

Integrative fish stock assessment by frequentist methods: confidence distributions and likelihoods for bowhead whales*

TORE SCHWEDER

Department of Economics, University of Oslo, Box 1095 Blindern, 0317 Oslo, Norway.
E-mail: tore.schweder@econ.uio.no

SUMMARY: Frequentist concepts of confidence distribution, reduced likelihood and likelihood synthesis are presented as tools for integrative statistical analysis, leading to distributional inference. Bias correction, often achieved by simulations, is emphasised. The methods are illustrated on data concerning bowhead whales off Alaska, and they are briefly compared with Bayesian methods.

Key words: bowhead whales, likelihood synthesis, confidence distribution, reduced likelihood, meta-analysis.

INTRODUCTION

Fish stock assessments are often based on data from different and independent sources. Methods for integrating diverse data have been somewhat neglected in the frequentist statistical tradition. There, emphasis has usually been on one set of data without much regard to previous relevant data, or to the future use of the data in question. Many fisheries scientists have therefore turned to Bayesian methods. Representing the results in distributional terms, e.g. posterior distributions, adds attraction to the Bayesian approach. There are, however, frequentist methods available for integrative analysis, and distributional inference. The purpose of this paper is to present the basic concepts of confidence distributions, reduced likelihoods and likelihood synthesis. The methodology is illustrated by some recent applications to the assessment of bowhead whales off

Alaska. Bayesian methods are also briefly compared with the frequentist methodology in the context of fish stock assessment.

In the frequentist tradition, the properties of a statistical method are analysed in probabilistic terms *ex ante*, before data are observed. The main property of a method of confidence interval estimation is, for example, the coverage frequency in repeated applications. *Ex post*, one is left with an observed confidence interval. The given degree of confidence is a property of this interval, a property that is inherited from the coverage probability of the interval estimation method *ex ante*. Wade (1999) discusses and compares frequentist, Bayesian and likelihood methods in the context of fish stock assessment. He comments that what is lacking from the frequentist methodology is a “strength of evidence” type argument, and he is critical of the widespread pre-specification of the level of significance in hypothesis testing and of the degree of confidence in interval estimation. The confidence distribution provides,

*Received December 6, 2000. Accepted January 9, 2002.

however, the “strength of evidence” type argument that Wade (1999) is missing. The confidence distribution is a generalisation and unification of confidence intervals and p-values. It can be regarded as the frequentist analogue of the Bayesian posterior distribution. Confidence distributions are not uncommon in fisheries assessment (Gavaris, 1999; Patterson *et al.*, 2001), but often half-baked as bootstrap distributions.

Patterson *et al.* (2001) discuss briefly the merits of the Bayesian approach to fish stock assessment. They also discuss frequentist methods aiming at confidence distributions representing the knowledge and uncertainty that the data allow for. Patterson *et al.* (2001) mention that the standard frequentist methodology does not offer a structured framework to incorporate prior information. They do mention the method of likelihood synthesis (Schweder and Hjort, 1996; 2002), but remark correctly that this approach is not in common use.

CONFIDENCE DISTRIBUTIONS

Fisher (1930) introduced the concept of fiducial probability distributions. Fisher could not accept the use of Bayesian inference based on flat priors, and his fiducial distribution was meant as a replacement for the Bayesian posterior distribution. Efron (1998) and others prefer the term “confidence distribution” rather than “fiducial probability distribution”. The defining property of a confidence distribution is that its quantiles span all possible confidence intervals. Confidence distributions and fiducial probability distributions are essentially the same thing.

Consider a statistical model for the data X . The model consists of a family of probability distributions for X , indexed by the vector parameter (ψ, χ) , where ψ is a scalar parameter of primary interest, and χ is a nuisance parameter (vector).

Definition 1

A univariate data-dependent distribution for ψ , with cumulative distribution function $C(\psi; X)$ and with quantile function $C^{-1}(\alpha; X)$ is an exact confidence distribution if

$$P_{\psi\chi}(\psi \leq C^{-1}(\alpha; X)) = P_{\psi\chi}(C(\psi; X) \leq \alpha) = \alpha$$

for all $\alpha \in (0,1)$ and for all probability distributions in the statistical model.

