Frequentist Analogues of Priors and Posteriors

Tore Schweder
Department of Economics, University of Oslo

Nils Lid Hjort
Department of Mathematics, University of Oslo

April 1999

Abstract

Independent data are efficiently integrated by adding their respective log-likelihoods. Instead of Bayesian updating of information, we propose to use the likelihood directly as a vehicle for coherent learning. If past data are summarised in a likelihood, it might be termed a prior likelihood component when integrated with new data. In the frequentist tradition, statistical reporting is often done in the format of confidence intervals. The confidence distribution, with quantiles specifying all possible confidence intervals provides a more complete report than a 95% interval, say. The concept of confidence distribution is discussed, and a new version of the Neyman–Pearson lemma is provided. Confidence distributions based on prior data represent frequentist analogues to Bayesian priors. These confidence distributions need to be converted to likelihoods before they can be integrated with the new data likelihood. This can be done if their probability bases are reported. Confidence distributions resulting from the integrated analysis, along with their probability bases, represent the frequentist analogue to the Bayesian posterior distributions.

Key Words: *confidence distributions and densities, frequentist posteriors and priors, integrating information, likelihood, Neyman–Pearson lemma, pivots, whales*

1 Introduction

Classical statistics of the Fisher–Neyman breed has a tendency to focus on the new data in isolation. Other sources of information about the parameters of the model might enter the analysis in the format of structural assumptions in the model, possibly as restrictions on certain parameters, or as hypotheses to be tested, etc. Coherent learning in the Bayesian sense of updating distributional knowledge of a parameter in the light of new data has no parallel in the non-Bayesian likelihood- or frequentist tradition. However, the likelihood function is the pre-eminent tool for integrating diverse data, and provided the distributional information concerning a parameter has a likelihood representation, it could simply be integrated with the likelihood of the new independent data by multiplication.

In the purist likelihood tradition of Edwards (1992) and Royall (1997), the likelihood function itself is the primitive. In this tradition, the likelihood function of the interest parameter is reported and interpreted without recourse to confidence intervals, p-values or the like. The parallel to Bayesian updating is thus straightforward. This process of updating a likelihood we term *likelihood updating*. When the current knowledge of a parameter is represented by a likelihood function, this likelihood is updated when combined with the likelihood of new data. The ratio between the updated and the previous likelihood represents the new information.

The majority of statisticians and econometricians are frequentists. Their favourite format of statistical reporting is the confidence interval, the p-values for certain hypotheses of interest, or point estimates accompanied with measures of uncertainty like the standard error. The likelihood function is, of course, a central concept in the frequentist tradition, but is mainly used as a tool to obtain efficient statistics with (asymptotic) frequentist interpretations. Unfortunately, the statistics reported by frequentist statisticians are seldom sufficient as input to likelihood updating. To achieve this, the reported statistics have to be converted to a likelihood function.

We will mainly be concerned with the reporting of frequentist statistics related to a scalar parameter, and their conversion to a likelihood component. This does not preclude us from discussing integrative likelihood analysis in multi-parameter models. As in many Bayesian situations, prior information is assumed to come in independent packages, one for each one-dimensional parameter. As the Bayesian, we will also assume that the prior information comes in the format of a distribution. Instead of regarding the prior distribution as a probability distribution, we will, however, assume that the distribution represents confidence intervals by its quantiles. Such distributions are called confidence distributions, and are studied in some detail in Section 2.

First we dwell on the distinction between confidence distributions and probability distributions. The well-known duality between hypothesis testing and confidence intervals emerges as a relationship between the p-value for a sequence of tests and the cumulative confidence distribution function. In Section 3 a version of the Neyman–Pearson lemma is provided, explaining the frequentist optimality of the confidence distribution in one-parameter models with monotone likelihood ratio. This also leads to optimal constructions of confidence distributions in parametric families of the exponential kind, via conditioning on ancillary statistics. These confidence distributions become uniformly most reliable in a sense made precise in Section 3.

We hold that for parameters of primary interest the complete confidence distribution should be reported, not only a pair of quantiles like the endpoints of the 95% confidence interval. We argue in particular that reporting the confidence density for interest parameters, the derivative of the cumulative confidence distribution, is an effective way of summarising results from statistical analyses of data, both presentationally and inferentially. This post-data parameter density curve shares some of the features and appealing aspects of the Bayesian posterior, but is purely frequentistic, and should also be non-controversial conceptually.

It is desirable to develop methods for obtaining approximate confidence distributions in situations where exact constructions either become too intricate or do not exist. In Section 4 we discuss various approximations, the simplest of which being based on the traditional delta method for asymptotic normality. Better versions emerge via corrections of various sorts. In particular we develop an acceleration and bias corrected bootstrap percentile interval method for constructing improved confidence densities. It has an appealing form and is seen to perform well in terms of accuracy.

For large samples, the asymptotic normality of regular statistics allows the confidence distribution to be turned into a likelihood function through its normal scores. This likelihood, called the normal-based likelihood, agrees with the so-called implied likelihood of Efron (1993). However, if the information content in the data is of small to moderate weight, the normal-based likelihood might be misleading. An example is provided in Section 5 to show that a given confidence distribution might relate to many different likelihood functions, dependent on the sampling situation behind the confidence distribution. For this reason it is advisable to supplement the reported confidence distribution with sufficient information concerning its probability basis to enable future readers to recover an acceptable likelihood function related to the confidence distribution.

Section 6 develops theory for confidence and likelihoods in models with exact or approximate pivots, while Section 7 considers ways of bootstrapping a given confidence distributions. An illustration is given for the problem of assessing the stock of bowhead whales outside Alaska.

Finally supplementing remarks and discussion are found in Section 8.

There is considerable current interest in building principles for and evaluating applications of combining different information sources in nontrivial situations. Application areas range from economics to assessments of whale stocks. Schweder and Hjort (1997) introduced likelihood synthesis to counter problems encountered with earlier attempts at Bayesian synthesis, and reviewed part of the literature. Recent articles on this theme include Berger, Liseo and Wolpert (1998) on eliminating nuisance parameters via integrated likelihoods as well as Poole and Raftery (1998) on 'Bayesian melding'; see also comments in Sections 7 and 8 below. Some discussion of bootstrap likelihoods and likelihoods based on confidence sets can be found in Davison and Hinkley (1997, Ch. 10).

2 Confidence distributions

Before relating confidence distributions to likelihoods, it is worthwhile having a closer look at the concept as a format of reporting statistical inference.

2.1 Confidence and statistical inference

Our context is a parametric model with an interest parameter ψ for which inference is sought. The interest parameter is assumed to be scalar, and to belong to a finite or infinite interval on the real line. The space of the parameter is thus linearly ordered. 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.

As the name indicates, the confidence distribution is related to confidence intervals, which are interval statements with the confidence fixed *ex ante*, and with endpoints calculated from the data. A one-sided confidence interval with (degree of) confidence $1-\alpha$ has right endpoint the corresponding quantile of the confidence distribution. If C is the cumulative confidence distribution calculated from the data, the left-sided confidence interval is $(-\infty, C^{-1}(1-\alpha))$. A right-sided confidence interval $(C^{-1}(\alpha), \infty)$ has confidence $1-\alpha$, and a two-sided confidence interval $[C^{-1}(\alpha), C^{-1}(\beta)]$ has confidence $\beta-\alpha$. Two-sided confidence intervals are usually equitailed in the sense that $\alpha = 1-\beta$.

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

Lemma 1 *The confidence of the statement '$\psi \leq \psi_0$' is the cumulative confidence distribution function value $C(\psi_0)$, and is equal to the p-value of a test of $H_0\colon \psi \leq \psi_0$ versus the alternative $H_1\colon \psi > \psi_0$.*

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 '$\psi > \psi_0$'.

Some care is needed when calculating and interpreting the confidence for statements determined *ex ante*. When ψ_0 is fixed, the statement '$\psi \neq \psi_0$' should, preferably, have confidence given by one minus the p-value when testing $H_0\colon \psi = \psi_0$. This can be calculated from the observed confidence distribution, and is $1 - 2\min\{C(\psi_0), 1 - C(\psi_0)\}$. It is, however, hard to see how ψ_0 could be calculated *ex post* making the statement '$\psi \neq \psi_0$' have any non-trivial confidence fixed in advance.

Confidence intervals are invariant w.r.t. monotone transformations. This is also the case for confidence distributions.

