Combining independent Bayesian posteriors into a confidence distribution, with application to estimating climate sensitivity

Nicholas Lewis[1]
Bath, United Kingdom

Manuscript accepted on 11 July 2016 by Journal of Statistical Planning and Inference for publication in a Special Issue on Confidence Distributions © 2016.

[1] *Corresponding author address*: Bath, United Kingdom.
E-mail: nhlewis@btinternet.com

Abstract

Combining estimates for a fixed but unknown parameter to obtain a better estimate is an important problem, but even for independent estimates not straightforward where they involve different experimental characteristics. The problem considered here is the case where two such estimates can each be well represented by a probability density function (PDF) for the ratio of two normally-distributed variables. Two different statistical methods – objective Bayesian and frequentist likelihood-ratio – are employed and compared. Each probabilistic estimate of the parameter value is represented by a fitted three-parameter Bayesian posterior PDF providing a close approximation to the ratio of two normals, that can legitimately be factored into a likelihood function and a noninformative prior distribution. The likelihood functions relating to the parameterised fits to the probabilistic estimates are multiplicatively combined and a prior is derived that is noninformative for inference from the combined evidence. An objective posterior PDF that incorporates the evidence from both sources is produced using a single-step approach, which avoids the order-dependency that would arise if Bayesian updating were used. The frequentist signed root likelihood-ratio method is also applied. The probability matching of credible intervals from the posterior distribution and of approximate confidence intervals from the likelihood-ratio method is tested, showing that both methods provide almost exact confidence distributions. The approach developed is applied in the important case of the Earth's equilibrium climate sensitivity, by combining an estimate from instrumental records with an estimate representing largely independent paleoclimate proxy evidence, resulting in a median estimate of 2.0°C and a 5–95% confidence/credibility interval of (1.1, 4.5) °C.

Keywords: combining evidence; confidence distribution; objective Bayesian; likelihood ratio; Bayesian updating; Jeffreys prior

1. Introduction

Often, existing estimates of a fixed but unknown parameter are poorly constrained and combining evidence from them offers the most obvious route to obtaining a less imprecise estimate. But, even assuming uncertainties in two estimates are independent, in most cases statistical theory does not provide a unique, optimal method of combining them. Multiplicatively combining the likelihood functions underlying the two estimates, if they are available, provides one obvious starting point, as that is a standard method for combining parametric information from independent sources, applicable in both frequentist and Bayesian paradigms. In the frequentist paradigm, a simple and straightforward approach, yielding approximate confidence distributions, is to apply standard likelihood-ratio methods (Pawitan 2001, p. 36–37) to the combined likelihood. If there are unwanted (nuisance) parameters, the likelihoods can be reduced, for instance by forming a profile-likelihood, either before or after combining them. However, the standard Bayesian method for using likelihood information to combine evidence from independent sources (Bayesian updating) differs from frequentist approaches, and leaves unresolved the problem of what prior distribution to use.

Efron (1998) wrote of the "250-year search for a dependable objective Bayesian theory" and suggested that the development of confidence distributions and approximations thereto might hold a key to it, points echoed by Schweder and Hjort (2002) and Singh, Xie and Strawderman (2005). Schweder and Hjort (2002) also wrote that "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." Fraser et al. (2010) stressed that, in absence of information about how the parameter value was generated, to avoid Bayesian inference giving misleading results it was necessary that a prior that provided correct calibration of posterior probabilities to frequencies, at least approximately, be used. Such an approach, which reflects a so-called "objective Bayesian" way of thinking, suggests a possible convergence in many cases between the resulting Bayesian posterior distributions and frequentist confidence distributions.

Here I compare simple objective Bayesian and frequentist likelihood-ratio methods for combining independent probabilistic estimates of an uncertain but fixed parameter, that can each be well represented by a (differing) ratio-normal distribution (that of the ratio of two normally-distributed random quantities, independent if not stated otherwise). In doing so, I take advantage of a transformed-normal approximation to ratio-normal distributions (Raftery and Schweder 1993). I show that the two methods provide almost identical inference, approximating closely to confidence distributions. I also show that the standard Bayesian methodology for information updating produces different results from those using the proposed objective Bayesian approach, with much worse probability matching and results that, inconsistently, depend on the order in which information is analysed. The methods developed here are exemplified by application to the important case of estimating the Earth's equilibrium climate sensitivity (ECS).

The rest of this paper is organized as follows. Section 2 discusses Bayesian and likelihood-ratio parameter inference and the methods used in this study. Section 3 discusses the physical relevance of the ratio-normal distribution, selects a suitable parameterized approximate distribution to use in this case and tests inference using it. Section 4 explains the methodology for combining the information embodied in two independent such parameterised estimates. Section 5 applies the methods developed to combine evidence regarding ECS. Section 6 summarises, and discusses issues raised.

2. Bayesian vs likelihood-ratio based parameter inference

2.1. Bayesian inference

Bayes theorem (Bayes 1763), for a univariate continuous parameter θ on which observed data $\boldsymbol{y}$ depend, states that the (posterior) probability density function (PDF), $p_\theta(\theta \mid \boldsymbol{y})$, for θ is proportional to the probability density of the data $p_y(\boldsymbol{y} \mid \theta)$ (the "likelihood" when considered as a function of θ, with $\boldsymbol{y}$ fixed) multiplied by the density of a prior distribution (prior) for θ, $p_\theta(\theta)$:

$$p_\theta(\theta \mid \boldsymbol{y}) \propto p_y(\boldsymbol{y} \mid \theta)\, p_\theta(\theta) \qquad (1)$$

(the subscripts indicating which variable density is for). The constant of proportionality is such that the posterior PDF integrates to unit probability. Under subjective Bayesian interpretations, the prior (and hence the posterior) represents the researcher's own degree of belief regarding possible parameter values, prior and posterior reflecting respectively relevant prior knowledge and that knowledge updated by the observed data. If the data are weak, an informative prior is likely to strongly influence parameter estimation, and the resulting posterior cumulative distribution function (CDF) is unlikely to approximate a confidence distribution.

Objective Bayesian approaches eschew use of a prior that reflects beliefs regarding the parameter value, and are usually used for inference in the absence of generally-agreed existing knowledge about parameter values. The aim is for the results to be – as for frequentist results – a function only of the data from which they are derived and the experimental setup, which determines both the model and the sampling plan. In order to achieve this a "noninformative prior", which is mathematically derived from the assumed statistical model and has no probabilistic interpretation, must be used (Bernardo and Smith 1994, p. 298 and p. 306; Kass and Wasserman 1996; Bernardo 2011). A noninformative prior is merely a tool for the generation of the desired posterior PDFs for the parameter(s) of interest (Bernardo and Smith 1994, p.306), and may be viewed as an appropriate mathematical weight function (Fraser et al. 2010). It is common to judge the merits of a noninformative prior by its "probability matching", i.e. how closely the resulting posterior probabilities agree with repeated sampling frequencies (Berger and Bernardo 1992, p. 36; Kass and Wasserman 1996). Although in some common cases parameter inference using a noninformative prior produces a posterior PDF that is an exact confidence density, in general this is not possible (Lindley 1958).

Noninformative priors were originally developed using invariance considerations (Jeffreys 1946). Jeffreys prior is the square root of the (expected) Fisher information – or of its determinant in multi-parameter cases. Fisher information – the expected value of the negative second derivative of the log-likelihood function with respect to the parameters – is a measure of the amount of information that the data, on average, carries about the parameter values. The more sophisticated reference analysis approach (Bernardo 1979; Berger and Bernardo 1992) uses information-theoretic concepts to derive a minimally-informative prior. In the univariate parameter continuous case, Jeffreys prior is normally the reference prior and is known to be the best noninformative prior, in the sense that, asymptotically, Bayesian posterior distributions generated using it are closer to exact confidence distributions than those yielded by any other prior (Welch and Peers 1963; Hartigan 1965).

A noninformative prior depends on the experiment-specific relationships the data values have with the parameter(s) and on the data-error characteristics, and thus on the form of the likelihood function; it is a mapping of probability reflecting relative volumes in data space (where the likelihood exists) and in parameter space, and their metric relationship (Kass 1989). Two experiments estimating the same parameter may therefore give rise to quite different noninformative priors.

