Combining Information Across Diverse Sources: The II-CC-FF Paradigm

Céline Cunen* Nils Lid Hjort†

Abstract
We introduce and develop a general paradigm for combining information across diverse data sources. In broad terms, suppose ϕ is a parameter of interest, built up via components $\psi_1, \ldots, \psi_k$ from data sources $1, \ldots, k$. The proposed scheme has three steps. First, the Independent Inspection (II) step amounts to investigating each separate data source, translating statistical information to a confidence distribution $C_j(\psi_j)$ for the relevant focus parameter ψ_j associated with data source j. Second, Confidence Conversion (CC) techniques are used to translate the confidence distributions to confidence log-likelihood functions, say $\ell_{c,j}(\psi_j)$. Finally, the Focused Fusion (FF) step uses relevant and context-driven techniques to construct a confidence distribution for the primary focus parameter $\phi = \phi(\psi_1, \ldots, \psi_k)$, acting on the combined confidence log-likelihood. In simpler setups, the II-CC-FF strategy amounts to versions of meta-analysis, but its potential lies in applications to harder problems. Illustrations are presented, related to actual applications.

Key Words: combining information, confidence distributions, confidence likelihoods, focused fusion, meta-analysis

1. Combining Information and the II-CC-FF Scheme

Our paper concerns the statistical task of combining information across different and perhaps very diverse data sources. This is of course a long-standing theme in statistics, with papers going back to Karl Pearson (cf. Simpson & Pearson (1904)); see Schweder & Hjort (2016, Ch. 13) for background, a general discussion of themes traditionally sorted under the bag-word meta-analysis, along with further basic references. The present paper aims at proposing and developing a certain paradigm, which we call the II-CC-FF method, meant to be powerfully applicable for ranges of situations far beyond the usual simpler setups. We will explain the role and nature of the Independent Inspection (II), Confidence Conversion (CC), Focused Fusion (FF) steps below.

A special case worth considering first is the textbook setup where $y_1, \ldots, y_k$ are independent estimators of the same quantity ψ, and where $y_j \sim \mathrm{N}(\psi, \sigma_j^2)$, with known standard deviations σ_j. An easy exercise in minimising variances shows that the optimally balanced overall estimator is

$$\widehat{\psi} = \frac{\sum_{j=1}^k y_j/\sigma_j^2}{\sum_{j=1}^k 1/\sigma_j^2} \sim \mathrm{N}\Big(\psi, \Big(\sum_{j=1}^k 1/\sigma_j^2\Big)^{-1}\Big). \tag{1}$$

A natural extension, though harder to analyse to full satisfaction, is when $y_j \sim \mathrm{N}(\psi_j, \sigma_j^2)$, with the individual means ψ_j differing according to a $\mathrm{N}(\psi_0, \tau^2)$. Here one wishes clear inference strategies for both overall mean ψ_0 and level of variation τ. We return to this particular problem in Section 5.1.

Many problems of modern statistics involving combining information are much more complicated than the situations sketched above, however. Sometimes one needs to combine 'hard' data, with clear measurements from controlled experiments, etc., with 'soft' data, associated with information more loosely connected to the parameters of primary interest,

*PhD candidate, Department of Mathematics, University of Oslo, Norway; clmcunen@math.uio.no
†Professor, Department of Mathematics, University of Oslo, Norway; nils@math.uio.no

perhaps via measurement errors or surrogate variables. In addition there might be prior distributions available, via subject matter experts, but only for some of the parameters at play, not enough to make it into a clear Bayesian analysis.

In reasonably general terms, assume there is a parameter ϕ of clear interest, related to parameters $\psi_1, \ldots, \psi_k$ via a function $\phi = \phi(\psi_1, \ldots, \psi_k)$. Suppose further that data source y_j provides information pertaining to ψ_j. Our II-CC-FF approach for reaching inference statements for the overall focus parameter ϕ can then be schematically set up as follows:

- ⋄ II, *Independent Inspection*: Data source y_j is used, via appropriate models and analyses, to yield a confidence distribution $C_j(\psi_j)$ for the main interest parameter associated with study j.
- ⋄ CC, *Confidence Conversion*: The confidence distribution is converted into a log-likelihood function for this main parameter of interest for study j, say $\ell_{c,j}(\psi_j)$.
- ⋄ FF, *Focused Fusion*: The combined confidence log-likelihood function $\ell^*(\psi_1, \ldots, \psi_k) = \sum_{j=1}^k \ell_{c,j}(\psi_j)$ is used to reach focused fusion inference for $\phi = \phi(\psi_1, \ldots, \psi_k)$.

The extent to which some or all of these steps will be relatively straightforward or rather complicated to carry out depends to a high degree on the special features of the given source combination problem. The steps are not 'isolated' or fully separated, but often related. In situations where the statistician has all the raw data and the particular models used for analysing the different sources of information, the CC step is in a conceptual sense not difficult, as the required profile log-likelihood parts may be worked out from first principles. In various situations confronting the modern statistician this is rather more difficult, however, as one might be content to base one's analysis on summary measures, directly or indirectly given via other people's work, reports and publications. In such cases the II-CC-FF paradigm looks fruitful.

We start in Section 2 with a brief review of confidence distributions, which are essential for the Independent Inspection (II) part of the programme. We then proceed with giving details related to the basics of Confidence Conversion (CC) in Section 3 and Focused Fusion (FF) in Section 4. The three-step machinery is then seen in action through four applications laid out in Section 5, followed by a brief discussion section rounding off our article.

