Indirect and Direct Likelihoods and Their Synthesis

Tore Schweder[1] and Nils Lid Hjort[2]

Department of Economics[1] and Department of Mathematics[2], University of Oslo

August 1997

ABSTRACT. This paper proposes and discusses ways of constructing an indirect likelihood function to summarise information in indirect data. The indirect likelihood can then be combined with the likelihood of the direct statistical data to form the basis of an ordinary likelihood analysis. We call this Likelihood Synthesis. In the case of a scalar parameter, the indirect log-likelihood is related to the cumulative confidence distribution F_c that yields canonical confidence intervals, that is, the distribution having the property that $[F_c^{-1}(\frac{1}{2}\alpha), F_c^{-1}(1-\frac{1}{2}\alpha)]$ is the canonical $1-\alpha$-interval. The indirect likelihood is transformation invariant, unlike analysis based on a Bayesian prior. It equals the confidence density only in the normal case. The indirect likelihood is intended for statistical reporting and for integrating results from other studies with new direct data.

Likelihood Synthesis is also a modelling strategy for likelihood analysis. By first developing a purely statistical model with likelihood L^f in a possibly high-dimensional parameter ϕ, and then restricting this to a sub-manifold, usually by a model function $\phi = \Phi(\theta)$ in a basic parameter θ, the likelihood to analyse is $L(\theta) = L^f(\Phi(\theta))$. Likelihood for interest parameters are found from L as profile likelihoods.

The theory is illustrated via two examples. In the first, the Fieller method of drawing inference for a quotient parameter is extended and illustrated through survey data for minke whales. The second example concerns bowhead whales, and presents a Likelihood Synthesis analysis that could replace the Bayesian Synthesis analysis of Raftery, Poole and Zeh (1996).

KEY WORDS: *Bayesian Synthesis, cumulative science, indirect likelihood, input and output parameters, Likelihood Synthesis, population dynamics, whale counting*

1. Introduction and summary.

1.1. THE LIKELIHOOD PRINCIPLE AND COHERENT LEARNING. The likelihood principle (Birnbaum, 1962) states that parametric inference should be based on the likelihood function if the principles of sufficiency and conditionality are observed. The concepts of likelihood and sufficiency and the method of maximum likelihood date back to papers of R.A. Fisher, from 1920 and 1922 and even to 1912; see the discussion in Geisser (1992), Edwards (1997), and Efron (1997). Many authors have discussed the likelihood principle and its consequences; a recent contribution of interest is Bjørnstad (1996). Bayesians have argued that accepting the likelihood principle should lead to Bayesian methodology. Berger and Wolpert (1984, p. 124) sought to demonstrate that "sensible use of the likelihood function seems possible only through Bayesian analysis". And Savage, in the discussion of Birnbaum (1962), said that

> "... I suspect that once the likelihood principle is widely recognized, people will not long stop at that halfway house but will go forward and accept the implications of personalistic probability for statistics".

In this paper we propose the concept of *indirect likelihood* as a means of completing what Savage called the halfway house of the likelihood of the direct data, in the presence of prior distributional information (indirect data). The Bayesian integrates his prior distributional information (personalistic or otherwise) with the information contained in the direct data by using Bayes' formula. His parameters are stochastic variables, and he modifies their distribution in light of new data. In the Bayesian tradition, no formal distinction is made between probabilities related to judgements or beliefs on one hand, and probabilities related to the outcome of an experiment on the other hand. This is unfortunate. Belief-probabilities have a different basis than experiment-probabilities. Our proposal is to reserve the term probability for the objective sampling probability, and to use the term 'confidence' or 'likelihood' for degree of belief.

To explain what we shall mean with an indirect likelihood, let us first quote Efron (1997) as he discusses Fisher's ideas and attitude:

> "... Fisher's main tactic was to logically reduce a given inference problem, sometimes a very complicated one, to a simple form where everyone should agree that the answer is obvious."

And Fisher's favourite target for the 'obvious' was the situation where a single X is observed from a $N(\beta, \sigma^2)$ with unknown mean and known variance. Here everyone agrees, say Fisher and Efron, that the correct $1 - \alpha$ confidence interval (to use terminology Fisher disliked) for β is $\widehat{\beta} \pm z_{1-\alpha/2}\sigma$, where $\widehat{\beta} = X$ (the equally obvious best estimate) and z_p denotes the standard normal p-quantile. This may also be written

$$\{\beta\colon (\beta - \widehat{\beta})^2/\sigma^2 \le z^2_{1-\alpha/2}\} = \{\beta\colon 2(\ell(\widehat{\beta}) - \ell(\beta)) \le \chi^2_{1,1-\alpha}\}, \tag{1.1}$$

in terms of the log-likelihood $\ell(\beta)$ for this experiment, where $\chi^2_{\nu,1-\alpha}$ is the upper α point of the χ^2 with ν degrees of freedom. Efron goes on to credit Fisher for his astounding resourcefulness at reducing complicated problems to the simple form above. One such area of importance is that of large-sample maximum likelihood theory, where we know that the recipe of the right hand side of (1.1) leads to more accurate intervals than the simpler $\widehat{\beta} \pm z_{1-\alpha/2}\widehat{\sigma}$.

These remarks also reflect our own attitude and aspiration when working towards a workable definition of and methods utilising the intended indirect likelihood. In the above situation, the confidence distribution function $F_c(\beta) = N((\beta - \widehat{\beta})/\sigma)$, where N is the standard normal cumulative, neatly and canonically summarises all correct confidence intervals; $[F_c^{-1}(\frac{1}{2}\alpha),\ F_c^{-1}(1 - \frac{1}{2}\alpha)]$ is the same as the interval above. There are other families of confidence intervals, say skewed ones, corresponding to other confidence distributions, but the F_c just given is the 'inferentially correct one' in this situation. This enables us to define an indirect log-likelihood $\ell_{\rm ind}$, by extension and analogy, in nice models, in terms of a given confidence distribution. This is elaborated on from both a theoretical and practical viewpoint in Section 3, which also deals with multiparameter situations. For the case of models with a single parameter β, the approach is to let $\ell_{\rm ind}(\beta)$ be the log-likelihood that would have generated the same set of central confidence intervals (leaving out confidence $\alpha/2$ at either end) by the standard recipe

$$\{\beta\colon 2(\ell_{\rm ind}(\widehat{\beta}_{\rm ind}) - \ell_{\rm ind}(\beta)) \le \chi^2_{1,1-\alpha}\}, \tag{1.2}$$

where $\widehat{\beta}_{\text{ind}}$ is the point estimate, defined as the 'confidence median', the remaining point of the confidence interval when the confidence level is sent to zero. This determines the indirect likelihood, up to an additive constant, as

$$\ell_{\text{ind}}(\beta) = -\tfrac{1}{2}\{z_{F_c(\beta)}\}^2 = -\tfrac{1}{2}\{N^{-1}(F_c(\beta))\}^2 \tag{1.3}$$

in terms of the cumulative confidence distribution F_c. This is further discussed and developed in Sections 2 and 3.

Edwards (1992, Chapter 3) defines *support* as the logarithm of the likelihood ratio, and, somewhat implicitly, prior support for one hypothesis against another as the log-likelihood ratio in an imaginary experiment. Edwards' prior support is the log-likelihood ratio of imaginary data from the experiment which gives weight to one hypothesis over the other that matches the prior preference. This concept of prior support corresponds to our indirect likelihood. We stress, however, the need to base the indirect likelihood on data; but, of course, some judgement will always enter an analysis and both indirect and direct likelihoods will be somewhat flavoured by subjectivity. Edward's verbal formula

$$\text{posterior support } = \text{prior support} + \text{experimental support}$$

could be rephrased in terms of total log-likelihood being equal to indirect plus direct log-likelihoods, which is the crucial point. We prefer the terminology of likelihoods, however; the term 'support' also conveys the spirit of the concept, but appears superfluous.

Efron (1993, 1994, 1996) also uses the term 'confidence distribution', denoting such a representation of a family of confidence intervals for the parameter. Our indirect likelihood is derived from the confidence distribution, in cases when the indirect data are summarised by confidence intervals. Our indirect likelihood is not, however, the same as the density of the confidence distribution. When the confidence intervals are constructed by Efron's ABC method, correcting for acceleration and bias, the likelihood that Efron terms the implied likelihood is precisely the likelihood we propose (Efron 1993); see Section 3. If the family of confidence regions is more complex than a nested set of intervals, the indirect likelihood gets more complicated. If the indirect likelihood has two local maxima, for example, the set of confidence regions will, for an interval of α values, consist of unions of two disjoint intervals. This system might be represented by two confidence distributions, and the indirect data can be defined in terms of these.

Bayesian inference is known to be invariant with respect to transformations of the parameter. The direct likelihood is of course transformation invariant, and when the Jacobian is carried along, the posterior is transformation invariant with respect to probabilities of events in the parameter space. When the Bayesian is in lack of indirect data on which to base his prior distribution, he has to invent some (perhaps implicitly or subconsciously). To keep away from subjectivity, he might use a non-informative prior. For some models, arguments can sometimes be given for particular priors to be non-informative. If the parameter space, Ω, has finite volume, a uniform distribution is often regarded as a non-informative prior. Also when the volume is infinite, Bayesians often revert to a flat prior, which then does not represent a proper prior distribution. Generally, the property of non-informativity is not transformation invariant, whether the distribution is proper or not. The Jacobian of the transformation will distroy the flatness. As other likelihoods, the indirect likelihood is, of course, transformation invariant. And the non-informative indirect likelihood is uniquely defined (up to an additive constant); put $\ell_{\text{ind}}(\beta)$ equal to zero on Ω and $-\infty$ outside.

The Bayesian stand (Berger and Wolpert, 1984) is that the only coherent methodology that is in accordance with the likelihood principle is the Bayesian way of characterising the likelihood function by averaging. The methodology we present, allowing indirect data to enter the likelihood function, is certainly coherent in the Bayesian sense, and it is no less in accordance with the likelihood principle than the Bayesian method. Nuisance parameters might, however, be difficult to handle in given situations (as for the Bayesian). The concept of profile likelihood, and various ways of approximating or modifying it (Barndorff-Nielsen and Cox, 1994) are, however, available and useful tools.