Lemma 2 *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*

$$C^\rho(\rho) = C^\psi(r^{-1}(\rho)).$$

To a large extent statistical inference is being carried out as follows. From optimality or structural considerations, an estimator of the parameter of interest, and possibly of the remaining (nuisance) parameters in the model, is determined. Then, the sampling distribution of the estimator is calculated, possibly by bootstrapping. Finally, statements of inference, e.g. confidence intervals, are extracted from the sampling distribution and its dependence on the parameter.

A sharp distinction should be drawn between the (estimated) sampling distribution and the confidence distribution. The sampling distribution of the estimator is the *ex ante* probability distribution of the statistic under repeated sampling, while the confidence distribution is calculated *ex post* and distributes the confidence 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

$$S(\psi) = \Pr\{\psi^* \leq \psi \mid \widehat{\psi}\} = F_{\widehat{\psi}}(\psi).$$

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 reasonable statistic and the hypothesis $H_0\colon \psi \leq \psi_0$ is suspect when $\widehat{\psi}$ is large, the p-value is $\Pr\{\psi^* > \widehat{\psi} \mid \psi_0\}$. The cumulative confidence distribution is then

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

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.

The confidence densities we deduce or approximate in the following would presumably be equivalent to the infamous fiducial distributions in the sense of Fisher, at least in cases where Fisher would have considered the mechanism behind the confidence limits to be inferentially correct; see the discussion in Efron (1998, Section 8). In view of old and on-going controversies and confusion surrounding this theme of Fisher, and the fact that such fiducial distributions sometimes have been put forward in ad hoc fashions and with vague interpretation, we emphasise that our distributions of confidence are actually derived from certain principles in a rigorous framework, and with a clear interpretation. Our work can perhaps be seen as being in the spirit of Neyman (1941). We share the view expressed in Lehmann (1993) that the distiction between the Fisherian and the Neyman–Pearson tradition is unfortunate. 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 3, in particular, work towards establishing confidence distributions that are inferentially correct.

Example 1. Consider the exponentially distributed variate T with probability density $f(t\,|\,\psi) = (1/\psi)\exp(-t/\psi)$. The cumulative confidence distribution function for ψ is $C(\psi\,|\,t) = \exp(-t/\psi)$. The confidence density is thus $c(\psi\,|\,t) = (\mathrm{d}/\mathrm{d}\psi)C(\psi\,|\,t) = t\psi^{-2}\exp(-t/\psi)$, which not only has a completely different interpretation from the sampling density of the maximum likelihood estimator, T, but also has a different shape. ■

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 $\widehat{\sigma}_j^2 = \sigma_j^2 W_j/\nu_j$ are available, where W_j is a $\chi^2_{\nu_j}$. 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 $K = K_{\nu_2,\nu_1}$ is the distribution function for the F statistic $(W_2/\nu_2)/(W_1/\nu_1)$. 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 $k = k_{\nu_2,\nu_1}$. See also Section 3.2 for an optimality result of the confidence density used here, and Section 4.3 for a very good approximation based on bootstrapping. ■

The calculation of the confidence distribution is easy when a pivot statistic for ψ is available. The random variable $\text{piv}(X, \psi)$ is a pivot (Barndorff-Nielsen and Cox, 1994) in a model with nuisance parameter χ and data X if the probability distribution of $\text{piv}(X, \psi)$ is the same for all (ψ, χ), and, secondly, the function $\text{piv}(x, \psi)$ is monotone and increasing in ψ for almost all x.

When the confidence distribution is based on a pivot, and F is the cumulative distribution function of the pivot, the confidence distribution is

$$C(\psi) = F(\text{piv}(X, \psi)). \tag{2}$$

This can also be turned around. If, in fact, $C(\psi; X)$ is a cumulative confidence distribution based on data X, then it is a pivot since at ψ it is uniformly distributed. Thus, a confidence distribution based on a sufficient statistic exists if and only if there is a pivot based on the sufficient statistic. And the cumulative confidence distribution function is simply the probability transform of the pivot.

2.2 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 $\nu = n - p$ degrees of freedom. Letting $t_\nu(\alpha)$ be the quantiles of this t-distribution, the confidence quantiles for β_j are the familiar $\widehat{\beta}_j + t_\nu(\alpha) S_j$. The cumulative confidence distribution function for β_j is seen from this to become

$$C(\beta_j \,|\, \text{data}) = 1 - G_\nu((\widehat{\beta}_j - \beta_j)/S_j) = G_\nu((\beta_j - \widehat{\beta}_j)/S_j),$$

where G_ν is the cumulative t-distribution with ν degrees of freedom. Note also that the confidence density $c(\beta_j \,|\, \text{data})$ is the t_ν-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 $\text{SSR}/\sigma^2 = \nu S^2/\sigma^2$ is a χ^2_ν, and the cumulative confidence distribution is found to be

$$C(\sigma \,|\, \text{data}) = \Pr\{\chi^2_\nu > \text{SSR}/\sigma^2\} = 1 - \Gamma_\nu(\nu S^2/\sigma^2),$$

where Γ_ν is the cumulative distribution function of the chi-square with density γ_ν. The confidence density becomes

$$c(\sigma \,|\, \text{data}) = \gamma_\nu\Big(\frac{\nu S^2}{\sigma^2}\Big)\frac{2\nu S^2}{\sigma^3} = \frac{S^\nu}{2^{\nu/2}\Gamma(\frac{1}{2}\nu)}\sigma^{-(\nu+1)}\exp(-\tfrac{1}{2}\nu S^2/\sigma^2),$$

which again is different from the likelihood. The likelihood, for the SSR part of the data, is the density of $\text{SSR} = \sigma^2\chi^2_\nu$, which is proportional to

$$L(\sigma) = \sigma^{-\nu}\exp(-\tfrac{1}{2}\nu S^2/\sigma^2).$$

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.

2.3 Ratio parameters and the Fieller solution

Ericsson, Jansen, Kerbesian and Nymoen (1998) discuss the empirical basis for the weights in monetary conditions indices, MCI. With R being the interest rate and e the exchange rate, $\text{MCI} = \psi(R - R_0) + (e - e_0)$. The relative weight of the interest rate, ψ, is estimated as the ratio of two regression coefficients, $\widehat\psi = \widehat\beta_1/\widehat\beta_2$ in a linear regression. Both regression parameters are assumed non-negative, and thus, $0 \le \psi \le \infty$. For Norway, Ericsson et al. (1998) found $[0, \infty]$ as the 95% confidence interval for ψ, and for the United States $[-\infty, \infty]$, since the point estimate is negative. It is an embarrassing conceptual problem to have a 95% confidence interval covering the whole range of the parameter, which certainly should have had confidence 100% and not 95%. This is known as the Fieller–Cressy problem, see e.g. Koschat (1987) and Dufour (1997).

Assume $(\widehat\beta_1, \widehat\beta_2)^{\text{tr}}$ to be $\text{N}((\beta_1, \beta_2)^{\text{tr}}, \Sigma)$, with $\Sigma = \sigma^2\Sigma_0$ estimated at df degrees of freedom. A confidence distribution for the quotient $\psi = \beta_1/\beta_2$ is found from inverting the normal test of $H_0\colon \psi = \psi_0$ versus $H_1\colon \psi > \psi_0$. The hypothesis is first reformulated to $H_0\colon \beta_1 - \psi_0\beta_2 = 0$, and the resulting p-value and cumulative confidence distribution function is that of Fieller (1940, 1953);

$$C(\psi) = G_{\text{df}}\left(\frac{\widehat\beta_1 - \psi\widehat\beta_2}{\widehat\sigma\sigma_0(\psi)}\right), \quad \sigma_0^2(\psi) = (1, \psi)\Sigma_0(1, \psi)^{\text{tr}}.$$

Again, G_{df} is the cumulative t-distribution function with df degrees of freedom.

Now, the Fieller test statistic is $\widehat\beta_1 - \psi_0\widehat\beta_2$ and orders data differently from the more natural test statistic $|\widehat\psi - \psi_0|$. One might therefore suspect that something is lost by the Fieller solution. However, Koschat (1987) found that no reasonable method other than the Fieller solution provides confidence intervals with exact coverage probabilities when $\Sigma_0 = I$. This was shown for the angle related to the ratio. Consequently, if something is to be gained in power by another ordering of data sets, it must be at the price of only obtaining approximate null behavior.

