© Board of the Foundation of the Scandinavian Journal of Statistics 2002. Published by Blackwell Publishers Ltd, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA Vol 29: 309–332, 2002

Confidence and Likelihood*

TORE SCHWEDER and NILS LID HJORT

University of Oslo

ABSTRACT. Confidence intervals for a single parameter are spanned by quantiles of a confidence distribution, and one-sided *p*-values are cumulative confidences. Confidence distributions are thus a unifying format for representing frequentist inference for a single parameter. The confidence distribution, which depends on data, is exact (unbiased) when its cumulative distribution function evaluated at the true parameter is uniformly distributed over the unit interval. A new version of the Neyman–Pearson lemma is given, showing that the confidence distribution based on the natural statistic in exponential models with continuous data is less dispersed than all other confidence distributions, regardless of how dispersion is measured. Approximations are necessary for discrete data, and also in many models with nuisance parameters. Approximate pivots might then be useful. A pivot based on a scalar statistic determines a likelihood in the parameter of interest along with a confidence distribution. This proper likelihood is reduced of all nuisance parameters, and is appropriate for meta-analysis and updating of information. The reduced likelihood is generally different from the confidence density.

Confidence distributions and reduced likelihoods are rooted in Fisher–Neyman statistics. This frequentist methodology has many of the Bayesian attractions, and the two approaches are briefly compared. Concepts, methods and techniques of this brand of Fisher–Neyman statistics are presented. Asymptotics and bootstrapping are used to find pivots and their distributions, and hence reduced likelihoods and confidence distributions. A simple form of inverting bootstrap distributions to approximate pivots of the abc type is proposed. Our material is illustrated in a number of examples and in an application to multiple capture data for bowhead whales.

Key words: abc correction, bootstrapping likelihoods, capture-recapture data, confidence distributions and densities, frequentist posteriors and priors, integrating information, Neyman–Pearson lemma, pivots, reduced likelihood

1. Introduction

This article presents core material in the frequentist tradition of parametric statistical inference stemming from R.A. Fisher and J. Neyman. Confidence distributions are the Neymanian interpretation of Fisher's fiducial distributions. Confidence distributions are often found from exact or approximate pivots. Pivots might also provide likelihoods reduced of nuisance parameters. These likelihoods are called reduced likelihoods, and date back to Fisher. Part of the material in our paper is old wine in new bottles; see Fraser (1996, 1998) and Efron (1998). In addition, we present a new and extended version of the Neyman–Pearson lemma, and also some new material on approximate confidence distributions and reduced likelihoods. Our article attempts to demonstrate the power of frequentist methodology, in the tapping of confidence distributions and reduced likelihoods. The theory is illustrated by a number of examples and is briefly compared to the Bayesian approach.

Confidence intervals and *p*-values are the primary formats of statistical reporting in the frequentist tradition. The close relationship between *p*-values and confidence intervals allows a unification of these concepts in the confidence distribution. Let the one-dimensional parameter

*This paper was presented by Nils Lid Hjort as an Invited Lecture at the 18th Nordic conference on Mathematical Statistics, Grimstad, Norway, June 2000.

of interest be ψ. A confidence distribution for ψ is calculated from the data within the statistical model. The cumulative confidence distribution function, C, provides $C(\psi_0)$ as the p-value when testing the one-sided hypothesis $H_0: \psi \leq \psi_0$ whatever value ψ_0 takes. Any pair of confidence quantiles constitutes, on the other hand, a confidence interval $(C^{-1}(\alpha), C^{-1}(\beta))$ with degree of confidence $\beta - \alpha$.

Distributions are eminent for presenting uncertain information. Much of the attraction of the Bayesian approach is due to the use of distributions as the format of presenting information, e.g. prior and posterior distributions. Fisher (1930) introduced fiducial probability distributions as an alternative to the Bayesian posterior distribution as a format for presenting what has been learned from the data in view of the model; see Fisher (1973) for his final understanding of fiducial probability and the fiducial argument. Quantiles of a fiducial distribution are endpoints of fiducial intervals. Following Neyman rather than Fisher in understanding fiducial intervals as confidence intervals, we use the term confidence distribution, in agreement with Efron (1998) and others. Section 2 gives basic definitions and properties.

The confidence density is not a probability density in the frequentist sense, and is only in special cases a likelihood function reduced of nuisance parameters. The connection between likelihood and confidence distributions is discussed in section 3. There we also explain how reduced likelihoods are obtained from exact or approximate pivots, following Fisher (1930). When the pivot is exact, the reduced likelihood is either a marginal likelihood or a conditional likelihood. We have settled for the wider term reduced likelihood, which also includes approximate likelihoods.

Confidence distributions and reduced likelihoods are usually obtained from pivots for the parameter of interest. In the presence of nuisance parameters, exact pivots are essentially available only in exponential models. In such models, reduction by sufficiency and conditioning might lead to canonical inference. We present in section 4 a new version of the Neyman–Pearson lemma for confidence distributions, along with further optimality results. It essentially says that the confidence distribution based on the canonical statistic, possibly via conditioning, is less dispersed than any other confidence distribution obtained from the data. In more complicated models where no canonical frequentist inference can be identified, one might be satisfied with confidence distributions and reduced likelihoods obtained from approximate pivots. These approximate pivots might be found from parametric bootstrapping. The pivot behind the acceleration and bias corrected bootstrap percentile interval method leads to confidence distributions. It has an appealing form and is seen to perform well in terms of accuracy in various cases. It also leads to good approximations for reduced likelihoods. This is the theme of section 5.

Our apparatus is illustrated in section 6 on data from a capture-recapture study of bowhead whales based on photo-identification. Finally some supplementing remarks and discussion are provided in section 7. We suggest that the confidence density is a very useful summary for any parameter of interest. The confidence distribution may serve as the frequentist analogue of the Bayesian's posterior density, and together with the reduced likelihood a frequentist apparatus for information updating is available as a competitor to the Bayesian methodology. Differences from the Bayesian paradigm are discussed, and some advantages and limitations are noted. We end by placing frequentist inference in terms of confidence distributions and reduced likelihood in the grand context of what Hald (1998) terms the three (so far) revolutions in parametric statistical inference.

2. Confidence distributions

2.1. Confidence and statistical inference

Our context is a given parametric model for data Y with probability distribution $P_{\psi,\chi}$, where ψ is a scalar parameter of primary interest, belonging to a finite or infinite interval on the real line, and χ the remaining (nuisance) parameters. The focus is on inference for ψ; issues of model selection and validation are thus (at the outset) outside our scope. With inference we shall understand statements of the type "$\psi > \psi_0$", "$\psi_1 \leq \psi \leq \psi_2$", etc., where ψ_0, ψ_1 etc. are values usually computed from the data. To each statement, we would like to associate how much confidence the data allow us to have in the statement.

Definition 1
A (one-dimensional) distribution for ψ, depending on the data Y, is a *confidence distribution for* ψ when its cumulative distribution function evaluated at the true value of ψ, $C(\psi; Y)$, has a uniform distribution under $P_{\psi,\chi}$, whatever the true value of (ψ, χ).

For simplicity, we write $C(\psi)$ for $C(\psi; Y)$ etc., keeping in mind that confidence, confidence quantiles, confidence density and other related objects are stochastic variables or realizations thereof. By the above definition, the (stochastic) confidence quantiles are endpoints of confidence intervals with degree of confidence given by the stipulated confidence. For one-sided intervals $(-\infty, \psi_\alpha)$, where $\psi_\alpha = C^{-1}(\alpha)$, the coverage probability is, in fact, $P_{\psi,\chi}\{\psi \leq \psi_\alpha\} = P_{\psi,\chi}\{C(\psi) \leq C(\psi_\alpha)\} = P_{\psi,\chi}\{C(\psi) \leq \alpha\} = \alpha$, while a two-sided confidence interval $(\psi_\alpha, \psi_\beta)$ has coverage probability $\beta - \alpha$.

Being an invertible function of the interest parameter, and having a uniform distribution independent of the full parameter, $C(\psi)$ is a pivot (Barndorff-Nielsen & Cox, 1994). On the other hand, whenever a pivot $\mathrm{piv}(Y, \psi)$ is available, taken to be increasing in ψ, and having cumulative distribution function F independent of the parameter,

$$C(\psi) = F(\mathrm{piv}(Y, \psi)) \tag{1}$$

is uniformly distributed and is thus the c.d.f. of a confidence distribution for ψ. If the natural pivot is decreasing in ψ, then $C(\psi) = 1 - F(\mathrm{piv}(Y, \psi))$.

Example 1. Consider the exponentially distributed variate Y with probability density $f(y; \psi) = (1/\psi)\exp(-y/\psi)$. The cumulative confidence distribution function for ψ is $C(\psi) = \exp(-Y/\psi)$. For each realisation y_{obs} of Y, $C(\psi)$ is a c.d.f. At the true value, $C(\psi_{\mathrm{true}})$ has the right uniform distribution. The confidence quantiles are the familiar $C^{-1}(\alpha) = Y/(-\log \alpha)$, and the confidence density is $c(\psi) = (\partial/\partial\psi)C(\psi; y_{\mathrm{obs}}) = y_{\mathrm{obs}}\psi^{-2}\exp(-y_{\mathrm{obs}}/\psi)$. The realized confidence density not only has a different interpretation from the likelihood function $L(\psi) = \psi^{-1}\exp(-y_{\mathrm{obs}}/\psi)$, but also a different shape.

Exact confidence distributions represent valid inference in the sense of statistical conclusion validity (Cook & Campbell, 1979). The essence is that the confidence distribution is free of bias in that any confidence interval $(\psi_\alpha, \psi_\beta)$ has exact coverage probability $\beta - \alpha$. The power of the inference represented by C is basically a question of the spread of the confidence distribution. We return to the issues of confidence power in section 4.

Hypothesis testing and confidence intervals are closely related. Omitting the instructive proof, this relation is stated in the following lemma.

Lemma 2

The confidence of the statement "$\psi \leq \psi_0$" is the degree of confidence $C(\psi_0)$ for the confidence interval $(-\infty, C^{-1}(C(\psi_0))]$, and is equal to the p-value of a test of H_0: $\psi \leq \psi_0$ vs the alternative H_1: $\psi > \psi_0$.

The opposite statement "$\psi > \psi_0$" has confidence $1 - C(\psi_0)$. Usually, the confidence distributions are continuous, and "$\psi \geq \psi_0$" has the same confidence as "$\psi > \psi_0$".

Confidence intervals are invariant w.r.t. monotone transformations. This is also the case for confidence distributions. Confidence distributions based essentially on the same statistic are invariant with respect to monotone continuous transformations of the parameter: If $\rho = r(\psi)$, say, with r increasing, and if C^ψ is based on T while C^ρ is based on $S = s(T)$ where s is monotone, then $C^\rho(\rho) = C^\psi(r^{-1}(\rho))$.