By definition, the stochastic interval $(-\infty, C^{-1}(\alpha; X))$ covers ψ with probability α , and is a one-sided confidence interval method with coverage probability α . The interval $(C^{-1}(\alpha; X), C^{-1}(\beta; X))$ will for the same reason cover ψ with probability $\beta - \alpha$, and is a confidence interval method with this coverage probability. When data have been observed as $X = x$, the realised numerical interval $(C^{-1}(\alpha; x), C^{-1}(\beta; x))$ will either cover or not cover the unknown true value of ψ . The degree of confidence $\beta - \alpha$ that is attached to the realised interval is inherited from the coverage probability of the stochastic interval. The confidence distribution has the same dual property. *Ex ante* data, the confidence distribution is a stochastic entity with probabilistic properties. *Ex post* data, however, the confidence distribution is a distribution of confidence that can be attached to interval statements. For simplicity, I will speak of confidence instead of “degree of confidence”.

The realised confidence $C(\psi_0; x)$ is the p-value of the one-sided hypothesis $H_0 : \psi \leq \psi_0$ versus $\psi > \psi_0$ when data have been observed to be x . The *ex ante* confidence, $C(\psi_0; X)$ is from the definition uniformly distributed. The p-value is just a transformation of the test statistic to the common scale of the uniform distribution (*ex ante*). The realised p-value when testing the two-sided hypothesis $H_0 : \psi = \psi_0$ versus $\psi \neq \psi_0$ is $2 \min \{C(\psi_0), 1 - C(\psi_0)\}$.

Confidence distributions are easily found when pivots (Barndorff-Nielsen and Cox, 1994) can be identified.

Definition 2

A function of the data and the interest parameter, $p(X, \psi)$, is a pivot if the probability distribution of $p(X, \psi)$ is the same for all (ψ, χ) , and the function $p(x, \psi)$ is increasing in ψ for almost all x .

If based on a pivot with cumulative distribution function F , the cumulative confidence distribution is

$$C(\psi; X) = F(p(X, \psi)).$$

From the definition, a confidence distribution is exact if and only if

$$C(\psi; X) \sim U$$

is a uniformly distributed pivot. The symbol \sim means “distributed as”.

Example: Linearity and normality

In linear normal models, a linear parameter, μ , has a normally distributed estimator, $\hat{\mu}$, and an independent estimator of standard error, $\hat{\sigma}$. Then,

$$\frac{\mu - \hat{\mu}}{\hat{\sigma}} \sim t_v,$$

meaning that the left hand side has a Student t-distribution with v degrees of freedom, independent of the parameters μ and σ , and is thus a pivot. With F_v being the cumulative t-distribution,

$$C(\mu) = F_v\left(\frac{\mu - \hat{\mu}}{\hat{\sigma}}\right)$$

is the student confidence distribution.

Fisher's fiducial argument is in this case in short hand:

$$\frac{(\mu - \hat{\mu})}{\hat{\sigma}} \sim t_v \Rightarrow \mu \sim \hat{\mu} + \hat{\sigma}t_v,$$

meaning that the confidence distribution is the t-distribution scaled by $\hat{\sigma}$, and located at $\hat{\mu}$. This is a specification *ex ante*, with $\hat{\mu}$ and $\hat{\sigma}$ being stochastic, and *ex post* with observed values for these variates.

The sampling distribution for the standard error estimator is related to the chi-square distribution χ_v^2 . In short hand, it is given by the pivot

$$\frac{\hat{\sigma}}{\sigma} \sim \sqrt{\chi_v^2 / v}.$$

The confidence distribution is therefore

$$C(\sigma) = \hat{\sigma} \sqrt{v / \chi_v^2}.$$

If, say, the data result in the estimate $\hat{\sigma} = 2$, the realised confidence distribution is

$$C(\sigma) = 2 \sqrt{v / \chi_v^2}.$$

Sampling distributions are different from confidence distributions. The sampling distribution of an estimator $\hat{\psi}$ at the observed estimate $\psi = \hat{\psi}_{obs}$ has cumulative distribution function

$$S(\psi) = P_{\hat{\psi}_{obs}}(\hat{\psi} \leq \psi) = F_{\hat{\psi}_{obs}}(\psi).$$

The cumulative confidence distribution is, however, a p-value:

$$C(\psi) = P_{\psi}(\hat{\psi} > \hat{\psi}_{obs}) = 1 - F_{\psi}(\hat{\psi}_{obs}).$$

Sampling distributions are often estimated by bootstrap distributions. To obtain distributional results for the parameter, bootstrap distributions

must therefore be converted to confidence distributions. The difference between the sampling distribution and the confidence distribution is illustrated for the standard error in the above example.