To facilitate coherent and cumulative learning, it is important that studies are reported on in a sufficiently complete fashion, to allow the results from one study to be used by the next study. This requires both the likelihood of the new direct data to be reported and also the combined likelihood (with proper reference to the sources for indirect data) in sufficient detail to allow them to be reconstructed for use in the next study. It is imperative to avoid that the same indirect data are being used with multiple weight. This applies to Bayesian methods as well as to likelihood methods.

1.2. LINKS TO RECENT LITERATURE. The concept of indirect likelihood and the ensuing method of *Likelihood Synthesis* have been developed as an alternative to Bayesian Synthesis as presented in Raftery, Givens and Zeh (1995). One of the present authors (T.S.) was a discussant when that paper was presented to the Joint Statistical Meeting in Toronto in 1994. His remarks flattering Bayesian Synthesis came in an awkward light when R. Wolpert pointed out that Bayesian Synthesis is plagued with a serious lack of invariance under reparametrisation; see the discussion following Raftery, Givens and Zeh (1995). This lack of invariance is related to the original Bayesian Synthesis being a case of calculating the conditional probability density given an event of probability zero. This was elaborated in Schweder and Hjort (1996), presented to the Scientific Committee of the International Whaling Commission to argue against the use of the original version of the Bayesian Synthesis method for assessing whale stocks. Indirect likelihood and Likelihood Synthesis were briefly sketched and discussed in Schweder and Hjort (1996) as likelihood alternatives to the Bayesian prior distribution and synthesis.

There is much current interest in establishing general ways of combining different information sources in nontrivial situations. Raftery and coworkers have made an attempt to overcome the problems of Bayesian Synthesis which is different from the the Likelihood Synthesis methods of Schweder and Hjort (1996) and the present paper; see Raftery, Poole and Givens (1996), and also comments by Bravington (1996). Efron (1996) discusses empirical Bayes methods for combining likelihoods, again different in spirit from the present paper. For a good partial review of relevant problems and methods, prior to the last decade, see Genest and Zidek (1986) and its discussion contributions.

The aim of the present paper is to develop the concept of indirect likelihood further and to discuss Likelihood Synthesis as a modelling strategy for accommodating diverse data and complex deterministic models in a single likelihood function. In the end this permits analysis by ordinary methods. Section 2 provides the initial general framework for our intended indirect likelihood and its role in Likelihood Synthesis. Section 3 discusses the indirect likelihood construction in more concrete terms, and delineates some general practical strategies. Also a characterisation result is reached involving normal confidence densities. Then Section 4 takes up computational and other practical aspects of the Likelihood Synthesis. This leads to illustrating applications, concerned with minke and bowhead whales studies, presented in Section 5. One of the applications involves an extended Fieller type analysis for quotient parameters. Finally supplementing remarks are offered in Section 6.

2. Indirect likelihood and Likelihood Synthesis: general considerations.

2.1. Indirect likelihood. Bayesians have an advantage over classical inference in their use of prior distributions as a vehicle for updating quantitative indirect information in the light of new direct data. The distinction between direct and indirect data has been discussed by Bravington (1996) and others, and is essentially that indirect data are based on analogy and past summary observations, while direct data are subject to an observational plan for gathering new data and possibly experimental manipulation that would give structure to the parameter of primary interest. There is inevitably an element of subjectivity in judging indirect data to be relevant, and thus to be included in the analysis together with the direct data when drawing inference concerning a parameter of primary interest. The pure objectivist position that often is taken by statisticians of the classical frequentist breed is in our view extreme, almost as extreme as the pure subjectivist position often taken by Bayesians. There are, in our experience, many cases where there are relevant indirect data and prior knowledge that have a role to play in the statistical analysis, even when carried out by classical methods.

We will argue that there often is uncertainty associated with indirect data. A concept allowing uncertain prior information to enter the likelihood analysis is therefore needed. To stay within the framework of likelihood analysis, the prior information should be embodied in a likelihood function to be combined with the likelihood function of the direct data. To distinguish the likelihood function based on indirect data from the ordinary likelihood function, and to allude to the element of subjectivity involved in judging the indirect data to be relevant, we have chosen the term *indirect likelihood* for the former. It is essential that there is scientific confidence in the indirect data, and that their relevance is well argued.

In mainstream classical statistics, prior knowledge and indirect data enter the analysis through the assumptions made concerning the structural form of the likelihood function including its support. These assumptions form the basis for the analysis, and the validity of the results rests with the validity of these assumptions. There is usually no quantified uncertainty associated with assumptions made based on prior information. By allowing the information in the indirect data to be presented in the format of a likelihood function, the inference drawn from the combined likelihood would account for this uncertainty in addition to the sampling variability in the direct data. To include all possible sources of uncertainty is, however, an impossible task.

When the indirect data are summarised in an indirect likelihood, it is in principle possible to test the relevance of these data by an ordinary likelihood ratio test. This would be analogous to a Bayesian testing whether his prior distribution is consistent with the direct likelihood. It is standard Bayesian practice to show the prior and posterior distributions in the same plot. When the two essentially agree in location, the Bayesian has gained precision in his information through the direct data. If the posterior differs markedly from the prior in location, the Bayesian might, however, be equally happy, claiming that he has learned something new and that he has corrected previous biases in his knowledge. We would normally require the indirect likelihood to be consistent with the direct data in the sense that a hypothesis of homogeneity in parameters between the two is not rejected. If the parameter describing the indirect data is significantly different from that describing the direct data, the indirect data are not directly relevant for drawing inference for the current value of the parameter of interest; and other explanations or models must be sought.

Although there is an element of subjectivity in the indirect likelihood, as there is in most assumptions leading to appropriate and tractable likelihood functions, our main concern is not to allow subjectivity per se in likelihood analysis. The focus is rather on methods for accumulating scientific information. When integrating information concerning an important parameter from different sources, efficiency is gained by having the relative weighting correct. It also helps to have the information presented in statistics that allow efficient integration. In smooth models, adding log-likelihoods of independent data components is an efficient way to integrate statistical data. This is also the case when integrating indirect and direct data when they are independent. In Schweder (1988), a version of the Neyman–Pearson lemma was used to show that independent test information is most efficiently integrated by way of likelihood functions.

The use of indirect likelihood is twofold. It can be used to present indirect data in a format that allows ordinary statistical analysis to be carried out on the likelihood function integrating direct and indirect data. The result of this analysis would typically be a profile likelihood in the parameters of primary interest, with consequent confidence intervals and test conclusions. The other use of indirect likelihoods is to report the main finding of a study in a format that invites subsequent studies to integrate the results with new direct data. The reported result-likelihood would typically be the profile likelihood. To prevent misuse of the reported likelihood, it is essential that its sources are made clear. It is easy to imagine that several studies could use the same indirect data. Their result likelihoods would then not constitute independent sources of information. To prevent such confusion, we support the suggestion made by among others Spiegelhalter, Freedman and Parmar (1996, p. 98) in the Bayesian case, that the direct likelihood is fully reported in the results section of the publication, while results based on combining direct and indirect data should be presented in the discussion section. See also our Section 5.2 below.

The frequentist interpretation of confidence intervals and p-values etc. is made slightly difficult when the inference is partially based on an indirect likelihood. If, however, the indirect likelihood itself is based on various pieces of past direct data, the frequentist interpretation stands, but hypothetical replicates of the study would then involve replication of the whole suite of studies that led to the indirect and the direct data. We prefer to not place too much emphasis on this frequentist interpretation. What matters is the log-likelihood function. Due to the frequentist interpretation in linear normal models, or smooth large-sample models, we characterise the log-likelihood function by its contours, as indexed by the

χ^2 distribution. In such nice models, minus twice the log-likelihood (normalised to value zero at the maximum) is χ^2 distributed with degrees of freedom equal to the dimension ν of the space of free parameters, and the region within the $-\frac{1}{2}\chi^2_{\nu,1-\alpha}$ is a $1-\alpha$ confidence region in the classical sense. We will use confidence region synonymous with likelihood contour region. Thus the 95% confidence region calculated from the available data is the set of parameter values inside the $-\frac{1}{2}\chi^2_{\nu,0.95}$ contour of the log-likelihood function, rather than a region calculated by a method that in a hypothetical replication of the experiment would cover the true value with probability 0.95.

This re-definition of degree of confidence, from coverage probability to contour index of the log-likelihood function, is valuable, we think, for several reasons. In scientific applications it is the data at hand that matter, and not hypothetical replications. In parametric models, the direct and indirect data are summarised by the likelihood function, and all that can be learned from the data are aspects of the likelihood. The log-likelihood is uniquely determined by its indexed contours, and these contours are thus primary aspects. To separate the concepts of probability on the one hand and likelihood and confidence on the other is also important, both for educational and interpretational purposes. Most important, however, is that by linking the confidence regions to the log-likelihood it becomes reasonably clear how indirect data presented in the format of a set of confidence intervals, say at levels 95%, 90%, 75%, 50% and 0% (the point estimate) can be integrated with new direct data in a likelihood analysis. Full information of a one-dimensional likelihood is, as we shall see, provided in the confidence distribution; see also Efron's (1993) discussion.

The re-definition is not primarily intended to change the way in which confidence intervals are calculated. It is rather the interpretation that, in the applied scientific context, should be changed. But statistical practice should, perhaps, be changed with respect to which confidence intervals are calculated and how they are presented. The point here is that sufficient information should be reported to allow the results from the reported study to be integrated with future or other data in an optimal way.

That confidence intervals often allow a frequentist interpretation is a good thing. It is through the frequentist interpretation of confidence as coverage probability in Gaußian or large-sample smooth models that the χ^2 is found to be the obvious index of the log-likelihood contours. The benefit of having an agreed method for indexing the log-likelihood contours is substantial, and outweighs the inconvenience of not always allowing a precise frequentist interpretation.