The sampling distribution of the estimator is the *ex ante* probability distribution of the statistic under repeated sampling, while the realized confidence distribution is calculated *ex post* and distributes the confidence that the observed data allow to be associated with different statements concerning the parameter. Consider the estimated sampling distribution of the point estimator $\widehat{\psi}$, say as obtained from the parametric bootstrap. If ψ^* is a random estimate of ψ obtained by the same method, the estimated sampling distribution is the familiar $S(\psi) = \Pr\{\psi^* \leq \psi; \widehat{\psi}\} = F_{\hat{\psi}}(\psi)$.

The confidence distribution is also obtained by (theoretically) drawing repeated samples, but now from different distributions. The interest parameter is, for the confidence distribution, considered a control variable, and it is varied in a systematic way. When $\widehat{\psi}$ is a statistic for which the hypothesis H_0: $\psi \leq \psi_0$ is suspect when $\widehat{\psi}$ is large, the p-value is $\Pr\{\psi^* > \widehat{\psi}_{\text{obs}}; \psi_0\}$. The cumulative confidence distribution is then

$$C(\psi) = \Pr\{\psi^* > \widehat{\psi}; \psi\} = 1 - F_\psi(\widehat{\psi}). \tag{2}$$

The sampling distribution and the confidence distribution are fundamentally different entities. The sampling distribution is a probability distribution, while the confidence distribution, *ex post*, is not a distribution of probabilities but of confidence—obtained from the probability transform of the statistic used in the analysis.

Fiducial distributions were introduced by Fisher (1930) by the quantiles of $C(\psi)$ determined by (1). Fiducial distributions for simple parameters are thus confidence distributions. Neyman (1941) gave the frequentist interpretation of the fiducial distribution, which Fisher partly disliked. Efron (1998) discusses this controversy, and supports the use of fiducial distributions as confidence distributions. We emphasize that our distributions of confidence are actually derived from certain principles in a rigorous framework, and with a clear interpretation. The traditional division between confidence intervals on the one hand and fiducial distributions on the other has in our view been overemphasized, as has the traditional division between significance testing (Fisher) and hypothesis testing (Neyman); see Lehmann (1993). The unity of the two traditions is illustrated by our version of the Neyman–Pearson lemma as it applies to Fisher's fiducial distribution (confidence distribution). Note also that we, in section 4 in particular, work towards determining canonically optimal confidence distributions.

Example 2. Suppose the ratio $\psi = \sigma_2/\sigma_1$ between standard deviation parameters from two different data sets are of interest, where independent estimates of the familiar form $\widehat{\sigma}_j^2 = \sigma_j^2 W_j / v_j$ are available, where W_j is a $\chi^2_{v_j}$. The canonical intervals, from inverting the optimal tests for single-point hypotheses $\psi = \psi_0$, take the form

$$[\widehat{\psi}/K^{-1}(1-\alpha)^{1/2}, \widehat{\psi}/K^{-1}(\alpha)^{1/2}],$$

where $\widehat{\psi} = \widehat{\sigma}_2/\widehat{\sigma}_1$ and $K = K_{v_2,v_1}$ is the distribution function for the F statistic $(W_2/v_2)/(W_1/v_1)$. Thus $C^{-1}(\alpha) = \widehat{\psi}/K^{-1}(1-\alpha)^{1/2}$. This corresponds to the confidence distribution function $C(\psi;\text{data}) = 1 - K(\widehat{\psi}^2/\psi^2)$, with confidence density

$$c(\psi;\text{data}) = k(\widehat{\psi}^2/\psi^2)2\widehat{\psi}^2/\psi^3,$$

expressed in terms of the F density $k = k_{v_2,v_1}$. See also section 4.1 for an optimality result of the confidence density used here, and section 5.2 for a very good approximation based on bootstrapping.

Example 3: linear regression. In the linear normal model, the n-dimensional data Y of the response is assumed $\text{N}(X\beta, \sigma^2 I)$. With SSR being the residual sum of squares and with $p = \text{rank}(X)$, $S^2 = \text{SSR}/(n-p)$ is the traditional estimate of the residual variance. With S_j^2 being the mean-unbiased estimator of the variance of the regression coefficient estimator $\widehat{\beta}_j$, $V_j = (\widehat{\beta}_j - \beta_j)/S_j$ is a pivot with a t-distribution of $v = n - p$ degrees of freedom. Letting $t_v(\alpha)$ be the quantiles of this t-distribution, the confidence quantiles for β_j are the familiar $\widehat{\beta}_j + t_v(\alpha)S_j$. The cumulative confidence distribution function for β_j is seen from this to become

$$C(\beta_j;\text{data}) = 1 - G_v((\widehat{\beta}_j - \beta_j)/S_j) = G_v((\beta_j - \widehat{\beta}_j)/S_j),$$

where G_v is the cumulative t-distribution with v degrees of freedom. Note also that the confidence density $c(\beta_j;\text{data})$ is the t_v-density centred at $\widehat{\beta}_j$ and with the appropriate scale.

Now turn attention to the case where σ, the residual standard deviation, is the parameter of interest. Then the pivot $\text{SSR}/\sigma^2 = vS^2/\sigma^2$ is a χ^2_v, and the cumulative confidence distribution is found to be $C(\sigma;\ \text{data}) = \Pr\{\chi^2_v > \text{SSR}/\sigma^2\} = 1 - \Gamma_v(vS^2/\sigma^2)$, where Γ_v is the cumulative distribution function of the chi-square with density γ_v. The confidence density becomes

$$c(\sigma;\text{data}) = \gamma_v\left(\frac{vS^2}{\sigma^2}\right)\frac{2vS^2}{\sigma^3} = \frac{2(v/2)^{v/2}S^v}{\Gamma(\frac{1}{2}v)}\sigma^{-(v+1)}\exp\left(-\frac{1}{2}vS^2/\sigma^2\right).$$

2.2. *Approximate confidence distributions for discrete data*

To achieve exact, say 95%, coverage for a confidence interval based on discrete data is usually impossible without artificial randomization. The same difficulty is encountered when constructing tests with exactly achieved significance level. Confidence distributions based on discrete data can never be exact. Since the data are discrete, any statistic based on the data must have a discrete distribution. The confidence distribution is a statistic, and $C(\psi)$ cannot have a continuous uniform distribution. Half-correction is a simple device to achieve an approximate confidence distribution. When T is the statistic on which p-values and hence the confidence distribution is based, half-correction typically takes the form

$$C(\psi) = \Pr_\psi\{T > t_{\text{obs}}\} + \tfrac{1}{2}\Pr_\psi\{T = t_{\text{obs}}\}.$$

For an illustration, let T be Poisson with parameter ψ. Then the density of the half-corrected confidence distribution simplifies to

$$c(\psi) = \frac{1}{2}\left\{\frac{\psi^{t_{\text{obs}}-1}}{(t_{\text{obs}}-1)!}\exp(-\psi) + \frac{\psi^{t_{\text{obs}}}}{t_{\text{obs}}!}\exp(-\psi)\right\} \qquad \text{provided } t_{\text{obs}} \geq 1.$$

Although the confidence distribution has a discrete probability distribution *ex ante*, it is a continuous distribution for ψ *ex post*.

A confidence distribution depends on the model, not only on the observed likelihood. The Bayesian posterior distribution depends on the other hand only on the observed likelihood (and the prior). Thus, while likelihoodists (Edwards, 1992; Royall, 1997, etc.) and Bayesians observe the likelihood principle, inference based on the confidence distribution, as with some other forms of frequentist inference (Barndorff-Nielsen & Cox, 1994), is in breach of the likelihood principle. This is illustrated by the following.

Example 4. Let T_x be the waiting time until x points are observed in a Poisson process with intensity parameter ψ, and let X_t be the number of points observed in the period $(0, t)$. The two variables are respectively gamma-distributed with shape parameter x and Poisson distributed with mean ψt. In one experiment, T_x is observed to be t. In the other, X_t is observed to be x. The observed log-likelihood is then identical in the two experiments, namely $\ell(\psi) = x \log(t\psi) - t\psi$. From the identity $\Pr\{T_x > t\} = \Pr\{X_t < x\}$, and since ψT_x is a pivot, the confidence distribution based on T_x has c.d.f. $C_t(\psi) = 1 - F(x-1; \psi t)$ where F is the c.d.f. of the Poisson distribution with mean ψt. This is not an exact confidence distribution if the experiment consisted in observing X_t. It is, in fact, stochastically slightly smaller than it should be in that case. As noted above, no non-randomized exact confidence distribution exists in the latter experiment.

3. Likelihood related to confidence distributions

To combine past reported data with new data, and also for other purposes, it is advantageous to recover a likelihood function or an approximation thereof from the available statistics summarizing the past data. The question we ask is whether an acceptable likelihood function can be recovered from a published confidence distribution, and if this is answered in the negative, how much additional information is needed to obtain a usable likelihood.

Frequentist statisticians have discussed at length how to obtain confidence distributions for one-dimensional interest parameters from the likelihood of the data in view of its probability basis. Barndorff-Nielsen & Cox (1994) discuss methods based on adjusted likelihoods and other modified likelihoods obtained from saddle-point approximations. Efron & Tibshirani (1993) and Davison & Hinkley (1997) present methods based on bootstrapping and quadratic approximations. The problem of finding a likelihood for a single parameter, reduced of nuisance parameters, has also received attention. When a marginal or conditional likelihood exists in the statistics behind the canonical confidence distribution, the problem is solved. Fisher (1930) not only introduced the fiducial distribution, but he also pointed to the likelihood associated with it through the underlying pivot. We follow his lead.

Example 3 (continued). The likelihood for σ is based on the pivot SSR/σ^2 or directly on the SSR part of the data. It is the density of $\mathrm{SSR} = \sigma^2\chi^2_\nu$, which is proportional to $L(\sigma) = \sigma^{-\nu}\exp(-\frac{1}{2}\nu S^2/\sigma^2)$. This is the two-stage likelihood for σ of Fisher (1922), and is the marginal likelihood. We will also term it the reduced likelihood for σ since it is obtained through the pivot, as specified in proposition 3 below. Taking logarithms, the pivot is brought on an additive scale, $\log S - \log\sigma$, and in the parameter $\tau = \log\sigma$ the confidence density is proportional to the likelihood. The log-likelihood also has a nicer shape in τ than in σ, where it is less neatly peaked. It is of interest to note that the improper prior $\pi(\sigma) = \sigma^{-1}$, regarded as the canonical "non-informative" prior for scale parameters like the present σ, yields when combined with the likelihood L the confidence distribution as the Bayes posterior distribution. See also the more general comment in section 7.4.

The likelihood is invariant under monotone transformations of the parameter. This is not the case for confidence densities, due to the presence of the Jacobian. A given reduced likelihood can thus relate to several confidence distributions. Realized confidence distributions can also relate to many likelihood functions (depending upon the model), as seen from the following example.

Example 5. Consider a uniform confidence distribution for ψ over (0,1). It is based on the statistic T with observed value $t_{\text{obs}} = \frac{1}{2}$. We shall consider three different models leading to this confidence distribution, and we calculate the likelihood function in each case.