Confidence distributions are unbiased by definition. The coverage probabilities of their confidence intervals are correct, and the confidence median is a median unbiased point estimator. Confidence distributions also have power properties. The Neyman-Pearson lemma specifies the conditions for the likelihood ratio tests to be most powerful. A parallel theorem exists for confidence distributions. An *ex ante* confidence distribution based on the likelihood ratio statistic is, in fact, less dispersed in a stochastic sense (has smallest variance, and is also less dispersed for non-quadratic measures of dispersion) than any confidence distribution based on other statistics for the same data (Schweder and Hjort, 2002).

Example: Multiple capture-recapture data

Consider a closed population of N differently marked individuals. As an example, take the population component of marked immature bowhead whales off Alaska. These whales were subject to photo surveys in the summer and autumn of both 1985 and 1986, leading to X_t unique animals being sampled in survey $t = 1, \dots, 4$, and X unique captures in the pooled survey. From these data, N is to be estimated. Assume that each animal has an unknown probability of being captured in a given survey, and that captures are independent across animals and surveys. The observed data are presented in Table 1. The multiple capture-recapture model of Darroch (1958) leads to a confidence distribution based on the conditional distribution of X given $\{X_t\}$ (Schweder, 2003).

Calculating the half-corrected p-value for $H_0: N \leq N_0$ for a number of values of N_0 , a very good normal approximation in $1/\sqrt{N}$ emerge:

$$C(N) = P(X > 62) + \frac{1}{2}P(X = 62) \sim \Phi(5.134 - 87.307/\sqrt{N}) \quad (1)$$

TABLE 1. – Number of individual captures in each of four photo surveys of bowhead whales in summer and autumn of 1985 and 1986, and total number of individual captures. Source: da Silva *et al.* (2000).

	X_1	X_2	X_3	X_4	X
Marked immature	15	32	9	11	62
Marked mature	44	20	49	7	113

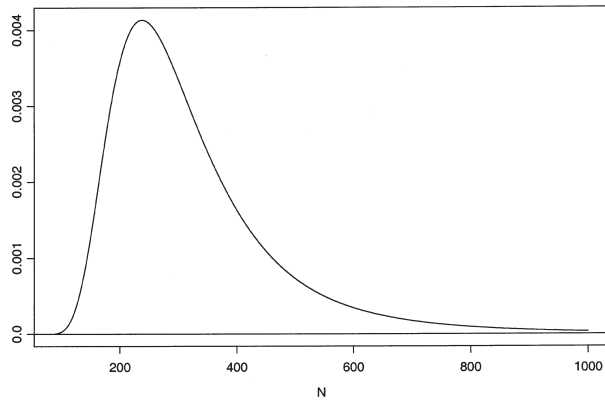


FIG. 1. – Confidence density of number of marked immature bowhead whales off Alaska in 1986.

These probabilities are conditional given $\{X_i\} = (15, 32, 9, 11)$, and Φ is the cumulative normal distribution function. The resulting confidence density is skewed, with a long tail to the right (Fig. 1).

REDUCED LIKELIHOODS

In integrative fish stock assessment, and in other situations where different pieces of data are joined in an integrated analysis, the challenge is often to find appropriate weights for the various data components. This problem is solved if well-founded likelihood functions exist for the various independent pieces of data. The likelihood function is the eminent tool for data integration whether the various data are collected in the same investigation, or whether some pieces of data are old and some are new. Old data might be problematic when they only exist in the format of published summaries. The challenge is then to establish a likelihood function based on these summaries. This is possible when sufficient information is provided, or even better, when the author in addition to the summaries and the inference provides what is called the reduced likelihood (Schweder and Hjort, 2002).

Let us, as we did for confidence distributions, restrict attention to cases in which exact reduced likelihoods are available and can be constructed from a pivot p . Let the pivot be a function of a one-dimensional statistic T providing the cumulative confidence distribution

$$C(\psi; T) = F(p(T, \psi)).$$

Let the pivot have density $f = F'$. Then

Definition 3

The reduced likelihood function is given by

$$L_{reduced}(\psi; X) = f\left(F^{-1}(C(\psi; T))\right) \left| \frac{dp}{dT} \right|.$$

It is remarkable that Fisher (1930) not only presented his fiducial distribution, which here is called the confidence distribution, but in passing he also mentioned the basics for the reduced likelihood. He did not use the term, but he clearly understood its significance.

