here, the maximum likelihood estimate, $\hat{\mu}$, can never lie in the interval, $[B,A]$, unless $n = n^*$. For a suitable choice of k, the reported likelihood at $\hat{\mu} \notin [B,A]$ will be arbitrarily larger than that for any $\mu \in [B,A]$. In particular, if $A = B = 0$, it can be made arbitrarily larger than that for the true value, if $\mu = 0$, with probability one, without specifying n^*.

In this example, no posterior distribution, derived in ignorance of the stopping rule, can reasonably take into account the 'information' in this reported likelihood. A closed set of values of the parameter can be excluded from the region of plausible values based on this reported likelihood, with probability close to or equal to one, simply by the 'scientist' appropriately choosing the values of A, B, k, and n^* and not telling the statistician. Kerridge (1963) provides bounds for errors in similar conditions, but where the number of possible models is finite; unfortunately, this cannot help us here.

This example illustrates clearly that the stopping rule can modify the likelihood function. With such a stopping rule and n very large, but less than n^* (the statistician only knows the value of n, not whether $n = n^*$, nor even that n^* exists), the reported likelihood for $\mu \in [B,A]$ is not comparable to that for $\mu \notin [B,A]$. Does this reported likelihood still have meaning? This stopping rule reduces the information available about values of $\mu \in [B,A]$: if $\hat{\mu} > A$, values in $[(A + B)/2, A]$ are more plausible than those in the other half of the exclusion interval, without providing any more precise relative weighting (and vice versa for $\hat{\mu} < B$). A large n supports the relative plausibility of $\mu \in [B,A]$ as opposed to $\mu \notin [B,A]$, with no further information as to where in the interval it might be located. As n becomes very large, all of the information is contained in $(n, \text{sign}[\hat{\mu} - (A+B)/2])$, so that $\hat{\mu}$ contains no further information about μ. The shape of the reported likelihood outside of the excluded interval must become distorted as n increases. How can this information be used? At what point is an observed n so large that it is more likely, from the reported data alone, that μ lies in rather than outside the exclusion interval?

A solution to this problem does not appear to have been given in the literature. We shall apply our approach in the last section below.

3. Conditions on the stopping rule

From these examples, we can derive a series of conditions on the stopping rule which cannot be ascertained simply by inspecting the reported likelihood.

Condition 1: *The rule must be independent of the unknown parameter.* In their discussion of stopping rules, statisticians often state that these must not depend on the unknown parameter. For example: 'stop for whatever reasons, which (conditional on the data) do not depend on θ' (Berger and Wolpert, 1988, p. 77) or 'we have a 'stopping rule' depending in some way on the data currently accumulated but not on further information about the unknown parameter' (Cox and Hinkley, 1974, p. 40). However, this lack of dependence seems never to be explicitly defined. For example, how does the first quotation permit one to distinguish between the two stopping rules suggested above in Example 3 for Bernoulli trials?

Such a condition cannot mean that the distribution of the sample size, given the observed responses, $\Pr(N=n|\boldsymbol{t})$, does not depend on the unknown parameter, where $\boldsymbol{t}$ is the minimal sufficient statistic in the corresponding fixed sample size case, because this would have the consequence that $\boldsymbol{t}$ must also be sufficient for the parameter in the sequential case. We saw in Example 1 that this will not be true even for simple sequential sampling. (It will in cases where $\boldsymbol{t}=\boldsymbol{y}$, so that there is no reduction; then, the vector length gives n.)

Condition 2: *The rule must be a martingale stopping time.* This arises from Example 3. It serves to eliminate unscrupulous experimenters; for other relevant examples, see Dawid and Dickey (1977).

This condition is closely related to the conditions whereby a missing data mechanism need not be included in the likelihood function (Rubin, 1976): the reason for stopping or for missingness must be non-informative or ignorable, *only* depending on the observed, and reported, data.

Condition 3: *The rule must not exclude the possibility of experiments leading to high likelihoods in certain regions of the parameter space.* What information can an experiment supply about the parameter if values with relatively high likelihood could not possibly occur in a given region of the parameter space? This condition states that the stopping rule must not constrain the possible *positions* of the likelihood function with respect to the parameter space.

When Condition 3 is not fulfilled, what information is there in the reported likelihood? Should it be declared to correspond to no possible likelihood function, as when

Condition 2 is broken? As in that case, the reported likelihood, by itself, can provide very misleading information. However, in Example 3, which breaks Condition 2, incorporation of the discarded information will yield a usable likelihood. On the other hand, even with complete information on the stopping rule of Example 4, how can useful information be extracted from the reported likelihood for estimation of the unknown parameter?