Where an estimated PDF is obtained by applying Bayes theorem using one set of data, and other data become available, Bayes theorem is often re-applied, using the original posterior PDF as the prior (Bayesian updating). This is the standard Bayesian method for combining evidence, and is seen by most Bayesian statisticians (including "objective" ones) as the only Bayesian way to update parameter estimation to reflect evidence from new data. In subjective Bayesian approaches, Bayesian updating provides an identical final posterior PDF whichever order the two data-based likelihood functions are used, since the prior is chosen independently of the experimental setup and multiplication of the likelihood functions is commutative. Such order-independence is essential for inferential consistency. However, where an objective Bayesian approach is used, Bayesian updating generally results in order-dependency. Where two independent estimates of a parameter value are derived from experiments with differing characteristics, the priors that are noninformative for each experiment separately will generally differ. Therefore, Bayesian updating would produce a different final posterior according to the order in which Bayes' theorem was applied to the two datasets. Although this issue is usually not mentioned (e.g., Bernardo and Smith 1994) it has been pointed out as a problem for objective Bayesian approaches (e.g., Kass and Wasserman 1996) and it has even been given as a reason to dismiss the objective Bayesian method entirely (Seidenfeld 1979). However, inconsistency can be avoided when taking an objective Bayesian approach, by applying Bayes theorem once only, to the multiplicatively-combined joint likelihood function for the two experiments taken together, with a single noninformative prior being computed for inference therefrom. This is the approach I adopt here; it is equivalent to the modification of Bayesian updating set out in Lewis (2013a). I show that combining information from different data in this way leads to better frequentist coverage properties than the use of standard Bayesian updating. The validity of standard Bayesian updating has also been questioned elsewhere (e.g., Williamson 2009). I treat the term 'objective Bayesian' as including the approach I adopt notwithstanding that it departs from most other such approaches in explicitly rejecting use of standard Bayesian updating.

2.2. Likelihood-ratio inference

Another way of objectively combining two independent probabilistic estimates for an unknown parameter is to use frequentist likelihood-ratio methods, which directly provide estimated confidence distributions. Independent likelihoods may be multiplicatively combined and a likelihood-ratio method applied to their product. Likelihood-ratio methods (including profile likelihood, used when there are nuisance parameters to eliminate) are based on asymptotic normal approximations. They produce asymptotic confidence distributions that are exact only when the underlying distributions are normals or normals after transform, but even basic methods typically provide reasonable approximations in other cases (Pawitan 2001, p. 61). Modified profile-likelihood methods can provide greater accuracy, at the expense of greater complexity (e.g., Cox and Reid 1987; Schweder and Hjort 2016, ch. 7); they have close links to objective Bayesian inference and often give closely corresponding results (Bernardo and Smith 1994, p. 343 and p. 481).

2.3. Parameter inference methods in this study

The main aim here is to present practical and objective (in the sense of depending only on data and experimental setup) methods for combining, using alternative objective Bayesian and frequentist statistical frameworks, independent probabilistic estimates of a fixed but unknown parameter. Both approaches involve first deriving likelihood functions from the input probabilistic estimates and multiplicatively combining them, rather than directly combining the probabilistic estimates themselves. The objective Bayesian method is based on deriving and using a approximate Jeffreys prior appropriate for use with the combined likelihood function; the frequentist method is based on applying a standard basic likelihood-ratio method to it. The likelihood-ratio method used is based on the signed square root of twice the log of the ratio of the maximum likelihood to the likelihood at any value of the parameter, θ, having, under the model, an asymptotic standard normal distribution. Thus, asymptotically, $\mathrm{sign}(\hat{\theta}-\theta)[2\{l(\hat{\theta})-l(\theta)\}]^{1/2} \sim N(0,1)$, where $\hat{\theta}$ is the value of θ maximising the log-likelihood $l(\theta)$ (e.g., Brazzale et al. 2007, p. 6).

The statistical methods employed in this paper require – assuming non-availability of actual likelihood functions and related noninformative priors – that suitable parameterized standard form distributions be fitted to the actual PDFs (or representative percentile points) for each of the available independent estimates. The standard form used – which need not necessarily be the same for each estimate – is chosen to meet four requirements:- first, that it has a small number of adjustable parameters; secondly that it can provide an accurate approximation to the estimated PDF; thirdly, that its functional form is compatible with the source of uncertainties in the problem being considered; and fourthly, that the fitted PDF can be conveniently represented in objective Bayesian terms as the product of an identifiable likelihood function and a prior distribution that is noninformative, given that likelihood function: it has the necessary "separability" property, as will be illustrated below Eq. (2). The method is illustrated for the case of combining two independent estimates, but readily generalises to combine any number of independent estimates.

3. Estimation where a ratio-normal distribution applies

In many physical and other problems an unknown parameter of interest, θ, conceptually linearly relates the value of two variables that can be more directly estimated (the "observable parameters", denoted by ψ_1 and ψ_2), such that $\theta = \psi_1 / \psi_2$. Estimation of ECS (section 5) is one example of such a problem. Typically, estimates $(\tilde{\psi}_1, \tilde{\psi}_2)$ of the true values of ψ_1 and ψ_2 are obtained, errors therein being taken as independent and normally distributed. Assume that these errors have known standard deviations, respectively σ_1 and σ_2. In such cases, the posterior PDF for the parameter, θ, should, assuming posterior PDFs for the two observable parameters are derived using uniform priors (thereby achieving probability matching), have a ratio-normal distribution. Three parameters are required to specify a ratio-normal distribution, since scaling all of $(\tilde{\psi}_1, \tilde{\psi}_2, \sigma_1, \sigma_2)$ pro-rata leaves the distribution unchanged. Satisfactory parameter estimation requires the denominator normal to have almost all its probability on one side of zero, otherwise very large positive or negative parameter values cannot be sufficiently excluded or distinguished. A simple but accurate analytical approximation to ratio-normal distributions is then available (henceforth "the ratio-normal approximation"), which has a form that can easily be factored into an identifiable likelihood function and a noninformative prior (Raftery and Schweder 1993).

The ratio-normal approximation is obtained by integrating out in the bivariate normal posterior density for (ψ_1, ψ_2), obtained from estimates $(\tilde{\psi}_1, \tilde{\psi}_2)$ by using the uniform Jeffreys prior to infer ψ_1 and ψ_2, which, as they are location parameters, is completely noninformative here. Raftery and Schweder (1993) show that on this basis the posterior distribution of the univariate variable

$-z(\theta\,|\,\tilde{\psi}_1,\tilde{\psi}_2) \equiv -\dfrac{\tilde{\psi}_1-\theta\tilde{\psi}_2}{\sqrt{V(\theta)}}$ ~ $N(0,1)$ to a good approximation, where $V=\sigma_1{}^2+\theta^2\sigma_2{}^2$, subject to $|\,\tilde{\psi}_2\,|/\sigma_2$ being sufficiently large for almost all probability for ψ_2 to lie on one side of zero. Following Raftery and Schweder (1993), I assume without loss of generality that $\tilde{\psi}_2$ is positive. If it is negative, multiplying both $\tilde{\psi}_1$ and $\tilde{\psi}_2$ by -1 before applying the equations yields the correct results. Changing variable from *z* to θ and differentiating provides an approximate posterior density for θ, as:

$$p_\theta\left(\theta\,|\,D\right) = \frac{\left(\tilde{\psi}_2\sigma_1{}^2+\tilde{\psi}_1\sigma_2{}^2\theta\right)}{V(\theta)^{3/2}}\,\phi\left[-\frac{\left(\tilde{\psi}_1-\theta\tilde{\psi}_2\right)}{\sqrt{V(\theta)}}\right] \qquad (2)$$

($\phi()$ representing the *N*(0,1) density function, and *D* the data). The first factor is the Jacobian derivative for the inverse transformation. Where $\text{sign}(\tilde{\psi}_1)\,\theta < -\text{sign}(\tilde{\psi}_1)\,\tilde{\psi}_2\sigma_1{}^2/\tilde{\psi}_1\sigma_2{}^2$, which occurs at some point when θ is on the other side of zero from where the posterior density is maximum, the thus-calculated posterior density is negative and meaningless; the relationship between θ and *z* becomes non-monotonic at that point. At the monotonicity limit $z=\sqrt{\tilde{\psi}_1^2/\sigma_1^2+\tilde{\psi}_2^2/\sigma_2^2}$, and $\phi(-z)$ is usually negligibly small. Provided $|\,\sigma_1/\tilde{\psi}_1\,|$ and $|\,\sigma_2/\tilde{\psi}_2\,|$ are both less than 0.5, or either is less than 0.35, $\phi(-z)$ will be under 0.0075. However, beyond the monotonicity limit $\phi(-z)$ rises again and, where the denominator normal does not have almost all its probability on one side of zero (aliasing between large positive and large negative values of θ thus arising), the calculated posterior density can reach non-negligible negative levels. Potential resulting inferential inaccuracies can largely be avoided by, beyond the monotonicity limit, fixing $\phi(-z)$ at its level there and setting the unreal negative posterior density values to zero; results in this paper are based on so doing. A ratio-normal distribution, and the approximation thereto, does not have finite posterior moments of any positive order, but its physical appropriateness in the circumstances outlined makes it the natural choice.