In polar coordinates, the angle is $\theta = \tan^{-1}(\psi)$ and the radial distance $\delta = (\beta_1^2 + \beta_2^2)^{1/2}$, with empirical counterparts $\widehat\theta$ and $\widehat\delta$. The Fieller confidence quantile function is

$$\theta \sim \widehat\theta \pm \begin{cases} \pi/2 & \text{if } \widehat\delta < \widehat\sigma t, \\ \arcsin(\widehat\sigma t/\widehat\delta) & \text{if } \widehat\delta \ge \widehat\sigma t, \end{cases} \tag{3}$$

where t is the appropriate quantile from the t-distribution with df degrees of freedom. When σ is known, df $= \infty$, which is the case we look at first. This formulation of the Fieller confidence distribution is given by Koschat (1987).

With support on $(0, \pi/2)$ the confidence distribution for θ could be transformed to a confidence distribution for $\psi = \tan(\theta)$ since the tan function is monotonously increasing over the first quarter sector. However, the confidence distribution for θ will have support outside $(0, \pi/2)$, and the discontinuity of the tan function at $\theta = \pi/2$ is an obstacle for the confidence distribution for ψ. Invoking the restriction $0 \le \psi \le \infty$, the confidence distribution for ψ based on the Fieller solution has mass less than 1 on the interior of its range, $(0, \infty)$ and point-mass at the boundary points.

3 Confidence level and confidence reliability

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 uniformly probability distribution. This is an *ex ante* statement. Before the data have been gathered, the confidence distribution is a stochastic element and $C(\psi_{\text{true}})$ is a random variable. If

$$\Pr_\psi\{C(\psi) \le c\} = c \quad \text{for } 0 \le c \le 1, \tag{4}$$

assuming the probability distribution to be continuous, the coverage probability of $(-\infty, C^{-1}(\alpha)]$ is

$$\Pr_\psi\{\psi \le C^{-1}(\alpha)\} = \alpha,$$

which conventionally is called the confidence level of the interval. The confidence distribution is exact if (4) holds exactly, and thus the coverage probability of a confidence interval obtained from C equals the nominal confidence level.

Confidence distributions provide point estimates, the most natural being the confidence median, $\widehat{\psi} = C^{-1}(0.5)$. When the confidence distribution is exact, this point estimator is median-unbiased. This property is kept under monotone transformations of the parameter.

The choice of statistic on which to base the confidence distribution is unambiguous only in simple cases. Barndorff-Nielsen and Cox (1994) are in agreement with Fisher when emphasising 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 have been on inductive behaviour, and the goal has been to find methods with optimal frequentist properties. In nice models like exponential families it turns out that methods favoured on structural and logical grounds also are favoured on grounds of optimality. This agreement between the Fisherian and Neyman–Wald schools is encouraging and helps to reduce the distinction between the two schools. This core of statistical theory needs to be reformulated in terms of confidence distributions.

3.1 Reliability and power

A method is *reliable* when it leads to similar conclusions for repeated samples. The more reliable, the less variability in results. A method that is both exact and reliable gives results that vary little, and which are centred at the truth. A cumulative confidence distribution is monotone: at $\psi > \psi_{\text{true}}$, one should have $C(\psi) \ge C(\psi_{\text{true}})$, etc. When C is exact, $C(\psi_{\text{true}}) \sim U$ (uniform on the unit interval), and above the true value, $C(\psi)$ must be stochastically larger than U (have cumulative distribution function less than that of U). Since $1 \geq C(\psi)$, the more the *ex ante* probability distribution of $C(\psi)$ is shifted towards its upper limit, the less variability it has in repeated samples. For $\psi < \psi_{\text{true}}$, it is desirable to have the probability distribution of $C(\psi)$ concentrated as much as possible towards low values.

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*.

In the standard Neyman–Pearson theory, the focus is on spread-measures of the indicator type, $\Gamma(t) = I(t > \psi_{\text{alt}})$ etc. When testing $H_0\colon \psi = \psi_0$ versus $H_1\colon \psi > \psi_0$, one rejects at level α if $C(\psi_0) < \alpha$. The power of the test is $\Pr\{C(\psi_0) < \alpha\}$ 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 3 (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:*

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

Now, every natural measure of spread in C around the true value of the parameter, ψ_0, can be expressed as a functional $\gamma(C) = \int_{-\infty}^{\infty} \Gamma(\psi - \psi_0)\, C(\mathrm{d}\psi)$, where $\Gamma(0) = 0$, Γ is non-increasing to the left of zero, and non-decreasing to the right. Here $\Gamma(t) = \int_0^t \gamma(\mathrm{d}u)$ is the integral of a signed measure γ.

Agree to say that a confidence distribution C^S is uniformly more reliable in expectation than C^T if

$$\mathrm{E}_{\psi_0} \gamma(C^S) \leq \mathrm{E}_{\psi_0} \gamma(C^T)$$

holds for all spread-functionals γ and at all parameter values ψ_0. With this definition, the Neyman–Pearson lemma yields the following.

Proposition 4 (Neyman–Pearson for mean-reliability) *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 reliable in expectation.*

Proof. By partial integration,

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

By the Neyman–Pearson lemma, $\mathrm{E}C^S(\psi + \psi_0) \leq \mathrm{E}C^T(\psi + \psi_0)$ for $\psi < 0$ while $\mathrm{E}(1 - C^S(\psi + \psi_0)) \leq \mathrm{E}(1 - C^T(\psi + \psi_0))$ for $\psi > 0$. Consequently, since both $(-\gamma)(\mathrm{d}\psi)$ and $\gamma(\mathrm{d}\psi) \geq 0$,

$$\mathrm{E}_{\psi_0} \gamma(C^S) \leq \mathrm{E}_{\psi_0} \gamma(C^T).$$

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 reliable than any other confidence distribution. ■

The Neyman–Pearson argument for confidence distributions can be strengthened. Say that a confidence distribution C^S is uniformly most reliable if, *ex ante*, $\gamma(C^S)$ 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.

Proposition 5 (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 reliable.*

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 $\mathrm{LR}(\psi_0, \psi; S) = \mathrm{LR}(\psi; S)$. By conditioning, and using the sufficiency of S, $C^T(\psi) = 1 - \mathrm{E}_\psi F_0(T \mid S) = 1 - \mathrm{E}_0\left[F_0(T \mid S)\mathrm{LR}(\psi; S)\right]$. Thus, from (6),

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

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

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

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 $\gamma(C^T)$. ■

3.2 Uniformly most powerful reliability 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 and Cox (1994). Presently we indicate how and why also the most reliable confidence distributions are of such conditional nature.

Proposition 6 *Let ψ be the scalar parameter and χ the nuisance parameter vector in an exponential model, with density of the form*

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

with respect to a fixed measure, for data vector x in a sample space region not dependent upon the parameters. Assume (ψ, χ) is contained in an open $(p + 1)$-dimensional parameter set. Then, for ψ and hence for all monotone transforms of ψ, there exist exactly valid confidence distributions, and the uniformly most reliable of these takes the conditional form

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

Here S_{obs} and A_{obs} denote the observed values of S and A.

Strictly speaking the above formula holds in the case of continuous distributions; a minor discontinuity correction amendment is called for in case of a discrete distribution.

The proof of this proposition, along with extensions and applications, as well as considerations of equivariance optimality, will appear elsewhere. One key ingredient here is that A is a sufficient and complete statistic for χ when $\psi = \psi_0$ is fixed. Note that the distribution of S given $A = A_{\rm obs}$ depends on ψ but not on $\chi_1, \ldots, \chi_p$.

Example 3. 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

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

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 uniformly most reliable confidence distribution, used here with a half-correction for discreteness, takes the simple form

$$\begin{aligned} C_{S\,|\,A}(\psi) &= {\rm Pr}_\psi\{S > S_{\rm obs} \,|\, A_{1,\rm obs}, \ldots, A_{m,\rm obs}\} + \tfrac12 {\rm Pr}_\psi\{S = S_{\rm obs} \,|\, A_{1,\rm obs}, \ldots, A_{m,\rm obs}\} \\ &= 1 - {\rm Bin}\Bigl(S_{\rm obs} \Bigm| \sum_{j=1}^m A_{j,\rm obs}, \frac{\psi}{1+\psi}\Bigr) + \tfrac12 {\rm bin}\Bigl(S_{\rm obs} \Bigm| \sum_{j=1}^m A_{j,\rm obs}, \frac{\psi}{1+\psi}\Bigr), \end{aligned}$$

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

4 Approximate confidence distributions

Uniformly most reliable exact inference is only possible in nice models. In a wider class of models, exact confidence distributions are available. The estimate of location based on the Wilcoxon statistic is for example exact 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, with finite data, 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.

Bootstrapping, simulation and asymptotics are useful tools in calculating approximate confidence distributions and in characterising 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. This will lead to first order accuracy properties in smooth models. A more refined method to obtain confidence distributions is via acceleration and bias corrections on bootstrap distributions, as developed below. This method, along with several other venues for refinement, will usually provide second order accuracy properties.