2.2. Likelihood Synthesis. The likelihood function is often more difficult to construct than what appears from statistical textbooks. The ideal is to capture the statistical variability in the direct data, the regularities of the subject matter theory and the information contained in the available indirect data, in a statistical model with its likelihood function. When the subject matter theory may be approximated with a deterministic model, as in the population dynamics model of Section 5.2 below, it is sometimes possible to obtain the likelihood of interest from the likelihood of an overparametrised model. This is the essence of Likelihood Synthesis, and as such it is more a strategy for statistical modelling than a (new) method of analysis per se. We will, in fact, rely entirely on ordinary likelihood analysis, with the profile likelihood as a central concept.

Likelihood Synthesis was sketched in Schweder and Hjort (1996) as an alternative to Bayesian Synthesis. Our aim was to carry the attractive features of Bayesian Synthesis over to the framework of likelihood analysis, in order to avoid problems connected with Bayesian Synthesis. The particularly attractive feature of Bayesian Synthesis is that the statistician is allowed to formulate descriptive statistical models for the various independent pieces of direct data. If indirect data are available, the information and uncertainty is captured in prior distributions. These likelihoods and priors are constructed without regards to the possible complex parametric structure that is imposed by the subject matter theory. In Bayesian Synthesis, the Bayesian statistician would thus be allowed to do what he is best at: establishing likelihood functions for the direct data that reflect sampling variability, and prior distributions for the parameters for which there is indirect information or expert knowledge available. The subject matter scientist would, on the other hand, be allowed to construct his theoretical model without having to cast it in tractable statistical terms. The synthesis would then be to filter the statistical information through the deterministic model to obtain inference for parameters of primary interest.

In Likelihood Synthesis, assume that there is a deterministic theoretical model driven by the input parameter θ. There are K pieces of independent direct or indirect data available. For data piece k, construct a likelihood, possibly an indirect likelihood, in the parameter ϕ_k. That this piece of data is relevant is reflected in ϕ_k being functionally related to θ, through the theoretical model. The likelihood to work with for $\phi = (\phi_1, \ldots, \phi_K)$ is thus of the form

$$L^f(\phi) = \prod_{k=1}^{K} L_k(\phi_k).$$

This is usually an overparametrised likelihood. We write L^f to indicate that this is the full likelihood constructed to represent the statistical information, without regards to the theoretical deterministic model

$$\phi = (\phi_1, \ldots, \phi_K) = (\Phi_1(\theta), \ldots, \Phi_K(\theta)) = \Phi(\theta).$$

The likelihood for θ is thus simply

$$L(\theta) = L^f(\Phi(\theta)). \tag{2.1}$$

From this likelihood, we would like to draw inference on the interest parameter γ related to the basic parameter through $\gamma = \Gamma(\theta)$. If $\dim(\gamma) < \dim(\theta)$, Likelihood Synthesis consists of calculating the profile likelihood for γ,

$$L_{\text{prof}}(\gamma) = \max\{L(\theta) \colon \Gamma(\theta) = \gamma\}. \tag{2.2}$$

Likelihood Synthesis may also be an effective modelling strategy in pure statistical problems. If, say, the parameter of interest is related to the quotients of pairs of other parameters, as in the extended Fieller example of Section 5.1, it pays to construct an overparametrised likelihood in the independent pieces of data first, and then to restrict this likelihood to a subspace determined by the quotient relationships.

3. Constructing the indirect likelihood.

If science were linear and accumulative and if scientists did report their findings in the format of (profile) likelihoods, there would be no need for the next scientist to construct an indirect likelihood for the parameter of interest. He could just use the last update of the likelihood for the parameter. When the indirect likelihood is not directly available for the indirect data found relevant for inclusion in the analysis, the problem is how to naturally summarise these indirect data and how to turn these summary statistics into an indirect likelihood.

3.1. THE ONE-PARAMETER CASE. Assume first that the parameter of interest and for which there is 'background' or indirect information, say β, is one-dimensional. The parameter of primary interest in the study might well be many-dimensional, and there might be independent indirect information on other components than β. By independence, the indirect log-likelihood of all indirect data is the sum of its independent components.

Assume that the information in β is given in terms of a nested set of confidence intervals. Generalising slightly, these intervals may be presented by a confidence distribution $F_c(\beta)$, such that $[F_c^{-1}(\frac{1}{2}\alpha), F_c^{-1}(1-\frac{1}{2}\alpha)]$ is the canonical $1-\alpha$ confidence interval for β; see the discussion around (1.1) and (1.2). For the moment, take F_c to be continuous. Note that F_c is interpreted in terms of degree of confidence and not coverage probability; regardless of this F_c certainly satisfies the requirements of a cumulative distribution function.

In classical statistics, a distinction is made between degree of confidence on one hand and coverage probability on the other. We keep this distinction. The confidence distribution is just one that summarises a partially nested set of confidence intervals. A Bayesian statistician does not distinguish between confidence and probability. This is, in our view, often unfortunate, since then degree of belief is mixed with probability related to sampling variability. There is thus an important difference in interpretation between indirect likelihood and prior density. The two approaches are also different in practice, except for the very important Gaußian case, as we show below.

The Bayesian would take $f_c(\beta) = F_c'(\beta)$ as a prior probability distribution for β. This density is in general not equal to the indirect likelihood function. The indirect likelihood is, on the contrary, the likelihood function that would have led to the confidence intervals generated by likelihood analysis. The $1-\alpha$-confidence interval obtained by inverting the likelihood ratio statistic based on a given log-likelihood, $\ell_{\rm ind}(\beta)$, using the traditional first-order approximation to its distribution (in the frequentist sense), is

$$[F_c^{-1}(\tfrac{1}{2}\alpha), F_c^{-1}(1-\tfrac{1}{2}\alpha)] = \{\beta\colon 2(\ell_{\rm ind}(\widehat\beta_{\rm ind}) - \ell_{\rm ind}(\beta)) \le \chi^2_{1,1-\alpha}\}, \qquad (3.1)$$

involving the upper α point of the χ^2 distribution with 1 degree of freedom. The confidence distribution thus leads to the indirect log-likelihood $\ell_{\rm ind}$ with its mode at the median $\widehat\beta_{\rm ind} = F_c^{-1}(\frac{1}{2})$ and with

$$\ell_{\rm ind}(\beta) = \begin{Bmatrix} -\frac{1}{2}\chi^2_{1,1-2F_c(\beta)} & \text{if } \beta \le \widehat\beta_{\rm ind} \\ -\frac{1}{2}\chi^2_{1,2F_c(\beta)-1} & \text{if } \beta \ge \widehat\beta_{\rm ind} \end{Bmatrix} = -\tfrac{1}{2}\{N^{-1}(F_c(\beta))\}^2 \qquad (3.2)$$

(defined up to an additive constant), writing N for the standard normal cumulative distribution function.

The right hand side of (3.1) represents 'correct' confidence intervals based on ℓ_{ind} when the indirect likelihood is Gaußian. Confidence intervals are, however, transformation invariant, and since any unimodal ℓ_i can be transformed to Gaußian shape, the right hand side represents 'correct' intervals for any unimodal likelihood; cf. comments about the frequentist interpretation of confidence intervals at the end of Section 2.1.

For multimodal likelihoods, the relationship between confidence intervals and log-likelihood is more complex. It is outside the scope of this paper to discuss this in detail. For our purposes it is sufficient to assume that there exists a likelihood function summarising the indirect data, and that this function is represented in one format or another. One possible format of representation is the family $\{l, R_l\}$ where l is the height to the top of the log-likelihood from anywhere on the contour R_l. The contours might be thought of as defining confidence regions, and in nice models the degree of confidence for R_l would be the cumulative chi square probability $K_1(2l)$. Thus if the contours are indexed by the degree of confidence, the (3.1) recipe yields

$$-2\ell_{\text{ind}}(\beta) \le \chi^2_{1,1-\alpha} \Longleftrightarrow \beta \in R_{1-\alpha}.$$

As a simple example, let $F_c(\beta) = N((\beta - \widehat{\beta}_{\text{ind}})/\sigma)$. Then

$$\ell_{\text{ind}}(\beta) = -\tfrac{1}{2}(\beta - \widehat{\beta}_{\text{ind}})^2/\sigma^2 \tag{3.3}$$

and $L_{\text{ind}}(\beta) = \exp(\ell_{\text{ind}}(\beta)) \propto f_c(\beta)$ in this case. This is actually the only situation with the indirect likelihood being proportional to the confidence density. This is seen by inverting (3.2). For $\beta < \widehat{\beta}_{\text{ind}}$, the inverse of (3.2) is $F_c(\beta) = N(-(-2\ell_{\text{ind}}(\beta))^{1/2})$, while for $\beta > \widehat{\beta}_{\text{ind}}$ it is $N((-2\ell_{\text{ind}}(\beta))^{1/2})$. Differentiation yields

$$f_c(\beta) = (2\pi)^{-1/2} L_{\text{ind}}(\beta)\,|\ell'_{\text{ind}}(\beta)|/(-2\ell_{\text{ind}}(\beta))^{1/2}.$$

Thus, f_c is proportional to L_{ind} if and only if $\ell'_{\text{ind}}(\beta) = (-2\ell_{\text{ind}}(\beta))^{1/2}/\sigma$ for some positive σ. The solution to this differential equation is the Gaußian log-likelihood (3.3). This proves the following.