The first model is a shift-uniform model with pivot $\psi - T + \frac{1}{2} = U$ where U has a uniform probability distribution over (0,1). Thus, $C(\psi) = \psi$ for $0 \le \psi \le 1$ representing the uniform confidence distribution. Further, T is uniform over $(\psi - \frac{1}{2}, \psi + \frac{1}{2})$ and the likelihood is $L_{\text{shift}}(\psi) = I_{(0,1)}(\psi)$, the indicator function. Second, consider the scale model with pivot $\frac{1}{2}\psi/T = U$. Again, the confidence distribution is the uniform. The probability density of T is easily found, and the likelihood based on $T = \frac{1}{2}$ comes out as $L_{\text{scale}}(\psi) = 2\psi I_{(0,1)}(\psi)$. The third model is based on a normally distributed pivot, $\Phi^{-1}(\psi) - \Phi^{-1}(T) = Z$, where Z has a standard normal distribution with c.d.f. Φ. For the observed data, the confidence distribution is the same uniform distribution. Calculating the probability density of T, we find the likelihood of the observed data $L_{\text{norm}}(\psi) = \exp[-\frac{1}{2}(\Phi^{-1}(\psi))^2]$.

These three possible log-likelihoods consistent with the uniform confidence distribution are shown in Fig. 1. Other log-likelihoods are also possible.

3.1. *Confidence and likelihoods based on pivots*

Assume that the confidence distribution $C(\psi)$ is based on a pivot piv with cumulative distribution function F and density f. Since ψ is one-dimensional, the pivot is often a function of a one-dimensional statistic T in the data X. The probability density of T is then

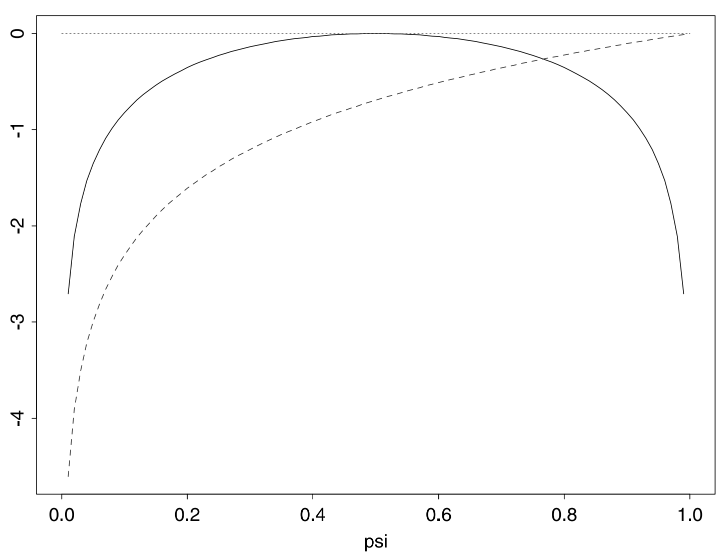

Fig. 1. Three log-likelihoods consistent with a uniform confidence distribution over [0, 1]. "Many likelihoods informed me of this before, which hung so tottering in the balance that I could neither believe nor misdoubt." – *Shakespeare*.

$$f^T(t;\psi) = f(\mathrm{piv}(t;\psi))\left|\frac{\partial \mathrm{piv}(t;\psi)}{\partial t}\right|;$$

see also Fisher (1930). Since $\mathrm{piv}(T;\psi) = F^{-1}(C(\psi))$ we have the following.

Proposition 3
When the statistical model leads to a pivot $\mathrm{piv}(T;\psi)$ *in a one-dimensional statistic* T, *increasing in* ψ, *the likelihood is*

$$L(\psi;T) = f(F^{-1}(C(\psi)))\left|\frac{\partial \mathrm{piv}(T;\psi)}{\partial T}\right|.$$

The confidence density is also related to the distribution of the pivot. Since one has $C(\psi) = F(\mathrm{piv}(T;\psi))$, $c(\psi) = f(\mathrm{piv}(T;\psi))|\partial \mathrm{piv}(T;\psi)/\partial\psi|$. Thus, the likelihood is in this simple case related to the confidence density by

$$L(\psi;T) = c(\psi)\left|\frac{\partial \mathrm{piv}(T;\psi)}{\partial T}\right| \Big/ \left|\frac{\partial \mathrm{piv}(T;\psi)}{\partial \psi}\right|. \tag{3}$$

There are important special cases. If the pivot is additive in T (at some measurement scale), say $\mathrm{piv}(T;\psi) = \mu(\psi) - T$ for a smooth increasing function μ, the likelihood is $L(\psi;T) = f\big(F^{-1}(C(\psi))\big)$. When furthermore the pivot distribution is normal, we will say that the confidence distribution has a normal probability basis.

Proposition 4 (Normal-based likelihood)
When the pivot is additive and normally distributed, the reduced log-likelihood related to the confidence distribution is $\ell(\psi) = -\frac{1}{2}\{\Phi^{-1}(C(\psi))\}^2$.

The normal-based likelihood might often provide a good approximate likelihood. Note that classical first order asymptotics leads to normal-based likelihoods. The conventional method of constructing confidence intervals with confidence $1-\alpha$, including those ψ for which $2\{\ell(\widehat{\psi}) - \ell(\psi)\}$ is less than the $1-\alpha$-quantile of the χ^2_1, where $\widehat{\psi}$ is the maximum likelihood estimate, is equivalent to assuming the likelihood to be normal-based. The so-called ABC confidence distributions of Efron (1993), concerned partly with exponential families, have asymptotic normal probability basis, as have confidence distributions obtained from Barndorff-Nielsen's r^* (Barndorff-Nielsen & Wood, 1998). Efron (1993) used a Bayesian argument to derive the normal-based likelihood in exponential models. He called it the implied likelihood.

In many applications, the confidence distribution is found by simulation. One might start with a statistic T which, together with an (approximate) ancillary statistic A, is simulated for a number of values of the interest parameter ψ and the nuisance parameter χ. The hope is that the conditional distribution of T given A is independent of the nuisance parameter. This question can be addressed by applying regression methods to the simulated data. The regression might have the format

$$\mu(\psi) - T = \tau(\psi)V \tag{4}$$

where V is a scaled residual distributed independently of (ψ, χ) in the parameter region of interest. Then $\mathrm{piv}(T;\psi) = \{\mu(\psi) - T\}/\tau(\psi)$, and the likelihood is $L(\psi) = f\big(F^{-1}(C(\psi))\big)/\tau(\psi)$. The scaling function τ and the regression function μ might depend on the ancillary statistic.

Example 6. Let T be Poisson with mean ψ. The half-corrected cumulative confidence distribution function is

$$C(\psi) = 1 - \sum_{j=0}^{t_{\mathrm{obs}}} \exp(-\psi)\psi^j/j! + \tfrac{1}{2}\exp(-\psi)\psi^{t_{\mathrm{obs}}}/t_{\mathrm{obs}}!.$$

Here $V = 2(\sqrt{\psi} - \sqrt{T})$ is approximately N(0,1) and is accordingly approximately a pivot for moderate to large ψ. From a simulation experiment, one finds that the distribution of V is slightly skewed, and has a bit longer tails than the normal. By a little trial and error, one finds that $\exp(V/1000)$ is closely Student distributed with df = 30. With Q_{30} being the upper quantile function of this distribution and t_{30} the density, the log-likelihood is approximately $\ell_s(\psi) = \log t_{30}(Q_{30}(C(\psi))$. Examples are easily made to illustrate that the $\ell_s(\psi)$ log-likelihood quite closely approximates the real Poisson log-likelihood $\ell(\psi) = t_{\mathrm{obs}} - \psi + t_{\mathrm{obs}}\log(\psi/t_{\mathrm{obs}})$. Our point here is to illustrate the approximation technique; when the exact likelihood is available it is of course preferable.

The confidence density depends on the parametrization. By reparametrization, the likelihood can be brought to be proportional to the confidence density. This parametrization might have additional advantages. Let $L(\psi)$ be the likelihood and $c(\psi)$ the confidence density for the chosen parametrization, both assumed positive over the support of the confidence distribution. The quotient $J(\psi) = L(\psi)/c(\psi)$ has an increasing integral $\mu(\psi)$, with $(\partial/\partial\psi)\mu = J$, and the confidence density of $\mu = \mu(\psi)$ is $L(\psi(\mu))$. There is thus always a parametrization that makes the observed likelihood proportional to the observed confidence density. When the likelihood is based upon a pivot of the form $\mu(\psi) - T$, the likelihood in $\mu = \mu(\psi)$ is proportional to the confidence density of μ.

Example 7. Let $\widehat{\psi}/\psi$ be standard exponentially distributed. Taking the logarithm, the pivot is brought on translation form, and $\mu(\psi) = \log\psi$. The likelihood and the confidence density is thus $c(\mu) \propto L(\mu) = \exp\{\widehat{\mu} - \mu - \exp(\widehat{\mu} - \mu)\}$. The log-likelihood has a more normal-like shape in the μ parametrization than in the canonical parameter ψ.

When the likelihood equals the confidence density, the pivot is in broad generality of the translation type. The cumulative confidence distribution function is then of translation type, with $C = F(\mu - \widehat{\mu})$, and so is the likelihood, $L = c = f(\mu - \widehat{\mu})$.

3.2. Bootstrapping a reduced likelihood

Bootstrapping and simulation have emerged as indispensable tools in statistical inference. One typically observes aspects of the bootstrap distribution of an estimator, and uses these to reach accurate inference. In our context it is also useful to focus on bootstrap distributions of the full log-likelihood curves or surfaces. Thus in example 3 (continued) above, the bootstrapped log-likelihood curve takes the form $\ell^*(\sigma) = -v\log\sigma - \frac{1}{2}v(S^*)^2/\sigma^2$, where S^* is drawn from the $S(\chi_v^2/v)^{1/2}$ distribution.

In a parametric bootstrapping experiment, where the data set comprises "old" and "new" parts, all the data need to be simulated, also the parts relating to past data and which enter the analysis through a reduced likelihood. Consider the situation where the distribution of the maximum likelihood estimate $\widehat{\psi} = \arg\max[\max_\chi\{L_{\mathrm{old}}(\psi)L_{\mathrm{new}}(\psi,\chi)\}]$ is sought for various values $(\widetilde{\psi},\widetilde{\chi})$ of the interest parameter and the nuisance parameter. From simulated old and new data, bootstrapped likelihood functions, $L^*_{\mathrm{old}}(\psi)$ and $L^*_{\mathrm{new}}(\psi,\chi)$ are obtained, and by maximization, $\widehat{\psi}^*$.