To see the amount of information which may be contained in n, the reader can calculate the series of possible reported likelihoods for Bernoulli trials, based on Eq. (2), with a Wald-type stopping rule analogous to Eq. (4) (see Armitage, 1958, for examples where $A = B = 0$). One can calculate $\hat{\pi}$ from n alone, without being informed of the number of successes!

We conclude that the stopping rule principle must be severely conditioned in order to be tenable. In the same way, the likelihood principle, in its strict form, is questionable. Conditions 2 and 3 must be fulfilled. Although all information about an unknown parameter, obtainable from an experiment, may be contained in the appropriate reported likelihood, the corresponding likelihood function must be determined through some minimal knowledge of the sample design. The likelihood must correspond to an appropriate probability model describing the sequential nature of the experiment, i.e. its sample path. It is not sufficient that the usually reported likelihoods be proportional to one another for the experiments to contain the same information about the unknown parameter.

4. Performable experiments

In order to understand better the implications of an experiment being performable, let us now examine what minimal knowledge an experimenter must have in order to complete an experiment. We suppose that the experimenter has complete control over the conduct of the experiment, although not necessarily of the point in time at which each result becomes available, choosing a design to maximize the required information produced and to minimize effort and expense. For simplicity, we look again at Bernoulli trials, to ascertain the probability of heads of coins.

If the trials are not sequential, there is one basic way in which they can be performed.

Experiment 1: The experimenter receives a container of indistinguishable coins, without necessarily being told how many, and throws them all simultaneously.

Here, the total number of trials is fixed before the experiment, and is not random, whether the experimenter is aware of it or not before performing the experiment. Two pieces of relevant information become available: the number of trials and the number of heads. Because the coins are indistinguishable and are thrown in the same way, the only reasonable model is the usual binomial likelihood. However, it only contains information about the probability of heads for an average of all coins in the container, supposed to be sufficiently identical for this to be meaningful; the coins are assumed to be *exchangeable*. This is typical of the usual cross-sectional survey or experiment. (The only other possibility would be to assume all coins to be distinct, in which case all Bernoulli probabilities are estimated to be either zero or one — not very useful.)

If the intention is to obtain information about the probability of heads for one given coin, the experiment must be sequential, a stochastic process. Exchangeability will now refer to interchanging the order of the results, which is only possible if the trial outcomes are assumed to be independent and identically distributed. Suppose, first, that the experimenter, who will decide when to stop, tosses the same coin each time. However, the result recorded cannot be known to the experimenter until all trials are completed. What are the possible stopping rules, the *performable* experiments?

Experiment 2: The experimenter fixes the total number of trials in advance.

Experiment 3: The experimenter fixes the total time of the experiment in advance.

Experiment 4: The experimenter stops when some external random (in the sense of not being predictable within the context of the experiment) event occurs.

Under the stated conditions, a negative binomial experiment cannot be performed, because that would require the unavailable knowledge of the cumulative number of heads at each step. Thus, only two relevant pieces of information directly concerning the performance of the experiment are available sequentially in the history or filtration of the experiment: the time passed and the number of trials. A third, the number of heads, is only available after termination.

Now, consider the negative binomial experiment.

Experiment 5: The experimenter stops after $Y_1 = y_1$ heads.

Here, the experimenter *must* know the result of each trial at each step (except for the first y_1 trials, for which the total number of heads after these is sufficient). If this design is chosen, the experimenter's intentions have changed from the previous experiments; otherwise, the additional effort in registering each individual result would not be warranted. The trials are *not* exchangeable; one-half of the permutations of a given observed result yield unperformable experiments. The sample path is additional information, *necessarily* provided by the experiment, as part of the filtration, in order for it to be performable. This is relevant information, as expressed in the experimenter's intentions, which cannot be discarded from the likelihood function, in a similar way to the extra trials violating the martingale stopping time for Example 3 not being discarded. Thus, Condition 2 is not sufficient to prevent important information from being excluded from the likelihood function.