2. Independent Inspection: Confidence Distributions

Suppose Y denotes a set of random observations, stemming from a model with parameter θ, typically multidimensional, and with $\psi = \psi(\theta)$ a one-dimensional focus parameter. A *confidence distribution* $C(\psi, y)$ for this focus parameter has the properties (i) it is a cumulative distribution function (c.d.f.) in ψ, for each y, and (ii) at the true value θ_0, with associated true value $\psi_0 = \psi(\theta_0)$, the distribution of $C(\psi_0, Y)$ is uniform on the unit interval. From this follows, under the standard continuity and monotonicity assumptions, that

$$\mathrm{Pr}_{\theta_0}\{C^{-1}(0.05, Y) \le \psi_0 \le C^{-1}(0.95, Y)\} = 0.90,$$

etc., i.e. $[C^{-1}(0.05, y_{\rm obs}), C^{-1}(0.95, y_{\rm obs})]$ is a 90% confidence interval for ψ, where $y_{\rm obs}$ denotes the observed dataset. Thus the confidence distribution $C(\psi, y_{\rm obs})$, qua random c.d.f., is a compact and convenient representation of confidence intervals at all levels, and indeed a powerful inference summary. A close relative is the *confidence curve*, which we tend to prefer as a post-data graphical summary of information for focus parameters, defined as

$$\mathrm{cc}(\psi, y_{\rm obs}) = |1 - 2\,C(\psi, y_{\rm obs})|. \tag{2}$$

It points to its cusp point, the median confidence point estimate $\widehat\psi_{0.50} = C^{-1}(\frac{1}{2}, y_{\rm obs})$, and the two roots of the equation $C(\psi, y_{\rm obs}) = \alpha$ form a confidence interval with this confidence level. Degrees of asymmetry are easier to spot and to convey using the confidence curve than with the cumulative confidence distribution itself; cf. illustrations in Section 5. We also note that the random ${\rm cc}(\psi, Y)$ has a uniform distribution, at the true position in the parameter space, since $|1-2\,U|$ is uniform when U is. Indeed

$$\Pr_{\theta_0}\{{\rm cc}(\psi_0, Y) \le \alpha\} = \alpha, \quad \text{for each } \alpha, \tag{3}$$

at the true parameters of the model. The confidence curve is arguably a more fundamental concept than the confidence distribution, as there are cases where a natural ${\rm cc}(\psi, Y)$ may be constructed, with a valid (3), even when confidence regions are formed by disjoint intervals (as with multimodal log-likelihood functions).

For an extensive treatment of confidence distributions, their constructions in different types of setup, properties and uses, see Schweder & Hjort (2016), and the review paper Xie & Singh (2013), with ensuing discussion contributions. The scope and broad applicability of confidence distributions are also demonstrated in a collection of papers published in the special issue *Inference With Confidence* of the journal *Journal of Statistical Planning and Inference*, 2017. Here we shall merely point to two important and broadly useful ways of constructing a confidence distribution, for a focus parameter ψ, based on data from a model with a multidimensional parameter θ. The first is to rely on an approximately normally distributed estimator, if available, say $\widehat\psi \sim {\rm N}(\psi, \kappa^2)$, and with standard deviation well estimated with an appropriate $\widehat\kappa$. Then, with $\Phi(\cdot)$ as usual denoting the c.d.f. of the standard normal,

$$C(\psi, y) = \Phi((\psi - \widehat\psi)/\widehat\kappa) \tag{4}$$

is an approximately correct confidence distribution, first-order large-sample correct under weak regularity conditions. In particular the estimator used can be the maximum likelihood one, say $\widehat\psi_{\rm ml}$, but other estimators are allowed too in this simple construction. The second is based on the profiled log-likelihood function $\ell_{\rm prof}(\psi) = \max\{\ell(\theta)\colon \psi(\theta) = \psi\}$, which leads to the deviance function

$$D(\psi) = 2\{\ell_{\rm prof}(\widehat\psi_{\rm ml}) - \ell_{\rm prof}(\psi)\} = 2\{\ell_{\rm prof,max} - \ell_{\rm prof}(\psi)\}. \tag{5}$$

As laid out in Schweder & Hjort (2016, Chs. 2, 3), the Wilks theorem with variations then lead naturally to

$${\rm cc}(\psi, y) = \Gamma_1(D(\psi)), \tag{6}$$

with $\Gamma_\nu(\cdot)$ denoting the c.d.f. of a χ^2 with degrees of freedom ν.

Typically, the second method (6) leads to a better calibrated confidence curve than the the first method (4). Further fine-tuning methods are developed, illustrated and discussed in Schweder & Hjort (2016, Chs. 7, 8).

3. Confidence Conversion: From Confidence to Likelihoods

Several well-explored methods, with appropriate variations and amendments, lead from likelihood functions to confidence distributions and confidence curves; cf. again several chapters of Schweder & Hjort (2016). Sometimes the CC step comes almost for free, in cases where the statistician can compute say log-likelihood profiles from raw data and given models. But in general the CC step of the II-CC-FF paradigm requires methods for going the other way, from confidence distributions or confidence curves to log-likelihood information, and this is more involved. Among the complications is that different experimental protocols, with ensuing different confidence distributions, might be having the same log-likelihood functions, so the link between confidence and likelihood is not one-to-one.