Focusing on the old data, the reduced likelihood $L_{\text{old}}(\psi)$ is assumed to depend on the scalar statistic T through a pivot $\text{piv}(T, \psi) = V$. Proposition 3 specifies the construction. A draw V^* from the known pivotal distribution produces a bootstrapped value T^* by solving $\text{piv}(T^*, \widetilde{\psi}) = V^*$. This produces in turn a bootstrapped version of the reduced likelihood of the old data,

$$L^*(\psi; \widetilde{\psi}) = f(\text{piv}(T^*, \psi))|\partial \text{piv}(T^*, \psi)/\partial T^*|.$$

In the location and scale model (4), for example, $T^* = \mu(\widetilde{\psi}) + \tau(\widetilde{\psi})V^*$, and

$$L^*(\psi; \widetilde{\psi}) = f\left(\frac{\mu(\widetilde{\psi}) - \mu(\psi) + \tau(\widetilde{\psi})V^*}{\tau(\psi)}\right)\frac{1}{\tau(\psi)}.$$

When the reduced likelihood is normal-based, the parametric bootstrap of the log-likelihood is $\ell^*(\psi; \widetilde{\psi}) = -\frac{1}{2}\{\mu(\widetilde{\psi}) - \mu(\psi) + Z^*\}^2$, where $Z^* \sim N(0, 1)$. In the case of i.i.d. data, one typically has $\mu(\psi) = \sqrt{n}\mu_1(\psi)$, and $\ell^*(\psi; \widetilde{\psi}) = -\frac{1}{2}n\{\mu_1(\widetilde{\psi}) - \mu_1(\psi) + Z^*/\sqrt{n}\}^2$. Modifications of the normal-based likelihood and its bootstrap are called for in cases where $\tau(\psi)$ is not constant in ψ.

4. Confidence power

Let $C(\psi)$ be the cumulative confidence distribution. The intended interpretation of C is that its quantiles are endpoints of confidence intervals. For these intervals to have correct coverage probabilities, the cumulative confidence at the true value of the parameter must have a uniform probability distribution. This is an *ex ante* statement. Before the data have been gathered, the confidence distribution is a statistic with a probability distribution, often based on another statistic through a pivot.

The choice of statistic on which to base the confidence distribution is unambiguous only in simple cases. Barndorff-Nielsen & Cox (1994) are in agreement with Fisher when emphasizing the structure of the model and the data as a basis for choosing the statistic. They are primarily interested in the logic of statistical inference. In the tradition of Neyman and Wald, emphasis has been on inductive behaviour, and the goal has been to find methods with optimal frequentist properties. In exponential families and in other models with Neyman structure (see Lehmann, 1959, ch. 4), it turns out that methods favoured on structural and logical grounds usually also are favoured on grounds of optimality. This agreement between the Fisherian and Neyman–Wald schools is encouraging and helps to reduce the division between the two schools.

4.1. The Neyman–Pearson lemma for confidence distributions

The tighter the confidence intervals are, the better, provided they have the claimed confidence. *Ex post*, it is thus desirable to have as little spread in the confidence distribution as possible. Standard deviation, inter-quantile difference or other measures of spread could be used to rank methods with respect to their discriminatory power. The properties of a method must be assessed *ex ante*, and it is thus the probability distribution of a chosen measure of spread that would be relevant. The assessment of the information content in a given body of data is, however, another matter, and must clearly be discussed *ex post*.

When testing H_0: $\psi = \psi_0$ vs H_1: $\psi > \psi_0$, one rejects at level α if $C(\psi_0) < \alpha$. The power of the test is $\Pr\{C(\psi_0) < \alpha; \psi_1\}$ evaluated at a point $\psi_1 > \psi_0$. Cast in terms of p-values, the power distribution is the distribution at ψ_1 of the p-value $C(\psi_0)$. The basis for test-optimality is monotonicity in the likelihood ratio based on a sufficient statistic, S,

$$\mathrm{LR}(\psi_1, \psi_2; S) = L(\psi_2; S)/L(\psi_1; S) \text{ is increasing in } S \text{ for } \psi_2 > \psi_1. \tag{5}$$

From Schweder (1988) we have the following.

Lemma 5 (Neyman–Pearson for p-values)
Let S be a one-dimensional sufficient statistic with increasing likelihood ratio whenever $\psi_1 < \psi_2$. Let the cumulative confidence distribution based on S be C^S and that based on another statistic T be C^T. In this situation, the cumulative confidence distributions are stochastically ordered:

$$C^S(\psi) \overset{ST(\psi_0)}{\leq} C^T(\psi) \text{ at } \psi < \psi_0 \quad \text{and} \quad C^S(\psi) \overset{ST(\psi_0)}{\geq} C^T(\psi) \text{ at } \psi > \psi_0.$$

In the standard Neyman–Pearson sense, the focus is on rejection vs non-rejection of a hypothesis, which amounts to using a spread measure for C of the indicator type. More generally, every natural measure of spread in C around the true value ψ_0 of the parameter can be expressed as a functional $\gamma(C) = \int_{-\infty}^{\infty} \Gamma(\psi - \psi_0) C(d\psi)$, where $\Gamma(0) = 0$, Γ is non-increasing to the left of zero, and non-decreasing to the right. Here $\Gamma(t) = \int_0^t \gamma(du)$ is the integral of a signed measure γ.

Proposition 6 (Neyman–Pearson for power in the mean)
If S is a sufficient one-dimensional statistic and the likelihood ratio (5) *is increasing in S whenever $\psi_1 < \psi_2$, then the confidence distribution based on S is uniformly most powerful in the mean; that is, $\mathrm{E}_{\psi_0}\gamma(C^S) \leq \mathrm{E}_{\psi_0}\gamma(C^T)$ holds for all spread-functionals γ and at all parameter values ψ_0.*

Proof. By partial integration,

$$\gamma(C) = \int_{-\infty}^{0} C(\psi + \psi_0)(-\gamma)(d\psi) + \int_0^{\infty} (1 - C(\psi + \psi_0))\gamma(d\psi). \tag{6}$$

By lemma 5, $\mathrm{E}_{\psi_0} C^S(\psi + \psi_0) \leq \mathrm{E}_{\psi_0} C^T(\psi + \psi_0)$ for $\psi < 0$ while $\mathrm{E}_{\psi_0}(1 - C^S(\psi + \psi_0)) \leq \mathrm{E}_{\psi_0}(1 - C^T \ (\psi + \psi_0))$ for $\psi > 0$. Consequently, since both $(-\gamma)(d\psi)$ and $\gamma(d\psi) \geq 0$, $\mathrm{E}_{\psi_0}\gamma(C^S) \leq \mathrm{E}_{\psi_0}\gamma(C^T)$. This relation holds for all such spread measures that have finite integral, and for all reference values ψ_0. Hence C^S is uniformly more powerful in the mean than any other confidence distribution.

The Neyman–Pearson argument for confidence distributions can be strengthened.

Proposition 7 (Neyman–Pearson for confidence distributions)
If S is a sufficient one-dimensional statistic and the likelihood ratio (5) *is increasing in S whenever $\psi_1 < \psi_2$, then the confidence distribution based on S is uniformly most powerful; that is, $\gamma(C^S)$ is stochastically less than or equal to $\gamma(C^T)$ for all other statistics T, for all spread-functionals γ, and with respect to the probability distribution at all values of the true parameter ψ_0.*

Proof. Let S be probability transformed to be uniformly distributed at the true value of the parameter, set at $\psi_0 = 0$ for simplicity. Write $\mathrm{LR}(0,\psi;S) = \mathrm{LR}(\psi; S)$ and let $F_\psi(t)$ be the c.d.f. of T. Now,

$$F_\psi(t) = \mathrm{E}_\psi F_\psi(t|S) = \mathrm{E}_\psi F_0(t|S) = \mathrm{E}_0[F_0(t|S)\ \mathrm{LR}(\psi; \mathrm{S})],$$

where the expectation is w.r.t. to the sufficient statistic S. With $C^T(\psi) = 1 - F_\psi(T)$, we have from (6) that

$$\gamma(C^T) = \mathrm{E}_0\left[(1 - F_0(T|S)\int_{-\infty}^{0} \mathrm{LR}(\psi;S)(-\gamma)(d\psi)\right] + \mathrm{E}_0\left[F_0(T|S)\int_{0}^{\infty} \mathrm{LR}(\psi;S)\gamma(d\psi)\right]$$

provided these integrals exist. Now, from the sign of γ and from the monotonicity of the likelihood ratio, $h_-(S) = \int_{-\infty}^{0} \mathrm{LR}(\psi;S)(-\gamma)(d\psi)$ is decreasing in S while $h_+(S) = \int_0^\infty \mathrm{LR}(\psi;S)\gamma(d\psi)$ is increasing in S. The functions φ_- and φ_+ of S that stochastically minimize

$$\mathrm{E}_0\{\varphi_-(S)h_-(S) + \varphi_+(S)h_+(S)\}$$

under the constraint that both $\varphi_-(S)$ and $\varphi_+(S)$ are uniformly distributed at $\psi_0 = 0$, are $\varphi_-(S) = 1 - S$ and $\varphi_+(S) = S$. This choice corresponds to the confidence distribution based on S, and we conclude that $\gamma(C^S)$ is stochastically no greater than $\gamma(C^T)$.

4.2. Uniformly most powerful confidence for exponential families

Conditional tests often have good power properties in situations with nuisance parameters. In the exponential class of models it turns out that valid confidence distributions must be based on the conditional distribution of the statistic which is sufficient for the interest parameter, given the remaining statistics informative for the nuisance parameters. That conditional tests are most powerful among power-unbiased tests is well known, see e.g. Lehmann (1959). There are also other broad lines of arguments leading to constructions of conditional tests, see e.g. Barndorff-Nielsen & Cox (1994). Presently we indicate how and why also the most powerful confidence distributions are of such conditional nature.

Proposition 8

Let ψ be the scalar parameter and χ the nuisance parameter vector in an exponential model, with a density w.r.t. Lebesgue measure of the form

$$p(y) = \exp\{\psi S(y) + \chi_1 A_1(y) + \cdots + \chi_p A_p(y) - k(\psi, \chi_1, \ldots, \chi_p)\},$$

for data vector y in a sample space region not dependent upon the parameters. Assume (ψ, χ) is contained in an open $(p+1)$-dimensional parameter set. Then

$$C_{S|A}(\psi) = \mathrm{Pr}_{\psi,\chi}\{S > S_{\mathrm{obs}} | A = A_{\mathrm{obs}}\}$$

is exact and uniformly most powerful. Here S_{obs} and A_{obs} denote the observed values of S and A.

Proof. The claim essentially follows from previous efforts by a reduction to the one-dimensional parameter case, and we omit the details. A key ingredient is that A is a sufficient and complete statistic for χ when $\psi = \psi_0$ is fixed; this parallels the treatment of Neyman–Pearson optimality of conditional tests for the exponential family, as laid out e.g. in Lehmann (1959). Note that the distribution of S given $A = A_{\mathrm{obs}}$ depends on ψ but not on $\chi_1, \ldots, \chi_p$.