One might argue that the sample path is irrelevant because we are studying *independent* Bernoulli trials. However, this can only be an hypothesis, a *model*. We can see the statistician, believer in the stopping rule principle, muttering silently 'I shall stop after y_1 heads and only record the number of tails' as the following sequence unfolds:

$$HTTTTHTTTTHTTTTHTTTTHTTTTHTTTT\cdots \qquad (5)$$

The sample path does provide information about the model of independent, constant probability, Bernoulli trials: whether the *hypotheses* of independence and constant probability are reasonable or not (Barnard et al. 1962, give a similar example to make the same point). This is not information necessary for point or interval estimation of the Bernoulli probability, assuming the model to be true, but it is a relevant information about that probability. This information is necessarily available, determining the experimenter's intentions, in contrast to the first four experiments; it must be included in the likelihood function. Thus, the negative binomial distribution is the correct formula to describe independent, constant probability, Bernoulli trials stopped after y_1 heads, but is not the correct likelihood function for such an experiment, because it discards relevant information about the Bernoulli probabilities being studied. The latter function must display the sample path necessarily observed for the experiment. It can be modelled by standard statistical procedures for stochastic processes, for example, by a Markov chain.

We now return to experiments with more than one coin.

Experiment 6: n different coins are to be used, one at each trial performed sequentially, with each result immediately available.

The coins must be distinguishable, so that the filtration now also contains which coin corresponds to each result in the sample path, but its individual Bernoulli probability will be estimated as zero or one. For useful inferences about the probability of heads, one must return to the same exchangeability hypothesis as in Experiment 1.

Finally,

Experiment 7: For each of n coins, the experimenter stops after a series of trials, each time with $Y_1 = y_1$ heads.

We can now, in this repeated measurement situation, reasonably compare probabilities of heads for different coins, i.e. *check* that exchangeability hypothesis. We necessarily have available a distinct sample path for each coin. We could collapse observations over coins or over sample paths; both would entail discarding information about the Bernoulli probabilities, necessarily produced by the experiment. One corresponds to assuming exchangeability of coins, the other to assuming independent and constant probabilities among trials. Is it, a priori, more reasonable to collapse over one rather than the other? And yet the classical negative binomial experiment would only collapse over trials, to obtain a product, over coins, of reported binomial likelihoods.

The choice of experiment, including the stopping rule, has direct implications for the definition of the likelihood function. Our final condition to define a valid likelihood function is

Condition 4: *A likelihood function must contain all the information in the minimal filtration necessary for the experiment to be performable.* This is closely related to the idea of informative stopping rules: 'to be informative, a stopping rule must carry information about θ additional to that available in $\boldsymbol{X}^N$' (Berger and Wolpert, 1988, p. 90), the solution being to 'consider *all* available observational information as part of the data' (Berger and Wolpert, 1988, p. 89). However, we have gone further by showing that many common sample designs, not usually considered to be informative, contain relevant available information and by defining precisely what must be included in the likelihood function.

5. Discussion

What these results tell us is that the relevant information in an experiment for models of interest is not just that which allows estimation of the parameters in those models, but also must include any available information about goodness of fit. If the likelihood function is to contain all relevant information, it cannot be restricted to a function only of these models of direct interest.

We can now look again at the Wald-type experiment of Example 4. For the likelihood function to contain the sample path, the model might be some general diffusion process. Suppose that the true mean, μ, lies somewhere in the interval, $[B, A]$. Then, the sequence of maximum likelihood estimates, which are necessarily available from the minimal filtration, should hover near the true value, with increasing precision, for some time, before wandering off, with probability (close to) one, over the nearest boundary, at which time the experiment is stopped. Although the final, 'stopped', likelihood does not allow estimation of the parameter, this series of values, and the full likelihood function for this process, do inform us as to what has happened.

Frequentist theory has long held that goodness of fit of a model can be judged from $\Pr(y|\boldsymbol{t})$, where $\boldsymbol{t}$ is the minimal sufficient statistic. Box (1980) has proposed a Bayesian interpretation of this. Similar possibilities are available directly from the likelihood function. (For examples of the direct likelihood approach to goodness of fit, see Lindsey (1974a, b, 1995) and Lindsey and Mersch (1992)).

Our results establish an important area of commonality between Bayesian and frequentist inference, through direct likelihood inference. Likelihood functions which are proportional contain the same information about the model, as the likelihood principle states, but these likelihoods must contain all relevant information from the sample design, information which a frequentist has always considered pertinent.

One can argue that this proposed likelihood is unnecessary if the trials are independent. But, independence is only an *hypothesis*; information about the model is being discarded. All other models are being given prior probability zero. If a sequence like Eq. (5) starts to appear, how would such a prior be suitably updated? This cannot be done from the previous posterior probability in the sequence; once a model has zero probability under the application of Bayes' theorem, it always does. A Bayesian cannot make scientific discoveries using strictly Bayesian inference procedures; the completely unforeseen has zero prior probability. The current prior cannot be modified by any form of updating using Bayes' formula. How, then, does a realistic Bayesian accomplish this updating in the face of an unexpected event?