Schweder & Hjort (2016, Ch. 10) develop and discuss this topic at some length. For the present purposes we shall be content with what we call the *chi-squared inversion*, associated with (6) above. It consists in using

$$\ell_c(\psi) = -\tfrac{1}{2}\Gamma_1^{-1}(\mathrm{cc}(\psi, y)) \tag{7}$$

as the profiled confidence log-likelihood contribution associated with a given confidence curve. When the confidence curve is constructed via $\mathrm{cc}(\psi, y) = |1 - 2\,C(\psi, y)|$, this is also equivalent to the *normal conversion* $\ell_c(\psi) = -\frac{1}{2}\{\Phi^{-1}(C(\psi, y))\}^2$. A relevant point here is that one often constructs a confidence curve $\mathrm{cc}(\psi, y)$ directly, not always via (2), making (7) a more versatile tool. The normal conversion confidence likelihood is also what Efron (1993) proposed, for coming from confidence to likelihood, via different arguments and for different purposes; see also Efron & Hastie (2016, Ch. 11).

One may work through various examples, to see how well the chi-squared inversion method (7) manages to approximate the real profiled log-likelihood. Both are guaranteed to be close to the negative quadratic $-\frac{1}{2}(\psi - \widehat{\psi}_{\rm ml})^2/\widehat{\kappa}^2$, for the appropriate $\widehat{\kappa}$, by arguments associated with large-sample calculus – including asymptotic normality of the maximum likelihood estimator and indeed the Wilks theorem, see Schweder & Hjort (2016, Ch. 2 and Appendix). The results are typically good and promising also when the data information volume is small, as long as the underlying models are smooth in their parameters.

4. Focused Fusion: From Full Likelihood to Focus Parameter

In this section we outline how the Focused Fusion step typically may be carried out, via profiling of the combined confidence log-likelihood. We first discuss a method for combining confidence distributions developed and used by Singh et al. (2005) and others, valid when all confidence components relate to a common focus parameter. We then point out that there are sometimes more powerful methods, using a machinery for loss and risk functions developed in Schweder & Hjort (2016, Chs. 5, 7, 8), before we tend to the focused profiling.

4.1 The back-and-forth linear combination transformation method

Suppose now that independent information sources $y_1, \ldots, y_k$ give rise to confidence distributions for the same parameter, say $C_1(\psi, y_1), \ldots, C_k(\psi, y_k)$. A general way of combining these into a single overall confidence distribution has been proposed and worked with by Singh et al. (2005), later on applied in various contexts by Xie & Singh (2013), Liu et al. (2014, 2015), and others. The starting point is that under the true state of affairs, the $\Phi^{-1}(C_j(\psi, Y_j))$ are independent standard normals, from the basic properties of confidence distributions; here $\Phi(\cdot)$ as usual denotes the c.d.f. for the standard normal. Hence $\sum_{j=1}^k w_j \Phi^{-1}(C_j(\psi, Y_j))$ is also standard normal, when the weights w_j are such that $\sum_{j=1}^k w_j^2 = 1$. This again implies that

$$\bar{C}(\psi, y) = \Phi\Big(\sum_{j=1}^k w_j \Phi^{-1}(C_j(\psi, y_j))\Big) \tag{8}$$

is a confidence distribution for ψ, using the combined dataset $y = (y_1, \ldots, y_k)$. The idea generalises to other basic distributions than the normal, but then the required convolutions become less tractable.

For the prototype situation associated with (1), the individual confidence distributions take the form $C_j(\psi, y_j) = \Phi((\psi - y_j)/\sigma_j)$, and the general (8) recipe yields

$$\bar{C}(\psi, y) = \Phi\Big(\sum_{j=1}^{k} w_j(\psi - y_j)/\sigma_j\Big).$$

Some considerations then lead to the best of these linear combinations, with weights w_j proportional to $1/\sigma_j$ and $\sum_{j=1}^{k} w_j^2 = 1$. This indeed agrees with the standard method (8).

Recipe (8) requires nonrandom weights w_j, and these could in various cases be fruitfully taken as proportional to $1/\sqrt{m_j}$, with m_j the sample size associated with data source y_j. In many other situations the balance is more delicate, however, perhaps demanding nonrandom weights, of the type $\widehat{w}_j$ estimating an underlying optimal but not observable $w_{j,0}$. Problems worked with in Liu et al. (2014, 2015) are of this type. In such cases recipe (8) is not entirely appropriate and is rather to be seen as an approximation, associated with confidence intervals with approximate levels of confidence. A better strategy would often be to work with the actual distribution, say H, of

$$Z^* = \sum_{j=1}^{k} \widehat{w}_j Z_j, \quad \text{with } Z_j = \Phi^{-1}(C_j(\psi, Y_j)).$$

The appropriate generalisation of the recipe above is then

$$\bar{C}(\psi, y) = H\Big(\sum_{j=1}^{k} \widehat{w}_j \Phi^{-1}(C_j(\psi, y_j))\Big), \tag{9}$$

perhaps with H evaluated or estimated via simulations. In situations with increasing data volume the estimated weights $\widehat{w}_j$ would come close in probability to the underlying $w_{j,0}$, and H would tend in distribution to Φ, hence with (9) leading back to (8). In yet other words, method (8) remains correct to the first-order large-sample degree, even though more careful versions of (9) would tend to work better for smaller samples.