The reduced likelihood is often a marginal or a conditional likelihood (Barndorff-Nielsen and Cox, 1994). I prefer the term “reduced likelihood” since its main property is to be a likelihood for the interest parameter ψ , with the nuisance parameter χ reduced away. In less clear-cut situations, when no exact pivot exists, one can find approximate pivots leading to approximate reduced likelihoods.

The confidence density is only appropriate as a likelihood when the pivot is linear in ψ and additive in the parameter and the data. The relationship between the confidence density and the reduced likelihood is generally

$$c(\psi; X) = L_{reduced}(\psi; T) \left| \frac{dp}{d\psi} / \frac{dp}{dT} \right|.$$

If the pivot is non-linear, the use of the confidence density as a likelihood introduces an unwanted term. Note that the Bayesian approach is to multiply prior densities with the likelihood of the new data. When the confidence density is taken as the prior density, the product of the prior and the likelihood agrees with Fisher’s likelihood analysis only when the underlying pivot is linear and additive.

For large data, the pivot distribution is often normal. More precisely, the pivot can be transformed to a scale to make it additive and normally distributed. In this case the reduced log likelihood agrees with the implied likelihood of Efron (1993). The normal score of the cumulative confidence at ψ is $\Phi^{-1}(C(\psi))$. The reduced (and implied) likelihood is then

$$L(\psi) = \exp\left[-\frac{1}{2}\left(\Phi^{-1}(C(\psi))\right)^2\right] \quad (2)$$

and is called the normal score reduced likelihood.

Example: Bowhead whales from photo-id

From the confidence distribution of the number of marked mature whales, obtained as that for

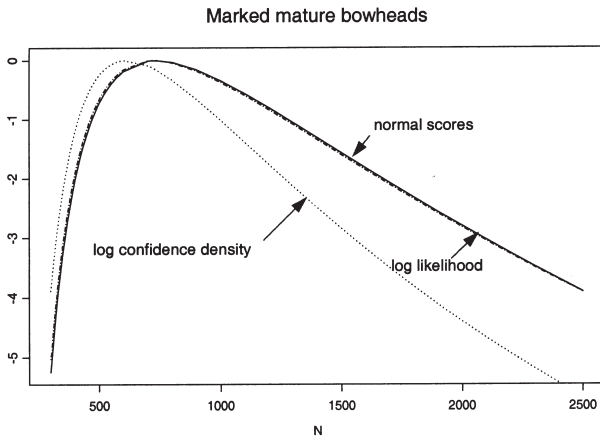


FIG. 2. – Log likelihood and log confidence density.

marked immature whales in (1), the normal score reduced likelihood is easily calculated. It is graphed in Figure 2. A graph of the exact log likelihood is also included. They agree well. The log confidence density is in disagreement with the two. This is due to non-linearity in the pivot. The Bayesian would thus introduce a negative bias in his analysis by using the confidence density as a prior to combine with new or other data, and he would over-state the precision of the photo-id data.

INTEGRATIVE STOCK ASSESSMENT: BOWHEAD WHALES

As an alternative to the Bayesian assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (International Whaling Commission, 1999), Schweder and Ianelli (2000) presented a frequentist assessment. In addition to data included in the Bayesian analysis, they also included age-reading data and data from the photo-identification study mentioned above. This is not the place to fully review the data and the model. For details, see Schweder and Ianelli (2000) and International Whaling Commission (1999), and the papers referred to in these two sources. The integrative stock assessment by frequentist means is briefly reviewed here in order to illustrate the main points of the analysis. Punt and Butterworth (2000) compared Bayesian and frequentist approaches to assessing the stock of bowhead whales.

The aim of the analysis is to estimate the current replacement yield for the purpose of setting catch limits for the aboriginal subsistence hunt in Alaska. A point estimate of replacement yield is not satisfactory. For reasons of precaution, a confidence dis-

tribution is needed for this parameter (the Bayesian analysis provided a posterior distribution), and the emphasis is on the lower tail of the distribution. To this end, the population dynamics is assumed to follow an age-specific Pella-Tomlinson model with density dependence in fecundity. Available data, including ‘soft data’ in the format of prior distributions, are used to estimate the parameters of the model, and consequently the current replacement yield.

The soft data are given as independent prior distributions on: Maximum sustainable yield rate; Maximum sustainable yield level; Carrying capacity; Adult mortality; Age at sexual maturity; and Maximum pregnancy rate, f . These prior distributions are partly based on general knowledge of baleen whales, on data from similar species of bowhead whales. The best supported prior distribution is in my view that on f . Inter birth intervals, $1/f$, are found to follow a uniform distribution over the interval 2.5 to 4 years. Consequently, the prior distribution for f has cumulative distribution function $C(f) = (8f-2)/(3f)$, $0.25 < f < 0.4$, which here is interpreted as a cumulative confidence distribution. The other prior distributions are also interpreted as confidence distributions. In the present brief presentation, they are not given.