Thus, unfortunately, for 'stopping rule Bayesians', if the possibility of anything unforeseen becomes apparent from the sample path, they, obliged to update, are incoherent (Dawid, 1982; Oakes, 1985; see also Shafer, 1985). Bayesians, by affirming the stopping rule principle, have placed themselves in a contradictory position, because their prior for any unforeseen event is irrevocably zero. An experiment is necessarily providing them with information which they cannot use without stepping outside the Bayesian paradigm, but which then may show their prior to be miscalibrated.

Hence, the likelihood principle is emptied of much of its import. How often will two reported likelihoods, from differently designed experiments, be proportional once the minimal filtration is included? Must we conclude that, for sequential trials, the current Bayesian approach is wrong, but that the only approach is a Bayesian one, based on a revised likelihood principle?

Acknowledgements

This work has arisen from research for books on Models for Repeated Measurements (Lindsey, 1993) and on Parametric Statistical Inference (Lindsey, 1996). It has profited greatly from a reading of the many papers of A.P. Dawid. The encouraging comments of George Barnard, Phil Dawid, David Hinkley, Philippe Lambert, Dennis Lindley, and Dave Sprott were greatly appreciated.

References

Anscombe, F.J. (1953). Sequential estimation. *J. Roy. Statist. Soc.* **B15**, 1–29.

Anscombe, F.J. (1963). Sequential medical trials. *J. Amer. Statist. Assoc.* **58**, 365–383.

Armitage, P. (1958). Numerical studies in the sequential estimation of a binomial parameter. *Biometrika* **45**, 1–15.

Armitage, P. (1963). Sequential medical trials: some comments on F.J. Anscombe's paper. *J. Amer. Statist. Assoc.* **58**, 384–387.

Barnard, G.A., G.M. Jenkins and C.B. Winsten (1962). Likelihood inference and time series. *J. Roy. Statist. Soc.* **A125**, 321–352.

Basu, D. (1975). Statistical information and likelihood. *Sankhya* **A37**, 1–71.

Berger, J.O. and R.L. Wolpert (1988). *The Likelihood Principle: A Review, Generalizations, and Statistical Implications.* Institute of Mathematical Statistics, Hayward.

Birnbaum, A. (1962). On the foundations of statistical inference. *J. Amer. Statist. Assoc.* **57**, 269–306.

Birnbaum, A. (1972). More on concepts of statistical evidence. *J. Amer. Statist. Assoc.* **67**, 858–861.

Box, G.E.P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *J. Roy. Statist. Soc.* **A143**, 383–430.

Cornfield, J. (1966). Sequential trials, sequential analysis and the likelihood principle. *Amer. Statist.* **20(2)**, 18–23.

Cox, D.R. and D.V. Hinkley (1974). *Theoretical Statistics.* Chapman & Hall, London.

Dawid, A.P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc.* **B41**, 1–31.

Dawid, A.P. (1982). The well-calibrated Bayesian. *J. Amer. Statist. Assoc.* **77**, 605–613.

Dawid, A.P. and J.M. Dickey (1977). Likelihood and Bayesian inference from selectively reported data. *J. Amer. Statist. Assoc.* **72**, 845–850.

Kerridge, D. (1963). Bounds for the frequency of misleading Bayes inferences. *Ann. Statist.* **34**, 1109–1110.

Lindsey, J.K. (1974a). Comparison of probability distributions. *J. Roy. Statist. Soc.* **B36**, 38–47.

Lindsey, J.K. (1974b). Construction and comparison of statistical models. *J. Roy. Statist. Soc.* **B36**, 418–425.

Lindsey, J.K. (1993). *Models for Repeated Measurements.* Oxford University Press, Oxford.

Lindsey, J.K. (1995). Fitting parametric counting processes using log linear models. *Appl. Statist.* **44**, 201–212.

Lindsey, J.K. (1996). *Parametric Statistical Inference.* Oxford University Press, Oxford.

Lindsey, J.K. and G. Mersch (1992). Fitting and comparing probability distributions with log linear models. *Comput. Statist. Data Anal.* **13**, 373–384.

Oakes, D. (1985). Self-calibrating priors do not exist. *J. Amer. Statist. Assoc.* **80**, 339–342.
Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
Shafer, G. (1985). Conditional probability. *Internat. Statist. Rev.* **53**, 261–277.
Wetherill, G.B. and K.D. Glazebrook (1986). *Sequential Methods in Statistics.* Chapman & Hall, London.