4.2 Confidence power and optimal methods

Strategies (8)–(9) are not always the most powerful, however. For situations where there are competing confidence distribution strategies for the same parameter, Schweder & Hjort (2016, Ch. 5) have developed a theory for loss and risk functions for these. Suppose again that ψ is the focus parameter and that the log-likelihood function at work, based on information sources $y_1, \ldots, y_k$, can be written in the form

$$\ell(\psi, \gamma_1, \ldots, \gamma_m) = \psi A + \gamma_1 B_1 + \cdots + \gamma_m B_m - d(\psi, \gamma_1, \ldots, \gamma_m) + h(y_1, \ldots, y_k), \tag{10}$$

where A and $B_1, \ldots, B_m$ are statistics, i.e. functions of the data collection, with observed values $A_{\rm obs}$ and $B_{1,\rm obs}, \ldots, B_{m,\rm obs}$, and with m often bigger than k. Then, under mild regularity conditions, there is an overall most powerful confidence distribution, namely

$$C^*(\psi, y) = \Pr_\psi\{A \ge A_{\rm obs} \,|\, B_1 = B_{1,\rm obs}, \ldots, B_m = B_{m,\rm obs}\}.$$

That this $C^*(\psi, y)$ indeed depends on ψ but not on the γ_j parameters is part of the result and the construction. Confidence power is measured via the risk function

$$r(C, \psi, \gamma) = {\rm E}_{\psi,\gamma} \int \Gamma(\psi_{\rm cd} - \psi)\, {\rm d}C(\psi_{\rm cd}, Y), \tag{11}$$

for any convex nonnegative $\Gamma(\cdot)$ with $\Gamma(0) = 0$. The random mechanism involved in the expectation here is a two-stage operation – first data y, governed by the (ψ, γ) held fixed, are used to generate the confidence distribution $C(\psi, y)$, and then $\psi_{\rm cd}$ is a random draw from this distribution.

Example. Suppose we are observing independent gamma distributed variables $Y_j \sim {\rm Gam}(a_j, \theta)$ for $j = 1, \ldots, k$, with densities proportional to $y_j^{a_j - 1}\exp(-\theta y_j)$, with known shape parameters a_j and unknown scale parameter θ. The canonical confidence distribution for data source y_j alone is

$$C_j(\theta, y_{j,{\rm obs}}) = {\rm Pr}_\theta\{Y_j \le y_{j,{\rm obs}}\} = G(\theta y_{j,{\rm obs}}, a_j, 1),$$

with $G(\cdot, a_j, 1)$ denoting the c.d.f. of the ${\rm Gam}(a_j, 1)$ distribution. This is indeed seen to be the optimal confidence distribution, based on y_j, by applying the theory of Schweder & Hjort (2016, Ch. 5), The (8) recipe leads to

$$\bar C(\theta, y) = \Phi\Big(\sum_{j=1}^k w_j \Phi^{-1}(G(\theta y_j, a_j, 1))\Big),$$

with appropriate weights; in particular, if the shape parameters a_j are the same, one ought to take $w_j = 1/\sqrt{k}$.

There is a better method, however. The log-likelihood function is $\ell(\theta) = \sum_{j=1}^k (a_j \log\theta - \theta y_j)$, with $y^* = \sum_{j=1}^k y_j$ as sufficient statistic. From $Y^* \sim {\rm Gam}(\sum_{j=1}^k a_j, \theta)$, its natural associated confidence distribution is

$$C^*(\theta, y) = G\Big(\theta\sum_{j=1}^k y_j, \sum_{j=1}^k a_j, 1\Big).$$

This may also be seen as coming out of the II-CC-FF schema. By the optimality theorem briefly explained above, the C^* method outperforms $\bar C$ and all other competitors, in terms of all risk functions of the indicated type. With $\Gamma(u) = |u|$ in (11), for example, so that performance is measured by the smallness of ${\rm E}_\theta|\theta_{\rm cd} - \theta|$, the improvement over the back-and-forth transformation method can amount to e.g. 20%, for smaller values of the a_j. For larger values of these a_j the risk difference is small. ■

4.3 Focused fusion via profiling

Suppose now that the II and CC steps have been successfully carried out, leading to confidence log-likelihood contributions $\ell_{c,j}(\psi_j)$ from information sources $j = 1, \ldots, k$. Assuming or taking these to be independent, the overall confidence log-likelihood function is

$$\ell_c^*(\psi_1, \ldots, \psi_k) = \sum_{j=1}^k \ell_{c,j}(\psi_j).$$

When focused inference is wished for, for a focus parameter $\phi = \phi(\psi_1, \ldots, \psi_k)$, the natural way forward is, again, via profiling:

$$\ell_{c,{\rm prof}}^*(\phi) = \max\{\ell_c^*(\psi_1, \ldots, \psi_k) \colon \phi(\psi_1, \ldots, \psi_k) = \phi\}.$$

By the Wilks theorem directly, or by variations of the arguments and details used to prove such theorems (cf. Schweder & Hjort (2016, Appendix)), the overall deviance function

$$D^*(\phi) = 2\{\max \ell_{c,{\rm prof}}^*(\phi) - \ell_{c,{\rm prof}}^*(\phi)\}$$

tends to a χ_1^2 with increasing information volume, at the true parameter values. Hence

$$\mathrm{cc}^*(\phi) = \Gamma_1(D^*(\phi))$$

is the outcome of the three-step II-CC-FF machine, a confidence curve for the focus parameter. Various fine-tuning techniques may be applied to improve on this first-order approximation method; cf. Schweder & Hjort (2016, Chs. 7, 8). In situations where the ψ_j represent the same focus parameter, common across sources, the scheme above simplifies.

