Bayesian synthesis or likelihood synthesis – what does Borel's paradox say?

Tore Schweder[1] and Nils Lid Hjort[2]
Departments of Economics[1] and Mathematics[2]
University of Oslo

February 1996

Abstract

The relevance of the Borel paradox to Bayesian synthesis [4] is explained and illustrated by examples related to the assessment of bowhead whales. It is argued that the paradox is serious, and that if conditions for the paradox to be of minor concern are observed, the attraction of the Bayesian synthesis method will fade, since then the freedom of disregarding the constraints of the deterministic model when carrying out the initial statistical analysis is destroyed. A likelihood based synthesis is proposed as an alternative. To accommodate indirect data in the likelihood framework, the concept of confidence likelihood is proposed. The idea is that when graded expert opinion summarises the indirect data in a set of nested confidence regions, there is a likelihood function that would produce the same set of nested confidence regions. Likelihood synthesis is discussed in two examples.

KEY WORDS: *assessment, bowhead whales, modelling, prior knowledge, synthesis*

1 Introduction

The Bayesian synthesis method was developed by Adrian Raftery and coworkers for the purpose of obtaining a comprehensive assessment of the stock of bowhead whales that is subject to Inuit subsistence whaling off Alaska.

The method was extensively discussed in the Scientific Committee of the International Whaling Commission (IWC), and it was used for the intended purpose at the Annual IWC meeting in 1994. The method has been well documented and reported on. Raftery, Givens and Zeh [4] have been eager to expose the method to a wider scientific audience than the Scientific Committee. In August 1994, the method was presented as the applied paper of the Journal of the American Statistical Association to the Joint Statistical Meeting in Toronto. In the discussion of this paper, Wolpert warned against the method: "the authors' imaginative departure from the Bayesian paradigm is dangerous and would be unnecessary, with a small amount of additional thought and modelling", see [4]. The root of the problem is that the post-model distribution as defined by RGZ (Raftery, Givens and Zeh) is the conditional distribution given a set of probability zero. Wolpert alluded to the so-called Borel paradox.

In this note we will explain the Borel paradox, and we will argue that Bayesian synthesis, as outlined by RGZ, is a specific way of calculating an entity which really is a 0/0-fraction. Such a fraction, determined as the limit of the fraction as both numerator and denominator go to 0, depends critically on how these limits are approached. In Bayesian synthesis, the choice of parametrisation affects how the limits of the numerator and the denominator are approached. The results from Bayesian synthesis will therefore depend on the formalities of the parametrisation, even when the information content in the pre-model distribution is kept constant. This dependency might be detrimental, as shown by constructed population assessment examples.

In the Bayesian tradition, the prior distribution is established by analysing or judging the relevant indirect data (background knowledge, analogies etc., see [1]), while the direct data enter the likelihood that is integrated with the prior through Bayes' formula to yield the posterior distribution. In Bayesian synthesis, this work is done first, to produce the pre-model distribution. If the pre-model distribution does not satisfy the deterministic model in the sense that the constraints of the model are met with pre-model probability one, the pre-model distribution needs to be "synthesised" by finding the conditional distribution given the model. This is done by filtering the pre-model distribution through the deterministic model to yield the post-model distribution, which is the end result. If the dimensionality of the pre-model distribution is the same as the dimensionality of the deterministic model, Bayesian synthesis is no different from ordinary Bayesian analysis, and the Borel paradox does not apply.

After having discussed some aspects of the relationship between the Bayesian synthesis method and the ordinary Bayesian analysis, we turn to the likelihood analogue of Bayesian synthesis. Since likelihoods are invariant with respect to parametric transformations, the Borel paradox does not affect the likelihood method.