The optimality of the conditional confidence distribution, and thus of conditional tests and confidence intervals, hinges on the completeness of the statistic A for each fixed ψ_0. By completeness, there cannot be more than one exact confidence distribution based on the sufficient statistic. The conditional confidence distribution is exact, and is thus optimal since it is the only exact one. The conditioning statistic A identifies what Fisher (1973) calls recognizable subsets, which in his view leads to conditional inference given A (whether the model is for continuous or discrete data).

Example 8. Consider pairs (X_j, Y_j) of independent Poisson variables, where X_j and Y_j have parameters λ_j and $\lambda_j\psi$, for $j = 1, \ldots, m$. The likelihood is proportional to

$$\exp\left\{\sum_{j=1}^{m} y_j \log\psi + \sum_{j=1}^{m}(x_j + y_j)\log\lambda_j\right\}.$$

Write $S = \sum_{j=1}^{m} Y_j$ and $A_j = X_j + Y_j$. Then $A_1, \ldots, A_m$ become sufficient and complete for the nuisance parameters when ψ is fixed. Also, $Y_j|A_j$ is a binomial $(A_j, \psi/(1+\psi))$. It follows from the proposition above that the (nearly) uniformly most powerful confidence distribution, used here with a half-correction for discreteness, takes the simple form

$$\begin{aligned} C_{S|A}(\psi) &= \Pr_\psi\{S > S_{\text{obs}}|A_{1,\text{obs}}, \ldots, A_{m,\text{obs}}\} + \tfrac{1}{2}\Pr_\psi\{S = S_{\text{obs}}|A_{1,\text{obs}}, \ldots, A_{m,\text{obs}}\} \\ &= 1 - \text{Bin}\Big(S_{\text{obs}}\Big|\sum_{j=1}^{m} A_{j,\text{obs}}, \frac{\psi}{1+\psi}\Big) + \tfrac{1}{2}\text{bin}\Big(S_{\text{obs}}\Big|\sum_{j=1}^{m} A_{j,\text{obs}}, \frac{\psi}{1+\psi}\Big), \end{aligned}$$

where $\text{Bin}(\cdot|n, p)$ and $\text{bin}(\cdot|n, p)$ are the cumulative and pointwise distribution functions for the binomial.

4.3. Large-sample optimality

Consider any regular parametric family, with a suitable density $f(x;\theta)$ involving a p-dimensional parameter θ. Assume data $X_1, \ldots, X_n$ are observed, with consequent maximum likelihood estimator $\widehat{\theta}_n$. Let furthermore θ_0 denote the true value of the parameter. It is well known that $\widehat{\theta}_n$ is approximately distributed as a normal, centred at θ_0, for large n. The following statement is loosely formulated, but may be made precise in various ways. The above situation, for large n, is approximately the same as that of observing a sample from the model with density $\exp\{\sum_{j=1}^{p}\theta_j u_j(x) - B(\theta)\}$, where $u_j(x) = \partial\log f(x;\theta_0)/\partial\theta$ and $B(\theta)$ the appropriate normalization constant. This goes to show that the inference situation is approximated with the form described in proposition 8. This can also be seen through normal approximations of the likelihood. Thus, broadly speaking, the normal confidence distribution constructed by the standard maximum likelihood machinery becomes asymptotically optimal. The following section offers additional insight into this first order large-sample result, along with second order correction recipes.

5. Approximate confidence and reduced likelihoods

Uniformly most powerful exact inference is in the presence of nuisance parameters only available in regular exponential models for continuous data and other models with Neyman structure. Exact confidence distributions exist in a wider class of models, but need not be canonical. The estimate of location based on the Wilcoxon statistic has for example an exact known distribution in the location model where only symmetry is assumed. In more complex models, the statistic upon which to base the confidence distribution might be chosen on various grounds: the structure of the likelihood function, perceived robustness, asymptotic properties, computational feasibility, perspective and tradition of the study. In the given model, it might be difficult to obtain an exact confidence distribution based on the chosen statistic. There are, however, various techniques available to obtain approximate confidence distributions and reduced likelihoods.

Bootstrapping, simulation and asymptotics are useful tools in calculating approximate confidence distributions and in characterizing their power properties. When an estimator, often the maximum likelihood estimator of the interest parameter, is used as the statistic on which the confidence distribution is based, bootstrapping provides an estimate of the sampling distribution of the statistic. This empirical sampling distribution can be turned into an approximate confidence distribution in several ways.

The simplest and most widely used method of obtaining approximate confidence intervals is the delta method. In a sample of size n, let the estimator $\widehat{\theta}_n$ have an approximate multinormal distribution centred at θ and with covariance matrix of the form S_n/n, so that $(n^{1/2})S_n^{-1/2}(\widehat{\theta}_n - \theta) \to_d \mathrm{N}(0, I)$. By the delta method, the confidence distribution for a parameter $\psi = h(\theta)$ is based on linearizing h at $\widehat{\theta}$, and yields

$$C_{\text{delta}}(\psi) = \Phi((\psi - \widehat{\psi})/\widehat{\sigma}_n) \tag{7}$$

in terms of the cumulative standard normal. The variance estimate is $\widehat{\sigma}_n^2 = \widehat{g}^{\text{tr}} S_n \widehat{g}/n$ where $\widehat{g}$ is the gradient of h evaluated at $\widehat{\theta}$. Again, this estimate of the confidence distribution is to be displayed post data with $\widehat{\psi}$ equal to its observed value $\widehat{\psi}_{\text{obs}}$.

This confidence distribution is known to be first order unbiased under weak conditions. That $C_{\text{delta}}(\psi)$ is first order unbiased means that the coverage probabilities converge at the rate $n^{-1/2}$, or that $C_{\text{delta}}(\psi_{\text{true}})$ converges in distribution to the uniform distribution at the $n^{1/2}$ rate. Note also that the confidence density as estimated via the delta method, say $c_{\text{delta}}(\psi)$, is simply the normal density $\mathrm{N}(\widehat{\psi}, \widehat{\sigma}_n^2)$. The additivity of the asymptotically normal pivot implies that the reduced likelihood is Gaussian and actually identical to the confidence density $c_{\text{delta}}(\psi)$.

More refined methods for obtaining confidence distributions are developed below.

5.1. The t-bootstrap method

For a suitable increasing transformation of ψ and $\widehat{\psi}$ to $\gamma = h(\psi)$ and $\widehat{\gamma} = h(\widehat{\psi})$, suppose

$$t = (\gamma - \widehat{\gamma})/\widehat{\tau} \quad \text{is an approximate pivot,} \tag{8}$$

where $\widehat{\tau}$ is a scale estimate. Let R be the distribution function of t, by assumption approximately independent of underlying parameters (ψ, χ). The approximate confidence distribution for γ is thus $C_0(\gamma) = R((\gamma - \widehat{\gamma})/\widehat{\tau})$, yielding in its turn $C(\psi) = R((h(\psi) - h(\widehat{\psi}))/\widehat{\tau})$ for ψ, with appropriate confidence density $c(\psi) = C'(\psi)$. Now R would often be unknown, but the situation is saved via bootstrapping. Let $\widehat{\gamma}^* = h(\widehat{\psi}^*)$ and $\widehat{\tau}^*$ be the result of parametric bootstrapping from the estimated model. Then the R distribution can be estimated arbitrarily well as $\widehat{R}$, say, obtained via bootstrapped values of $t^* = (\widehat{\gamma} - \widehat{\gamma}^*)/\widehat{\tau}^*$. The confidence distribution reported is then as above but with $\widehat{R}$ replacing R:

$$C_{t\text{boot}(\psi)} = \widehat{R}\Big((h(\psi) - h(\widehat{\psi}))/\widehat{\tau}\Big).$$

Example 9. Figure 2 illustrates the t-bootstrap method for the case of the correlation coefficient in the binormal family, using Fisher's zeta transformation $h(\rho) = \frac{1}{2}\log\{(1+\rho)/(1-\rho)\}$ and a constant for $\widehat{\tau}$. The density $c_{t\text{boot}}(\rho)$ is shown rather than its cumulative, and has been computed via numerical derivation. We note that the exact confidence distribution for ρ involves the distribution of the empirical correlation coefficient $\widehat{\rho}$, which however is quite complicated and is available only as an infinite sum (see Lehmann, 1959, p. 210).

This t-bootstrap method applies even when t is not a perfect pivot, but is especially successful when it is, since t^* then has exactly the same distribution R as t. Note that the method automatically takes care of bias and asymmetry in R, and that it therefore aims at

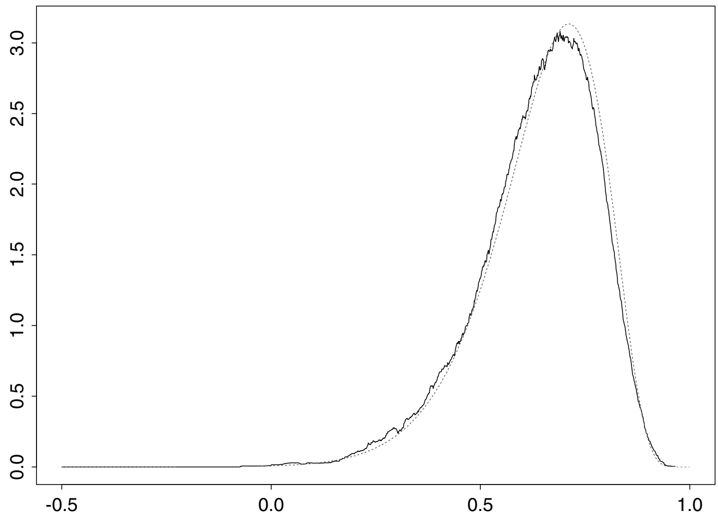

Fig. 2. Approximate confidence density for a binormal correlation coefficient, having observed $\hat{\rho} = 0.667$ from $n = 20$ data pairs, via the t-bootstrap method. The confidence density curve is computed via numerical derivation of the $C_{t\mathrm{boot}}(\rho)$ curve, using 5000 bootstrap samples. The smoother curve is the simpler version which takes a standard normal for the distribution R.

being more precise than the delta method above, which corresponds to zero bias and a normal R. The problem is that an educated guess is required for a successful pivotal transformation h and the scale estimate $\hat{\tau}$. Such might be facilitated and aided by regression analysis of the bootstrapped statistics. Furthermore, the t-bootstrap interval is not invariant under monotone transformations. The following method is not hampered by this shortcoming.

5.2. The acceleration and bias corrected bootstrap method

Efron (1987) introduced acceleration and bias corrected bootstrap percentile intervals, and showed that these have several desirable aspects regarding accuracy and parameter invariance. Here we will exploit some of these ideas, but "turn them around" to construct accurate bootstrap-based approximations to confidence distributions and reduced likelihoods.