5. Applications

Below we illustrate the capacity for the II-CC-FF paradigm to solve problems in rather different application settings. We emphasise that its scope of applicability is broader, also when it comes to combining 'hard' with 'soft' data and with partial expert opinions, and aim at demonstrating this in an upcoming journal paper.

Table 1: Skulls: For each of the five time epochs, the table gives the estimate $\widehat{\psi}$ and its estimated standard deviation $\widehat{\sigma}$. See Section 5.1 and Figure 1.

epoch	$\widehat{\psi}$	$\widehat{\sigma}$
−4000	2.652	0.562
−3300	2.117	0.442
−1850	1.564	0.337
−200	2.914	0.621
150	1.764	0.374

5.1 Skullometrics

In their fascinating anthropometrical study of the inhabitants of Upper Egypt, from the earliest prehistoric times to the Mohammedan Conquest, Thomson & Randall-Maciver (1905) report on skull measurements for more than a thousand crania. A subset of their data is reported on and analysed in Claeskens & Hjort (2008, Chs. 1 and 9), see in particular their Figures 1.1 and 9.1. This pertains to four cranium measurements, say $y = (y_1, y_2, y_3, y_4)^{\mathrm{t}}$, for 30 skulls, from each of five time Egyptian epochs, corresponding to $-4000, -3300, -1850, -200, 150$ on our A.D. scale. We model these vectors as

$$Y_{t,j} \sim \mathrm{N}_4(\xi_t, \Sigma_t) \quad \text{for } j = 1, \ldots, n_t,$$

for each of the five epochs t. There is a variety of parameters worth recording and analysing, where the emphasis is on identifying the necessarily small changes over time; see also Schweder & Hjort (2016, Example 3.10). One might add that such questions, pertaining to the anthropometric evolution over millennia, also touching the demographic history of emigration and immigration in ancient Egypt, do not touch the first or second waves of controversy in the wake of Gould (1981). For the present illustration we choose to focus on the variance matrices, not the means, and consider

$$\psi = \{\max \mathrm{eigen}(\Sigma)\}^{1/2}/\{\min \mathrm{eigen}(\Sigma)\}^{1/2},$$

the ratio of the largest root-eigenvalue to the smallest root-eigenvalue of the variance matrix of the four skull measurements. This is the ratio of the largest to the smallest standard deviations of linear combinations a^tY of the four skull measurements, normalised to have coefficient vector length $\|a\| = 1$. This parameter is one of several natural measures of degree to which the skull distribution is 'stretched'. The question is whether the ψ parameter has changed over time. We assess the degree of change, if any, via the spread parameter τ in the natural model taking $\psi_1, \ldots, \psi_5 \sim \mathrm{N}(\psi_0, \tau^2)$. Rather than merely providing a test of the implied hypothesis $H_0\colon \psi_1 = \cdots = \psi_5$, which is equivalent to $\tau = 0$, with its inevitable p-value and a yes-no answer as with a traditional one-way layout type test, we aim at giving a full confidence distribution for τ, again applying the II-CC-FF scheme.

Table 1 gives point estimates

$$\widehat{\psi}_t = \{\max \mathrm{eigen}(\widehat{\Sigma}_t)\}^{1/2} / \{\min \mathrm{eigen}(\widehat{\Sigma}_t)\}^{1/2}$$

for the five time epochs, along with estimated standard deviations σ_t for these estimators, the latter obtained via bootstrapping from the estimated multinormal distributions. For our present purposes the underlying distributions for the estimators are approximately normal, with the standard deviations σ_t approximately known. Figure 1 displays point estimates with 0.90 confidence intervals (left panel), for the five epochs. The log-likelihood for these five estimates, under the implied $\mathrm{N}(\psi_0, \sigma_j^2 + \tau^2)$ model, writing k for the number of data sources involved, is

$$\ell(\psi_0, \tau) = -\tfrac{1}{2} \sum_{t=1}^{k} \Big\{ \log(\sigma_t^2 + \tau^2) + \frac{(\widehat{\psi}_t - \psi_0)^2}{\sigma_t^2 + \tau^2} \Big\}.$$

The ensuing profiled log-likelihood is

$$\ell_{\mathrm{prof}}(\tau) = -\tfrac{1}{2} \sum_{t=1}^{k} \Big[\log(\sigma_t^2 + \tau^2) + \frac{\{\widehat{\psi}_t - \widetilde{\psi}_0(\tau)\}^2}{\sigma_t^2 + \tau^2} \Big], \text{ with } \widetilde{\psi}(\tau) = \frac{\sum_{t=1}^{k} \widehat{\psi}_t / (\sigma_t^2 + \tau^2)}{\sum_{t=1}^{k} 1/(\sigma_t^2 + \tau^2)}.$$