In likelihood analysis, there might also be indirect data available. Traditionally, such data are only used in specifying constraints on the parameter of interest. Often, a more graded approach is called for. If the indirect data either lead directly to a likelihood function summarising the prior information, or can be used to construct a nested set of confidence regions that in turn leads to a likelihood function, this likelihood can be combined with the likelihood of the direct data in the ordinary way. The likelihood summarising the indirect data will be called the *confidence likelihood*, and it plays the same role in likelihood analysis as the prior distribution plays in Bayesian analysis. We will term likelihood analysis of direct and indirect data in the presence of a deterministic model *likelihood synthesis*. Thus, in situation where Bayesian synthesis could be used, likelihood synthesis is also applicable – but then without being hampered by the lack of invariance that is illustrated by the Borel paradox. We will therefore argue that likelihood synthesis is a better approach to integrating multiple sources of direct and indirect data through a deterministic model, than Bayesian synthesis.

2 The Borel paradox and Bayesian synthesis

Following RGZ, Bayesian synthesis is a method of integrating information from various sources through a deterministic model as follows. The various data and expert judgements determine a pre-model distribution for the pair of parameters (θ, ϕ). This joint distribution, $p(\theta, \phi)$, is obtained without any regards to a deterministic model, Φ, which relates the input parameter θ to the output ϕ by $\phi = \Phi(\theta)$. The post-model distribution proposed by RGZ is

$$\pi(\theta) = \frac{p(\theta, \Phi(\theta))}{\int p(t, \Phi(t))\, dt},$$

where π is understood as the conditional density, given the model manifold $M = \{(\theta, \phi) \colon \phi = \Phi(\theta)\}$, which is the surface of input-output parameter points (θ, ϕ) consistent with the model.

The problem demonstrated by the Borel paradox is that π is not the only conditional density of θ given M. Denote by $\theta \in d\theta$ the event that θ lies in an infinitesimal region $d\theta$, and $A_\gamma(\varepsilon)$ the event $|\phi - \Phi(\theta)| \leq \varepsilon\gamma(\theta)$ for a given positive real function $\gamma(\theta)$. Here $A_\gamma(\varepsilon)$, for $\varepsilon > 0$, is a system of neighbourhoods around the model manifold M. By the definition of conditional probability, the conditional probability density we seek, π, is determined as the limit of

$$P(\theta \in d\theta \mid A_\gamma(\varepsilon)) = \frac{P(\theta \in d\theta \cap A_\gamma(\varepsilon))}{P(A_\gamma(\varepsilon))}. \tag{1}$$

As $\varepsilon \downarrow 0$, both numerator and denominator at the right hand side tend to zero, and the fraction tends to a limit $\pi_\gamma(\theta)\, d\theta$ by l'Hospital's rule. By direct calculation,

$$\pi_\gamma(\theta) \propto \gamma(\theta)\pi(\theta),$$

and the conditional density is thus seen to depend critically on the choice of γ. The root of the problem is that the event $(\theta, \phi) \in M$ has pre-model probability zero.

The indeterminacy can also be seen as lack of invariance to a change of variables. Let $\theta \to \theta$ and $\phi \to \psi = (\phi - \Phi(\theta))/\gamma(\theta)$. By calculating the Jacobian of the transformation (the Jacobian is the determinant of the matrix of first derivatives $\partial\phi/\partial\psi$), the joint density of (θ, ψ), $p^{[\theta,\psi]}(\theta, \psi)$, is proportional to $\pi_\gamma(\theta) = \gamma(\theta)\pi(\theta)$ for $\psi = 0$, i.e. for $(\theta, \phi) \in M$. The conditional density found by the standard ratio calculation is thus $\pi_\gamma(\theta)$.

The conclusion is therefore that an arbitrary probability density π' defined on the support of π, is just as much a version of the conditional density of θ given M as is π. This is obtained by simply choosing $\gamma = \pi'/\pi$. Wolpert [4] points to one of the core lessons of probability: conditioning on null-events is indeterminate. Conditioning must be subject to a σ-algebra (which essentially is a system of neighbourhoods like $A_\gamma(\varepsilon)$), or with respect to a null-event involving a stochastic variable like ψ. There are wider classes of transformations than those spanned by the functions γ. It is actually possible to obtain discrete distributions as versions of the conditional distribution of θ given the null-event M, even when (θ, ϕ) has a continuous distribution. We will not pursue this further.