Suppose that on some transformed scale, from ψ and $\hat{\psi}$ to $\gamma = h(\psi)$ and $\hat{\gamma} = h(\hat{\psi})$, one has

$$(\gamma - \hat{\gamma})/(1 + a\gamma) - b \sim \mathrm{N}(0, 1) \tag{9}$$

to a very good approximation, for suitable constants a (for acceleration) and b (for bias). Both population parameters a and b tend to be small; in typical setups with n observations, their sizes will be $O(n^{-1/2})$. Assuming $a\hat{\gamma} > -1$, the pivot in (9) is increasing in γ and $C(\gamma) = \Phi\big((\gamma - \hat{\gamma})/(1 + a\gamma) - b\big)$ is the confidence distribution for γ. Thus

$$C(\psi) = \Phi\left(\frac{h(\psi) - h(\hat{\psi})}{1 + ah(\psi)} - b\right) \tag{10}$$

is the resulting confidence distribution for ψ. This constitutes a good approximation to the real confidence distribution under assumption (9). It requires h to be known, however, as well as a and b.

To get around this, look at bootstrapped versions $\hat{\gamma}^* = h(\hat{\psi}^*)$ from the estimated parametric model. If assumption (9) holds uniformly in a neighbourhood of the true parameters, then also

$$(\widehat{\gamma}^* - \widehat{\gamma})/(1 + a\widehat{\gamma}) \sim \mathrm{N}(-b, 1)$$

with good precision. Hence the bootstrap distribution may be expressed as

$$\widehat{G}(t) = \Pr_*\{\widehat{\psi}^* \le t\} = \Pr_*\{\widehat{\gamma}^* \le h(t)\} = \Phi\left(\frac{h(t) - \widehat{\gamma}}{1 + a\widehat{\gamma}} + b\right),$$

which yields $h(t) = (1 + a\widehat{\gamma})\{\Phi^{-1}(\widehat{G}(t)) - b\} + \widehat{\gamma}$. Substitution in (10) is seen to give the abc formula

$$\widehat{C}_{\mathrm{abc}}(\psi) = \Phi\left(\frac{\Phi^{-1}(\widehat{G}(\psi)) - b}{1 + a(\Phi^{-1}(\widehat{G}(\psi)) - b)} - b\right), \tag{11}$$

since $\Phi^{-1}(\widehat{G}(\widehat{\psi})) = b$. Note that an approximation $\widehat{c}_{\mathrm{abc}}(\psi)$ to the confidence density emerges too, by evaluating the derivative of $\widehat{C}_{\mathrm{abc}}$. This may sometimes be done analytically, in cases where $\widehat{G}(\psi)$ can be found in a closed form, or may be carried out numerically.

The reduced abc likelihood is from (9) equal to $L(\gamma) = \phi((\gamma - \widehat{\gamma})/(1 + a\gamma))/(1 + a\gamma)$, which yields the log-likelihood

$$\ell_{\mathrm{abc}}(\psi) = -\frac{1}{2}\{\Phi^{-1}(\widehat{C}_{\mathrm{abc}}(\psi))\}^2 - \log[1 + a\{\Phi^{-1}(\widehat{G}(\psi)) - b\}], \tag{12}$$

since the unknown proportionality factor $1 + a\widehat{\gamma}$ appearing in $h(t)$ is a constant proportionality factor in $L_{\mathrm{abc}}(h(\psi))$.

It remains to specify a and b. The bias parameter b is found from $\widehat{G}(\widehat{\psi}) = \Phi(b)$, as noted above. The acceleration parameter a is found as $a = \frac{1}{6}$skew, where there are several ways in which to calculate or approximate the skewness parameter in question. Extensive discussions may be found in Efron (1987), Efron & Tibshirani (1993, ch. 14, 22) and in Davison & Hinkley (1997, ch. 5), for situations with i.i.d. data. One option is via the jackknife method, which gives parameter estimates $\widehat{\psi}_{(i)}$ computed by leaving out data point i, and use a equal to $(6n^{1/2})^{-1}$ multiplied by the skewness of $\{\widehat{\psi}_{(\cdot)} - \widehat{\psi}_{(1)}, \ldots, \widehat{\psi}_{(\cdot)} - \widehat{\psi}_{(n)}\}$. Here $\widehat{\psi}_{(\cdot)}$ is the mean of the n jackknife estimates. Another option for parametric families is to compute the skewness of the logarithmic derivative of the likelihood, at the parameter point estimate, inside the least favourable parametric subfamily; see again Efron (1987) for more details.

Note that when a and b are close to zero, the abc confidence distribution becomes identical to the bootstrap distribution itself. Thus (11) provides a second order non-linear correction of shift and scale to the immediate bootstrap distribution. We also point out that results similar in nature to (12) have been derived in Pawitan (2000), in a somewhat different context of empirical likelihood approximations.

Example 10. Consider again the parameter $\psi = \sigma_2/\sigma_1$ of example 2. The exact confidence distribution was derived there and is equal to $C(\psi) = 1 - K(\widehat{\psi}^2/\psi^2)$, with $K = K_{\nu_2,\nu_1}$. We shall see how successful the abc apparatus is for approximating the $C(\psi)$ and its confidence density $c(\psi)$.

In this situation, bootstrapping from the estimated parametric model leads to $\widehat{\psi}^* = \widehat{\sigma}_2^*/\widehat{\sigma}_1^*$ of the form $\widehat{\psi}F^{1/2}$, where F has degrees of freedom ν_2 and ν_1. Hence the bootstrap distribution is $\widehat{G}(t) = K(t^2/\widehat{\psi}^2)$, and $\widehat{G}(\widehat{\psi}) = K(1) = \Phi(b)$ determines b. The acceleration constant can be computed exactly by looking at the log-derivative of the density of $\widehat{\psi}$, which from $\widehat{\psi} = \psi F^{1/2}$ is equal to $p(r;\psi) = k(r^2/\psi^2)2r/\psi^3$. With a little work the log-derivative can be expressed as

$$\frac{1}{\psi}\left\{-\nu_2 + (\nu_1 + \nu_2)\frac{(\nu_2/\nu_1)\widehat{\psi}^2/\psi^2}{1 + (\nu_2/\nu_1)\widehat{\psi}^2/\psi^2}\right\} =_d \frac{\nu_1 + \nu_2}{\psi}\left\{\mathrm{Beta}\left(\frac{1}{2}\nu_2, \frac{1}{2}\nu_1\right) - \frac{\nu_2}{\nu_1 + \nu_2}\right\}.$$

Calculating the three first moments of the Beta gives a formula for its skewness and hence for a. (Using the jackknife formula above, or relatives directly based on simulated bootstrap estimates, obviates the need for algebraic derivations, but would give a good approximation only to the a parameter for which we here found the exact value.)

Trying out the abc machinery shows that $\widehat{C}_{\rm abc}(\psi)$ is amazingly close to $C(\psi)$, even when the degrees of freedom numbers are low and imbalanced; the agreement is even more perfect when v_1 and v_2 are more balanced or when they become larger. The same holds for the densities $\widehat{c}_{\rm abc}(\psi)$ and $c(\psi)$; see Fig. 3.

5.3. *Comparisons*

The t-bootstrap method and the abc method remove bias by transforming the quantile function of the otherwise biased normal confidence distribution, $\Phi(\psi - \widehat{\psi})$. The delta method simply corrects the scale of the quantile function, while the abc method applies a shift and a non-linear scale change to remove bias both due to the non-linearity in ψ as a function of the basic parameter θ as well as the effect on the asymptotic variance when the basic parameter is changed. The t-bootstrap method would have good theoretical properties in cases where the $\widehat{\psi}$ estimator is a smooth function of sample averages. Theorems delineating suitable second-order correctness aspects of both the abc and the t-bootstrap methods above can be formulated and proved, with necessary assumptions having to do with the quality of approximations involved in (8) and (9). Methods of proof would, for example, involve Edgeworth or Cornish–Fisher expansion arguments; see e.g. Hall (1992). Such could also be used to add corrections to the delta method (7).

There are still other methods of theoretical and practical interest for computing approximate confidence distributions, cf. the broad literature on constructing accurate confidence intervals. One approach would be via analytic approximations to the endpoints of the abc interval, under suitable assumptions; the arguments would be akin to those found in

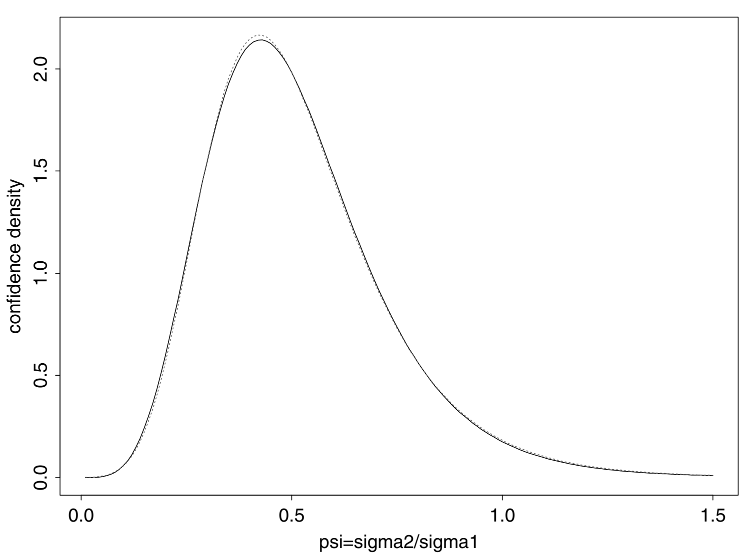

Fig. 3. True confidence density along with abc-estimated version of it (dotted line), for parameter $\psi = \sigma_2/\sigma_1$ with 5 and 10 degrees of freedom. The parameter estimate in this illustration is $\widehat{\psi} = 0.50$. The agreement is even better when v_1 and v_2 are closer or when they are larger.

DiCiccio & Efron (1996) and Davison & Hinkley (1997, ch. 5) regarding "approximate bootstrap confidence intervals". Another approach would be via modified profile likelihoods, following work by Barndorff-Nielsen and others; see Barndorff-Nielsen & Cox (1994, chs. 6, 7), Barndorff-Nielsen & Wood (1998), and Tocquet (2001). Clearly more work and further illustrations are needed the better to sort out which methods have the best potential for accuracy and transparency in different situations. At any rate the abc method (11) appears quite generally useful and precise.