RESULT. *The indirect likelihood is proportional to the probability density of the confidence distribution if and only if the distribution is normal.*

The likelihood function is known to be transformation invariant. That is, if $\eta = \eta(\beta)$ is an increasing function of β, the likelihood for the parameter β is the transformed likelihood for η; $L(\beta) = L^\eta(\eta(\beta))$. When constructed from a confidence distribution, the indirect likelihood is indeed transformation invariant:

$$\ell_{\text{ind}}(\beta) = -\tfrac{1}{2}\{N^{-1}(F_c(\beta))\}^2 = -\tfrac{1}{2}[N^{-1}\{F_c^\eta(\eta(\beta))\}]^2.$$

The invariance is due to the indirect likelihood being defined in terms of the cumulative distribution and not through the density. The Bayesian prior is, of course, defined in terms of the density, and when the parameter is transformed, the Jacobian of the transform must be multiplied in. The presence of the Jacobian implies that the notion of 'non-informativity' is slightly problematic in Bayesian analysis. If the indirect likelihood is the indicator function for a subset of the parameter space, the elements of the subset are equally likely. If the prior density is uniform over an interval, the Bayesian would only be able to say that under the particular parametrisation, the parameter is uniformly distributed.

Efron(1993) introduced 'implied likelihood' based on a system of confidence intervals. His likelihood is related to the confidence density through a data-doubling argument. In the one-parametric case, Efron's implied likelihood agrees with our likelihood obtained from (3.1) when the confidence intervals are constructed by Efron's ABC method, and when the additive bias has been removed. Theorem 2 of Efron (1993) states that the likelihood is second order efficient in the sense that the ratio between the 'true' likelihood underlying the ABC-intervals and the indirect likelihood is of the order $1/n$.

Discrete and improper confidence distributions give rise to indirect likelihoods of interest. Consider the two-point distribution with mass $\frac{1}{2}$ in β_1 and β_2. By (3.2),

$$\ell_{\text{ind}}(\beta) = \begin{cases} 0 & \text{if } \beta_1 \le \beta \le \beta_2, \\ -\infty & \text{outside the interval.} \end{cases}$$

This is the appropriate indirect likelihood when $\beta_1 \le \beta \le \beta_2$ is known to be true, but no indirect data are available to discriminate between values inside the interval. If smoothness is desirable, the mixed normal distribution

$$F_c(\beta) = \tfrac{1}{2}N((\beta - \beta_1)/\sigma) + \tfrac{1}{2}N((\beta - \beta_2)/\sigma),$$

with σ small, will result in an indirect likelihood representing the inequality constraint $\beta_1 \le \beta \le \beta_2$.

A half-open inequality constraint, but with say the left interval being uncertain, could be represented by the improper confidence distribution $F_c(\beta) = \frac{1}{2}N((\beta - \beta_1)/\sigma)$, with σ reflecting the amount of uncertainty concerning the left endpoint, β_1. The $1-\alpha$-confidence intervals related to this improper distribution are the half-open intervals $[\beta_1 + \sigma N^{-1}(\alpha), \infty)$.

When F_c is only partially determined, one might want to find a smooth indirect likelihood that is reasonably consistent wt the summary information on F_c.

EXAMPLE 3.1: AN EXPONENTIAL TILTING TRANSFORM. To account for non-normality and skewness, suppose the confidence distribution can be represented as

$$\beta = \widehat{\beta}_{\text{ind}} + \sigma\{\exp(aZ) - 1\}/a, \quad \text{where } Z \sim N(0,1).$$

The possible range for β is from $\widehat{\beta}_{\text{ind}} - \sigma/a$ to infinity for a positive, and vice versa for a negative. When a goes to zero this includes the special case of a $N(\widehat{\beta}_{\text{ind}}, \sigma^2)$. With some manipulations one finds

$$\ell_{\text{ind}}(\beta) = -\tfrac{1}{2}\{a^{-1}\log(1 + a(\beta - \widehat{\beta}_{\text{ind}})/\sigma)\}^2.$$

Again, a going to zero corresponds to the normal log-likelihood. With knowledge of two or more quantiles, or two or more moments, the scale and acceleration parameters σ and a may be found in a given situation. Confidence intervals here are of the form $\widehat{\beta}_{\text{ind}} + \sigma\{\exp(\pm a z_{1-\alpha/2}) - 1\}/a$.

EXAMPLE 3.2: A POWER TRANSFORM OF THE NORMAL. Another three-parameter family with skewed confidence intervals, but with the full real line as support, is obtained from

$$F_c(\beta) = \{N((\beta - \widehat{\beta}_{\text{ind}})/\sigma)\}^{\lambda},$$

where $\lambda > 0$ is the power parameter. The indirect likelihood is then

$$\ell_{\rm ind}(\beta) = -\tfrac{1}{2}[N^{-1}\{N((\beta - \widehat{\beta}_{\rm ind})/\sigma)\}^{\lambda}]^2.$$

The related confidence intervals are

$$[\widehat{\beta}_{\rm ind} + \sigma N^{-1}((\tfrac{1}{2}\alpha)^{1/\lambda}),\ \widehat{\beta}_{\rm ind} + \sigma N^{-1}((1 - \tfrac{1}{2}\alpha)^{1/\lambda})].$$

If only the estimator $\widehat{\beta}_{\rm ind}$ and, say, a 90% confidence interval is given, σ and λ could be calculated. With more than three quantiles of F_c, a normal probability plot for various values of λ would be useful.

Example 3.3: Quantile smoothing. Assume that the confidence distribution is summarised by a series of (approximate) quantiles, say those representing left and right endpoints of confidence intervals $[F_c^{-1}(\frac{1}{2}\alpha),\ F_c^{-1}(1 - \frac{1}{2}\alpha)]$ for confidence degrees $1 - \alpha$ equal to 0.95, 0.90, 0.75, 0.50 and 0. Let the indirect data be $F_c(t_j) = p_j$ for $j = 1, \ldots, J$. A quantile–quantile plot of F_c against different distributions such as various transformed normals, various gamma or beta distributions, or Pearson type IV distributions, may lead to a confidence distribution that fits the indirect quantile data well.

If the search among standard distributions is unsuccessful, a smoothing spline may be fitted to the normal probability plot $(N^{-1}(F_c(t_j)),\ t_j)$. The smoother, $S(t)$, may also be constructed in a more careful fashion to reflect different uncertainty levels when setting the $t_j = F_c^{-1}(p_j)$ values. When found it leads to the indirect log-likelihood $\ell_{\rm ind}(\beta) = -\frac{1}{2}S(\beta)^2$.

3.2. Several parameters. The multiparameter case is more difficult, both operationally and interpretationally. The simplest situation is the one where the indirect information can be summarised in terms of a multivariate normal distribution. If the format is a log-likelihood function, the work is done. If, however, the format is a point estimate $\widehat{\beta}_{\rm ind}$ with an approximate covariance matrix Σ, then the standard elliptic confidence region with degree of confidence α is

$$T_\alpha = \{\beta\colon (\beta - \widehat{\beta}_{\rm ind})^{\rm t}\Sigma^{-1}(\beta - \widehat{\beta}_{\rm ind}) \le \chi^2_{\nu,\alpha}\},$$

with $\nu = \dim(\beta)$. The corresponding indirect log-likelihood is

$$\ell_{\rm ind}(\beta) = -\tfrac{1}{2}(\beta - \widehat{\beta}_{\rm ind})^{\rm t}\Sigma^{-1}(\beta - \widehat{\beta}_{\rm ind}). \tag{3.4}$$

We stress that this is not only an unrealistically simple textbook case; in situations with a reasonable amount of data supporting the indirect knowledge one would have approximate sufficiency and normality of the indirect point estimator $\widehat{\beta}_{\rm ind}$. This also motivates a general indirect likelihood definition as follows: If $T_\alpha = \{\beta\colon g(\beta) \le \chi^2_{\nu,\alpha}\}$ denotes the nested set of confidence sets, define $\ell_{\rm ind}(\beta) = -\frac{1}{2}g(\beta)$. Note in this connection that a function is completely characterised by its set of contours.

In non-Gaußian situations we recommend investing efforts in finding parameters, possibly with the help of transformations, for which the background information is independent from that of the others. Then one may concentrate on the several one-parameter problems in turn; the simultaneous indirect log-likelihood is simply the sum over the independent components.

There is a rich and recent literature both on parameter transformations towards orthogonality and on various higher-order corrections that could be made on the simple (3.4) construction. Key words include multidimensional skewness corrections, Bartlett identities and Edgeworth expansions; see Cox and Reid (1989) and Barndorff-Nielsen and Cox (1989, 1994). Some of these techniques are relevant in the present context too, but exploring this further takes us out of the natural bounds of this paper.

Throughout this section β has been used to denote an arbitrary statistical parameter. In typical applications of Likelihood Synthesis the β might be a component or a subvector of the full vector ϕ of descriptive parameters.

4. Likelihood Synthesis: algorithms and practical considerations.

Likelihood Synthesis is ordinary likelihood analysis of an overparametrised model restricted to a lower-dimensional manifold of the full parameter space.

4.1. Basic ideas. Let $\phi \in \Omega$ be the full K-dimensional parameter, with Ω an open connected region in $I\!R^K$, and let $L^f(\phi)$ be the full likelihood. The L^f function will of course depend on the chosen parameterisation. If indirect data are included in the format of a indirect likelihood, independence yields the full log-likelihood

$$\ell^f(\phi) = \ell_{\text{dir}}(\phi) + \ell_{\text{ind}}(\phi). \qquad (4.1)$$

The structural model, on top of the structure embedded in ℓ^f, is a q-dimensional sub-manifold $\mathcal{M} \in \Omega$. The log-likelihood to analyse is thus

$$\ell^f: \mathcal{M} \to I\!R.$$

In many cases, $\mathcal{M}$ represents a functional relationship $\phi = \Phi(\theta)$ where Φ is the model function (and not the normal cumulative). Since $\mathcal{M}$ is q-dimensional, so is θ. Let Θ in $I\!R^q$ be the parameter space for θ. The log-likelihood function may then be written

$$\ell(\theta) = \ell^f(\Phi(\theta)): \Theta \to I\!R. \qquad (4.2)$$

In some cases there will be canonical or at least quite natural choices for the basic parameter θ and the model function Φ. This will typically be the case when $\mathcal{M}$ represents a dynamical model, say in ecology or economy. Such models are often directed in time, and more easily calculated forwards than backwards. As an example, consider the simple population dynamics model in Section 5.2 below. From the natural input parameter θ, consisting of stock size in 1848, θ_1, and maximum sustainable yield rate, θ_2, output parameters like ϕ, consisting of stock size in 1993 and replacement yield that year, are easily calculated. In principle, however, θ could be solved for ϕ, and we could have cast the likelihood in terms of ϕ, or in some other related two-dimensional parameter.