Example 1.

Suppose there were two independent posterior distributions for the stock abundance of bowhead whales for a given year. Since the same stock is estimated twice, the model is simply $\theta = \phi$, where both θ and ϕ denotes stock abundance. It might be a bit surprising to frame this situation as one of Bayesian synthesis, but from a formal point of view it really is an extremely simple case. If θ is based on data observed prior to those leading to ϕ, the relevance of Bayesian synthesis might come more easily to mind. In this example, the model function is Φ = identity. By independence the pre-model distribution is the product $p^{[\theta]}(\theta)p^{[\phi]}(\phi)$. For simplicity, assume that the two posterior distributions are identical, and gamma:

$$p^{[\theta]}(\theta) = p^{[\phi]}(\theta) \propto \theta^{\alpha-1}\exp(-\beta\theta) \tag{2}$$

with mean $\alpha/\beta = 7800$ and variance $\alpha/\beta^2 = 700^2$. This agrees with the pre-model distribution of P_{1988} of RGZ. By choosing $\gamma = \theta^{2a}\exp(-2b\theta)$, for $a > -\alpha$ and $b > -\beta$ in (1),

$$\pi_\gamma(\theta) \propto \theta^{2(\alpha+a)-1}\exp(-2(\beta+b)\theta).$$

This is a gamma distribution with mean $\mu = (\alpha+a)/(\beta+b)$ and variance $\sigma^2 = \frac{1}{2}(\alpha+a)/(\beta+b)^2$. The choice $a = 0$ and $b = 0$ gives $\gamma = 1$, which perhaps is the most natural choice. Then $\mu = \alpha/\beta = 7800$ and $\sigma = 700/\sqrt{2}$. In the continuation of this example presented below, it is seen that the choice of $\gamma = 1$ corresponds to the likelihood synthesis solution.

From a mathematical point of view, other choices are equally good as solutions to the Bayesian synthesis problem. By choosing a and b appropriately, all possible gamma-distributions can be obtained. If, for example, $a = -\frac{5839}{49}$ and $b = \frac{167}{4900}$, the post-model abundance distribution is gamma, with a mean of $\mu = 100$ bowhead whales in the Alaska stock, and with standard deviation $\sigma = \sqrt{1000}$. These numbers are not offered as serious abundance estimates, but rather as a taste for what sort of results one can obtain by using the Bayesian synthesis method according to its definition as a conditional density given the model manifold.

Example 2.

Consider now the bowhead case, as presented in RGZ. Assume that the unit of measurement for the output variables is changed from stock size in 1988, P_{1988}, to $\phi_1 = MSYR{*}P_{1988}$. Here, $MSYR$ is the maximum sustainable yield rate, i.e. the yield rate that is sustained at the "optimal" stock level $MSYL$. A possible rationale for this rescaling is that since the stock level is well below $MSYL$, ϕ_1 gives a rough estimate of a "safe" replacement yield in 1988, provided the production function is concave. All other inputs and outputs are kept unaltered.

The Jacobian of this transformation is simply $MSYR^{-1}$. The joint pre-model distribution of inputs and outputs is thus $p(\theta, \phi) = p_{RGZ}(\theta, \phi)/MSYR$ where p_{RGZ} is the joint pre-model distribution of RGZ evaluated at (θ, ϕ). The post-model distribution resulting from the reformulation is

$$\pi(\theta) \propto \pi_{RGZ}(\theta)/MSYR. \tag{3}$$

This change in the post-model distribution is only due to the change $P_{1988} \rightarrow MSYR * P_{1988}$. It is in all reality based on the same information (pre-model distribution), but the post-model marginal of $MSYR$ will, from (3), be shifted to the left relative to the RGZ result.