The second factor in the RHS of (2) represents the likelihood function. It is identical both to the profile likelihood for θ obtained from the bivariate distribution in (ψ_1,ψ_2) space, and to the implied likelihood derived using Efron's "data-doubling" method (Efron 1993) – see Appendix A. We thus have for the ratio-normal approximation the derived likelihood function:

$$L\left(\theta\right)=\phi\left[-\frac{\left(\tilde{\psi}_1-\theta\tilde{\psi}_2\right)}{\sqrt{V\left(\theta\right)}}\right] \qquad (3)$$

and, since the posterior density is the product of prior and likelihood, by implication the derived prior

$$\pi_\theta^{JP}(\theta) = \frac{\left(\tilde{\psi}_2\sigma_1{}^2 + \tilde{\psi}_1\sigma_2{}^2\theta\right)}{V(\theta)^{3/2}} \tag{4}$$

which is identical to the implied prior per Efron (1993). It is identical to the Jacobian PDF conversion factor for the parameter transformation from z to θ. The prior (4) is noninformative since Raftery and Schweder (1993) uses an objective Bayesian approach, the prior used to infer the Bayesian posterior for (ψ_1, ψ_2) as the first step in deriving the approximate ratio-normal posterior being noninformative.

Frequentist analysis is also performed using the likelihood-ratio method with the likelihood function (3). Provided the factorisation of the ratio-normal approximation is correct, its likelihood is (ignoring non-monotonicity issues) a transformed normal, for which the likelihood-ratio method provides exact confidence estimation.

The accuracy with which the objective Bayesian posterior CDF given by the ratio-normal approximation, and the distribution obtained by applying the likelihood-ratio method to its derived likelihood function, provide estimates matching a confidence distribution for the ratio of two normals is pertinent. I test this by random sampling from numerator and denominator normals in two hypothetical estimation cases. The first case, experiment A, involves a much larger standard deviation, relative to the mean, in the numerator normal distribution than in the denominator normal. In the second case, experiment B, the relative uncertainty in the denominator normal greatly exceeds that in the numerator normal. The parameter θ and the true value of the denominator variable ψ_2 (and hence also that of the numerator variable, ψ_1) are initially fixed, so as to isolate the effect of varying uncertainty in the numerator and denominator standard deviations. For each of experiments A and B, separately, 15,000 samples were drawn from each of unit mean numerator and denominator normal distributions (representing pairs of possible estimates of the observable parameters). Negative central estimates for the denominator normal, a minute proportion of the samples, were changed to +0.001 so that those instances are, appropriately, treated as producing a very large positive parameter estimate. The results (Figure 1) are plotted in terms of the match, at each percentage point of one-sided lower credible or confidence intervals for the parameter, between the percentage involved and the proportion of samples for which the true value of θ fell below the applicable percentile of the posterior CDF or likelihood-ratio derived upper confidence bound. In panel (a), the numerator and denominator normals have standard deviations of respectively 0.4 and 0.1 (Experiment A). In panel (b) they are respectively 0.1 and 0.4 (Experiment B). Solid lines are for objective Bayesian inference, credible intervals being taken from the CDF of the ratio-normal approximation. They are partly obscured by the dashed lines for frequentist inference, confidence intervals being calculated using the likelihood-ratio method on the derived likelihood function underlying the ratio-normal approximation posterior. Black dotted lines show perfect probability matching (frequentist coverage).

For both experiments, an essentially perfect match (represented by a straight line from the origin to the 100%–100% point) is provided by both methods. Numerical coverage at various percentiles is given in Table 1 (upper section). A perfect match implies that the distribution provides exact lower-sided confidence intervals at all confidence levels, the standard definition of a confidence distribution (CD) (Schweder and Hjort 2002; Singh, Xie and Strawderman 2005). In the absence of non-monotonicity in the transformation between θ and z, the likelihood-ratio method and the objective Bayesian method using the prior in (4) would provide identical inference for θ, as a result of that prior equating to the Jacobian PDF conversion factor for the transformation from z to θ (and to a normal model applying in z-space).

The matching is likewise almost perfect (not illustrated) when either ψ_1 and ψ_2, or θ and ψ_2, are randomly selected for each draw. In both these cases, the same two combinations of standard deviations were used as before. In order to avoid a significant proportion of the resulting samples of $\tilde{\psi}_2$ having a material part of their probability below zero, resulting in the ratio-normal approximation being unsuitable, in both cases the interval from which ψ_2 is drawn is bounded below at 1; the intervals from which θ and ψ_1 are drawn are bounded below at 0. The upper limits of the intervals (4 in the first case and 2 in the second case, for both variables) are of little significance.

The near exact probability matching provided by the posterior CDF shows that the ratio-normal approximation to the distribution of the ratio of two normals is extremely good in both cases considered. The similarly almost-exact matches for the likelihood-ratio derived confidence bounds provide further confirmation that the factorisation of the ratio-normal approximation posterior into likelihood function and noninformative prior is correct, in the sense that it leads to precise and correct confidence intervals. Since the ratio-normal approximation posterior represents a transformed normal distribution, and a normal posterior distribution corresponds to a normal likelihood when using a uniform prior (which is known to be completely noninformative in this case), if the ratio-normal approximation were perfect then applying the likelihood-ratio method to its likelihood function would provide an exact confidence distribution.

In the univariate continuous case, Jeffreys prior gives rise to one-sided Bayesian credible intervals that are, asymptotically, closer in probability to confidence intervals than for any other prior (Hartigan 1965), with agreement exact for a location parameter or monotonic transformation thereof (Lindley 1958). Because it involves transforming a pivot variable, the ratio-normal approximation essentially meets the conditions for exact agreement, subject to negligible probability lying beyond the point where the transformation becomes non-monotonic. The essentially exact probability matching obtained here implies that the prior (4) is either the Jeffreys prior or a close approximation thereto. Jeffreys prior involves integrating over the relevant sample space, so for an exact ratio-normal distribution it cannot depend on $\tilde{\psi}_1$ or $\tilde{\psi}_2$. However, where inference is based on a pivot variable, the joint sample–parameter space is mapped 1–1 to the pivotal space, where the density function that applies is specified by a label that may depend on both parameters and observations (Barnard 1980). Therefore, integration over the pivotal quasi-sample space is appropriate when deriving Jeffreys prior. Where there are observations of two variables but a univariate pivot is used, the information available from the observations that is not embodied in that primary pivot is represented by a secondary pivot that provides an ancillary statistic. Inference then involves conditioning on the observed value of that ancillary (Barnard 1980), so integrating over its possible values is not appropriate when deriving a Jeffreys prior for inference using the univariate pivot. The ratio-normal approximation specifies an approximate, fixed, distribution for a univariate pivot variable, $z \equiv (\tilde{\psi}_1 - \theta\tilde{\psi}_2)/V(\theta)^{1/2}$. When a Jeffreys prior is derived by twice differentiating the log-likelihood for the pivot variable and then integrating over the 1D pivotal space to obtain the probability-weighted average, the result is the prior in (4). This indicates that (4) is Jeffreys prior for the pivot-based ratio-normal approximation, ignoring the restricted monotonicity range. I term (4) a "pivotal-Jeffreys prior", since it can be interpreted as a Jeffreys prior in the pivotal space, but (as it is data-dependent) it is not a Jeffreys prior for the exact ratio-normal distribution.