4.1 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 $\sqrt{n} S_n^{-1/2}(\widehat\theta_n - \theta) \to_d {\rm N}(0, I)$. By the delta method, the confidence distribution for a parameter $\psi = p(\theta)$ is based on linearising p at $\widehat{\theta}$, and yields

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

in terms of the cumulative standard normal. The variance estimate is $\widehat{\sigma}_n^2 = \widehat{g}^{\rm tr} S_n \widehat{g}/n$ where $\widehat{g}$ is the gradient of p 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 $\widehat{\psi}_{\rm obs}$.

This confidence distribution is known to be first order unbiased under weak conditions. That $C_{\rm delta}(\psi)$ is first order unbiased means that the coverage probabilities converge at the rate $n^{-1/2}$, or that $C_{\rm delta}(\psi_{\rm true})$ 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 $c_{\rm delta}(\psi)$, is simply the normal density $\mathrm{N}(\widehat{\psi}, \widehat{\sigma}_n^2)$.

4.2 The t-bootstrap method

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

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

where $\widehat{\tau}$ is proportional to an estimate of the standard deviation of $\widehat{\gamma}$. Let R be the distribution function of t, by assumption approximately independent of underlying parameters (ψ, χ). The canonical confidence intervals for γ then take the form $\widehat{\gamma} - R^{-1}(1-\alpha)\widehat{\tau} \le \gamma \le \widehat{\gamma} + R^{-1}(\alpha)\widehat{\tau}$, which backtransform to intervals for ψ, with

$$C^{-1}(\alpha) = h^{-1}(\widehat{\gamma} - R^{-1}(1-\alpha)\widehat{\tau}).$$

Solving for α leads to the confidence distribution $C(\psi) = 1 - R((h(\widehat{\psi}) - h(\psi))/\widehat{\tau})$, with appropriate confidence density $c(\psi) = C'(\psi)$. Now R would often be unknown, but the situation is saved via bootstrapping. Let $\widehat{\gamma}^* = h(\widehat{\theta}^*)$ 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:

$$C_{\rm tboot}(\psi) = 1 - \widehat{R}((h(\widehat{\psi}) - h(\psi))/\widehat{\tau}). \tag{9}$$

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 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 that the interval is not invariant under monotone transformations. The following method is not hampered by these shortcomings.

4.3 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.

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

$$(\widehat{\gamma} - \gamma)/(1 + a\gamma) \sim \mathrm{N}(-b, 1) \tag{10}$$

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, as further commented upon below. Assumption (10) can be written $\widehat{\gamma} - \gamma = (1 + a\gamma)(Z - b)$, where Z is standard normal. This leads to

$$1 + a\widehat{\gamma} = (1 + a\gamma)(1 + a(Z - b))$$

and a canonically correct interval for γ, and hence ψ, as explained in a minute. The essentials of the arguments below are that (10) describes a pivotal model on a transformed scale and that the apparatus already established for deriving confidence distributions from pivots becomes applicable via the transformation lemma of Section 2 in conjunction with bootstrapping. We include a little more detail, however, to pinpoint the roles of various ingredients.

Start with $z^{(\alpha)} \le Z \le z^{(1-\alpha)}$, the symmetric interval including Z with probability $1 - 2\alpha$, writing $z^{(\varepsilon)}$ for $\Phi^{-1}(\varepsilon)$. This leads after some algebra to

$$h^{-1}\Big(\frac{\widehat{\gamma} - (z^{(1-\alpha)} - b)}{1 + a(z^{(1-\alpha)} - b)}\Big) \le \psi \le h^{-1}\Big(\frac{\widehat{\gamma} - (z^{(\alpha)} - b)}{1 + a(z^{(\alpha)} - b)}\Big).$$

Writing this interval as $[C^{-1}(\alpha), C^{-1}(1 - \alpha)]$ and solving $C^{-1}(\alpha) = \psi$ for α gives the confidence distribution

$$C(\psi) = \Phi\Big(\frac{h(\psi) - h(\widehat{\psi})}{1 + ah(\psi)} - b\Big). \tag{11}$$

This constitutes a good approximation to the real confidence distribution, say $C_{\mathrm{exact}}(\psi)$, under assumption (10). But it requires h to be known, as well as values of a and b.

To come around this, look at bootstrapped versions $\widehat{\gamma}^* = h(\widehat{\psi}^*)$ from the estimated parametric model. If assumption (10) holds uniformly in a neighbourhood of the true parameters, then also

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

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

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

It follows, again after some algebra, that the lower endpoint in the interval for ψ above satisfies

$$\widehat{G}(\psi_{\mathrm{low}}) = \Phi\Big(b - \frac{z^{(1-\alpha)} - b}{1 + a(z^{(1-\alpha)} - b)}\Big),$$

and similarly for $\widehat{G}(\psi_{\mathrm{up}})$. This gives firstly the so-called BC_a intervals of Efron (1987), say $[\widehat{C}^{-1}(\alpha), \widehat{C}^{-1}(1 - \alpha)]$, applying the $\widehat{G}^{-1}$ transformation here. Secondly it gives us an acceleration and bias corrected approximation to the confidence distribution found in (11), through solving $\widehat{C}^{-1}(\alpha) = \psi$ for α. The result is the abc formula

$$\widehat{C}_{\mathrm{abc}}(\psi) = \Phi\Big(\frac{\Phi^{-1}(\widehat{G}(\psi)) - b}{1 + a(\Phi^{-1}(\widehat{G}(\psi)) - b)} - b\Big). \tag{12}$$

Note that an approximation $c_{\mathrm{abc}}(\psi)$ 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.

It remains to specify a and b. The bias parameter b is found from $\widehat{G}(\widehat{\psi}) = \Phi(b)$. The acceleration parameter a is found as $a = \frac{1}{6}\text{skew}$, 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 and Tibshirani (1993, Chs. 14 and 22) and in Davison and Hinkley (1997, Ch. 5). One option is via the jackknife method, which gives parameter estimates $\widehat{\psi}_{(i)}$ computed by leaving out data point i, and use

$$a = (6\sqrt{n})^{-1}\text{skew}\{\widehat{\psi}_{(\cdot)} - \widehat{\psi}_{(1)}, \ldots, \widehat{\psi}_{(\cdot)} - \widehat{\psi}_{(n)}\}.$$

Here $\widehat{\theta}_{(\cdot)}$ 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. In typical setups, both a and b will in fact go to zero with speed of order $1/\sqrt{n}$ in terms of sample size n. Thus (12) provides a second order non-linear correction of shift and scale to the immediate bootstrap distribution.

Example 4. 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 $\widehat{\psi}$, 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

$$\frac{1}{\psi}\Big\{-\nu_2 + (\nu_1 + \nu_2)\frac{(\nu_2/\nu_1)\widehat{\psi}^2/\psi^2}{1 + (\nu_2/\nu_1)\widehat{\psi}^2/\psi^2}\Big\} =_d \frac{\nu_1 + \nu_2}{\psi}\Big\{\text{Beta}(\tfrac{1}{2}\nu_2, \tfrac{1}{2}\nu_1) - \frac{\nu_2}{\nu_1 + \nu_2}\Big\}.$$

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 $\widehat{C}_{\text{abc}}(\psi)$ is amazingly close to $C(\psi)$, even when the degrees of freedom numbers are low and imbalanced; the agreement is even more perfect when ν_1 and ν_2 are more balanced or when they become larger. The same holds for the densities $\widehat{c}_{\text{abc}}(\psi)$ and $c(\psi)$; see Figure 1. ■

4.4 Discussion

The delta 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, but has a couple of drawbacks compared to the abc method. It is for example not invariant under monotone transformations. Theorems delineating suitable second-order correctness aspects of both the abc and the t-bootstrap methods above can be formulated and proved, with