Based on Figure 11 of [3], the marginal post-model distribution for $MSYR$ is roughly estimated as a gamma distribution (2) with index parameter $\alpha = 4.38$ and scale parameter $\beta^{-1} = 0.00653$. In this gamma distribution the mean is 0.0286 and the mode is 0.022. In the re-formulated model with post-model density for the input parameters given by (3), the marginal post-model distribution of $MSYR$ is gamma with index parameter $\alpha = 3.38$ and scale parameter $\beta^{-1} = 0.00653$, and with mean 0.022 and mode 0.016. There is thus nearly a 30% reduction in $MSYR$ by this slight change in the formulation of the model, without altering the realities.

By also changing another output parameter, say from annual rate of increase to "safe" rate of increase in the replacement yield, $ROI \rightarrow MSYR * ROI$, the post-model marginal of $MSYR$ is shifted further toward low values:

$$\pi(MSYR) \propto \pi_{RGZ}(MSYR)/MSYR^2.$$

This example is only meant for illustration. We are not claiming that there really is a 30% or more positive bias in the estimate of $MSYR$ for the bowhead stock, the point is to demonstrate that Bayesian synthesis can potentially give grossly misleading results.

3 Bayesian synthesis in relation to standard Bayesian analysis

In standard Bayesian analysis, the prior is combined with the data likelihood by Bayes' formula. The posterior distribution is the conditional distribution, given the event that data are as observed. This event has probability zero when data are continuous. One could thus fear that the Borel paradox affects all Bayes analysis. This is, however, not the case, since the conditional density calculated through Bayes' formula is precisely that version of the conditional density that agrees with the Borel σ-algebra that is assumed in the formulation of the statistical model for the direct data.

In standard Bayes analysis, the dimensionality of the prior distribution (of its support) equals that of the parameter space. That will not be the case in Bayes synthesis. The pre-model distribution will typically consist of prior distributions for components without direct data, and of posterior distributions obtained from standard Bayesian analysis for components with relevant direct data. The pre-model distribution will then have dimensionality equal to the total number of input and output parameters. The model manifold will be a lower dimensional surface in this product space, and will thus have probability zero in the pre-model distribution. In Bayesian synthesis, one calculates a version of the conditional pre-model distribution given the null event that the model is satisfied. When no canonical parametrisation is available, or in other words, when there is no canonical σ-algebra of which the model manifold is a member, there is no unique post-model distribution. This is the normal state of affairs, and the Borel paradox will thus normally affect Bayes synthesis.

When the pre-model distribution has the same dimensionality as the model manifold, there is no problem. But then, there is really nothing to synthesise. We will only use the term Bayes synthesis when the dimensionality of the pre-model distribution exceeds that of the model manifold.

It has been suggested that if the pre-model distributions "agree well" for model inputs and outputs, according to the model, the Bayesian synthesis method provides an approximation to a full Bayes analysis, for which the Borel paradox does not apply, see [4, Authors' rejoinder]. In Example 1, model inputs and outputs had identical pre-model distributions. Since the model simply was the identity, this could be taken as perfect agreement. The paradox did, however, affect this model. Agreement in model inputs and outputs must therefore mean something different.

What probably is meant is that model output is stochastically dependent on input, and that the joint pre-model distribution has its mass in a band along the model manifold. If that is the case, a change of variables will have less effect, and the effect is smaller the more the joint distribution is concentrated along the model manifold.

However, the basic idea of Bayesian synthesis is that the empirical aspects of the problem should be analysed separately, and with no regard to the possibly complicated model. The various data and expert opinions should be described in terms of pre-model distributions for relevant parameters of the model, disregarding the deterministic relationships between model inputs and outputs imposed by the model. This freedom for the empiricist to carry out his work without taking on board the theory as expressed in the deterministic model, and the freedom of the theorist to develop the model without having to bother about estimation etc., is the great appeal of the method. If agreement relative to the deterministic model in inputs and outputs in the pre-model distribution is required, it seems to us that the bottom goes out of the proposed method.

The same remark applies to the suggestion that if the deterministic model was made only slightly stochastic, then to condition on the model would not be to condition on a null set, and hence the paradox would go away. But this presupposes that the pre-model distributions of inputs and outputs agree with this stochastic model formulation, something that would preclude separate empirical analysis of the different sets of data and expert opinions relevant to inputs and outputs.