4. Combining information from fitted ratio-normal approximation distributions

4.1 Deriving a noninformative prior for combined-evidence estimation

Since Jeffreys prior is the square root of Fisher information, and Fisher information pertaining to separate experiments with independent data is additive, Jeffreys prior applicable for inference from two independent sources of evidence combined is simply the sum in quadrature of the Jeffreys priors applicable to each of them separately. On the basis that the prior given by (4) is Jeffreys prior (or a sufficiently close approximation thereto for addition in quadrature to be appropriate), Jeffreys prior (or a close approximation thereto) for inference for parameter θ from the combination of two independent estimates A and B, each represented by a ratio-normal approximation, $\pi^{JP}_{\theta_{AB}}(\theta)$, is therefore:

$$\pi^{JP}_{\theta_{AB}}(\theta) = \sqrt{\frac{\left(\tilde{\psi}_{A2}\sigma_{A1}{}^2 + \tilde{\psi}_{A1}\sigma_{A2}{}^2\theta\right)^2}{\left(\sigma_{A1}{}^2 + \theta^2\sigma_{A2}{}^2\right)^3} + \frac{\left(\tilde{\psi}_{B2}\sigma_{B1}{}^2 + \tilde{\psi}_{B1}\sigma_{B2}{}^2\theta\right)^2}{\left(\sigma_{B1}{}^2 + \theta^2\sigma_{B2}{}^2\right)^3}} \qquad (5)$$

I term (5) a "Jeffreys pooled prior". This prior differs from that derived in the Bayesian melding approach proposed by Poole and Raftery (2000). Rather than corresponding to Jeffreys' prior based on the additively-pooled separate Fisher informations, their prior involved logarithmic pooling: multiplying the two separate priors, one raised to the power α and the other to the power $1 - \alpha$. Their argument for doing so related to how to combine subjective expert opinions, not to the informativeness of the data, and the value of α was left undetermined.

4.2 Probability matching for the combination estimate

The accuracy of the approximate confidence distributions obtained by applying the likelihood-ratio method applied to multiplicatively-combined ratio-normal approximation likelihood functions, and the objective Bayesian posterior CDF given by using the Jeffreys pooled prior (5) with the same combined likelihood, is tested using the previously drawn random samples. Since draws are from the numerator and denominator normals for each experiment, the fidelity of the ratio-normal approximation to actual ratio-normals (already shown to be excellent) is tested as well as the accuracy of each method of combining the available evidence. Results, shown in Figure 2, are plotted on the same basis as in Figure 1. The solid line for inference using the Jeffreys pooled prior (5) for the combined ratio-normal approximations for each experiment is partly obscured by the dashed likelihood-ratio based confidence line. The excellent matches for both methods at all percentage points show that each of them provides results equating closely to a confidence distribution.

Test results are also shown for the two posterior CDFs derived by Bayesian updating. The short-dashed and dash-dotted lines are for inference based on Bayesian updating of the ratio-normal approximation posterior for respectively experiment A or B, used as a prior, with the derived ratio-normal approximation likelihood function for the other experiment. That equates to single-step Bayesian inference, from the product of the two derived ratio-normal approximation likelihood functions, using the derived pivotal-Jeffreys prior for respectively experiment A or B. The much poorer, and differing, matches in the Bayesian updating cases demonstrate that this standard Bayesian method is unsatisfactory for combining individually probability-matching estimates where they involve different (noninformative) priors: the combination estimates are neither consistent nor provide good probability matching.

The characteristics of all the PDFs in terms of one-sided credible intervals (CDF points), and of all the likelihoods as regards one-sided confidence intervals derived using the likelihood-ratio method, are given in the lower section of Table 1. As previously, reported results are for testing at fixed θ and ψ_2 values, but are almost identical when θ and ψ_2 values are randomly selected for each draw.

5. Application to estimating the Earth's climate sensitivity

A good example of an application of the approach set out herein is estimation of the Earth's equilibrium climate sensitivity. ECS quantifies the increase in global mean surface temperature ("Global temperature") resulting, once the ocean has equilibrated, from a doubling of atmospheric carbon dioxide concentration (which increases the "effective radiative forcing" of the climate system). ECS is generally regarded as, within limits, a fixed property of the climate system that linearly relates changes in Global temperature to changes in effective radiative forcing. Despite considerable efforts over several decades to narrow estimates of ECS, the latest (fifth) assessment report of the Intergovernmental Panel on Climate Change (IPCC) was only able to put wide bounds on ECS, concluding that there was no more than 5% probability that ECS is under 1°C and no more than 10% probability that ECS exceeds 6°C (Collins et al 2014). Uncertainty regarding the poorly-constrained upper tail of the ECS distribution is of particular concern, since it significantly influences central estimates of the economic damage from future anthropogenic global warming. In principle, combining probabilistic estimates from different sources should enable uncertainty regarding ECS to be narrowed. Estimates based on instrumental data relating to climate change during the industrial period, although themselves highly correlated, should involve uncertainty that is largely independent of that for proxy-based estimates relating to earlier (paleoclimate) periods, so combining estimates in these two categories is particularly attractive. There have been various attempts to do so (Annan and Hargreaves 2006; Hegerl et al. 2006), but all using subjective Bayesian methods, with choice of initial prior distributions significantly influencing the results.

Observationally-based estimation of ECS typically involves, directly or indirectly, dividing an estimated change in Global temperature by the estimated causative change in effective radiative forcing involved (reduced appropriately where the ocean does not reach equilibrium and expressed relative to that applicable to a doubling of carbon dioxide concentration) ("Adjusted radiative forcing"). Uncertainties in both Δ Global temperature and Δ Adjusted radiative forcing (Δ indicating a change), arising from measurement error and internal climate system variability, are usually considered to be approximately normally distributed. Where instrumentally-measured data is used, generally uncertainty in Δ Global temperature as a proportion of its central estimate is moderate, but uncertainty in Δ Adjusted radiative forcing as a proportion of its central estimate is substantial. This gives rise to estimated distributions for ECS that are typically highly skewed, with a long upper tail. Where paleoclimate proxy-based data is used, generally uncertainty in Δ Global temperature as a proportion of the central estimate is much larger than during the industrial period, but uncertainty in Δ Adjusted radiative forcing as a proportion of its mean estimate can be lower. Therefore, ECS distributions estimated from paleoclimate data are typically only moderately skewed. Ratio-normal approximation posteriors were found to closely fit estimated posterior PDFs from a wide selection of instrumental and paleoclimate studies featured in the IPCC fifth assessment report. Whilst the sensitivity estimated by instrumental studies may differ slightly from ECS estimated by paleoclimate studies, the true value of the parameter being estimated is normally assumed to be identical in the two cases, as I assume.

In order to demonstrate how inference using the methods set out in this paper compares with other methods for combining instrumental and paleoclimate evidence regarding ECS, it is appropriate to select a median and uncertainty range to represent each type of evidence, with ratio-normal approximations then being fit thereto. The IPCC fifth assessment report gave an uncertainty range for ECS from paleoclimate evidence of 1–6°C, with no more than a 10% probability (in each case) that ECS was above 6°C or below 1°C. The median ECS estimates from the various individual paleoclimate studies it considered were concentrated around or slightly below 3°C. Accordingly, consistent with a slightly conservative interpretation of the IPCC fifth assessment report, the representative posterior distribution for ECS as estimated from paleoclimate evidence is taken as a ratio-normal approximation fit to a 10–90% range of 1–6°C, with a median of 3°C.

The IPCC fifth assessment report did not give an uncertainty range for ECS from instrumental evidence, but it did place weight on new studies of observed warming over the industrial period that suggested ECS might be lower than previously thought, and accordingly reduced the lower bound for ECS given in the previous IPCC report. It also significantly revised upwards the best estimate for the increase in radiative forcing during the industrial period. Based on the new radiative forcing estimates given in the IPCC fifth assessment report, Lewis and Curry (2014) obtained – for its preferred, best constrained, estimate – a median ECS value slightly below 1.7°C. The 5–95% uncertainty range for that estimate had lower and upper limits of slightly above respectively 1°C and 4°C. The method used did not provide a likelihood function for ECS, but the posterior PDF underlying its ECS estimate can be well matched by a ratio-normal approximation. The majority of recent (2012 onwards) ECS median estimates based on warming over the industrial period have also lain in the 1.5–2°C range, with 5% uncertainty bounds around 1°C but varying 95% uncertainty bounds. In order best to reflect the new IPCC fifth assessment report radiative forcing estimate distributions, the representative posterior distribution for ECS as estimated from instrumental evidence used here is based on the Lewis and Curry (2014) estimate, rounded conservatively to a 5–95% range of 1.0–4.5°C with a median of 1.7°C, fit by a ratio-normal approximation.