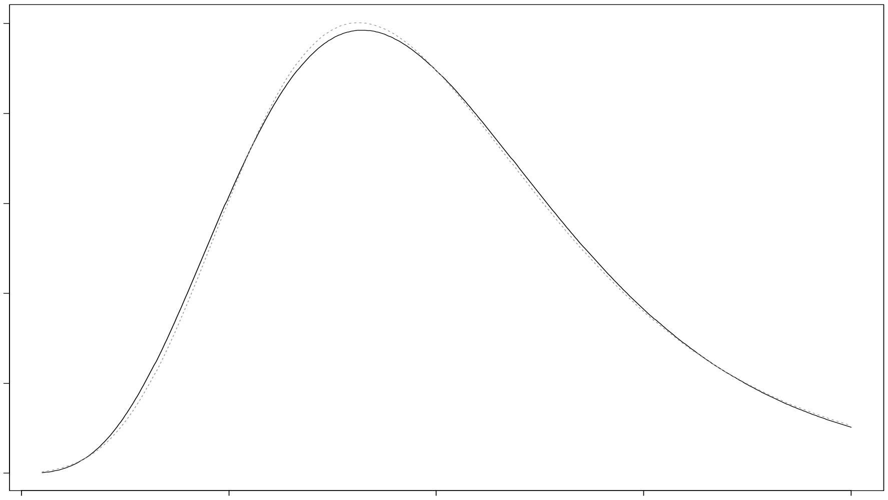

FIGURE 1: True confidence density along with abc-estimated version of it, for parameter $\psi = \sigma_2/\sigma_1$ with 4 and 9 degrees of freedom. The parameter estimate in this illustration is $\widehat{\psi} = 2.00$. The agreement is even better when ν_1 and ν_2 are closer or when they are larger.

necessary assumptions having to do with the quality of approximations involved in (8) and (10). Methods of proof would for example involve Edgeworth or Cornish–Fisher expansion arguments. Such could also be used to add corrections to the delta method (7).

Some asymptotic methods of debiasing an approximate confidence distribution involves a transformation of the confidence itself and not its quantile function. From a strict mathematical point of view there is of course no difference between acting on the quantiles or the confidence. But methods like the abc method above are most naturally viewed as a transformation of the confidence for each given value of the parameter.

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 DiCiccio and Efron (1996) and Davison and 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 and Cox (1994, Chs. 6 and 7) and Barndorff-Nielsen and Wood (1998). Clearly more work and further illustrations are needed to better sort out which methods have the best potential for accuracy and transparency in different situations. At any rate the abc method (12) appears quite generally useful and precise.

5 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 summarising 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. An example will show that a confidence distribution is in itself not sufficient to determine the likelihood of the reduced data, T, summarised by C. A given confidence distribution could, in fact, result from many different probability models, each with a specific 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 and Cox (1994) discuss adjusted likelihoods and other modified likelihoods based on saddle-point approximations, like the r^*. Efron and Tibshirani (1993) and Davison and Hinkley (1997) present methods based on bootstrapping and quadratic approximations. These methods are very useful, and in our context they are needed when inference based on all the available data combined in the integrated likelihood is done. To obtain the integrated likelihood, the likelihood components representing the (unavailable) data behind the confidence distributions are needed, however. To recover a likelihood from a confidence distribution is a problem that has not been addressed in the literature, as far as we know.

By definition, a likelihood is a probability density regarded as a function of the parameters, keeping the data at the observed value. A confidence distribution can not be interpreted as a probability distribution. It distributes confidence and not probability. The confidence density is therefore not usually a candidate for the likelihood function we seek. It is the probability distribution of the confidence distribution, regarded as the data, which matters. We will now demonstrate by means of a simple example that a given confidence distribution can relate to many different likelihoods, according to its different probability bases.

Consider the uniform confidence distribution over $[0.4, 0.8]$. The cumulative confidence distribution function is

$$C(\psi) = (\psi - 0.4)/0.4 \quad \text{for } 0.4 \leq \psi \leq 0.8. \tag{13}$$

This confidence distribution could have come about in many different ways, and the likelihood associated with the confidence distribution depends on the underlying probability models.

Shift model. In this model, the confidence distribution is based on a statistic with the sampling property

$$T = \psi - 0.2 + 0.4\,U,$$

where U is uniform (0, 1). The observed value is $T_{\text{obs}} = 0.6$, which indeed results in (13). The density of T is $f(t; \psi) = 2.5$ on $-0.2 < t < \psi + 0.2$, and the log-likelihood becomes zero on $(T_{\text{obs}} - 0.2, T_{\text{obs}} + 0.2)$ and $-\infty$ outside this interval.

Scale model. The confidence distribution is now based on

$$T = 0.4 + (\tfrac{1}{2}\psi - 0.2)/U.$$

The observed value $T_{\text{obs}} = 0.6$ leads to the uniform confidence distribution over $[0.4, 0.8]$. In this scale model the density of T is

$$f(t; \psi) = \frac{\frac{1}{2}\psi - 0.2}{(t - 0.4)^2} \quad \text{for } t > 0.2 + \tfrac{1}{2}\psi,$$

and the log-likelihood is $\log(\psi - 0.4)$ for $0.4 < \psi < 2T_{\text{obs}} - 0.4$.

Transformed normal model. Let now the confidence distribution be based on the statistic

$$T = 0.4\Big\{1 + \Phi\Big(Z + \Phi^{-1}\Big(\frac{\psi - 0.4}{0.4}\Big)\Big)\Big\} \quad \text{for } 0.4 \leq \psi \leq 0.8,$$

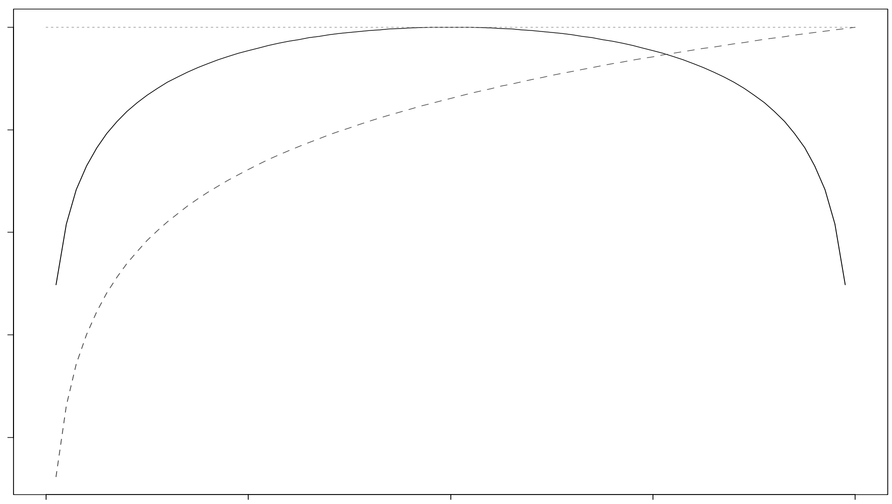

FIGURE 2: Three log-likelihoods consistent with a uniform confidence distribution over $[0.4, 0.8]$. *'Many likelihoods informed me of this before, which hung so tottering in the balance that I could neither believe nor misdoubt.'* – SHAKESPEARE.

where $Z \sim \mathrm{N}(0,1)$. Again, with $T_{\mathrm{obs}} = 0.6$, the confidence distribution is the uniform over the interval $[0.4, 0.8]$. The likelihood is invariant under a transformation of the data, say

$$V = \Phi^{-1}(T/0.4 - 1) = Z + \Phi^{-1}((\psi - 0.4)/0.4).$$

Since then $V_{\mathrm{obs}} = 0$, the log-likelihood is

$$l(\psi) = -\tfrac{1}{2}\{\Phi^{-1}((\psi - 0.4)/0.4)\}^2 \quad \text{for } 0.4 \le \psi \le 0.8.$$

Three possible log-likelihoods consistent with the uniform confidence distribution are shown in Figure 2. Other log-likelihoods are also possible.

6 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 typically a function of a one-dimensional statistic T in the data X. The probability density of T is then

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

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

Proposition 7 *When the probability basis for the confidence distribution is a pivot* $\mathrm{piv}(T;\psi)$ *in a one-dimensional statistic* T*, increasing in* ψ*, the likelihood is*

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

The confidence density is also related to the distribution of the pivot. Since one has $C(\psi) = F(\mathrm{piv}(T;\psi))$,

$$c(\psi) = f(\mathrm{piv}(T;\psi))\left|\frac{\mathrm{dpiv}(T;\psi)}{\mathrm{d}\psi}\right|.$$

Thus, the likelihood is in this simple case related to the confidence density by

$$L(\psi;T) = c(\psi)\frac{\mathrm{dpiv}(T;\psi)}{\mathrm{d}T}\left(\left|\frac{\mathrm{dpiv}(T;\psi)}{\mathrm{d}\psi}\right|\right)^{-1}.$$

There are important special cases. If the pivot is additive in T (at some measurement scale), say

$$\mathrm{piv}(T;\psi) = T - \mu(\psi) \tag{14}$$

for a smooth monotone 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 8 (Normal-based likelihood) *When the probability basis for the confidence distribution is an additive and normally distributed pivot, the log-likelihood related to the confidence distribution is*

$$l(\psi) = -\tfrac{1}{2}\left\{\Phi^{-1}(C(\psi))\right\}^2.$$

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 $1-\alpha$,

$$\{\psi\colon 2(l(\widehat{\psi}) - l(\psi)) < \Phi^{-1}(1-\tfrac{1}{2}\alpha)\}$$

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 and Wood, 1998).