To sum up this point, in traditional Bayesian methodology it is assumed that the statistician is able to formulate a likelihood function consistent with the substantive model. The likelihood summarises the information in all relevant direct data. This likelihood is then combined with the prior distribution summarising the information in the indirect data. See [1] for a discussion of direct and indirect data. In Bayesian synthesis, the main point is, however, that direct and indirect data can be summarised in an appropriate descriptive model, which often will be a posterior distribution, but with no regards to consistency to the substantive deterministic model. The synthesis is then to filter this (descriptive) posterior distribution through the deterministic model. By separating the statistical modelling from the substantive modelling, freedom is given to both empiricist and theoretician, and they can join forces to obtain a synthesis by the proposed method. Solutions to the Borel paradox problem that involve transforming the method to a more ordinary, but possibly very complicated, Bayesian method, are not attractive, in that one loses the intended freedom of modelling and analysis. In view of these remarks Bayesian synthesis appears to lose its envisaged appeal.

4 Likelihood synthesis

Inference based on the likelihood function is known to be invariant to reparametrisation. A likelihood-based method of synthesis should therefore not be affected by the Borel paradox or the like. General information on likelihood inference is found in [2]

From separate statistical analyses of the various independent sets of data, one obtains a likelihood $L(\theta, \phi)$ in the input and the output parameter. For the direct data, that is, the directly observed data, the likelihood is obtained by standard statistical analysis. A likelihood might also be required for the indirect data, say those based on analogy and those based on expert opinion. There is no tradition for specifying a likelihood for such indirect data.

We will, however, argue that it might be just as natural to specify a likelihood function that represent the information content in the indirect data, as it is to specify a prior distribution in the Bayesian framework. To keep such a likelihood apart from a likelihood based on direct data, and to allude to its judgemental nature, we propose the term *confidence likelihood.* Suppose there are indirect data available for the parameter θ but that these data are not already summarised in a likelihood function. Filtered by the background knowledge and theory of the subject field, the indirect data are summarised in a nested set of confidence regions T_α, where α is the degree of confidence. The set is nested when $T_{\alpha_1} \subset T_{\alpha_2}$ when $\alpha_1 \leq \alpha_2$. Typically, $T_0 = \{\widehat{\theta}_{\text{indirect}}\}$ consists of the point estimate based on the indirect data. The function defined by

$$l(\theta) \geq -\frac{1}{2}\chi^2_\nu(\alpha) \Leftrightarrow \theta \in T_\alpha,$$

where $\chi^2_\nu(\alpha)$ is the α-point of the χ^2-distribution with $\nu = \dim(\theta)$ degrees of freedom, is the confidence log likelihood related to the system of confidence regions. Since $l(\widehat{\theta}_{\text{indirect}}) = 0$, l is actually a log likelihood ratio, and treated as a genuine log likelihood ratio the nested system of confidence regions T_α is obtained from l by calculating the contours at the appropriate χ^2-quantiles.

When the prior information is purely judgemental, or based on analogy, it will be natural to start with a set of confidence regions and then summarise these in a confidence likelihood, as explained above. If, however, the prior information is based on (old) direct data, the information might well come in the form of a likelihood function. Whether based on confidence regions or directly obtained from previous studies, we prefer the name confidence likelihood. It is, actually, essential that there is scientific confidence in the prior information, whether used in a likelihood analysis or a Bayesian analysis.

When the joint likelihood of the relevant direct and indirect data has been established, the model $\phi = \Phi(\theta)$ acts on the likelihood simply by restricting its domain to M. That is, the post-model likelihood is

$$\Lambda(\theta) = L(\theta, \Phi(\theta))$$

The statistical problem is then the familiar, of estimation, confidence regions, hypothesis testing etc.

Example 1, continued.