Figure 3 demonstrates the application of the approach developed in this paper to combining these representative instrumental and paleoclimate probabilistic ECS estimates, the fitted ratio-normal approximation posterior PDFs for which are shown in panel (a). Panels (c) and (d) show the derived likelihood functions and noninformative priors respectively, for the individual instrumental and paleoclimate estimates and for the combination estimate. Panel (b) shows the resulting combination estimate posterior PDF, applying Bayes' theorem once using the derived Jeffreys pooled prior and the multiplicatively-combined likelihood function, and for comparison posterior PDFs from Bayesian updating of the instrumental estimate posterior PDF, used as a prior, with the derived likelihood function for the paleoclimate estimate, and vice versa. It also shows the posterior PDF produced using a uniform prior, the prior most commonly used for Bayesian ECS estimation. The box plots indicate the percentiles 5 and 95 (vertical bar at ends), 10 and 90 (box-ends), and 50 (vertical bar in box). Those in panels (a) and (b) show Bayesian credible intervals calculated from posterior PDFs and allow for probability that lies outside the *x*-axis ECS range. The box plots in panel (c) show confidence intervals based on the corresponding derived likelihood functions, using the likelihood-ratio method.

Table 2 gives the parameter values for the ratio-normal approximation fits to the instrumental and paleoclimate estimates plotted in Figure 3. It also gives various percentile points for the Bayesian posteriors, and for the likelihood-ratio based confidence distributions based on the derived likelihoods. All percentile points agree very closely between Bayesian inference using priors derived herein and likelihood-ratio inference from the derived likelihoods. For both combination estimates, the 5%, 50% and 95% percentile ECS values are respectively 1.1°C, 2.0°C and 4.5°C. The 97.5% percentile ECS values for the combination estimates are considerably lower, at 5.3°C, than those from either the instrumental or paleoclimate individual estimates. Inference using Bayesian updating produces considerably different results, with percentile point ECS values being up to 20% higher or 15% lower depending on the order of updating, and being higher still using the uniform prior – over 1.9x as high for the 97.5% percentile.

6. Discussion

6.1 Conclusions

This study sets out an objective Bayesian method for combining independent estimates with differing uncertainty characteristics, where they can be represented by parameterised posterior distributions of appropriate form that are factorisable into a likelihood function and noninformative prior. The method is developed for the case of combining two estimates, each of which corresponds to the ratio of two normally-distributed variables, using an approximation devised by Raftery and Schweder (1993). Probability matching is found to be extremely good in all cases tested. Inference is also carried out using a frequentist likelihood-ratio method. Confidence intervals derived using that method with the factorised likelihood agree exactly with credible intervals from the original ratio-normal approximation posterior distribution and, when the likelihoods are multiplied, agree almost exactly with credible intervals from the objective Bayesian combination method. Various non-Bayesian methods for combining estimates apart from simple (or modified) profile likelihood also exist (Schweder and Hjort 2002; Singh, Xie and Strawderman 2005; Xie, Singh and Strawderman 2011; Xie and Singh 2013), but are not investigated here.

The probability matching of posterior PDFs derived by combining evidence from two experiments involving quite different data–parameter relationships (and hence likelihood functions) using a noninformative Jeffreys prior computed, as here, from the sum of the Fisher information for the two experiments, was also tested in both Lewis (2013a) and in Lewis (2013b: Supplemental Information), for cases involving different forms of distribution from those here, and found to be accurate.

The methods developed are applied to estimation of equilibrium climate sensitivity, a central issue in climate science, resulting in a median estimate of 2.0°C and a 5–95% confidence/credibility interval of (1.1, 4.5) °C. Previous studies of the same problem have almost all used subjective Bayesian approaches, employing priors that significantly influenced the results.

Using standard Bayesian updating to combine evidence with differing characteristics is shown to produce results that vary according to ordering and disagree with those obtained using the objective Bayesian method proposed.

The near identical, very good performance of the objective Bayesian method using the Jeffreys pooled prior and the likelihood-ratio method when combining evidence in the cases considered is encouraging. One can obtain insight into why the two methods produce almost identical inference by considering inference in the monotonic region of the *z*-plane relating to the two ratio-normal approximation distributions involved. Since both z_A and z_B, the arguments of the likelihood functions (3) for the two estimates being combined, are functions of the parameter θ, as θ is varied a curve is traced out in the *z*-plane, with their joint density reaching a maximum at some point, say $(\hat{z}_A, \hat{z}_B)$ – the (constrained to the curve) maximum likelihood point. When applying the likelihood-ratio method to the multiplicatively-combined likelihood function, at every point on the θ-curve the standard normal on which the confidence distribution is based has the same density, up to proportionality, as does the original multiplicatively-combined likelihood function. That is because z_A and z_B also have standard normal distributions. It follows, on differentiating the confidence distribution, that if the rate at which the argument, *r*, of its standard normal increased with θ was proportional to the Jeffreys pooled prior everywhere along the θ-curve, then the confidence density and the objective Bayesian posterior density would be proportional, and therefore (since they both integrate to one or almost one) identical. Since the two terms under the square root in (5) are the squared derivatives of z_A and z_B with respect to θ, it can be seen that the Jeffreys pooled prior equals the rate of change of distance *l* along the θ-curve with θ, $\frac{\partial l}{\partial \theta}$. It turns out that the geometry involved in the *z*-plane is such that the ratio of $\frac{\partial l}{\partial \theta}$ to $\frac{\partial r}{\partial \theta}$ is fairly stable. It varies by no more than ±6% in the ECS estimation case, over the range of non-negligible likelihood, hence the near identical performance of the two methods. It is not clear whether a general phenomenon has been uncovered here, since it seems conceivable that in combined-estimates problems involving other functions of other one-parameter models, the corresponding curve has a sub-region in which these two rates of change are quite different.

It is also the case that the Jeffreys pooled prior equals up to proportionality the factor required to convert, into a PDF for the parameter, a joint PDF (along the curve traced out as the parameter varies) in the z-plane corresponding to the ratio-normal approximations representing the two separate estimates involved, provided non-monotonicity can be ignored (Mardia et al. 1979, Eq. 2.5.16; Mosegaard and Tarantola 2002, Eq. 46 – see Lewis 2013b, section S5).

6.2 Incorporating prior information

The methods developed can be used to incorporate, within an objective Bayesian (or a likelihood-ratio) approach, generally-agreed probabilistic existing knowledge about a fixed parameter where it can be represented as the product of the likelihood from a notional observation and a prior that is noninformative in relation thereto. That is also what happens in the 'luckiness MDL approach' , which was derived with completely different, non-Bayesian reasoning, and is advocated within the "minimum description length" literature, an information-theoretic approach to statistical inference (Grunwald 2007, Section 11.3.1). Where that is the case – and Hartigan (1965) suggests it normally is – then it follows that Bayesian inference based on combining an observational likelihood with a prior accurately representing probabilistic existing knowledge will generally not provide objectively-valid inference, since it effectively involves Bayesian updating. In such cases, as in this paper, the parameter is not a Kolmogorov random variable (Barnard 1994, p. 30), so the conditions for the conditional probability lemma to apply and Bayes theorem to be valid are not satisfied (Fraser et al. 2010).

6.3 Other comments

The approach used here is based on factorising a parameterised posterior density of physically-appropriate form that well-matches existing estimates; where actual likelihood functions and related noninformative priors underlying an estimated posterior density are available they may of course be used directly. There is no requirement for the same form of factorisable parameterised distribution to be used for each such estimate to be combined; whatever form is physically-appropriate for each experiment should be used. Additionally, the approach readily generalises to the combination of independent estimates from more than two sources. Because confidence distributions and objective Bayesian posteriors derived from the same evidence are often similar, in suitable cases it might be possible to use the approach set out here to combine confidence distributions, by treating them as if they were objective Bayesian posteriors.