6. Confidence inference in a multinomial recapture model

Assuming captures to be stochastically independent between occasions and letting all individuals having the same capture probability p_t on occasions $t = 1, 2, 3, 4$, say (see below), we have the multinomial multiple-capture model of Darroch (1958). The likelihood

$$L(N, p_1, \ldots, p_4) \propto \binom{N}{X} \prod_{t=1}^{4} p_t^{X_t} (1 - p_t)^{N - X_t}$$

factors in X, the number of unique individuals sampled over the four occasions, and $A = \{X_t\}$, which emerges as a conditioning statistic in inference for N. For fixed N, A is in fact sufficient and complete for the nuisance parameter $\{p_t\}$. The confidence distribution for N based on X in the conditional model given A is therefore suggested by the arguments that led to proposition 8. With half-correction due to discreteness, the c.d.f. of the confidence distribution is

$$C(N) = \Pr_N\{X > X_{\text{obs}} | A\} + \tfrac{1}{2}\Pr_N\{X = X_{\text{obs}} | A\}.$$

The conditional distribution is computed via the hypergeometric distribution. Let R_t be the number of recaptures on occasion t relative to previous captures. Set $R_1 = 0$. The total number of recaptures is $R = \sum_{t=1}^{4} R_t = \sum_{t=1}^{4} X_t - X$. Given the number of unique captures previous to t, $\sum_{i=1}^{t-1}(X_i - R_i)$, R_t has a hypergeometric distribution. In obvious notation, the conditional distribution is therefore

$$\Pr_N\{X = x | \{x_t\}\} = \sum_{r_2=0}^{r} \sum_{r_3=0}^{r-r_2} \prod_{t=2}^{4} \binom{N - \sum_1^{t-1}(x_i - r_i)}{x_t - r_t} \binom{\sum_1^{t-1}(x_i - r_i)}{r_t} \bigg/ \binom{N}{x_t}. \tag{13}$$

The present approach generalizes to an arbitrary number of capture occasions, but it assumes the population to be closed and homogeneous with respect to capturing, which is independent over capturing occasions.

Example 11. Bowhead whales in Alaska. In the summers and autumns of 1985 and 1986, photographs were taken of bowhead whales north of Alaska (see da Silva *et al.*, 2000; Schweder, 2002). We shall be concerned mainly with the immature component of the population that had natural marks on their bodies. The numbers of individuals identified in photographs taken on each of the four sampling occasions and in the pooled set of photographs were 15, 32, 9, 11, 62 respectively.

The confidence distribution for number of immature whales is $C(N) = \Pr_N\{X > 62\} + \frac{1}{2}\Pr_N\{X = 62\}$, calculated in the conditional distribution (13). The conditional probability provides a reduced likelihood for N, viz. $L(N) = \Pr_N\{X = 62\}$. The likelihood happens to be extremely close to the normal-based likelihood calculated from $C(N)$, see Fig. 4. This agreement is due to the underlying conditional pivot. With $\widehat{N}$ being the conditional maximum likelihood estimate and Z a standard normal variate, the pivot is approximately $\widehat{N}^{-1/2} - aN^{-1/2} = \sigma Z$. The coefficients a and σ and the power $-\frac{1}{2}$ are estimated from *q-q*

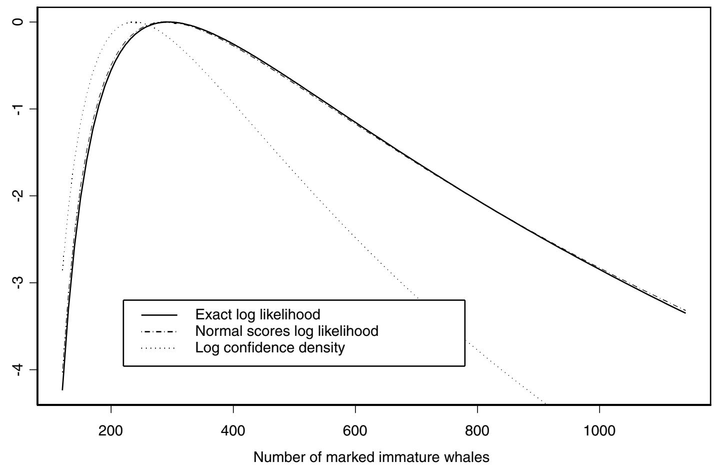

Fig. 4. The exact conditional log-likelihood is shown together with the normal scores log-likelihood and the log-confidence density for the number of marked immature bowhead whales.

plots of the confidence distribution (13) calculated from the observed data. To an amazing accuracy, we find $C(N) \approx \Phi(5.134 - 87.307N^{-1/2})$ over the relevant range in Fig. 4. The natural parameter is $\mu(N) = 1/N^{1/2}$. Due to the non-linearity in the natural parameter, the likelihood is different from the confidence density (taking N to be continuous); in this case the difference is actually substantial, see Fig. 4. (The same picture emerges for mature whales; see Schweder (2002).)

7. Discussion

The context of this paper is that of parametric statistical models with one or more parameters of primary interest. The model is assumed given. The very important questions of model selection, model diagnosis and model uncertainty are not touched upon here. Our concern is the assessment of uncertainty in conclusions due to sampling variability, given the model. Inference concerning a single parameter of primary interest made from a certain set of data has both direct and indirect use. Directly, the inference provides information about the state of nature or society. Indirectly, inference made in one study is used together with new data as part of the combined data for that study.

7.1. Confidence

The confidence distribution is a direct generalization of the confidence interval, and is a useful format of presenting statistical inference. Our impression is that the confidence interval is by far the most popular format of presenting frequentist inference, particularly since reporting an estimate and its standard error is typically viewed as an implicit way of presenting a confidence interval based on the normal distribution. The 95% confidence interval together with the point estimate will often be a natural summary of the confidence distribution. For parameters of great interest, such as the number of bowhead whales in the stock that currently is subject to whaling by native Alaskans, a more complete report of the confidence distribution might be desirable. The confidence distribution, sometimes called the risk profile, is actually used in fisheries stock assessment (Gavaris & Ianelli, 2002). The lower tail of the confidence distribution for stock size is of particular interest when a precautionary approach to management is taken.

When there is more than one interest parameter, simultaneous confidence distributions might be demanded. Unfortunately, the concept of multi-parameter confidence distributions is difficult. The guidance we give is only applicable in special cases, e.g. when there are separate and independent pivots for each of the parameters. The family of confidence are post ellipsoids available in linear models with normal data provides, of course, a multivariate confidence distribution. So does the family of confidence sets obtained by asymptotic considerations, say by the modified likelihood ratio statistic w^* or w^{**} (Skovgaard, 2001). Note, however, that a marginal of a multivariate confidence distribution is typically a confidence distributions for the corresponding parameter only when the parameter is linear in the full parameter vector. A general theory for multi-parameter confidence distributions and their reduction to confidence distributions for derived parameters has not been developed as far as we know.

Following Neyman, the number 0.95 associated with a realized 95% confidence interval is termed degree of confidence. We follow Neyman in this respect, but drop "degree of" for simplicity. The confidence distribution is simply a distribution of confidence *ex post*, and confidence is exactly what we are used to as the degree of trust in a reported confidence interval. Neyman and Fisher disagreed on this point. When Fisher's fiducial probability distribution exists, it will typically be identical to the confidence distribution, and the interval will then be fiducial probability interval. We side with Neyman, Efron and others on the use of "confidence" instead of "probability". And we side with Fisher and with Efron and others that the full distribution (of confidence) is of potential interest. Distributions for parameters are thus understood as confidence distributions and not probability distributions. The concept of probability is reserved for (hypothetically) repeated sampling, and is interpreted frequentistically.

7.2. Likelihood

The likelihood function is the ideal inferential format for indirect use of the results. It gives a sufficient summary of the data, and it allows the future user to give the past data the appropriate weight in the combined analysis. In situations with one interest parameter and several nuisance parameters, there is a need to be freed of the nuisance parameters. This could be done in the final analysis, with past data represented by the full likelihood, or it could be done by representing the past data with a reduced likelihood for the parameter of interest. Getting rid of nuisance parameters is often a taxing operation. A thorough understanding of the data and the underlying probability mechanism is required. For this reason, and also to avoid duplication of work, it will often be natural to carry out this reduction in the initial analysis. Since the reduced likelihood usually comes for free when the pivot behind the confidence distribution is identified, it certainly seems rational to present the reduced likelihood for parameters for which the estimate is suspected to be subject to later updating.

In meta-analysis, one often experiences more variability in estimates across studies than stipulated by the individual likelihoods or standard errors. In such cases, a latent variable representing unaccounted variability in the individual reduced likelihoods might be called for. Anyway, the reduced likelihoods are better suited for meta-analysis than only estimates accompanied by standard errors.

Bootstrapping or simulation is an indispensable tool in confidence estimation. The future user would thus like to simulate the past reduced data along with the new data. This is easily done when the pivot underlying the reduced likelihood representing the past data is known.

A pivot in several parameters and in the same number of statistics determines a reduced likelihood exactly as in section 3. It does also dictate how to simulate the data reduced to these statistics.

When data are weak, it is sometimes tempting to include a component in the likelihood that represents prior beliefs rather than prior data. This subjective component of the likelihood should then, perhaps, be regarded as a penalizing term rather than a likelihood term. Schweder & Ianelli (2000) used this approach to assess the status of the stock of bowhead whales subject to inuit whaling off Alaska. Instead of using the prior density as a penalizing term, they used the corresponding normal-based likelihood. In one corresponding Bayesian analysis, the prior density was, of course, combined multiplicatively with the likelihood of the observed data (Raftery *et al.*, 1995).

7.3. Reporting

To enable the best possible direct use of inference concerning a parameter of primary interest, it should, in our view, be reported as a confidence distribution. Along with the confidence distribution, a report of the underlying pivot and its distribution is helpful for future indirect use of the data. It will usually determine both the reduced likelihood, which could be explicitly reported, and how to carry out future parametric bootstrapping of the then past reduced data. There are, of course, other important things to report than the final conclusions, e.g. the confidence distribution for each of the parameters of primary interest.

The confidence distribution represents the frequentist analogue of the posterior marginal distribution of the interest parameter in a Bayesian analysis, and the reduced likelihood together with information on how to simulate the past reduced data corresponds to the Bayesian's use of a past marginal posterior as a prior for the new analysis. We hope that reporting of confidence distributions and reduced likelihoods (with information on the pivots) will have the same appeal to scientists as does Bayesian reporting.

7.4. The Bayesian paradigm vs the Fisher–Neyman paradigm

As Efron (1998) remarks, both Fisher and Neyman would probably have protested against the use of confidence distributions, but for quite different reasons. Neyman might also have protested against our use of reduced likelihoods. We will, nevertheless, call the approach to statistical inference lined out in the present paper the Fisher–Neyman paradigm.

It is pertinent to compare the frequentist Fisher–Neyman paradigm with the Bayesian paradigm of statistical inference. Most importantly, the two approaches have the same aim: to arrive at distributional knowledge in view of the data within the frame of a statistical model. The distribution could then be subject to further updating at a later stage, etc. In this sense, our approach could be termed "frequentist Bayesian" (a term both frequentists and Bayesians probably would dislike).

We would like to stress as a general point the usefulness of displaying the confidence density $c(\psi)$, computed from the observed data, for any parameter ψ of interest. This would be the frequentist parallel to the Bayesian's posterior density. We emphasize that the interpretation of $c(\psi)$ should be clear and non-controversial; it is simply an effective way of summarizing and communicating all confidence intervals, and does not involve any prior.