The observational data consist of: relative abundance estimates based on visual surveys 1978-1988; an abundance estimate (visual and acoustic) for 1993; an abundance estimate based on photo-id for 1985/86; stock composition (proportion of calves and adults); length and age readings of 45 whales (aspartic acid racemisation in eye balls, George *et al.*, 1999) together with the length distribution in the catch; catches from 1848 to present. These data are assumed independent. Each data set provides a likelihood component. Considerable work has been invested in obtaining these likelihood components. The reader is referred to the primary documents for details.

The frequentist analysis starts by constructing the likelihood for all the data. Schweder and Ianelli (2000) assumed all prior distribution to have arisen from normally distributed pivots, making the reduced log-likelihoods of the normal score type (2). The log-likelihood resulting from the prior distribution of f is thus

$$l(f) = -\frac{1}{2} \left(\Phi^{-1} \left(\frac{8f-2}{3f} \right) \right)^2.$$

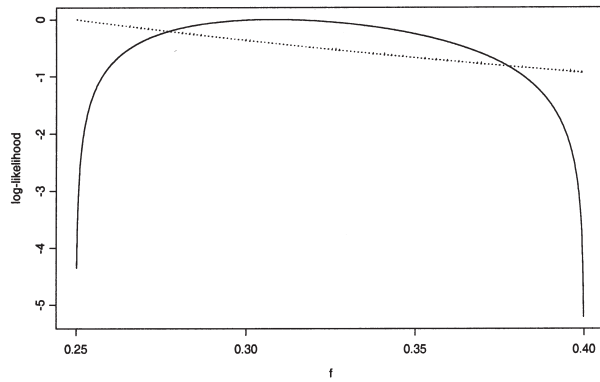


FIG. 3. – Log likelihood from the prior distribution of maximum pregnancy rate. The normalised log confidence density is drawn as a dotted line.

This log-likelihood is displayed in Figure 3. The confidence density is $2/3 f^{-2}$. The reduced likelihood favours intermediate values of f , and is quite different from the confidence density.

The underlying structural model can in general be parameterised in many different ways. A one-component Pella Tomlinson model, $P_{t+1} = P_t + \alpha P_t (1 - (P_t/K)^z)$ in the 3 parameters α , z , and K , might just as well have been parameterised in $MSYR$, $MSYL$ and K . In the first parameterisation, α and z are input parameters (Raftery *et al.*, 1995), while $MSYR$ and $MSYL$ are outputs, while their roles are interchanged in the second parameterisation. What are input parameters of the structural model and what are output parameters is a matter of choice. This choice should not, however, influence the results of the analysis. Likelihoods are invariant to parameterisation, which is an important property.

In the chosen parameterisation, some likelihood components are functions of input parameters, while others are functions of output parameters, and some might be functions of both. This applies to the prior likelihood components, as well as to the observational components. The likelihood component based on the important abundance estimate (visual and acoustic) in 1993 is, for example, only a function of the total stock size in 1993, which in the parameterisation chosen by Schweder and Ianelli (2000) is an output parameter. Issues of parameterisation have been discussed in connection with bowhead assessments (Schweder and Hjort, 1996; Punt and Butterworth, 1999).

When the various likelihood components have been constructed, they are multiplied together to form the total likelihood function. The model is then fitted to the data by maximising the total likelihood. Since some data are softer than others, the more

dubious priors are discarded provided the model can be fitted to the surviving data. Discarding a prior is equivalent to changing the corresponding likelihood component to the non-informative flat likelihood. Note that a constant likelihood is non-informative in the strict sense, while the concept of non-informative prior distributions is problematic in the Bayesian sense.

When the data on which the model is to be fitted has been identified, one might want to look for simplifications in the structural part of the model. This is ordinary model selection, and the guiding principle is to allow simplifications that make biological sense but do not inflict substantial (statistically significant) loss of fit as measured by the likelihood ratio.