In the cases considered, only the parameter of interest was estimated by multiple studies. In other cases, joint likelihood functions and joint priors would need to be derived for all parameters estimated in common, and the unwanted parameters integrated out once the combined joint posterior had been obtained (the likelihood-ratio method being applied to the profile of the combined joint likelihood).

The approach used here assumes that the uncertainties involved in the individual estimates are independent. Combining non-independent evidence is in general difficult and is outside the scope of this paper.

Acknowledgements

I thank Paul Kirwan, Tore Schweder and an anonymous reviewer for very helpful comments that significantly improved the manuscript.

Appendix A. Derivation of the implied likelihood

Efron (1993) considers a univariate parameter θ for which, based on data *x*, a one-sided level-α confidence interval with upper endpoint $\theta^x(\alpha)$ is derived. Let the inverse function of $\theta^x(\alpha)$, assumed differentiable, be $\alpha^x(\theta)$. Suppose one had observed two independent data sets, both of which equal the actual data, the combined data giving rise to upper endpoint $\theta^{xx}(\alpha)$, with inverse $\alpha^{xx}(\theta)$. Efron derives the implied likelihood $L^{\dagger x}(\theta)$:

$$L^{\dagger x}(\theta) = \frac{d\alpha^{xx} / d\theta}{d\alpha^x / d\theta} \tag{A1}$$

For the ratio-normal approximation

$$\alpha^x(\theta) \;= \Phi\left[-\frac{\left(\tilde{\psi}_1 - \theta\tilde{\psi}_2\right)}{\sqrt{\sigma_1^2 + \theta^2\sigma_2^2}}\right] \tag{A2}$$

($\Phi()$ representing the $N(0,1)$ cumulative distribution function). Following the standard rule that, with known standard deviation, uncertainty declines with the square root of the number of independent observations:

$$\alpha^{xx}(\theta) \;= \Phi\left[-\frac{\left(\tilde{\psi}_1 - \theta\tilde{\psi}_2\right)}{\sqrt{\sigma_1^2 / 2 + \theta^2\sigma_2^2 / 2}}\right] \tag{A3}$$

which, differentiating and applying (A1) and (A2), implies:

$$L^{\dagger x}(\theta) = \frac{\dfrac{\sqrt{2}}{\sqrt{\pi}} \dfrac{\left(\tilde{\psi}_2{\sigma_1}^2 + \tilde{\psi}_1{\sigma_2}^2\theta\right)}{\left(\sigma_1^2 + \theta^2\sigma_2^2\right)^{3/2}} \exp\left[-\dfrac{\left(\tilde{\psi}_1 - \theta\tilde{\psi}_2\right)^2}{\left(\sigma_1^2 + \theta^2\sigma_2^2\right)}\right]}{\dfrac{1}{\sqrt{2\pi}} \dfrac{\left(\tilde{\psi}_2{\sigma_1}^2 + \tilde{\psi}_1{\sigma_2}^2\theta\right)}{\left(\sigma_1^2 + \theta^2\sigma_2^2\right)^{3/2}} \exp\left[-\dfrac{\left(\tilde{\psi}_1 - \theta\tilde{\psi}_2\right)^2}{2\left(\sigma_1^2 + \theta^2\sigma_2^2\right)}\right]} \tag{A4}$$

$$\propto \phi\left[-\frac{\left(\tilde{\psi}_1 - \theta\tilde{\psi}_2\right)}{\sqrt{V(\theta)}}\right]$$

which is the second factor of the RHS of (2).

References

Annan, J. D., and J. C. Hargreaves, 2006. Using multiple observationally-based constraints to estimate climate sensitivity. *Geophys. Res. Lett*., **33**, L06704.

Barnard, G.A , 1980. Pivotal inference and the Bayesian controversy. *Trabajos de Estadistica Y de Investigacion Operativa*, February 1980, Volume 31, Issue 1, pp 295-318

Barnard, G. A., 1994. Pivotal inference illustrated on the Darwin Maize Data. In *Aspects of Uncertainty*, Freeman, PR and AFM Smith (eds), Wiley, 392pp.

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. *Philos. Trans. R. Soc. Lond*. **53** (1763) 370–418; 54 (1764) 269–325. Reprinted in Biometrika 45 (1958), 293-315.

Berger J. O. and J. M. Bernardo, 1992. On the development of reference priors (with discussion). In: Bernardo J. M., Berger J. O., Dawid A. P., Smith A. F. M. (eds) *Bayesian Statistics 4*. Oxford University Press, pp 35-60.

Bernardo J. M., 1979. Reference posterior distributions for Bayesian inference (with discussion). *J Roy Statist Soc Ser B* 41:113–147.

Bernardo, J.M. and A.F.M. Smith, 1994. *Bayesian Theory*. Wiley, 608pp.

Bernardo, J.M., 2011. Modern Bayesian Inference: Foundations and Objective Methods, 263–306. In *Philosophy of Statistics*, P. Bandyopadhyay and M. Forster, eds. North Holland, 1253 pp.

Brazzale, A.R., A.C. Davidson and N. Reid, 2007. *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge University Press, 236pp.

Collins, M., Knutti R. et al., 2014. Long-term Climate Change: Projections, Commitments and Irreversibility. *Climate Change 2013: The Physical Science Basis Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Cox, D. R. and Reid, N., 1987. Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc B*, 49, 1–39.

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.

Fraser, D. A. S., N. Reid, E. Marras and G. Y. Yi, 2010. Default priors for Bayesian and frequentist inference. *J. R. Statist. Soc.* B, **72**, 5, 631-654.

Grunwald, P., 2007. *The Minimum Description Length Principle*. MIT Press.

Hartigan, J.A., 1965: The Asymptotically Unbiased Prior Distribution, *Ann. Math. Statist*., **36**, 4, 1137-1152.

Hegerl G., T. C. Crowley, W. T. Hyde and D. J. Frame, 2006. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature*, **440**, 1029–1032, doi:10.1038/nature04679.

Jeffreys, H., 1946. An Invariant Form for the Prior Probability in Estimation Problems. *Proc. Roy. Soc. A*, **186**, 453–461.

Kass, R. E., 1989: The geometry of asymptotic inference. *Statistical Science*, **4**, 188-219.

Kass, R. E. and L. Wasserman, 1996. The Selection of Prior Distributions by Formal Rules. *J. Amer. Stat. Ass.*, **91**, 435, 1343-1370.

Lewis, N., 2013a. Modification of Bayesian Updating where Continuous Parameters have Differing Relationships with New and Existing Data. *arXiv*:1308.2791 [stat.ME].

Lewis, N., 2013b. An objective Bayesian improved approach for applying optimal fingerprint techniques to estimate climate sensitivity. *Journal of Climate*, **26**, 7414-7429.

Lewis N., Curry J.A., 2014. The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Clim. Dyn.* , **45** (2015), 1009–1023. DOI 10.1007/s00382-014-2342-y

Lindley, D. V., 1958. Fiducial Distributions and Bayes Theorem. *J. Roy. Statist, Soc. B*, **20**, 1, 102-107

Mardia, K. V., J. T. Kent and J. M. Bibby, 1979: *Multivariate Analysis.* Academic Press, 518pp.

Mosegaard, K. and A. Tarantola, 2002: Probabilistic Approach to Inverse Problems. *Int'l. Handbook of Earthquake and Engineering Seismology, Volume 81A*: Lee, W. H. K. et al., Int'l Assoc. Seismol. & Phys. Earth's Interior, Committee on Education, Academic Press, 933pp.

Pawitan, Y., 2001. *In all Likelihood: Statistical Modeling and Inference Using Likelihood* Ch. 3.4. Oxford Univ. Press, 514 pp.

Poole, D. and Raftery, A. E., 2000. Inference in deterministic simulation models: The Bayesian melding approach. *Journal of the American Statistical Association*, **95**, 1244–1255.

Raftery, A.E. and Schweder, T., 1993. Inference About the Ratio of Two Parameters, With Application to Whale Censusing. *The American Statistician*, **47**, 4, 259-264.

Schweder, T. & Hjort, N.L., 2002: Confidence and likelihood. *Scand. J. Stat.*, **29**, 309–332.