A normal-based likelihood equals the related confidence density if and only if the confidence distribution is normal. This generalises as follows: For a given continuous confidence distribution, there is a probability basis making the corresponding likelihood equal to the confidence density.

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

$$T = \mu(\psi) + \tau(\psi)V \tag{15}$$

where V is a scaled residual. Then $\mathrm{piv}(T;\psi) = (T-\mu(\psi))/\tau(\psi)$, 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 5. Let X be Poisson with mean ψ. The half-corrected cumulative confidence distribution function is

$$C(\psi) = 1 - \sum_{x=0}^{X} e^{-\psi} \frac{\psi^x}{x!} + \tfrac{1}{2} e^{-\psi} \frac{\psi^X}{X!} e^{-\psi}.$$

Here $Y = 2(\sqrt{\psi} - \sqrt{X})$ is approximately $\mathrm{N}(0,1)$ and is accordingly approximately a pivot for moderate to large ψ. From a simulation experiment, one finds that the distribution of Y is slightly skewed, and has a bit longer tails than the normal. By a little trial and error, one finds that $\exp(Y/1000)$ 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

$$l_s(\psi) = \log t_{30}(Q_{30}(C(\psi)) - \log t_{30}(0).$$

Examples are easily made to illustrate that the $l_s(\psi)$ log-likelihood quite closely approximates the real Poisson log-likelihood $l(\psi) = x - \psi + x \log(\psi/x)$. ■

Usually, the likelihood associated with a confidence distribution is different from the confidence density. The confidence density depends on the parametrisation. By reparametrisation, the likelihood can be brought to be proportional to the confidence density. This parametrisation might have additional advantages.

Let $L(\psi)$ be the likelihood and $c(\psi)$ the confidence density for the chosen parametrisation, 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 $(\mathrm{d}/\mathrm{d}\psi)\mu = J$, and the confidence density of $\mu = \mu(\psi)$ is $L(\mu(\psi))$. There is thus always a parametrisation that makes the likelihood proportional to the 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 6. 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 $c(\mu) \propto L(\mu) = \exp(\widehat{\mu} - \mu - \exp(\widehat{\mu} - \mu))$. Bootstrapping this confidence distribution and likelihood is achieved by adding the bootstrap residuals $\log V^*$ to $\widehat{\mu}$ above, where V^* is standard exponentially distributed. The log-likelihood has a more normal-like shape in the μ parametrisation than in the canonical parameter ψ. Also, being a translation family in μ, the likelihood and the confidence density are easily interpreted. ■

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, $C = F(\mu - \widehat{\mu})$ and so is the likelihood, $L = c = f(\mu - \widehat{\mu})$. In this case, bootstrapping amounts to drawing bootstrap values from the confidence distribution, and substituting these for the point estimate $\widehat{\mu}$.

7 Bootstrapping a confidence distribution and a related likelihood

7.1 From confidence to bootstraps

Bootstrapping has emerged as an indispensable tool in statistical inference. When working with normal-based implied likelihoods, it is often desirable to mimic the result of bootstrapping the original data used when calculating the confidence distribution. A bootstrap replicate would then result in a perturbed confidence distribution, and thus in a perturbed likelihood. As with confidence distribution, its related likelihood, as determined by its probability basis, can be perturbed by conditional bootstrapping. In the location and scale model (15), bootstrapping of the confidence distribution and the likelihood amounts to drawing observations V^* from the pivot distribution. This leads to the bootstrap cumulative confidence distribution function

$$C^*(\psi) = F\left(\frac{T^* - T_{\text{obs}}}{\tau(\psi)} + F^{-1}(C(\psi))\right).$$

When the probability basis is normal and the scale τ is constant (and then chosen as unity), the bootstrapped confidence distribution is

$$C^*(\psi) = \Phi(\Phi^{-1}(C(\psi)) + T^* - T_{\text{obs}}),$$

where T^* is a bootstrap replicate of the normal score of the original statistic, T. On the normal score scale, $T^* - T_{\text{obs}}$ is then normally distributed, and since bias has been removed through the confidence estimation, we may take $T^* - T_{\text{obs}} = Z^* \sim \mathrm{N}(0,1)$. In this case, the bootstrapped log-likelihood is

$$l^*(\psi) = -\tfrac{1}{2}\{\Phi^{-1}(C(\psi)) + Z^*\}^2. \qquad (16)$$

The bootstrapped confidence distribution and the likelihood obtained from (15) has the desirable property of having identical support to the original confidence distribution. From (16), the maximum bootstrap likelihood estimate is in the normal case

$$\psi^* = C^{-1}(\Phi(0)) = \text{median}(C^*),$$

and its bootstrap distribution is precisely the confidence distribution.

Example 7. Let the confidence distribution be uniform over $[0.4, 0.8]$, and assume that it has a normal probability basis. The bootstrapped confidence distribution is then

$$l^*(\psi) = -\tfrac{1}{2}\{\Phi^{-1}((\psi - 0.4)/0.4) + Z^*\}^2,$$

with Z^* from a standard normal. ■

Example 8. The population dynamics model used in the assessment of bowhead whales is age-structured (Schweder and Ianelli, 1998). The yearly natural survival probability, ψ, was assumed independent of age for adult whales. The prior distribution chosen for this parameter, which we for illustration will interpret as a confidence distribution, has cumulative distribution function

$$C(\psi) = \Phi\left(\frac{\psi - 0.99}{0.02}\right) \Big/ \Phi\left(\frac{0.995 - 0.99}{0.02}\right) \quad \text{for } \psi \leq 0.995.$$

The normal-based log-likelihood implied by this confidence distribution is shown in Figure 3, together with accompanying bootstrap replicates. ■

7.2 A simple population dynamics model for Bowhead whales

The management of fisheries and whaling rests on the quality of stock assessments. Within the International Whaling Commission, the assessment of the stock of bowhead whales subject to

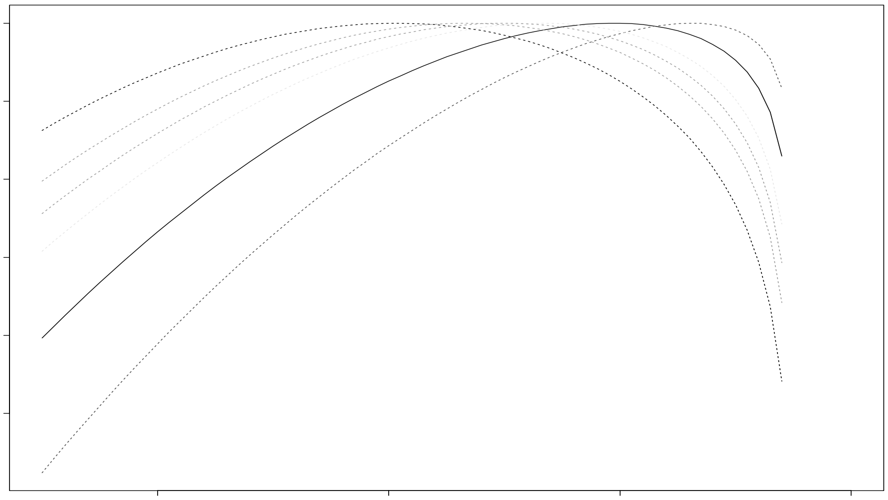

FIGURE 3: Normal-based log-likelihood for a truncated normal confidence distribution (line) and five replicated bootstrap log-likelihoods (dotted).

exploitation by Native Americans in Alaska has been discussed repeatedly. The most recent assessment was done by Bayesian methodology, see International Whaling Commission (1999) and Poole and Raftery (1998).

Raftery and his co-workers have developed the Bayesian synthesis approach, and in Poole and Raftery (1998) the most recent version of this approach is presented. In addition to using their method on the more complex age-specific population dynamics model used by the International Whaling Commission, they also illustrate their method on a simplified version of the model. To illustrate our approach, and to compare it to the Bayesian synthesis approach with 'Bayesian melding' to harmonise conflicting prior distributions, it is applied to the simplified population dynamics model with the same data as used by Poole and Raftery (1998, Section 3.6).