Finally, confidence distributions for one-dimensional interest parameters are obtained, usually by a simulation exercise. This should be done for one parameter at the time. The aim is to identify an approximately pivotal quantity and its distribution. The pivot will usually be based on an estimator or a test statistic for the parameter of interest, i.e. its maximum likelihood estimator. This simulation exercise might be seen as a parametric bootstrap, and bias-correction methods to obtain confidence distributions can be employed (Efron and Tibshirani, 1993; Schweder and Hjort, 2002)

The main findings concerning data, priors and the parameter of primary interest: current replacement yields for the bowhead stock, are as follows (Schweder and Ianelli, 2000).

1. The age data of George *et al.* (1999) are inconsistent with the survey data. The age data point towards a much slower dynamics than indicated by the survey data. The survey data and the photo-identification data are consistent within the Pella Tomlinson model.

2. The dubious priors on $MSYL$, $MSYR$ and K can safely be dropped.

3. Based on survey data, the photo-identification data, and the remaining prior data, the lower end of the confidence distribution for Replacement Yield in number of bowhead whales have quantiles given in Table 2.

TABLE 2. – Lower confidence quantiles of Replacement Yield in Alaskan bowhead whales.

Confidence	1%	5%	10%	25%	50%
Replacement Yield	63	86	116	140	155

DISCUSSION

The methodology I have sketched is, I believe, in the best Fisher-Neyman-Cox-Efron tradition (Efron 1998). Instead of reporting results in the format of test results or 95% confidence intervals, I suggest inference concerning parameters of primary interest to be reported in terms of confidence distributions. The spread and shape of a confidence distribution convey a picture of the information and the statistical uncertainty in the inference, in much the same way as a posterior distribution is interpreted in a Bayesian analyses.

Since information updating is done via the likelihood function, it is important to have all data to be integrated in the analysis represented as likelihood components. The likelihood of prior data, reduced of nuisance parameters, is in general different from the prior confidence density of the parameter. Thus, in order to allow results to be efficiently used in future meta analyses or other integrative analyses, reduced likelihoods for important parameters should be obtained and presented along with their confidence distributions.

Confidence distributions and reduced likelihoods are usually obtained via simulation experiments. Since simulations are likely to be used in future analyses where the reduced likelihood of the study is integrated with other likelihood components, one should also report how the resulting reduced likelihood is to be simulated. To avoid multiple uses of data, it is necessary to know the data basis for the published reduced likelihood.

Compared to the Bayesian tradition, the subjectivity involved in priors not based on data is avoided. In the frequentist approach, the Bayesian's problem with finding 'non-informative' priors does not exist. A flat likelihood is non-informative in every sense of the word. Also, the problem the Bayesian has when there are more prior distributions than there are free parameters (Schweder and Hjort, 1996; Pool and Raftery, 1998) does not affect the frequentist analysis.

By multiplying independent likelihood components, also for prior data, data are integrated in an optimal way. Integrating data solely through the likelihood function could be called *likelihood synthesis*.

It happens that different data are inconsistent with each other. This was the case for the age data and the survey data in the bowhead example. But this is then a scientific problem to be resolved, not a

problem of the method of analysis. Both Bayesian and frequentist methods are useful in model selection and data confrontations. The use of informative priors in the Bayesian approach might, however, obscure the confrontation of one set of data with another.

From the applied and pragmatic point of view, the most important asset of the frequentist methodology is, perhaps, that it strives towards unbiased inference. The very concept of bias is difficult for the Bayesian, particularly for the subjectivist. Bayes' formula is certainly a mathematically correct way to update a prior distribution in the light of new data summarized in the likelihood function. The subjectivist scientist can be wrong, but can he be biased? Bias understood of systematic error in repeated use of the method on data of the type specified in the statistical model is indeed a frequentist notion. A frequentist point of view is thus needed to declare a Bayesian posterior distribution free of bias. As is demonstrated by the example below, Bayesian posteriors can be badly biased, often due to non-linearities and nuisance parameters. If, however, the posterior distribution can be declared free of bias in the sense that its quantile spans intervals with given coverage probability in repeated use, the Bayesian posterior is a confidence distribution, and is totally acceptable from the frequentist point of view. Berger *et al.* (1999) report that the pragmatic Bayesian approach of using flat priors on well-chosen transforms of the basic parameter often leads to nearly unbiased posteriors.

Example

There are 10 independent normally distributed observations X_j , each with a separate mean μ_j , but with common variance 1. The interest parameter is $\psi = \mu_1^4 + \dots + \mu_{10}^4$. Taking $\mu_1 = \dots = \mu_{10} = 3$, and thus $\psi = 810$, the simulated data vector is

$$x = (3.2, 4.3, 4.2, 3.0, 4.1, 3.8, 1.2, 3.1, 2.7, 3.3).$$

The maximum likelihood estimate is $\hat{\psi} = 1568$.