A maximum likelihood estimate of abundance $\widehat{\theta} = 7800$, with variance $\mathrm{var}(\widehat{\theta}) = v = 700^2$, and with gamma-shaped likelihood, has log-likelihood

$$l = (\widehat{\theta}/v)(\widehat{\theta}\log\theta - \theta). \tag{4}$$

If the other abundance estimate has the same likelihood function, the joint likelihood is, by independence,

$$l = (\widehat{\theta}/v)\{\widehat{\theta}(\log\theta + \log\phi) - (\theta + \phi)\}.$$

On $M{:}\, \theta = \phi$, the pooled likelihood is simply $l = 2(\widehat{\theta}/v)(\widehat{\theta}\log\theta - \theta)$, and the pooled maximum likelihood estimate remains $\widehat{\theta} = 7800$, with standard error $(v/2)^{1/2} = 700/\sqrt{2}$. This estimate is invariant to transformations of θ and ϕ.

To illustrate the concept of confidence likelihood, assume that one of the abundance estimates, $\widehat{\theta}$, is based on indirect data, and that the expert opinion is summarised in a set of nested confidence intervals for θ. If these are

$$T_\alpha = \{\theta{:}\, (\widehat{\theta}/v)\{\widehat{\theta}\log(\theta/\widehat{\theta}) - (\theta - \widehat{\theta})\} \geq -\frac{1}{2}\chi^2(\alpha)\}$$

with $\chi^2(\alpha)$ being the α-quantile of the χ^2-distribution with 1 degree of freedom, the related confidence likelihood is exactly the gamma-shaped likelihood (4).

Example 2, continued.

Assume that the pre-model distribution of RGZ, or rather that of the Scientific Committee, could be reinterpreted as a likelihood. In that case the comprehensive assessment of the bowhead stock would no longer be hampered by the Borel paradox. Whether such a reinterpretation is acceptable is rather to the Scientific Committee itself to decide. The reinterpretation would involve some work required to obtain the nested systems of confidence regions based on indirect data, needed for the confidence likelihoods. This work would be unfamiliar, but hardly more difficult than the work required to land the necessary priors for the pre-model distribution:

- For the survival rates, the confidence regions would be the standard normal intervals $\widehat{s} \pm \sigma z_{\alpha/2}$, where $\widehat{s}$ is the point estimate and σ its assessed standard error and z_a are α-quantiles of the normal distribution. These assessments are based on data reported in "BAM" (see [3, Section 4]). The system of standard normal confidence intervals relates to a quadratic log confidence likelihood. Quadratic log confidence likelihoods agree with assessments made in the Scientific Committee for both the two survival rates s and s_0. The constraints $0 < s_0 \leq s \leq 1$ is at least as natural to impose on the confidence likelihood as on the pre-model distribution.

- Age at sexual maturity is taken as uniform over $10, 11, \ldots, 30$. This means a flat likelihood over these integers.

- MSYR was modelled as a gamma distribution. The confidence likelihood consistent with skewed confidence intervals, such as those in example 1, is the gamma-shaped likelihood.

The list should have been continued, and updated with the discussion of the Scientific Committee in 1994 and 1995.

RGZ and Wolpert use terms like "natural parametrisation" and "realistic description" to discriminate between more or less acceptable formulations of the same model. The parametrisation in RGZ was felt to be natural to many members of the Scientific Committee. We may speculate that this was because those member essentially interpreted these pre-model distributions as likelihoods.

The computational approach of [4] might be viable, also for likelihood synthesis. Suppose the pre-model likelihood is $L(\theta, \phi)$. If a scaling is found so that $p(\theta, \phi) = cL(\theta, \phi)$ is a probability density, the SIR algorithm of RGZ computes the post-model density $\pi(\theta) \propto p(\theta, \Phi(\theta)) = \Lambda(\theta)$. If, say, the likelihood is Gaussian, as would be the normal state of affairs when sufficient data are available, a scaling is available and the pre-model distribution is multinormal. The 5000 second-stage sample of RGZ is a sample from π.