Poole and Raftery (1998) understand their prior distributions as prior probability distributions to be used in a Bayesian integration. We re-interpret this prior distributions as prior confidence distributions. Instead of a Bayesian integration, this confidence distributions will be integrated with the likelihood of P_{1993} and ROI, after having been converted to likelihoods. To do this, their probability basis needs to be determined. As Bayesians, Poole and Raftery are not concerned with this issue, and no information is thus available. For simplicity, we chose to assume a normal probability basis for each of the three prior confidence distributions. This determines their likelihoods as follows.

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

$$P_{t+1} = P_t - C_t + P_t\, 1.5\,\text{MSYR}\{1 - (P_t/P_{1848})^2\},$$

where P_{1848} is taken as the carrying capacity of the stock, and MSYR (maximum sustainable yield rate) is the productivity parameter. Yankee whaling started in 1848. From that year on, the catch history $\{C_t\}$ is available, and is assumed to be exactly known. There are two free parameters in this model. We choose these to be MSYR and P_{1848}. Stock sizes for other years are also parameters, but are determined by MSYR and P_{1848}. Poole and Raftery (1998) list the following independent priors:

$$\begin{aligned} P_{1848} &\sim 6400 + \text{Gamma}(2.81, 0.000289), \\ \text{MSYR} &\sim \text{Gamma}(8.2, 372.7), \\ P_{1993} &\sim \text{N}(7800, 1300^2), \end{aligned}$$

where Gamma(a, b) denotes the gamma distribution with mean a/b and standard deviation $\sqrt{a}/b$.

In addition to prior information, there are two log-likelihood components based on recent survey data. One component concerns P_{1993}, and is the Gaußian

$$l_4(P_{1993}) = -\tfrac{1}{2}\{(P_{1993} - 8293)/626\}^2.$$

The source of this information is different from that of the prior distribution for P_{1993} above. The other component concerns the recent rate of increase in stock size ROI, which is defined through $P_{1993} = (1 + \text{ROI})^{15} P_{1978}$. It has likelihood

$$l_5(\text{ROI}) = \frac{9}{2} \log \left\{ 1 + \frac{1}{8} \left(\frac{\log(1 + \text{ROI}) - 0.0302}{0.0069} \right)^2 \right\} - \log(\text{ROI} + 1)$$

obtained from the t-distribution with 8 degrees of freedom, and an exponential transformation (Poole and Raftery, 1998).

The combined log-likelihood

$$l = l_1(P_{1848}) + l_2(\text{MSYR}) + l_3(P_{1993}) + l_4(P_{1993}) + l_5(\text{ROI})$$

is an extremely narrow curved ridge. See the bootstrap sample of maximum likelihood estimate in Figure 4. The maximum likelihood estimate is presented in the table together with quantiles of the confidence distributions for the various parameters obtained by the abc method, having employed 1000 bootstrap replicates.

The bootstrapping is determined by the probability bases of the likelihood components as follows:

$$\begin{aligned} l_1^*(P_{1848}) &= -\tfrac{1}{2}\left\{\Phi^{-1}(G_{2.81,0.000289}(P_{1848} - 6400)) + Z^*\right\}^2, \\ l_2^*(\text{MSYR}) &= -\tfrac{1}{2}\left\{\Phi^{-1}(G_{8.2,372.7}(\text{MSYR})) + Z^*\right\}^2, \\ l_3^*(P_{1993}) &= -\tfrac{1}{2}\{(P_{1993} - 7800)/1300 + Z^*\}^2, \\ l_4^*(P_{1993}) &= -\tfrac{1}{2}\{(P_{1993} - 8200)/564 + Z^*\}^2, \\ l_5^*(\text{ROI}) &= \frac{9}{2} \log \left\{ 1 + \frac{1}{8} \left(\frac{\log(1 + \text{ROI}) - 0.0302}{0.0069} + T_8^* \right)^2 \right\} - \log(\text{ROI} + 1). \end{aligned}$$

Here the four Z^*s are independently drawn from the standard normal distribution and independently from T_8^* which is drawn from the t_8-distribution. For each draw of these five 'bootstrap residuals', the perturbed likelihood is maximised, leaving us with the bootstrap estimate $\theta^* = (P_{1848}^*, \text{MSYR}^*)$.

Each bootstrap replicate θ^* induces bootstrap replicates of P_{1993} and ROI through the deterministic population dynamics model. The bootstrap distributions are nearly normal for both the two induced parameters, $P_{1993}^* \sim \text{N}(8117, 556^2)$ and $\text{ROI}^* \sim \text{N}(0.0249, 0.0059^2)$. By simple bias correction relative to the maximum likelihood estimates, we obtain the quantiles in the table. For

Parameter	MLE	Quantile	Prior	Posterior	Bayesian
P_{1848}		0.025	8263	10932	12057
	13152	0.5	14997	13017	14346
		0.975	30348	16085	17980
MSYR		0.025	0.0096	0.0157	0.0113
	0.0267	0.5	0.0211	0.0275	0.0213
		0.975	0.0394	0.0425	0.0333
P_{1993}		0.025		7027	7072
	8117	0.5		8117	8196
		0.975		9207	9322
ROI		0.025		0.0145	0.0105
	0.0255	0.5		0.0261	0.0204
		0.975		0.0377	0.0318

Table 1: Maximum likelihood estimates and quantiles for prior and posterior of frequentist and Bayesian distributions, for the parameters of interest.

the primary parameters, transformation is necessary to obtain normality for the bootstrap sample. This is achieved by a Box–Cox transformation of exponent −1: $10000/P^*_{1848} \sim \mathrm{N}(0.7525, 0.0748^2)$, while a square root transform normalises the bootstrap distribution of MSYR:

$$(\text{MSYR}^*)^{1/2} \sim \mathrm{N}(0.1611, 0.0206^2). \tag{17}$$

Kolmogorov–Smirnov tests yielded p-values around 0.5 in both cases. A simple bias correction on these scales leaves us with the confidence quantiles in the table.

The probability basis of the input confidence distributions leading to the log-likelihoods l_1, l_2 and l_3 provides a basis for the bootstrap, as explained above. For ROI, we assume the likelihood given by Poole and Raftery to be based on the estimate of ROI, appropriately transformed, being a T_8 distance from the transformed parameter, where T_8 is drawn from a t-distribution with 8 degrees of freedom.

The prior and the posterior confidence densities of MSYR are shown in Figure 5. The main reason for the posterior being shifted to the right of the prior confidence distribution is the influence of the data on ROI. The bootstrap density for MSYR is also shown. Note that the bias correction pushed the posterior confidence distribution towards higher values of MSYR. The bias correction is roughly +6% in this parameter.

The probability basis for the posterior confidence distribution for MSYR is normal. It is, in fact, based on (17) and an assumption of the posterior confidence distribution on this scale only being shifted by a constant amount relative to the bootstrap distribution, $\text{MSYR}^{1/2} \sim (\text{MSYR}^*)^{1/2} + b$. The bias correction b is estimated as

$$2\big[(\widehat{\text{MSYR}})^{1/2} - \text{mean}\{(\text{MSYR}^*)^{1/2}\}\big] = 0.0046,$$

where $\widehat{\text{MSYR}}$ is the maximum likelihood estimate of MSYR. With the probability basis being normal, the posterior log-likelihood of MSYR is

$$l_{\text{post}}(\text{MSYR}) = -\tfrac{1}{2}(\text{MSYR}^{1/2} - 0.166)^2/0.0206^2.$$

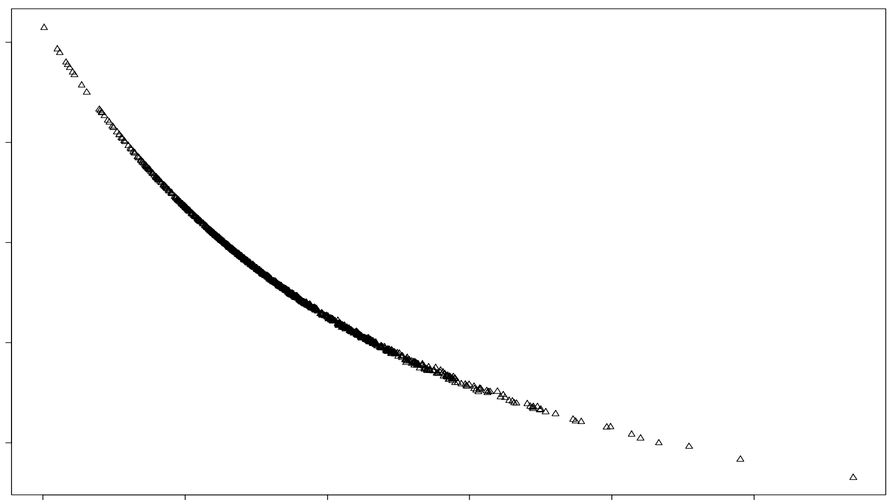

FIGURE 4: One thousand bootstrap estimates of P_{1848} and MSYR.