In other cases, there are no immediately obvious input and output parameters. In the Fieller example 5.1 below, either θ_1 and θ_2 could serve as input and the other as output. To fix ideas, let θ be the input parameter of the deterministic model and let $\phi = \Phi(\theta)$ be the output parameter for which there is direct or indirect data available. The log-likelihood to investigate is (4.2). Let further $\gamma = \Gamma(\theta)$ be the parameter of primary interest. If $\dim(\gamma) = q = \dim(\theta)$, the likelihood function for γ is simply the parametric curve

$$(\Gamma(\theta), \ell(\theta)).$$

If $\dim(\gamma) < q$, we would use the profile log-likelihood

$$\ell^{\gamma}(\gamma) = \ell_{\mathrm{prof}}(\gamma) = \max\{\ell(\theta): \Gamma(\theta) = \gamma\}. \tag{4.3}$$

This constitutes Likelihood Synthesis.

Barndorff-Nielsen and Cox (1994) discuss several aspects of profile likelihood analysis, and many of the points made by them are directly relevant and useful in the kind of applications we have in mind for Likelihood Synthesis.

4.2. CALCULATION. In nice models, such as the Fieller model of Section 5.1 below, the profile likelihood can be obtained by analytical methods. If possible, this is to be preferred over numerical calculations.

When numerical methods are necessary, one would normally start with calculating $\ell(\theta)$ for a large number of values in the high likelihood region. For each of these, calculate the interest parameter, $\gamma = \Gamma(\theta)$. Assuming γ to be scalar, a plot of the scatter $(\Gamma(\theta), \ell(\theta))$ would be instructive. Since ℓ^{γ} is the profile, the scatter would show an approximate upper boundary that would be an estimate of ℓ^{γ}. If this boundary appears to be convex, a convex envelope could be fitted to the scatter as an estimate of ℓ^{γ}.

In some cases, one would expect $\ell(\theta)$ to be approximately Gaußian. To the scatter $(\theta, \ell(\theta))$ one might then fit a quadratic by regression analysis,

$$\ell(\theta) = -\tfrac{1}{2}(\theta - \widehat{\theta})^{t}\Sigma^{-1}(\theta - \widehat{\theta}) + e(\theta).$$

If the residuals $e(\theta)$ are small, the fit is successful. Since $\ell(\theta)$ is smooth, $e(\theta)$ would also be smooth, and one should not expect the residuals to be without structure, as is usually the case in regression analysis. If $\widehat{\theta} = \arg\max \ell(\theta)$ is found, say by a Newton-like method, before the regression was carried out, the regression equation is linear in the $q(q+1)/2$ parameters of the symmetric information matrix Σ^{-1} (preferably restricted by non-negative definiteness of Σ^{-1}). Having obtained a successful Gaußian fit, the profile likelihood for γ is found from this fit, preferably by analytical methods.

4.3. THE GAUSSIAN CASE. With $\ell(\theta) = -\frac{1}{2}(\theta - \widehat{\theta})^{t}\Sigma^{-1}(\theta - \widehat{\theta})$ being quadratic, and $\gamma = \Gamma\theta$ being linear, the maximum likelihood estimate of γ is $\widehat{\gamma} = \Gamma\widehat{\theta}$ and the profile likelihood is simply

$$\ell_{\mathrm{prof}}(\gamma) = -\tfrac{1}{2}(\gamma - \widehat{\gamma})^{t}(\Gamma\Sigma\Gamma^{t})^{-1}(\gamma - \widehat{\gamma}).$$

It is of interest to note that the Bayesian would come to the same inference. Assume that both the direct and indirect likelihood components are Gaußian, and that both are likelihoods representing confidence distributions that the Bayesian would use for his pre-model posteriors to be synthesised. In the simplest case, there is a posterior based on the direct data and a prior of dimension $q = \dim(\theta)$, and an additional prior on θ from that part of the confidence distribution that was not used with the direct data. The problem of

Bayesian Synthesis is then to integrate these two independent pre-model distributions into a synthesised posterior density. Raftery, Poole and Givens (1996) suggest that the appropriate synthesis is the geometric mean of the two pre-model densities. The two densities are, in fact, the two likelihood components since both are Gaußian. The synthesis of the Bayesian densities is thus the likelihood $L(\theta)$. Furthermore, the Bayesian would use the marginal distribution for γ as the resulting posterior for the parameter of interest. And this marginal distribution corresponds precisely to the profile $\ell_{\rm prof}(\gamma)$. Finally, the Bayesian probability regions for γ are exactly matched by the confidence regions constructed from $\ell_{\rm prof}$.

4.4. MORE GENERAL CASES. Likelihood Synthesis, with or without an indirect likelihood, is just ordinary likelihood analysis, and as such subject to occasional difficulties and pitfalls. The likelihood $\ell(\theta)$ may turn out to be multimodal or ill-behaved in other respects. Also, the calculation of the profile likelihood for an interest parameter γ might be difficult due to nonlinearities in $\gamma = \Gamma(\theta)$ and lack of smoothness or concavity in $\ell(\theta)$.

The intention is in any case to draw on the general results and experience available for likelihood inference. Thus approximate standard errors can be found by taking square roots of the diagonal elements of the inverse estimated information matrix, for example. Fine-tuning and corrections might be called for, using technology of bootstrapping or higher-order likelihood analysis, see again Barndorff-Nielsen and Cox (1994), for example. Another idea from classical asymptotics theory which may be useful in the present context is that of LeCam's efficient alternative to the maximum likelihood method, see LeCam and Yang (1990, Ch. 5.3). The idea involves fitting a local quadratic function in the right area of the parameter space.

REMARK: INDIRECT AND PENALISED LIKELIHOODS. We have discussed various aspects of the combined log-likelihood construction (4.1). Yet another interpretation is to view this a penalised likelihood approach; to find the best estimate of ϕ one maximises $\ell_{\rm dir}(\phi)$ penalised by the added factor $\ell_{\rm ind}(\phi)$. In the penalised likelihood tradition this would usually be written

$$\ell^*(\phi) = \ell_{\rm dir}(\phi) + \lambda\,\ell_0(\phi), \tag{4.4}$$

say, with an additional smoothing parameter λ determining the strength or not of the penalisation. Here ℓ_0 would be determined by pragmatic considerations about desirable smoothing effects rather than from bona fide prior considerations or from real data. Nevertheless this is of the same form as (4.1). Method (4.4) can also be given a Bayesian interpretation in that it corresponds mathematically to having a prior of the form $L_0(\phi)^\lambda$ for ϕ.

5. Examples and illustrations.

5.1. AN EXTENDED FIELLER METHOD. Raftery and Schweder (1993) discussed non-Bayesian and Bayesian methods of estimating the quotient $\gamma = \phi_1/\phi_2$, in a situation where separate information on ϕ_1 and ϕ_2 is available. The non-Bayesian method known as the Fieller technique (see Fieller, 1940 and 1954) can be applied when the available estimator $(\widehat\phi_1, \widehat\phi_2)$ is approximately binormally distributed.

Assume for simplicity that $\widehat\phi_1$ and $\widehat\phi_2$ are independent with $\widehat\phi_j$ being $N(\phi_j, \sigma_j^2)$ where σ_1 and σ_2 are known. Thus the log-likelihood is

$$\ell(\phi_1, \phi_2) = -\tfrac{1}{2}\{(\widehat\phi_1 - \phi_1)^2/\sigma_1^2 + (\widehat\phi_2 - \phi_2)^2/\sigma_2^2\}.$$

For given γ, the maximum likelihood estimate of ϕ_2 is the weighted mean $\widehat{\phi}_2(\gamma) = (\widehat{\phi}_1\gamma/\sigma_1^2 + \widehat{\phi}_2/\sigma_2^2)/(\gamma^2/\sigma_1^2 + 1/\sigma_2^2)$, and the profile log-likelihood for γ becomes

$$\ell_{\text{prof}}(\gamma) = -\tfrac{1}{2}(\widehat{\phi}_1 - \gamma\widehat{\phi}_2)^2/(\sigma_1^2 + \gamma^2\sigma_2^2).$$

This agrees with the Fieller technique of using $\widehat{\gamma} = \widehat{\phi}_1/\widehat{\phi}_2$ as the point estimate and

$$\widehat{\gamma}\{1 \pm z_0(c_1^2 + c_2^2 - z_0^2 c_1^2 c_2^2)^{1/2}\}/(1 - z_0^2 c_2^2)$$

as the confidence interval, where $c_j = \sigma_j/\widehat{\phi}_j$ are the coefficients of variation and z_0 the appropriate normal quantile.

With additional independent information on γ, say with γ having the Gaußian indirect log-likelihood

$$\ell_{\text{ind}}(\gamma) = -\tfrac{1}{2}(\gamma - \widehat{\gamma}_{\text{ind}})^2/\widehat{\tau}^2,$$

the combined log-likelihood becomes

$$\ell(\gamma) = \ell_{\text{prof}}(\gamma) + \ell_{\text{ind}}(\gamma) = -\tfrac{1}{2}(\widehat{\phi}_1 - \gamma\widehat{\phi}_2)^2/(\sigma_1^2 + \gamma^2\sigma_2^2) - \tfrac{1}{2}(\gamma - \widehat{\gamma}_{\text{ind}})^2/\widehat{\tau}^2.$$

The Fieller technique can be extended to multivariate data. Let $\widehat{\phi}_1$ be a p-dimensional vector of observed numerators and $\widehat{\phi}_2$ a p-dimensional vector of denominators. Thus there are $K = 2p$ descriptive parameters. Assume the two $\widehat{\phi}_j$ to be $N_p(\phi_j, \Sigma_j)$ and independent. The structural model is $\phi_1 = \gamma(\theta) * \phi_2$ with '$*$' denoting element-wise multiplication, and with θ being a q-dimensional vector of parameters determining the vector of ratios $\gamma = \phi_1/\phi_2$. For given θ, the maximum likelihood estimate of ϕ_1 and ϕ_2 is readily obtained, and the profile log-likelihood for θ is found to be

$$\ell_{\text{prof}}(\theta) = -\tfrac{1}{2}(\widehat{\phi}_1 - \gamma(\theta) * \widehat{\phi}_2)^{\text{t}}(\Sigma_1 + D(\theta)\Sigma_2 D(\theta))^{-1}(\widehat{\phi}_1 - \gamma(\theta) * \widehat{\phi}_2). \qquad (5.1)$$

Here $D(\theta)$ is the diagonal matrix with elements $\gamma(\theta)$. The structural model might well be a transformed linear model with categorical or continuous covariates. A case of a log-linear model is discussed below. It is also possible to depart from the assumptions of numerators and denominators being independent, the scale of the covariance matrices being known or the distribution being normal.