Schweder, T. and Hjort, N.L. 2016. *Confidence, Likelihood, Probability*. Cambridge University Press. 500 pp.

Seidenfeld, T., 1979. Why I am not an objective Bayesian; Some reflections prompted by Rosenkrantz. *Theory and Decision*, **11**, 413–440.

Singh, K., Xie, M., and Strawderman, W. E., 2005. Combining Information From Independent Sources Through Confidence Distributions. *The Annals of Statistics*, **33**, 159–183.

Welch B.L. and H. W. Peers, 1963. On formulae for confidence points based on integrals of weighted likelihoods. *J. Roy. Soc. Ser. B*, **25**, 318-329

Williamson, J., 2009. Objective Bayesianism, Bayesian conditionalisation and voluntarism. *Synthese*, **178**, 67–85

Xie, M., K Singh and W E Strawderman, 2011. Confidence Distributions and a Unifying Framework for Meta-Analysis, *Journal of the American Statistical Association*, **106**, 320–333

Xie, M. and K. Singh, 2013. Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review [with discussion and a rejoinder]. *International Statistical Review* **81**, 3–39.

PDF concerned and relevant characteristic	Ratio-normal approximation numerator normal distribution that is sampled		Ratio-normal approximation denominator normal distribution that is sampled		Fraction of cases with true ratio < 5% point	Fraction of cases with true ratio < 10% point	Fraction of cases with true ratio < 50% point	Fraction of cases with true ratio < 90% point	Fraction of cases with true ratio < 95% point
	Mean	Standard deviation	Mean	Standard deviation					
Estimate A: Credible bound – Jeffreys prior (4)	1	0.4	1	0.1	0.050	0.100	0.496	0.906	0.955
Estimate A: Confidence bound – likelihood-ratio	*1*	*0.4*	*1*	*0.1*	*0.050*	*0.100*	*0.496*	*0.906*	*0.955*
Estimate B: Credible bound – Jeffreys prior (4)	1	0.1	1	0.4	0.050	0.103	0.502	0.901	0.949
Estimate B: Confidence bound – likelihood-ratio	*1*	*0.1*	*1*	*0.4*	*0.050*	*0.103*	*0.503*	*0.900*	*0.949*
Combined: Credible bound – Jeffreys prior (5)	-	-	-	-	0.050	0.103	0.504	0.901	0.950
Combined: Confidence bound – likelihood-ratio	-	-	-	-	*0.052*	*0.106*	*0.505*	*0.899*	*0.947*
Bayesian update PDFs									
Credible bound – PDF for A, likelihood for B	-	-	-	-	0.073	0.140	0.604	0.944	0.974
Credible bound – PDF for B, likelihood for A	-	-	-	-	0.025	0.058	0.408	0.867	0.931

Table 1. Simulation experiment. Characteristics of all the PDFs in terms of one-sided credible intervals (CDF points), and (in italics) confidence bounds given by the likelihood-ratio method. Upper panel, results for individual parameter estimates, A and B, represented by ratio-normal approximations with different skewness. Lower panel, results for estimates combining the information from estimates A and B. Credible intervals for the combined estimate based on Bayesian updating are shown at the bottom.

PDF and prior involved	Ratio-normal approximation numerator normal distribution		Ratio-normal approximation denominator normal distribution		ECS at particular credible or confidence points (top of lower one-sided intervals) [°C]					
	Mean	Standard deviation	Mean	Standard deviation	5%	10%	50%	90%	95%	97.5%
Instrumental estimate: fitted ratio-normal approximation posterior; prior per (4)	1.700	0.200	1 (fixed)	0.376	1.00	1.11	1.70	3.32	4.50	6.51
Per derived likelihood					*1.00*	*1.11*	*1.70*	*3.32*	*4.50*	*6.51*
Paleoclimate estimate: fitted ratio-normal approximation posterior; prior per (4)	3.00	1.533	1 (fixed)	0.295	0.47	1.00	3.00	6.00	7.38	9.01
Per derived likelihood					*0.47*	*1.00*	*3.00*	*6.00*	*7.38*	*9.01*
Combined – objective Bayesian; prior per (5)					1.10	1.24	2.00	3.73	4.50	5.33
Per multiplicatively combined likelihood					*1.10*	*1.23*	*1.98*	*3.70*	*4.48*	*5.31*
Bayesian update PDFs										
Instrumental PDF updated by paleo likelihood					1.07	1.20	1.86	3.23	3.85	4.52
Paleo PDF updated by instrumental likelihood					1.28	1.45	2.42	4.33	5.17	6.07
Using uniform prior & the derived likelihoods					1.33	1.52	2.68	5.53	7.28	10.18

Table 2. Combining information for two cases consisting of independent estimates of the Earth's equilibrium climate sensitivity based on different types of evidence, each represented by a ratio-normal approximation posterior PDF. Results are stated in terms of the ECS values corresponding to the tops of specified one-sided credible intervals (based on the stated prior) and confidence intervals (in italics, derived using the likelihood-ratio method). Lower panel, results from standard Bayesian updating.

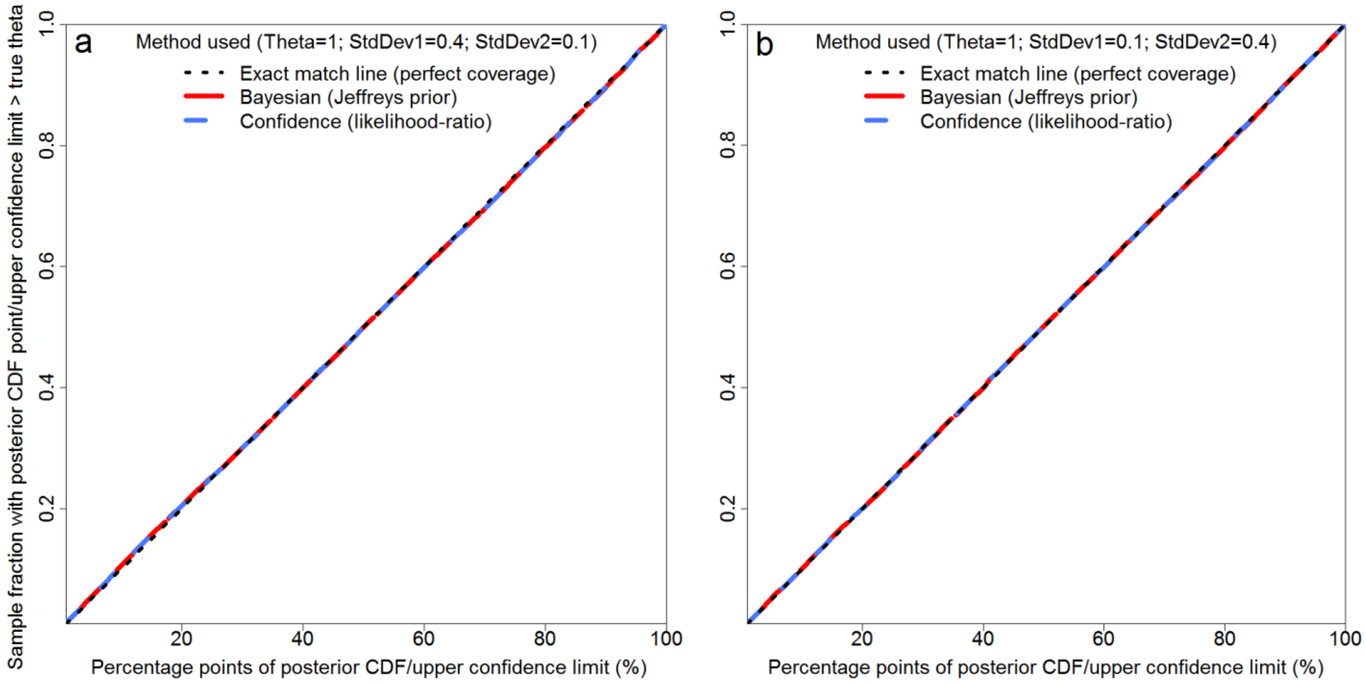

Figure 1. Probability matching – individual estimates. Estimation is of a parameter θ, the ratio of two observable parameters, and corresponds to estimating the ratio of the true means of two unit mean normally-distributed independent random variables. StdDev1 and StdDev2: numerator and denominator variable standard deviations.