With a flat prior on the vector μ , the joint Bayesian posterior distribution is the joint normal, $\mu \sim N_{10}(x, I)$. The posterior distribution for ψ is then the corresponding marginal distribution shown in Figure 4.

By parametric bootstrapping, I found the following local approximate sampling distribution

$$\sqrt[4]{\hat{\psi}} \sim .864 + .944\sqrt[4]{\psi} + .647Z$$

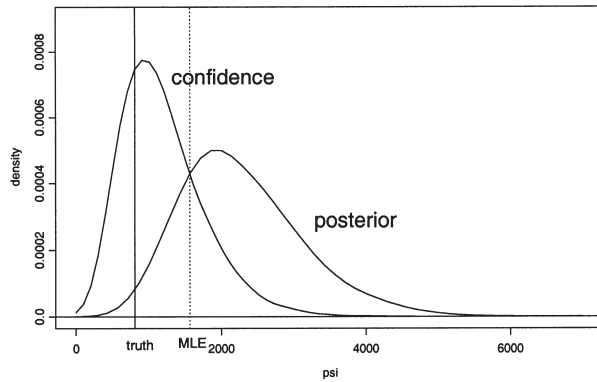


FIG. 4. – Bayesian posterior distribution and approximate confidence density for a non-linear parameter.

in the neighbourhood of $\hat{\mu} = x$. Here, Z is a standard normal variate. The approximate cumulative confidence distribution is then,

$$C(\psi) = \Phi \left(\frac{.864 + .944\sqrt{\psi} - 4\sqrt{\hat{\psi}}}{.647} \right) \quad (3)$$

with confidence density shown in Figure 4. As opposed to the Bayesian posterior, the confidence distribution automatically corrects for the inherent bias, at least approximately.

To illustrate the frequency behaviour of the Bayesian posterior, as well as the confidence density, 10 replicates of $\hat{\mu}$ were simulated. For each replicate, the Bayesian posterior density, and the confidence density calculated from (3) are displayed in Figure 5. This ends the example.

Often in fish stock assessments, the purpose is to predict the status of the stock at some future point in time. The parameters of the model are only of secondary interest. This problem is handled well in the Bayesian approach, particularly when a frequentist view is taken and potential bias is removed from the

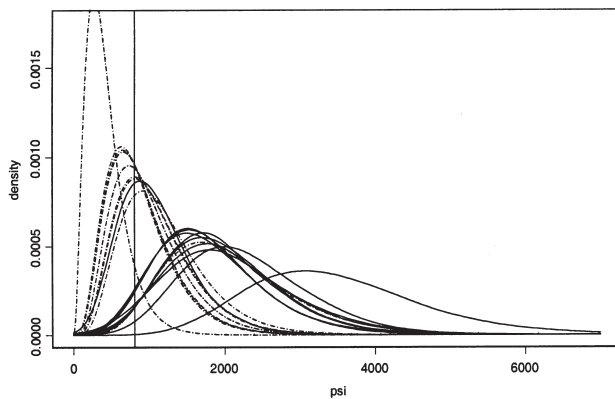


FIG. 5. – In ten simulated replicates, the Bayesian posterior density (solid lines) and confidence densities based on (3) (dashed lines) are displayed. The vertical line indicates the true value of the parameter.

prediction distribution. Hall *et al.* (1999) show how bootstrapping techniques can be used in a straightforward manner to obtain nearly unbiased prediction distributions. Their method is well suited for the frequentist approach outlined here.

The attractions of the type of ‘frequentist Bayesianism’ sketched above come at a price. A Bayesian analysis ends up in a joint posterior for the vector of basic parameters. From this joint distribution, one-dimensional posterior distributions are obtained for any scalar parameter of interest simply by calculating the appropriate marginal distribution. This is particularly easy to do when the joint posterior distribution is approximated by a (large) sample. In the frequentist approach, a joint confidence distribution cannot in general yield confidence distributions for scalar parameters of interest by computing its marginal distributions. Here, unbiasedness is of primary importance, and simulations with bias correction (Efron and Tibshirani, 1993) are usually needed to obtain approximately valid confidence distributions.

In most applications, the parameter is multi-dimensional. For each selection of an interest parameter, there is then a vector of nuisance parameters. The handling of nuisance parameters to obtain approximate priors with consequential approximate confidence distributions and reduced likelihoods can be difficult. This is an area of statistical research. Some guidelines are given in Schweder and Hjort (2002).