The ratio estimators are used in estimating animal abundance from line transect surveys. In a shipborn survey of minke whales in the northeastern Atlantic in 1995, vessels were outfitted with two independent observational platforms, see Schweder, Skaug, Dimakos, Langaas and Øien (1997). For each platform and survey area, the ratio of the number of whales seen to the estimated effective strip area is an estimate of the spatial whale density. Denoting by $\widehat{\phi}_{1A}$ the count for platform A, $\widehat{\phi}_{1B}$ the count for platform B and $\widehat{\phi}_1$ the vector of counts, and similarly for the estimated effective strip areas, a reasonable assumption is that $\widehat{\phi}_j$ is $N_2(\phi_j, \Sigma_j)$ and independent for $j = 1, 2$. Assume for simplicity that the covariance matrices are known. The correlation between the counts is due to the two platforms being exposed to the same whales throughout the cruise, and the correlation between the effective strip areas is due to the same transect being run for both platforms and the effective strip width being estimated from a common parametric model. The parameter representing whale density, θ, is related to the two mean vectors simply by $\phi_1 = \theta\phi_2$, since in this case $\gamma = (\theta, \theta)^t$.

In Schweder et al. (1997) the whale density was estimated by $\widetilde{\theta} = (\widehat{\phi}_{1A} + \widehat{\phi}_{1B})/(\widehat{\phi}_{2A} + \widehat{\phi}_{2B})$. The maximum likelihood estimate, $\widehat{\theta}$, is different from the simple quotient estimate $\widetilde{\theta}$. For survey area Gåsbanken in the eastern part of the Barents Sea, the counts and estimated effective search areas from the 1995 cruise are:

$$\widehat{\phi}_1 = \begin{pmatrix} 35 \\ 33 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 97.7 & 17.7 \\ 17.7 & 78.6 \end{pmatrix},$$
$$\widehat{\phi}_2 = \begin{pmatrix} 576 \\ 445 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 1329 & 881 \\ 881 & 1437 \end{pmatrix}.$$

Here the unit for ϕ_1 is number of minke whales seen under primary search effort, and that for ϕ_2 is km^2 under primary search effort.

The maximum likelihood estimate is $\widehat{\theta} = 0.0661$ as compared to $\widetilde{\theta} = 0.0666$, while the 95% confidence interval obtained from $\{\theta : 2(\ell_{\text{prof}}(\widehat{\theta}) - \ell_{\text{prof}}(\theta)) \leq \chi^2_{1;.95}\}$ is [0.0379, 0.0965] and the corresponding interval based on the profile likelihood for the sum counts over the sum effective area is [0.0382, 0.0972]. The profile likelihood based on (5.1) is for the above data, in fact, very similar to that based on the sum count and the sum area. With other data, for example with less balance between effective areas and whales seen, the simple Fieller method based on the sums might come out considerably different from the extended Fieller estimate based on (5.1).

With $p \geq 2$ it is possible to formulate hypotheses related to homogeneity or other structural aspects that could be tested by likelihood ratio methods. As a simple example, consider the above application, with structural model $\phi_1 = \gamma * \phi_2$ with no restriction on the $p = 2$-dimensional vector γ (other than its components being positive). The hypothesis of homogeneity, $\gamma_1 = \gamma_2 = \theta$, is then simply tested by the likelihood ratio statistic $-2\ell_{\text{prof}}(\widehat{\gamma})$ since the profile likelihood equals 0 at the optimum when ϕ_1 and ϕ_2 are allowed to be different. In the above numerical example, $-2\ell_{\text{prof}}(\widehat{\gamma}) = 0.33$, and the homogeneity hypothesis stands.

Time series of ratio estimates are sometimes used for estimating growth rates in animal populations. In Raftery and Zeh (1996), count and coverage was estimated for $p = 10$ years between 1978 and 1993. The log-linear model $\phi_1 = \exp(\theta_0 + \theta_1 t) * \phi_2$, where t is the vector of time points, was estimated by weighted linear regression of $\log(\widehat{\phi}_1/\widehat{\phi}_2)$ on t. The multivariate Fieller technique is, however, more efficient when say $\widehat{\phi}_1$ and $\widehat{\phi}_2$ are multivariate normal, as seems reasonable. The coverage, ϕ_2, was only independently estimated for 5 of the 10 years, and for the remaining years the average over these five years was used. This induces a correlation structure in $\widehat{\phi}_2$.

In the general case there might be a vector of nuisance parameters, σ, in addition to the interest parameters ϕ_1 and ϕ_2. The likelihood based on the direct data, $L^f(\phi_1, \phi_2; \sigma)$ will often factor since there will be independent data for the numerators and the denominators. With the structural model, $\phi_1 = \gamma(\theta) * \phi_2$, with γ possibly being a a transformed linear model in observed covariates with regression vector θ, the direct likelihood is $L_{\text{dir}}(\theta; \phi_2, \sigma) = L^f(\gamma(\theta)\phi_2, \phi_2; \sigma)$. If there also are indirect data available, with indirect likelihood $L_{\text{ind}}(\theta)$, the log-likelihood obtained by Likelihood Synthesis is

$$\ell(\theta) = \ell_{\text{dir}}(\theta; \widehat{\phi}_2(\theta), \widehat{\sigma}(\theta)) + \ell_{\text{ind}}(\theta).$$

Whether this is a well-behaved likelihood function depends primarily on the fit of the structural model to the direct data, which might be tested by a likelihood ratio test, and also the coherence between the indirect data and the direct data, which also might be tested by a likelihood ratio test. Only in special cases, as with normally distributed statistics $\widehat{\phi}_1$ and $\widehat{\phi}_2$ and with covariance matrices known up to a scale factor, will $\ell(\theta)$ have an analytic form.

5.2. BOWHEAD WHALES. This example illustrates indirect likelihood and Likelihood Synthesis in contrast to Bayesian priors and Bayesian Synthesis. The story behind it is interesting. In 1994 the Scientific Committee of the International Whaling Commission agreed on a stock assessment of bowhead whales migration and feeding north of Alaska and Northwestern Canada. This stock is of great interest from a conservationist point of view since it was heavily harvested by yankee whalers down to a very low level around 1900, and it is the only stock of bowhead whales that has recovered substantially from commercial whaling. The stock is also interesting from a whaling point of view since it is still harvested by Alaskan Eskimos.

The Bayesian Synthesis method was constructed to estimate the status of the stock of bowhead whales (International Whaling Commission, 1995), but management advice had to be reconsidered (International Whaling Commission, 1996) due to lack of transformation invariance (Schweder and Hjort 1996) and also some other questions including whether the 'forwards' or the 'backwards' variant of the method is appropriate (Punt and Butterworth, 1996).

Raftery, Poole and Zeh (1996) presented a new version of Bayesian Synthesis based on geometric pooling of the independent prior distributions that effectively would go into their synthesis. To investigate this method, they tried it in a simplified population dynamics model for bowhead whales. We use the same deterministic age-structured Leslie matrix model for the dynamics of the stock, and we re-interpret their prior probability densities as densities in confidence distributions yielding indirect likelihoods.

At the beginning of year t, there are P_t bowhead whales. With C_t being the catch in year t, the dynamical model is

$$P_{t+1} = P_t - C_t + 1.5\,\theta_2 P_t\{1 - (P_t/\theta_1)^2\},$$

where θ_1 is the carrying capacity and θ_2 the maximum sustainable yield rate. Set $t = 1$ for 1848 and assume $P_1 = \theta_1$, since the yankee whaling started in 1848. The catch series C_t for $t = 1, \ldots, 145$ (in 1992) is assumed to be exactly known. Raftery, Poole and Givens (1996) list the following independent priors for θ_1, θ_2 and $\phi = P_{1993}$:

$$\begin{aligned}\theta_1 &\sim 6400 + \mathrm{Gamma}(2.81, 0.000289),\\ \theta_2 &\sim \mathrm{Gamma}(8.2, 372.7),\\ \phi &\sim N(7800, 1300^2),\end{aligned}$$

where Gamma(a, b) denotes the distribution with density proportional to $t^{a-1}\exp(-bt)$ (it has mean value a/b and standard deviation $a^{1/2}/b$). In addition there was a fourth piece of independent information, to the effect that the value $\widehat{\phi} = 8293$ was observed from a $N(\phi, 626^2)$ distribution. The conceptual and practical problem Raftery, Poole and Givens wanted to handle was that of having three independent priors in a model of dimension two.