A confidence distribution for τ can be based on this, but a simpler and powerful alternative is to use

$$Q(\tau) = \sum_{t=1}^{k} \frac{\{\widehat{\psi}_t - \widetilde{\psi}(\tau)\}^2}{\sigma_t^2 + \tau^2} \quad \text{and} \quad C(\tau) = 1 - \Gamma_{k-1}(Q(\tau)),$$

the point being that $Q(\tau)$ for a given true value of τ has the χ^2_{k-1} distribution; see Schweder & Hjort (2016, Ch. 13). This confidence distribution is shown in the right panel of Figure 1, with a confidence point mass $C(0) = 0.221$ at zero. This is actually also a p-value for the hypothesis of equal means, and here not small enough to warrant a claim that this particular ψ parameter has changed over the four thousand years of Egyptian history – other skullometric parameters have however changed; see Claeskens & Hjort (2008, Section 9.1) and Schweder & Hjort (2016, Example 3.5). A 0.95 interval for τ, also indicated in the figure, is $[0, 1.266]$, and the median confidence estimate is 0.392.

5.2 Meta-analysis of two-by-two tables based on incomplete information

As mentioned earlier, the II-CC-FF paradigm covers many existing meta-analysis methods as special cases. In the case of meta-analysis of 2×2 tables, Schweder & Hjort (2016, Chs. 5, 13) provides optimal confidence distribution for inference about a fixed odds ratio parameter, both when the event counts are modelled as binomial pairs and as Poisson pairs. This partly involves the use of (10), via appropriate conditional distributions; see also Schweder & Hjort (2013a) and Cunen & Hjort (2015) for more details and further discussion. These optimal solutions can indeed be presented within the II-CC-FF framework.

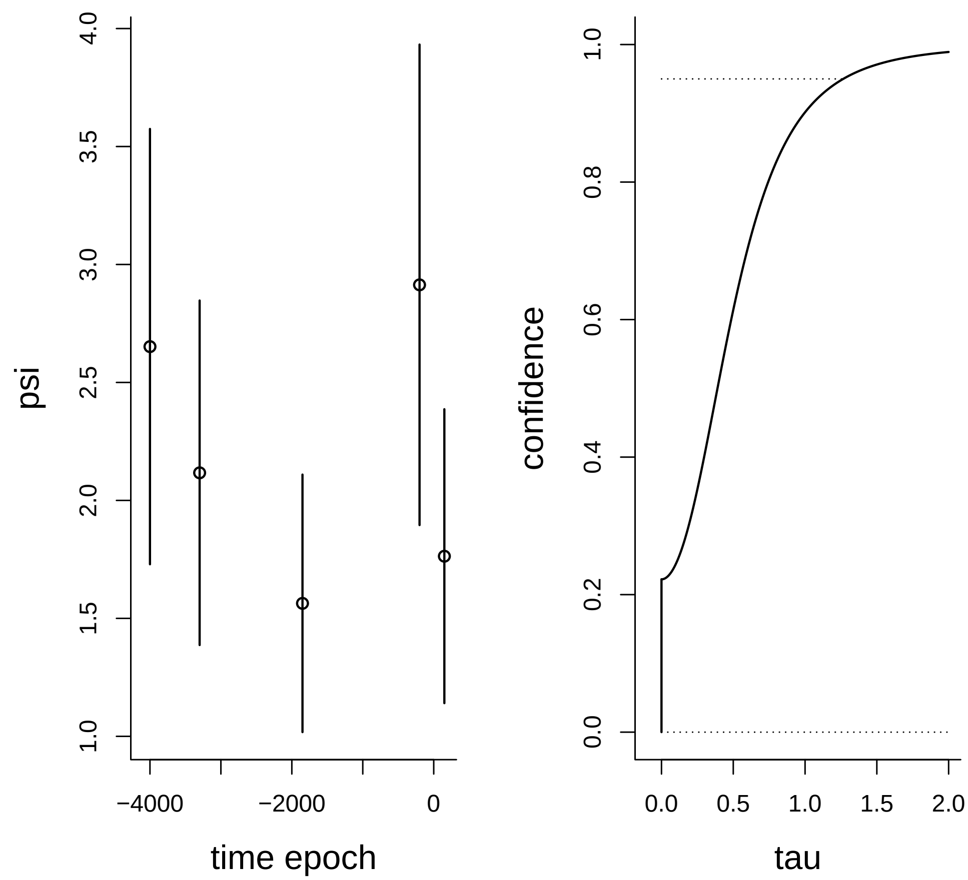

Figure 1: Left panel: Point estimates $\widehat{\psi}_t$ with 90% confidence intervals, for the skull stretch parameter ψ, across five time epochs. Right panel: Confidence distribution for the variability parameter τ. See Section 5.1 and Table 1.