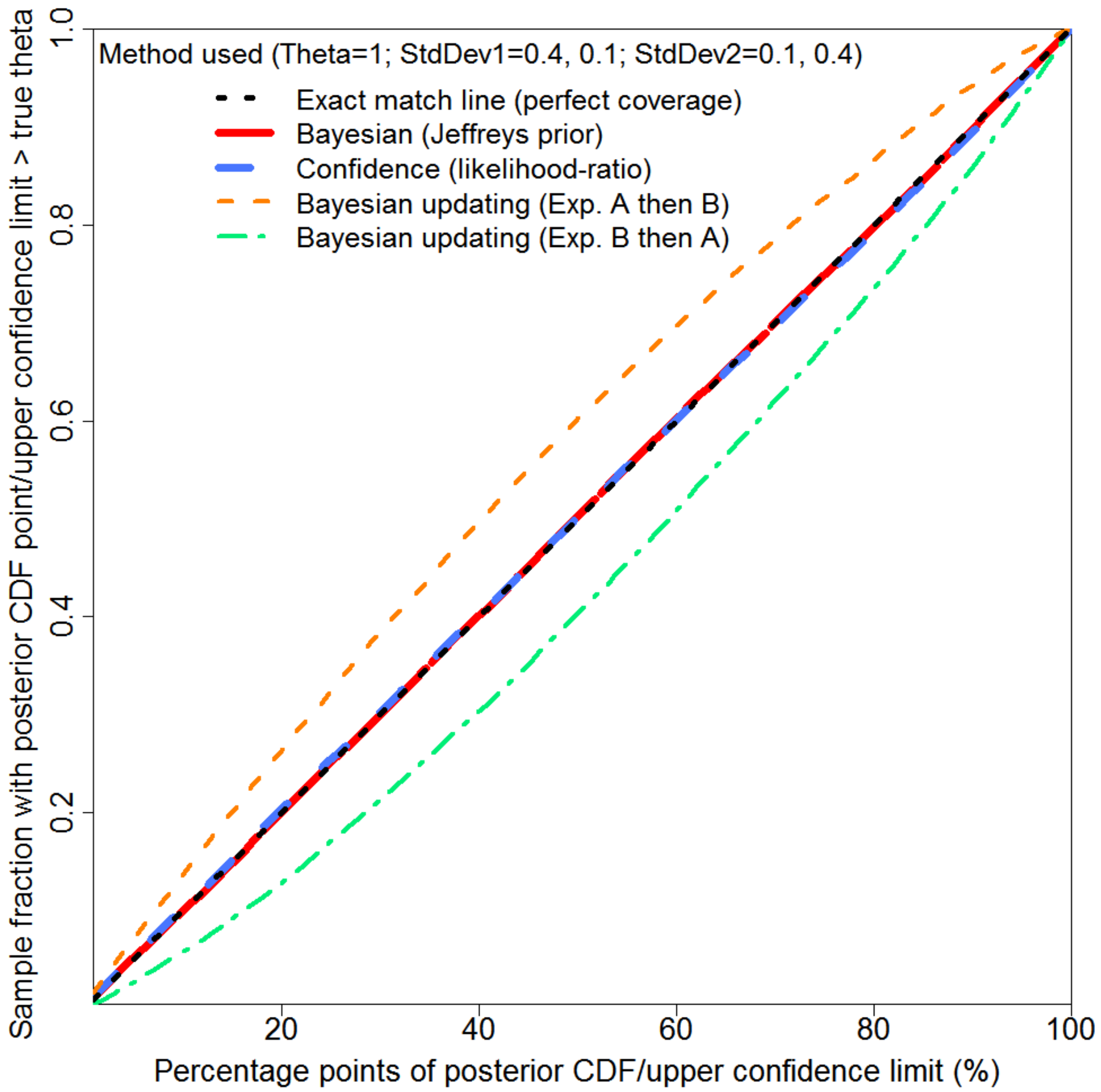

Figure 2. Probability matching – combined estimate. As in Figure 1, but with estimation based on pairs of independent samples (each comprising estimates for both numerator and denominator observable parameters) from both experiments A and B.

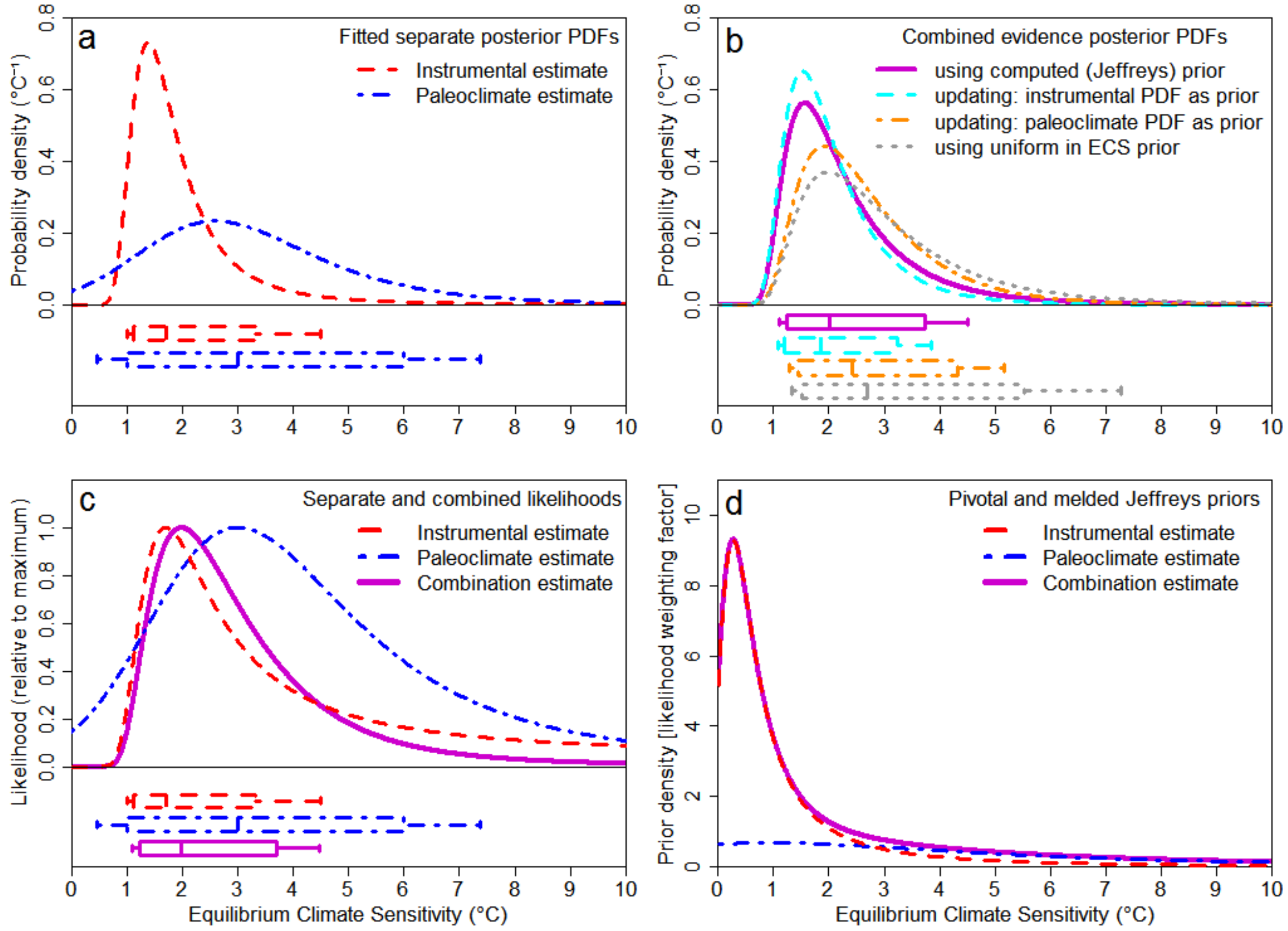

Figure 3. Inference based on PDFs for ECS selected to represent instrumental evidence, paleoclimate evidence and on evidence derived from combining them. The origins of the two fitted posterior PDFs in panel a are explained in section 5 and the parameters of the ratio-normal approximations involved are set out in Table 2.