For Likelihood Synthesis it is no problem with having more (or fewer) independent pieces of indirect data than there are dimensions in the model. Since independent components of the indirect likelihood are treated in the same way as independent components of the direct likelihood, it does not matter for the total likelihood whether a piece of data is regarded as direct of indirect. To illustrate, take the Gamma distribution for θ_2 as a confidence distribution yielding the indirect log-likelihood

$$\ell_{\text{ind}}(\theta_1, \theta_2) = -\tfrac{1}{2}\{N^{-1}(G_{8.2,372.7}(\theta_2))\}^2$$

(where $G_{a,b}$ is the cumulative Gamma distribution function), while the Gamma distribution for θ_1 and the two normal distributions for ϕ are taken as confidence distributions giving components of the direct log-likelihood

$$\ell_{\text{dir}}(\theta_1, \theta_2, \phi) = -\tfrac{1}{2}\left[\{N^{-1}(G_{2.81,0.000289}(\theta_1 - 6400))\}^2 + (\phi - 7800)^2/1300^2 + (\phi - 8293)^2/626^2\right].$$

The total log-likelihood is then

$$\ell(\theta_1, \theta_2) = \ell_{\text{ind}}(\theta_1, \theta_2) + \ell_{\text{dir}}(\theta_1, \theta_2, \Phi(\theta_1, \theta_2)),$$

where Φ is the deterministic population dynamics model result for $\phi = P_{1993}$.

The $\widehat{\phi} = 8293$ is the stock abundance estimate obtained from the 1993 survey. The other pieces of information, summarised as pre-1993 priors above, are also based on considerable amounts of data. The Gamma distribution for the productivity parameter θ_2, is, for example, based on the estimated growth rate in a time series of abundance estimates prior to 1993. We will investigate the profile log-likelihood for θ_2, as composed of the direct profile log-likelihood and the indirect one.

— *Figure (see page 24), to be placed around here* —

FIGURE. *Profile log-likelihoods for θ_2, maximum sustainable yield rate. Exact profile log-likelihoods are shown for the indirect (dotted), the direct (small circles, upper curve) and total (small circles, lower curve) cases. Three-parameter log-likelihood approximations to the direct and total ones, using the exponentially tilted normal family, are also shown.*

The figure shows the exactly calculated indirect, direct and total log-likelihoods; the first uses the Gamma assumption while the two latter were computed numerically for θ_2 values equal to $0.007, 0.008, \ldots, 0.050$. For the direct and total cases, log-likelihood approximations, using the exponentially tilted normal as in Example 3.1, are also shown, having been fitted by least squares for each component. The fit is remarkably good. To each fitted component, a 95% confidence interval is calculated from the formula

$$\widehat{\theta}_2 + \sigma\{\exp(\pm 1.96\, a) - 1\}/a.$$

For the indirect likelihood case the interval is computed exactly from the Gamma distribution. The results, together with the results of the two versions of Bayesian Synthesis (Raftery, Poole and Givens, 1996) are shown in the following table.

TABLE. *The parameters $\widehat{\theta}_2$, a, σ of the separately fitted exponentially tilted normal log-likelihoods to the indirect, direct and total log-likelihoods, respectively. For each fitted component, 95% confidence intervals (lower, upper) are given, and also two 95% Bayesian credibility intervals taken from Raftery, Poole and Givens (1996).*

	lower	*estimate*	*upper*	a	σ
indirect likelihood	0.0096	0.0211	0.0394		
direct likelihood	0.0000	0.0193	0.0896	0.650	0.01776
total likelihood	0.0102	0.0209	0.0373	0.220	0.00670
Bayesian, forwards	0.0077	0.0157	0.0295		
Bayesian, backwards	0.0069	0.0206	0.0464		

Our results agree more with the results from the backwards Bayesian Synthesis method than with the forwards variant, but our confidence interval is shorter. The most important difference however is that Likelihood Synthesis yields unique results and that these follow directly from the likelihood principle, with likelihoods constructed from confidence distributions.

When comparing the indirect and the direct components in the table and figure, it is evident that most of the information lies with the indirect component. If information is measured by square root length of the 95% confidence intervals, there is 1.8 times as much information in the indirect data as in the direct data. This is actually as expected since the indirect data summarise a series of earlier abundance estimates with respect to growth rate, which is the primary source of information on productivity parameters like θ_2.

6. Supplementing remarks.

REMARK 1. A strategy that has been used occasionally when dealing with complicated nuisance parameters is to first do the analysis conditional on such parameters, and then average with respect to their prior distribution. More specifically, if say $\widehat{\theta}_\gamma$ is the estimator obtained conditionally on an γ parameter, the suggestion is to use

$$\widehat{\theta}_{(1)} = \int \widehat{\theta}_\gamma \pi(\gamma)\,\mathrm{d}\gamma, \tag{6.1}$$

where π is the prior density. This appears to be the general proposal made in Restrepo, Hoenig, Powers, Baird and Turner (1992; see in particular their p. 742).

This is not a good strategy. In a full Bayesian framework the best estimator (with squared error loss) is the conditional mean, which can also be expressed via iterated expectations,

$$\widehat{\theta}_{(2)} = \mathrm{E}(\theta \,|\, \mathrm{data}) = \mathrm{E}\{\mathrm{E}(\theta \,|\, \gamma,\, \mathrm{data})\} = \int \widehat{\theta}_\gamma \pi(\gamma \,|\, \mathrm{data})\,\mathrm{d}\gamma. \tag{6.2}$$

One might, somewhat heuristically, take the last formula as a general strategy, that is, even for other $\widehat{\theta}_\gamma$ estimators than the Bayesian one. The argument suggests that it would at least be better to use the posterior and not the prior density for the nuisance parameter in (6.1).

A simple example demonstrating this is as follows. Let $x_1, \ldots, x_n$ be independent and normal (μ, σ^2), and assume there is a prior $N(\mu_0, \tau_0^2)$ for μ. The maximum likelihood estimator for the variance conditional on μ is $\widehat{\sigma}^2_\mu = n^{-1}\sum_{i=1}^n (x_i - \mu)^2$. The (6.1) suggestion leads to

$$\widehat{\sigma}^2_{(1)} = \mathrm{E}_{\mathrm{prior}}\, \widehat{\sigma}^2_\mu = n^{-1}\sum_{i=1}^n (X_i - \mu_0)^2 + \tau_0^2,$$

which has obvious drawbacks; it converges to $\sigma^2 + (\mu - \mu_0)^2 + \tau_0^2$ as n grows, for example, seriously overshooting the intended estimand. Suggestion (6.2) gives a more reasonable result,

$$\widehat{\sigma}^2_{(2)} = \mathrm{E}_{\mathrm{posterior}}\, \widehat{\sigma}^2_\mu = n^{-1} \sum_{i=1}^n (x_i - \widehat{\mu})^2 + \widehat{\tau}^2,$$

where $\widehat{\mu}$ and $\widehat{\tau}^2$ are posterior mean and variance for μ. We would typically have $\widehat{\mu}$ close to $\bar{x}$ and $\widehat{\tau}^2$ close to $\widehat{\sigma}^2/n$, showing that the estimate above is perhaps only slightly bigger than it should be. It is asymptotically equivalent to the full and correct maximum likelihood solution.

A more elaborate full-Bayes version of this example is as follows, utilising the traditional conjugate prior for (μ, σ) in the normal model. To describe the prior, let $\lambda = 1/\sigma^2$ be Gamma $(\frac{1}{2}a, \frac{1}{2}b)$, and let μ for given σ be a $N(\mu_0, \sigma^2\tau_0^2) = N(\mu_0, p^{-1}\lambda^{-1})$, writing $p = 1/\tau_0^2$ for this precision parameter. The full likelihood for $(\lambda, \mu, \mathrm{data})$ is proportional to

$$\lambda^{\frac{1}{2}(a+n)-1} \exp[-\tfrac{1}{2}\lambda\{b + n\widehat{\sigma}_0^2 + pn(\bar{x} - \mu_0)^2/(p+n)\}]\lambda^{1/2} \exp\{-\tfrac{1}{2}\lambda(p+n)(\mu - \widehat{\mu})^2\},$$

where $\widehat{\sigma}_0^2$ is the traditional $n^{-1}\sum_{i=1}^n (x_i - \bar{x})^2$. This shows that

$$\begin{aligned} \lambda \,|\, \mathrm{data} &\sim \mathrm{Gamma}\{\tfrac{1}{2}(a+n), \tfrac{1}{2}(b + n\widehat{\sigma}_0^2 + pn(\bar{x} - \mu_0)^2/(p+n))\}, \\ \mu \,|\, \lambda, \mathrm{data} &\sim N(\widehat{\mu}, (p+n)^{-1}\lambda^{-1}), \end{aligned}$$

in which $\widehat{\mu} = \mathrm{E}(\mu \,|\, \mathrm{data}) = (n\bar{x} + p\mu_0)/(n+p)$.

Since $\widehat{\mu}$ is independent of σ both strategies agree on $\widehat{\mu}$ as the estimate of μ. The situation is different for σ^2, however. In the μ-conditional framework, $\lambda \,|\, \mathrm{data}$ is a Gamma $\{\frac{1}{2}(a + n), \frac{1}{2}(b + n\widehat{\sigma}_0^2 + n(\bar{x} - \mu_0)^2\}$, with resulting

$$\widehat{\sigma}^2_\mu = \mathrm{E}(\lambda^{-1} \,|\, \mu, \mathrm{data}) = \{b + n\widehat{\sigma}_0^2 + n(\bar{x} - \mu)^2\}/(a + n - 2).$$

Strategy (6.2) gives the full Bayes estimate

$$\widehat{\sigma}^2_{(2)} = \frac{b + n\widehat{\sigma}_0^2 + pn(\bar{x} - \mu_0)^2/(p+n)}{a + n - 2},$$

which is quite close in probability to simply $\widehat{\sigma}_0^2$ as n becomes large. Method (6.1), on the other hand, gives a far too large variance estimator.

Remark 2. Bayesians have developed various strategies for obtaining prior distributions that are based on data and not on subjective beliefs, as a response to the criticism that science should be based on data, with as little subjective judgement involved as possible. Empirical Bayes is the most successful of these strategies. By allowing subjectivity in likelihood analysis by way of indirect likelihoods, the objectivists might be worried.