For the following application we will demonstrate that the II-CC-FF paradigm can provide a near-optimal solution, even if the analysis is only based on incomplete information. Normand (1999) provides a medical dataset with six studies investigating death rates among heart attack patients. The treatment group received the drug Lidocaine and the control group did not, and the number of deaths recorded in each group. Thus there are binomial pairs

$$y_{i,0} \sim \mathrm{binom}(m_{i,0}, p_{i,0}) \quad \text{and} \quad y_{i,1} \sim \mathrm{binom}(m_{i,1}, p_{i,1}),$$

with subscript '1' indicating treatment and '0' control, see Table 2. Since the probabilities are small we work here with the Poisson rate model version, where the death counts are seen as Poisson with rates $e_{i,0}\lambda_{i,0}$ and $e_{i,1}\lambda_{i,1}$, with exposure numbers $e_{i,0}$ and $e_{i,1}$ proportional to (or equal to) sample sizes. Our model takes

$$y_{i,0} \sim \mathrm{Pois}(e_{i,0}\lambda_{i,0}) \text{ and } y_{i,1} \sim \mathrm{Pois}(e_{i,1}\lambda_{i,1}), \quad \text{with} \quad \lambda_{i,1} = \gamma\lambda_{i,0}.$$

The model has $k + 1$ parameters, with $k = 6$ the number of studies being combined. Here γ is the focus parameter of interest, associated with the potential risk factor, the degree to

Table 2: Lidocaine data: Death rates for two groups of acute myocardial infarction patients, in six independent studies; see Section 5.2 and Figure 2.

m_1	m_0	y_1	y_0
39	43	2	1
44	44	4	4
107	110	6	4
103	100	7	5
110	106	7	3
154	146	11	4

which Lidocaine leads to higher risk for this population of myocardial infarction patients. Cunen & Hjort (2015) constructed the optimal confidence distribution for this common odds ratio parameter γ. In the present demonstration of the II-CC-FF scheme, let us assume that the data from the Lidocaine studies are not available and that we are only given the six different confidence curves for the odds ratio parameters in each of the studies (the coloured curves in Figure 2). We thus have incomplete information, since we behave as if we neither have the raw data of Table 2, nor information on how the confidence curves were computed. We will nonetheless use the II-CC-FF paradigm to combine the six confidence curves.

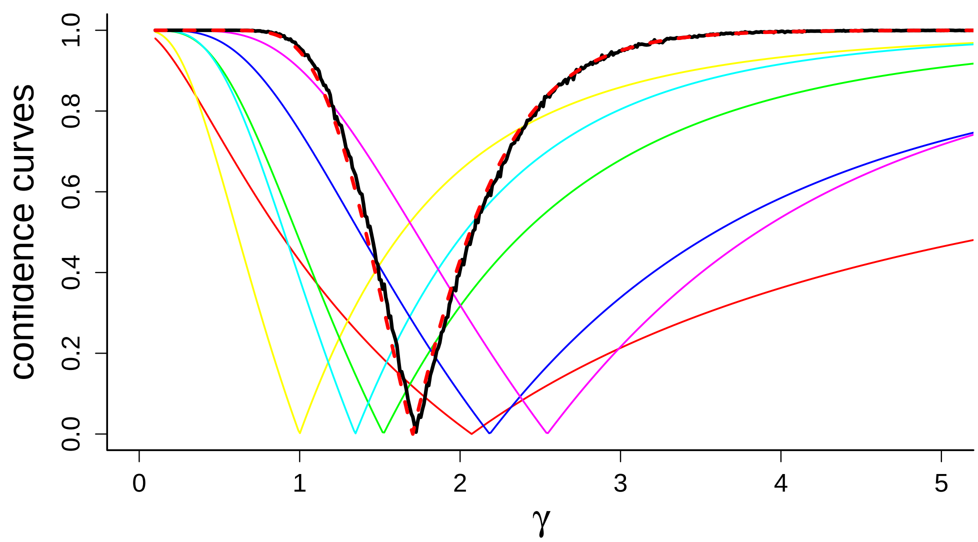

Figure 2: The coloured curves are the confidence curves for the odds ratio from each of the six studies. The thick black curve is the optimal combined confidence curve, while the red dashed curve is the combined confidence curve based on incomplete information. See Section 5.2 and Table 2.

The first step, II, is already taken care of, as we have the six confidence curves. The CC step is more interesting, however. Since we assume or pretend that we do not know how the confidence curves were constructed, we need to obtain the confidence log-likelihoods by approximate methods. Here we call on the chi-squared inversion method, as per Section 3, leading to

$$\text{CC:} \qquad \ell_{c,j}(\gamma) = -\tfrac{1}{2}\Gamma_1^{-1}(\mathrm{cc}_j(\gamma)) \quad \text{for } j = 1, \ldots, k,$$

where $\mathrm{cc}_j(\gamma)$ is the confidence curve from study j and Γ_1^{-1} is the quantile function of the χ_1^2 distribution. For the FF step, we sum the log-likelihoods, find the combined deviance function and apply the Wilks theorem, as per Section 4:

$$\begin{aligned} \text{FF:} \qquad \ell_c^*(\gamma) &= \sum_{j=1}^{6} \ell_{c,j}(\gamma), \\ D^*(\gamma) &= 2\{\ell_c^*(\widehat{\gamma}) - \ell_c^*(\gamma)\}, \\ \mathrm{cc}^*(\gamma) &= \Gamma_1(D(\gamma)). \end{aligned}$$

The resulting confidence curve closely matches the optimal confidence curve for γ that was obtained in Cunen & Hjort (2015), as seen in Figure 2.

5.3 Abundance of humpback whales

The II-CC-FF paradigm readily lends itself to combination of information from published sources, where we may not have access to the full data, but only summary measures. Paxton et al. (2009) provide estimates of the abundance of humpback whales in the North Atlantic in the years 1995 and 2001. The two estimates are based on different surveys and can be considered independent. The authors also provide 95% confidence intervals, via a somewhat complicated model involving aggregation of line transect data from different areas via spatial smoothing, and also involves bootstrapping. The available information is as presented in Table 3; note here that the natural 95% confidence interval is not at all symmetric around the point estimate, with an implied skewness to the right.