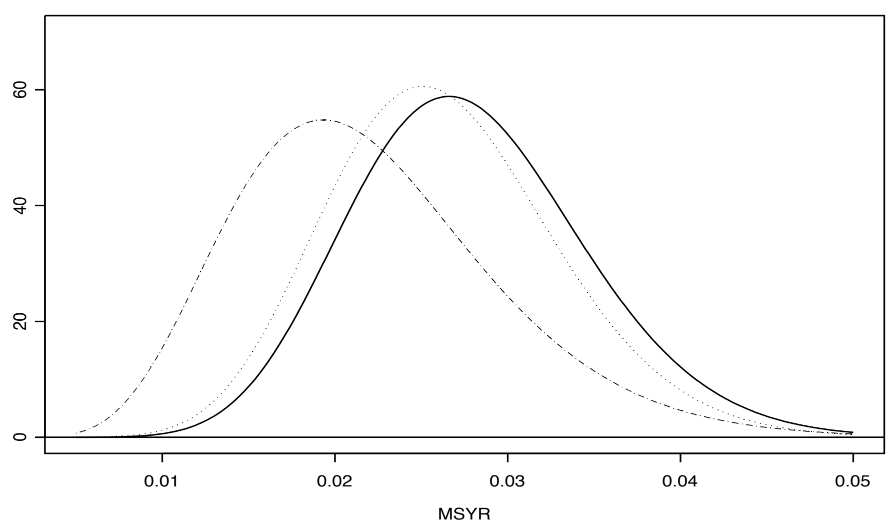

FIGURE 5: Prior confidence density (broken line), bootstrap density (dotted) and posterior confidence densities for MSYR.

8 Discussion

The confidence distribution is an attractive format for reporting statistical inference for parameters of primary interest. To allow future good use of the results it is desirable to allow a likelihood to be constructed from the confidence distribution. An alternative is to make the original data available, or to present the full likelihood. However, the work invested in reducing the original data to a confidence distribution for the parameter of interest would then be lost. To convert the posterior confidence distribution to a likelihood the probability basis for the confidence distribution must be reported.

Our suggestion is accordingly to extend current frequentist reporting practice from only reporting a point estimate, a standard error and a (95%) confidence interval for the parameters of primary interest. To help future readers, one should report the confidence distribution fully, and supplement it with information on its probability basis. This latter information will often be qualitative.

8.1 Advantages with our approach

The advantages of representing the information contained in a confidence distribution in the format of (an approximate) likelihood function are many and substantial.

By adding the log-likelihoods of independent confidence distributions for the same parameter, an integrated likelihood, and thus a combined confidence distribution is obtained. The merging of independent confidence intervals has attracted considerable attention, and the use of implied likelihoods presents a solution to the problem. One might, for example, wish to merge independent confidence intervals for the same parameter to one interval based on all the data. When the probability basis and the confidence distribution are known for each data set, the related log-likelihoods can be added, and an integrated confidence distribution (accompanied with its probability basis) is obtained.

A related problem is that of so-called meta-analyses. If independent confidence distributions are obtained for the same parameter, the information is combined by adding the implied log-likelihoods. A frequent problem in meta-analysis is, however, that the interest parameter might not have exactly the same value across the studies. This calls for a model that reflects this variation, possibly by including a random component. In any event, the availability of implied likelihood functions from the various studies facilitates the meta-analysis, whether a random component is needed or not.

Studies in fields like ecology, economics, geophysics etc. often utilise complex models with many parameters. To the extent results are available for some of these parameters, it might be desirable to include this information in the study. If these previous results appear in the format of confidence distributions accompanied by explicit probability bases, their related likelihoods are perfectly suited to carry this information into the combined likelihood of the new and the previous data. If a confidence distribution is used that is not based on (previous) data, but on subjective judgement, its related likelihood can still be calculated and combined with other likelihood components, provided assumptions regarding its probability basis can be made. This subjective component of the likelihood should then, perhaps, be regarded as a penalising term rather than a likelihood term.

Finally, being able to obtain the implied likelihood from confidence distributions, and being able to calculate confidence distributions from data summarised by a likelihood within a statistical model, a methodology parallel to and competing with Bayesian methodology emerges. This methodology is frequentist in its foundation. As the Bayesian methodology, it provides a framework for coherent learning and its inferential product is a distribution: a confidence distribution instead of a Bayesian posterior probability distribution.

8.2 Differences from the Bayesian paradigm

It is pertinent to compare our frequentist approach with the Bayesian approach to coherent learning. Most importantly, the two approaches have the same aim: to update distributional knowledge in the view of new data within the frame of a statistical model. The updated 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). There are, however, substantial differences between the two approaches. Compared to the Bayesian approach, we would like to emphasise the following.

Distributions for parameters are understood as confidence distributions and not probability distributions. The concept of probability is reserved for (hypothetically) repeated sampling, and is interpreted frequentistically. To update a confidence distribution it must be related to its probability basis, as the likelihood related to the confidence distribution. To update a distribution the frequentist needs more information than the Bayesian, namely its probability basis. On the other hand, the distinction between probability and confidence is basic in the frequentist tradition.

It is possible to start at scratch, without any (unfounded) subjective probability 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 she has to construct priors. The frequentist will, however, not have principle 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 example, there were three priors but only two free parameters. Without modifications of the Bayesian synthesis approach like the melding of Poole and Raftery (1998), the Bayesian gets into trouble. Due to the Borel paradox (Schweder and Hjort, 1997), the Bayesian synthesis will, in fact, be completely determined by the particular parametrisation. With more prior distributions than there are free parameters, Poole and 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, however, there are independent prior distributional information on a parameter, 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 $\sigma/\sqrt{k}$.

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 emphasised 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 in the Bayesian tradition. There has, however, been some recent interest in the matter. To obtain frequentist unbiasedness, the Bayesian will have to choose her prior with unbiasedness in mind. Is she then a Bayesian? Her prior distribution will then not represent prior knowledge of the parameter in case, but an understanding of the model. Our 'frequentist Bayesianism' solves this problem in principle. It takes as input (unbiased) prior confidence distributions and delivers (unbiased) posterior confidence distributions.

References

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

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

[3] Berger, J.O., Liseo, B. and Wolpert, R.L. (1999). Integrated likelihood methods for eliminating nuisance parameters. Technical report, Institute of Statistics and Decision Sciences, Duke University.

[4] DiCiccio, T.J. and Efron, B. (1996). Bootstrap confidence intervals (with discussion). *Statistical Science* **11**, 189–228.

[5] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press, Cambridge.

[6] Dufour, J.M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* **65**, 1365–1387.

[7] Edwards, A.W.F. (1992). *Likelihood* (expanded edition). John Hopkins University Press, Baltimore.

[8] Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association* **82**, 171–200.

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

[10] Efron, B. (1998). R.A. Fisher in the 21$^{\text{st}}$ century (with discussion). *Statistical Science* **13**, 95–122.

[11] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, London.

[12] Ericsson, N.R., Jansen, E.S., Kerbesian, N.A. and Nymoen, R. (1998). Interpreting a monetary condition index in economic policy. Technical report, Department of Economics, University of Oslo.

[13] Koschat, M.A. (1987). A characterisation of the Fieller solution. *The Annals of Statistics* **15**, 462–468.

[14] Lehmann, E.L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[15] Lehmann, E.L. (1993). The Fisher, Neyman–Pearson theories of testing hypotheses: One theory or two? *Journal of the American Statistical Association* **88**, 1242–1249.

[16] Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika* **32**, 128–150.

[17] Poole, D. and Raftery, A.E. (1998). Inference in deterministic simulation models: The Bayesian melding approach. Technical Report no. 346, Department of Statistics, University of Washington.

[18] Royall, R.M. (1997). *Statistical Evidence. A Likelihood Paradigm.* Chapman and Hall, London.

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

[20] Schweder, T. and Hjort, N.L. (1997). Indirect and direct likelihoods and their synthesis. Statistical Research Report #12, University of Oslo.

[21] Schweder, T. and Ianelli, J.N. (1998). Bowhead assessment by likelihood synthesis: methods and difficulties. Paper IWC/SC/50/AS2.