Although likelihood functions traditionally are not regarded as probability distributions, and thus the idea of having a sample from a likelihood function might seem strange, this should not be a problem when a scaling is available so that $\pi = c\Lambda$ indeed is a probability density. In the normal case, marginals of π are related to corresponding profile likelihoods of Λ.

There are other, and usually less expensive, computational approaches than the SIR algorithm. One is the traditional, to fit a quadratic to $l = \log(\Lambda)$, and to infer from this the approximate maximum likelihood estimate and its approximate covariance matrix.

5 Conclusion

We agree with Wolpert that Bayesian synthesis has serious drawbacks, and is not likely to be a very useful method. The basic idea of Bayesian synthesis is, however, valuable. There is a need for statistical methods for drawing inference from direct and indirect empirical data in the light of a deterministic model based on theory. Since population dynamics models, and also deterministic models in other fields, might be complicated, it is often difficult to specify an explicit statistical model that is consistent with the deterministic model. The problem could also be put the other way around: it is often difficult to formulate a deterministic (or probabilistic) model that is sufficiently realistic and acceptable from a theoretical point of view, and at the same time is sufficiently simple to allow an explicit and realistic statistical model. By allowing the analysis to be split in an empirical component where the pre-model distribution, or preferably the pre-model likelihood, is obtained, and a theoretical component where the deterministic model is formulated, this difficulty is overcome. This is the valuable idea basic to Bayesian synthesis. But due to the Borel paradox, and to some lesser degree other nuisances related to Bayesian methodology, Bayesian synthesis can not be recommended as a method for filtering the pre-model information through the deterministic model.

Likelihood synthesis appears to be a good alternative. By the use of confidence likelihoods, indirect data are summarised in likelihood terms. The pooled likelihood is obtained by multiplying together component likelihoods from independent sources, and the deterministic model is simply a restriction of the support of the likelihood function. As such, likelihood synthesis is just statistical routine inference based on a likelihood function.

It should be added that likelihood inference is not always simple and straightforward. Multiple local maxima, unboundedness, non-Gaussian curvature and flatness at the top are problems that can arise and seriously hamper a likelihood analysis. Such problems do occur in practice, particularly in complex models, and when the statistical model does not adequately describe the data. If, say, there is conflicting information on a particular parameter from the different sources of data, the likelihood surface might be multi-modal. Analysis of the likelihood surface by way of model diagnostics will be helpful in identifying problems with damaging incoherence in direct and indirect data relative to the model, and other possible deficiencies.

When there are more sources of indirect data than there are dimensions in the model manifold, synthesis is needed to restrict the post-model inference to be consistent with the deterministic model. Due to the Borel paradox, Bayesian synthesis will unfortunately be indeterminate, and the post-model inference will depend on more or less arbitrary choices in the parametrisation of the model. Likelihood synthesis is not hampered by the same lack of invariance. It does, in fact, handle multiple sources of indirect data exactly as it handles multiple direct data. The crucial question for likelihood synthesis is as in standard likelihood inference, whether the total likelihood obtained as the product of the independent likelihood components, and restricted to the effective parameter space (the model manifold), is well behaved. This will depend on the specifics of the model and the direct and indirect data, and cannot be answered in general terms.

6 Acknowledgements

Harald Goldstein and Arvid Raknerud, both at Department of Economics, University of Oslo, have read the paper and gave valuable suggestions.

References

[1] Bravington, M.V. 1995. An appraisal of Bayesian synthesis, with suggested modifications and diagnostics. IWC/SC/47/AS18.

[2] Barndorff-Nielsen, O.E. and Cox, D.R. 1994. *Inference and Asymptotics.* Chapman & Hall.

[3] Givens, G.H., Zeh, J.E. and Raftery, A.E. 1994. Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using the BALEEN II model in a Bayesian synthesis framework. *Rep. int. Whal. Commn*, **46**: 345–364.

[4] Raftery, A.E., Givens, G.H. and Zeh, J.E. 1995. Inference from a deterministic population dynamics model for bowhead whales. With discussion. *Journal of the American Statistical Association*, **90**, 402–430.