REFERENCES

- Barndorff-Nielsen, O.E. and D.R. Cox. – 1994. *Inference and Asymptotics*. Chapman & Hall, London.
- Berger, J.O., B. Liseo and R.L. Wolpert. – 1999. Integrated likelihood methods for eliminating nuisance parameters. *Statist. Sci.*, 14: 1-28.
- da Silva, C.Q., J. Zeh, D. Madigan, J. Laake, D. Rugh, L. Baraff, W. Koski and G. Miller. – 2000. Capture-recapture estimation of bowhead whale population size using photo-identification data. *J. Cetacean Res. Manag.*, 2: 45-61.
- Darroch, J.N. – 1958. The multiple-recapture census. I: Estimation of a closed population. *Biometrika*, 45: 343-359.
- Efron, B. – 1993. Bayes and likelihood calculations from confidence intervals. *Biometrika*, 80: 3-26.
- Efron, B. – 1998. R.A. Fisher in the 21st century (with discussion). *Statist. Sci.*, 13: 95-122.
- Efron, B. and R.J. Tibshirani. – 1993. *An Introduction to the Bootstrap*. Chapman & Hall.
- Fisher, R.A. – 1930. Inverse probability. *Proc. Cambridge Phil. Society*, 26: 528-535.
- Gavaris, S. – 1999. Dealing with bias in estimating uncertainty and risk. In: V. Restrepo (ed.), *Providing Scientific Advice to Implement the Precautionary Approach Under the Magnusson-Stevens Fishery Conservation and Management Act*, pp. 17-23. NOAA Technical Memorandum NMFS-F/SPO-40.
- George, J.C., J. Bada, J. Zeh, L. Scott, S.E. Brown, T. O’Hara and R. Suydam. – 1999. Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic acid racemization.

- Can. J. Zool.* 77: 571-580.
- Hall, P., L. Peng and N. Tajvidi. – 1999. On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika*, 86: 871-880.
- International Whaling Commission. – 1999. Report of the Scientific Committee, Annex G. Report of the Sub-Committee on Aboriginal Subsistence Whaling. *J. Cetacean Res. Manag.*, 1(SUPPL.): 179-94.
- Patterson, K., R. Cook, C. Darby, S. Gavaris, L. Kell, P. Lewy, B. Mesnil, A. Punt, V. Restrepo, D.W. Skagen and G. Stefansson. – 2001. Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries*, 2: 125-157.
- Poole, D. and A.E. Raftery. – 1998. Inference for Deterministic Simulation Models: The Bayesian Melding Approach. Technical Report no. 346, Department of Statistics, University of Washington (unpublished).
- Punt, A.E. and D.S. Butterworth. – 1999. On assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*) using a Bayesian approach. *J. Cetacean Res. Manag.*, 1: 53-71.
- Punt, A.E. and D.S. Butterworth. – 2000. Why do Bayesian and maximum likelihood assessments of the Bering-Chukchi-Beaufort Seas stock of bowhead whales differ? *J. Cetacean Res. Manag.*, 2: 125-133.
- Raftery, A.E., G.H. Givens and J.E. Zeh. – 1995. Inference from a deterministic population dynamics model for bowhead whales (with discussion). *J. Am. Statist. Ass.*, 90: 402-430.
- Schweder, T. – 2003. Abundance estimation from photo-identification data: confidence distributions and reduced likelihood for bowhead whales off Alaska. *Biometrics* (to appear).
- Schweder, T. and N.L. Hjort. – 1996. Bayesian synthesis or likelihood synthesis - what does the Borel paradox say? *Rep. int. Whal. Commn*, 46: 475-479.
- Schweder, T. and N.L. Hjort. – 2002. Confidence and likelihood. *Scandinavian J. Statist.*, 29: 309-332
- Schweder, T. and J.N. Ianelli. – 2000. Assessing the Bering-Chukchi-Beaufort Seas stock of bowhead whales from survey data, age-readings and photo-identifications using frequentist methods. Paper IWC/SC/52/AS13 presented to The International Whaling Commission Scientific Committee, June 2000, 179-194.
- Wade, P.R. – 1999. A comparison of statistical methods for fitting population models to data. In: G.W. Garner, S.C. Amstrup, J.L. Laake, B.F.J. Manly, L.L. McDonald and D.G. Robertson (eds.), *Marine Mammal Survey and Assessment Methods*, pp 249-270. AA Balkema, Rotterdam.