That indirect likelihood could be misused to introduce undue subjective beliefs or prejudice into the statistical analysis is not a valid critique. By insisting that each independent piece of the likelihood, both the direct and the indirect likelihood based on indirect data, is well documented and argued, the reader would know what the basis is for the inference. The main use of indirect likelihood is to allow indirect *data* to enter the analysis. There might occasionally be a place for expert judgements that are not directly based on specific reported data. That this can be accommodated by the indirect likelihood is, in our view, rather an asset of the concept than a weak point.

An important difference between Bayesian analysis and likelihood analysis is worth noting in this context. Let the likelihood of the direct data, ℓ_{dir}, be cast in the q-dimensional parameter θ. Without a prior distribution in q dimensions, the Bayesian will be unable to carry out his analysis. This will often force him to use priors with a weak basis in reported data. The likelihood analyst will, on the other hand, in principle be able to work with $\ell_{\text{dir}}(\theta) + \ell_{\text{ind}}(\theta)$ for any indirect likelihood that is well supported, even if ℓ_{ind} is non-informative in the likelihood sense in one or more dimensions. He will thus have no need to elicit (from experts or from himself) an indirect likelihood with weak support.

Acknowledgements. We are grateful to B. Efron and C.E. Priebe for comments leading to improvement of an earlier version of this paper.

References

Barndorff-Nielsen, O. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

Barndorff-Nielsen, O. and Cox, D.R. (1994). *Inference and Asymptotics.* Chapman and Hall, London.

Birnbaum, A. (1962). On the foundations of statistical inference (with discussion contributions). *Journal of the American Statistical Association* **57**, 269–306.

Bjørnstad, J. (1996). On the generalization of the likelihood function and the likelihood principle. *Journal of the American Statistical Association* **91**, 791–806.

Bravington, M. (1996). An appraisal of Bayesian Synthesis, with suggested modifications and diagnostics. *Rep. Int. Whal. Commn.* **46**, 531–540.

Cox, D.R. and Reid, N. (1989). Parameter orthogonality and approximate conditional inference (with discussion contributions). *Journal of the Royal Statistical Society* **B 49**, 1–39.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika* **80**, 3–26.

Efron, B. (1994). Missing data, imputation, and the bootstrap (with discussion contribution). *Journal of the American Statistical Association* **89**, 463–479.

Efron, B. (1996). Empirical Bayes methods for combining likelihoods (with discussion contributions). *Journal of the American Statistical Association* **91**, 538–565.

Efron, B. (1997). R.A. Fisher in the 21$^{\text{st}}$ century. Technical report, Department of Statistics, Stanford University.

Edwards, A.W.F. (1992). *Likelihood* (expanded edition; first edition 1972). Cambridge University Press.

Edwards, A.W.F. (1997). Three early papers on efficient parametric estimation (introduction to three papers by R.A. Fisher). *Statistical Science* **12**, 35–38.

Fieller, E.C. (1940). The biological standardisation of insulin. *Journal of the Royal Statistical Society*, Suppl. 7, 1–65.

Fieller, E.C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society* **B 16**, 175–185.

Geisser, S. (1992). Introduction to Fisher's 1922 paper, 'On the mathematical foundations of theoretical statistics'. In *Breakthroughs in Statistics* **I** (eds. S. Kotz and N.L. Johnson), 1–10.

Genest, C. and Zidek, J.V. (1986). Combining probability distributions: a critique and an annotated bibliography (with discussion contributions). *Statistical Science* **1**, 114–148.

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. *Reports of the International Whaling Commission* **46**, 345–364.

International Whaling Commission (1995). *Rep. Int. Whal. Commn.* **45**, p. 144.

International Whaling Commission (1996). *Rep. Int. Whal. Commn.* **46**, p. 74.

LeCam, L. and Yang, G.L. (1990). *Asymptotics in Statistics. Some Basic Concepts.* Springer-Verlag, New York.

Punt, A.E. and Butterworth, D.S. (1996). Further remarks on the Bayesian approach for assessing the Bering–Chukchi–Beaufort Seas stock of bowhead whales. *Rep. Int. Whal. Commn.* **46**, 481–491.

Raftery, A.E., Givens, G.H. and Zeh, J.E. (1995). Inference from a deterministic population dynamics model for bowhead whales (with discussion contributions). *Journal of the American Statistical Association* **90**, 401–430.

Raftery, A.E., Poole, D. and Givens, G.H. (1996). The Bayesian Synthesis assessment method: Resolving the Borel paradox and comparing the backwards and forwards variants. Paper SC/48/AS16 presented to the Scientific Committee of the International Whaling Commission.

Raftery, A.E. and Schweder, T. (1993). Inference about the ratio of two parameters, with application to whale censusing. *American Statistician* **47**, 259–264.

Raftery, A.E. and Zeh, J.E. (1996). Estimating bowhead whale *(Balanea mysticetus)* population size and rate of increase from the 1993 census. Technical Report no. 303, Department of Statistics, University of Washington.

Restrepo, V.R., Hoenig, J.M., Powers, J.E., Baird, J.W., and Turner, S.C. (1992). A simple simulation approach to risk and cost analysis, with applications to swordfish and cod fisheries. *Fishery Bulletin* **90**, 736–748.

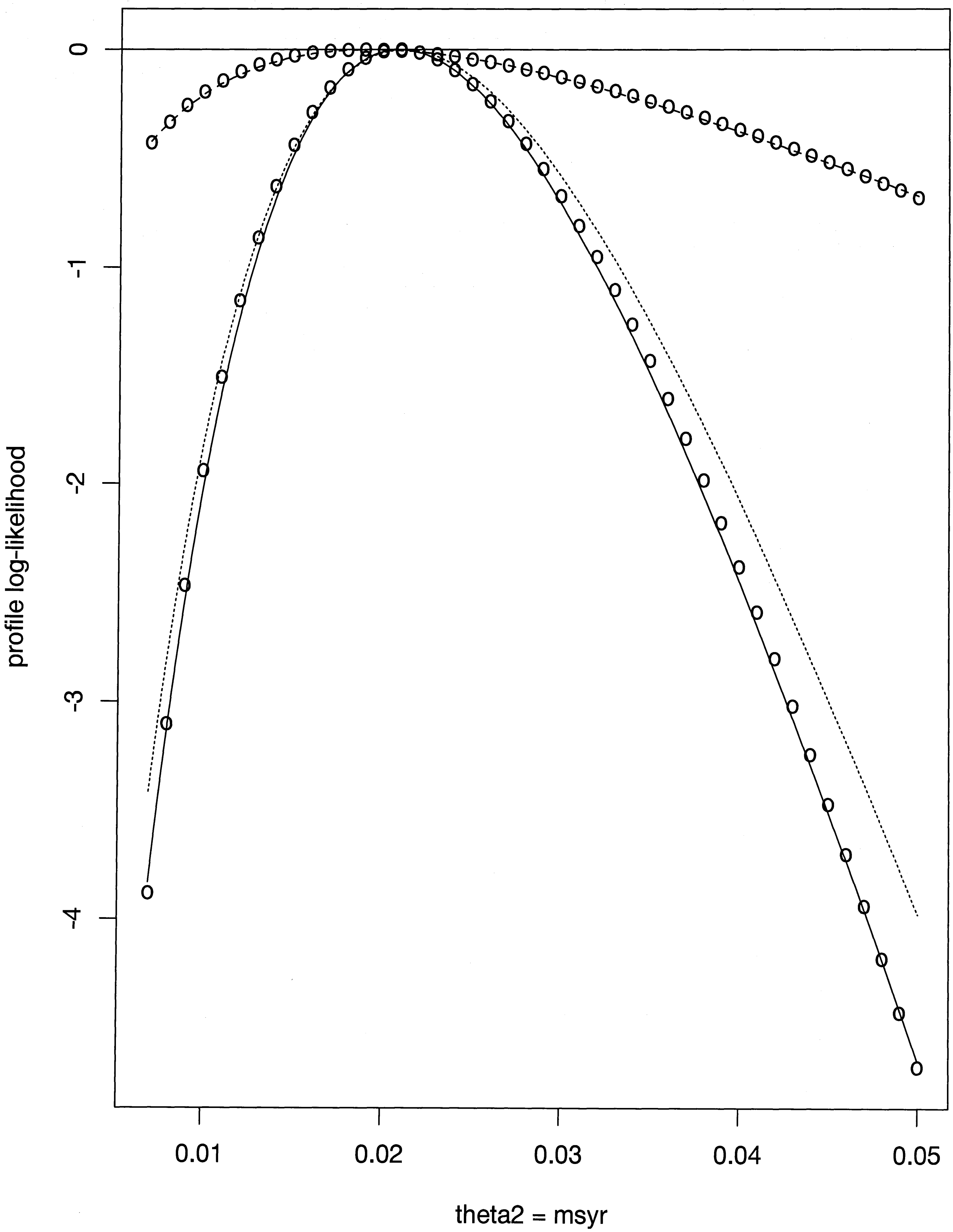
Savage, L.J.(1962). Discussion contribution to A. Birnbaum's paper. *Journal of the American Statistical Association* **57**.

Schweder, T. (1988). A significance version of the basic Neyman–Pearson theory for scientific hypothesis testing (with discussion contribution). *Scandinavian Journal of Statistics* **15**, 225–242.

Schweder, T. and Hjort, N.L. (1996). Bayesian Synthesis or Likelihood Synthesis – what does Borel's paradox say? *Rep. Int. Whal. Commn.* **46**, 475–479.

Schweder, T., Skaug, H.J., Dimakos, X.K., Langaas, M. and Øien, N. (1997). Abundance of northeastern Atlantic minke whales, estimates for 1989 and 1995. Paper SC/48/NA1 presented to the Scientific Committee of the International Whaling Commission. 84 pp., in press.

Spiegelhalter, D.J., Freedman, L.S. and Parmar, M.K.P. (1996). Bayesian approaches to randomized trials. In *Bayesian Biostatistics* (eds. D.A. Berry and D.K. Stangl), 67–108. Marcel Dekker.



0
-1
-2
-3
-4
profile log-likelihood
0.01
0.02
0.03
0.04
0.05
theta2 = msyr