One may ask when the $c(\psi)$ curve is identical to a Bayesian's posterior. This is clearly answered by (3) in the presence of a pivot; the confidence density agrees exactly with the Bayesian updating when the Bayesian's prior is proportional to

$$\pi_0(\psi) = \left|\frac{\partial \mathrm{piv}(T;\psi)}{\partial\psi}\right| \Big/ \left|\frac{\partial \mathrm{piv}(T;\psi)}{\partial T}\right|. \tag{14}$$

In the pure location case the pivot is $\psi - T$, and π_0 is constant. When ψ is a scale parameter and the pivot is ψ/T, the prior becomes proportional to ψ^{-1}. These priors are precisely those found to be the canonical "non-informative" ones in Bayesian statistics. In the correlation coefficient example of section 5.1, the approximate pivot used there leads to $\pi_0(\rho) = 1/(1-\rho^2)$ on (−1,1), agreeing with the non-informative prior found using the so-called Jeffreys' formula. Method (14) may also be used in more complicated situations, for example via abc or *t*-bootstrap approximations in cases where a pivot is not easily found. See also Schweder & Hjort (2002).

It is possible for the frequentist to start at scratch, without any (unfounded) subjective prior distribution. In complex models, there might be distributional information available for some of the parameters, but not for all. The Bayesian is then stuck, or has to construct priors. The frequentist will in principle, however, not have problems in such situations. The concept of non-informativity is, in fact, simple for likelihoods. The non-informative likelihoods are simply flat. Non-informative Bayesian priors are, on the other hand, a thorny matter. In general, the frequentist approach is less dependent on subjective input to the analysis than the Bayesian approach. But if subjective input is needed, it can readily be incorporated (as a penalising term in the likelihood).

In the bowhead assessment model (Schweder & Ianelli, 2000) there were more prior distributions than there were free parameters. Without modifications of the Bayesian synthesis approach like the melding of Poole & Raftery (1998), the Bayesian gets into trouble. Due to the Borel paradox (Schweder & Hjort, 1996), the Bayesian synthesis will, in fact, be completely determined by the particular parametrization. With more prior distributions than there are free parameters, Poole & Raftery (1998) propose to meld the priors to a joint prior distribution of the same dimensionality as the free parameter. This melding is essentially a (geometric) averaging operation. If there are independent prior distributional information on a parameter, however, it seems wasteful to average the priors. If, say, all the prior distributions happen to be identical, their Bayesian melding will give the same distribution. The Bayesian will thus not gain anything from k independent pieces of information, while the frequentist will end up with a less dispersed distribution; the standard deviation will, in fact, be the familiar $\sigma/k^{1/2}$.

Non-linearity, non-normality and nuisance parameters can produce bias in results, even when the model is correct. This is well known, and has been emphasized repeatedly in the frequentist literature. Such bias should, as far as possible, be corrected in the reported results. The confidence distribution aims at being unbiased: when it is exact, the related confidence intervals have exactly the nominal coverage probabilities. Bias correction has traditionally not been a concern to Bayesians. There has, however, been some recent interest in the matter. To obtain frequentist unbiasedness, the Bayesian will have to choose the prior with unbiasedness in mind. Is this, then, Bayesian? The prior distribution will then not represent prior knowledge of the parameter in case, but an understanding of the model. In the Fisher–Neyman paradigm, this problem is in principle solved. It takes as input (unbiased) prior confidence distributions converted to reduced likelihoods and delivers (unbiased) posterior confidence distributions.

There are several other issues that could have been discussed, such as practicality and ease of communicating results to the user. Let us only note that exact pivots and thus exact confidence distributions and reduced likelihoods are essentially available only in regular exponential models for continuous data. In other models one must resort to approximate solutions in the Fisher–Neyman tradition. Whether based on asymptotic considerations or simulations, often guided by invariance properties, an *ad hoc* element remains in frequentist inference. The Bayesian machinery will, however, in principle always deliver an exact posterior when the prior and the likelihood is given. This posterior might, unfortunately, be wrong from a frequentist point of view in the sense that in repeated use in the same situation it will tend to miss the target. Is it then best to be approximately right, and slightly arbitrary in the result, as in the Fisher–Neyman case, or arriving at exact and unique but perhaps misleading Bayesian results?

7.5. The grand perspective

Hald (1998) speaks of three revolutions in parametric statistical inference, due to Laplace in 1774 (inverse probability, Bayesian methods with flat priors), Gauss and Laplace in 1809–1812 and Fisher in 1922. This is not the place to discuss Fisher's revolution in any detail, other than to note that it was partly a revolt against the Laplacian Bayesianism. When discussing Neyman's 1934 paper on survey sampling (Neyman, 1934), Fisher stated, "All realised that problems of mathematical logic underlay all inference from observational material. They were widely conscious, too, that more than 150 years of dispute between the pros and the cons of inverse probability had left the subject only more befogged by doubt and frustration." To get around the problems associated with prior distributions, Fisher proposed the fiducial distribution as a replacement for the Bayesian posterior. Efron (1998) emphasizes the importance of the fiducial distribution, which he prefers reformulated to the confidence distribution discussed in the present paper. The fiducial argument is not without problems (see e.g. Brillinger, 1962; Wilkinson, 1977; Welsh, 1996) and has often been regarded as "Fisher's biggest blunder" (see Efron, 1998). By converting to the confidence formulation, as Neyman did in 1941 but which Fisher resisted, Efron holds that the method can be applied to a wider class of problems and that it might hold a key to "our profession's 250-year search for a dependable objective Bayes theory". We agree, and hope with Efron (1998) and also with Fraser when discussing Efron (1998) that fiducial or confidence distributions will receive renewed interest. By emphasizing the importance of the reduced likelihood associated with a confidence distribution, and by pointing out the role of the underlying (approximate) pivot for future parametric bootstrapping, a form of objective Bayes methodology has been sketched. Our form of 'frequentist Bayesianism' does not involve Bayes' formula, although we have nothing against using Bayesian techniques to produce confidence distributions with correct frequentist properties. But it seeks to deliver digested statistical information in the format of distributions, and it provides a method for rational updating of such statistical information.

Acknowledgements

We are grateful for helpful and constructive comments from referees, an associate editor and the editor, which contributed to improvements in content and presentation.

References

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

Barndorff-Nielsen, O. E. & Wood, T. A. (1998). On large deviations and choice of ancillary for p^* and r^*. *Bernoulli* **4**, 35–63.

Brillinger, D. R. (1962). Examples bearing on the definition of fiducial probability with a bibliography. *Ann. Math. Statist.* **33**, 1349–1355.

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation*. Houghton Mifflin, Boston.

Darroch, J. N. (1958). The multiple-recapture census. I: Estimation of a closed population. *Biometrika* **45**, 343–359.

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press, Cambridge.

DiCiccio, T. J. & Efron, B. (1996). Bootstrap confidence intervals (with discussion). *Statist. Sci.* **11**, 189–228.

Edwards, A. W. F. (1992). *Likelihood* (expanded edition). John Hopkins University Press, Baltimore.
Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.* **82**, 171–200.
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. & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall, London.
Fisher, R. A. (1922). On the mathematical foundation of theoretical statistics. *Philos. Trans. Roy. Soc. Lond.* **A 222**, 309–368.
Fisher, R. A. (1930). Inverse probability. *Proc. Cambridge Philos. Soc.* **26**, 528–535.
Fisher, R. A. (1973). *Statistical methods and scientific inference* (3rd edn). Hafner Press.
Fraser, D. A. S. (1968). *The structure of inference*. Wiley, New York.
Fraser, D. A. S. (1996). Some remarks on pivotal models and the fiducial argument in relation to structural models. *Internat. Statist. Rev.* **64**, 231–236.
Fraser, D. A. S. (1998). Contribution to the discussion of Efron's paper. *Statist. Sci.* **13**.
Gavaris, S. and Ianelli, J. N. (2002). Statistical issues in fisheries stock assessments. *Scand. J. Statist.* **29**, 245–271.
Hald, A. (1998). *A history of mathematical statistics from 1750 to 1930*. Wiley, New York.
Hall, P. (1992). *The bootstrap and Edgeworth expansions*. Springer-Verlag, Budapest.
Lehmann, E. L. (1959). *Testing statistical hypotheses*. Wiley, New York.
Lehmann, E. L. (1983). *Theory of point estimation*. Wiley, New York.
Lehmann, E. L. (1993). The Fisher, Neyman–Pearson theories of testing hypotheses: one theory or two? *J. Amer. Statist. Assoc.* **88**, 1242–1249.
Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection (with discussion). *J. Roy. Statist. Soc. Ser. B* **97**, 558–625.
Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika* **32**, 128–150.
Pawitan, Y. (2000). Computing empirical likelihood from the bootstrap. *Statist. Probab. Lett.* **47**, 337–345.
Poole, D. & Raftery, A. E. (1998). Inference in deterministic simulation models: The Bayesian melding approach. Technical Report No. 346, Department of Statistics, University of Washington.
Raftery, A. E., Givens, G. H. & Zeh, J.E. (1995). Inference from a deterministic population dynamics model for bowhead whales (with discussion). *J. Amer. Statist. Assoc.* **90**, 402–430.
Royall, R. M. (1997). *Statistical evidence. A likelihood paradigm*. Chapman & Hall, London.
Schweder, T. (1988). A significance version of the basic Neyman–Pearson theory for scientific hypothesis testing (with discussion). *Scand. J. Statist.* **15**, 225–242.
Schweder, T. (2002). Abundance estimation from photo-identification data: confidence distributions and reduced likelihood for bowhead whales off Alaska. *Biometrics* (to appear).
Schweder, T. & Hjort, N. L. (1996). Bayesian synthesis or likelihood synthesis—what does Borel's paradox say? *Rep. Internat. Whaling Comm.* **46**, 475–479.
Schweder, T. & Hjort, N. L. (2002). Frequentist analogues of priors and posteriors. In *The science of econometrics* (ed. B. Stigum) (to appear).
Schweder, T. & Ianelli, J. N. (2000). Assessing the Bering–Chukchi–Beaufort Seas stock of bowhead whales from survey data, age-readings and photo-identifications using frequentist methods. Paper IWC2000/SC/52/AS13 presented to International Whaling Commission (unpublished).
da Silva, C. Q., Zeh, J., Madigan, D., Lake, J., Rugh, D., Baraff, L., Koski, W. and Miller, G. (2000). Capture–recapture estimation of bowhead whale population size using photo-identification data. *J. Cetacean Res. Manag.* **2**, 45–61.
Skovgaard, I. M. (2001). Likelihood asymptotics. *Scand. J. Statist.* **28**, 3–32.
Tocquet, A. S. (2001). Likelihood based inference in non-linear regression models using the p^* and R^* approach. *Scand. J. Statist.* **28**, 429–443.
Welsh, A. H. (1996). *Aspects of statistical inference*. Wiley, New York.
Wilkinson, G. N. (1977). On resolving the controversy in statistical inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 119–171.

Received January 2001, in final form October 2001

Tore Schweder, Department of Economics, P.B. 1095, N–0317 Oslo, Norway.
E-mail: tore.schweder@econ.uio.no