Table 3: Abundance assessment of a humpback population, from 1995 and 2001, summarised as 2.5%, 50%, 97.5% confidence quantiles; from Paxton et al. (2009). See Section 5.3 and Figure 3.

	2.5%	50%	97.5%
1995	3439	9810	21457
2001	6651	11319	21214

For this illustration we shall assume that the underlying true abundance underlying these two studies has remained constant, with population size ψ. Under this assumption we wish to combine the two surveys in order to get a more precise point estimate, along with confidence statements. The first step, *Independent Inspection*, requires us to construct confidence distributions for ψ from each of the two surveys. In Schweder & Hjort (2016, Ch. 10), certain methods are proposed and developed for constructing confidence distributions based only on an estimate and a confidence interval. With a positive parameter, like abundance, one may use

$$\text{II:} \qquad C(\psi, y) = \Phi\Big(\frac{h(\psi) - h(\widehat{\psi})}{s}\Big)$$

with a power transformation $h(\psi, a) = \mathrm{sgn}(a)\psi^a$; see also Schweder & Hjort (2013b) for some more discussion of this approach (along with a different application, essentially also using the II-CC-FF paradigm). In order to estimate the power a and the scale s the following two equations must be solved,

$$\psi_1^a - \widehat{\psi}^a = -1.96\, s \quad \text{and} \quad \psi_2^a - \widehat{\psi}^a = 1.96\, s,$$

where $[\psi_1, \psi_2]$ is the 95% confidence interval and $\widehat{\psi}$ the median confidence point estimate. For the whale abundance, we find (a, s) equal to $(0.321, 2.798)$ for 1995 and $(0.019, 0.007)$ for 2001 (a small value of a indicates that the transformation is nearly logarithmic). The corresponding confidence curves are shown in Figure 3. In this case the confidence log-likelihoods in the *Confidence Conversion* step are easily obtained. For year j,

$$\text{CC:} \qquad \ell_{c,j}(\psi) = -\tfrac{1}{2}\{h_j(\psi) - h_j(\widehat{\psi})\}^2/s_j^2.$$

In the final *Focused Fusion* step, we sum the two confidence log-likelihoods, find the combined deviance function and construct an approximative combined confidence curve by the Wilks theorem, as per Section 2:

$$\text{FF:} \quad \ell_c^*(\psi) = \ell_{c,1}(\psi) + \ell_{c,2}(\psi), \quad D(\psi) = 2\{\ell_c^*(\widehat{\psi}_c) - \ell_c^*(\psi)\}, \quad \mathrm{cc}(\psi) = \Gamma_1(D(\psi)),$$

where $\widehat{\psi}_c$ is the combined maximum likelihood estimate. Here we obtain the red curve in Figure 3, with $\widehat{\psi}_c = 10847$ and a 95% confidence interval $[6692, 17156]$.

For this illustration our parameter of interest was ψ, the assumed common abundance assessed and estimated in the two surveys. Alternatively we could consider the two surveys as aiming for different parameters ψ_1 and ψ_2 and then define our parameter of interest to be for example the difference or the ratio between the two abundances. This is not a difficult exercise, using the general II-CC-FF schema, but we do not pursue this here. Instead we point to the following illustration, where the II-CC-FF is used for a similar problem.

5.4 Ratio of standard deviations

The application discussed above illustrates that we can combine information from different sources concerned with estimating a common parameter ψ. The II-CC-FF paradigm also lends itself to combination across information sources in a broader sense, where the different sources inform about different parameters $\psi_1, \ldots, \psi_k$ and where we are interested in a function of these parameters, say $\phi = \phi(\psi_1, \ldots, \psi_k)$. For the following illustration the parameter of interest is $\rho = \sigma_2/\sigma_1$, the ratio between two standard deviations, where one study informs us about σ_1 and another about σ_2.

Our II-CC-FF setup is very general, and can be applied with appropriate work and fine-tuning to handle various variations over the same theme, such as more difficult background information and more complex models behind the two assessments of σ_1 and σ_2. It might e.g. be the case that one of the estimates is more precise than the other, that one of them stems from simple direct data and the other from data more indirectly and less clearly yielding confidence for the relevant parameter. For the present illustration we are content to show how the story pans out in a simple enough setting, where the first source is an independent normal sample with $n_1 = 4$ and empirical standard deviation $\widehat{\sigma}_1 = 0.50$, and the second source is a similar sample but with $n_2 = 10$ and $\widehat{\sigma}_2 = 3.00$. With such i.i.d. normal data the confidence curve for the standard deviation from data source j is based on the natural pivot $\widehat{\sigma}_j^2/\sigma_j^2 \sim \chi^2_{\nu_j}/\nu_j$, which leads to

$$\text{II:} \qquad C_j(\sigma_j) = 1 - \Gamma_{\nu_j}(\nu_j \widehat{\sigma}_j^2/\sigma_j^2),$$

where $\nu_j = n_j - 1$ and Γ_{ν_j} is the c.d.f. of the χ^2 with ν_j degrees of freedom. See the two confidence curves for the standard deviations in the left panel of Figure 4. In general, for

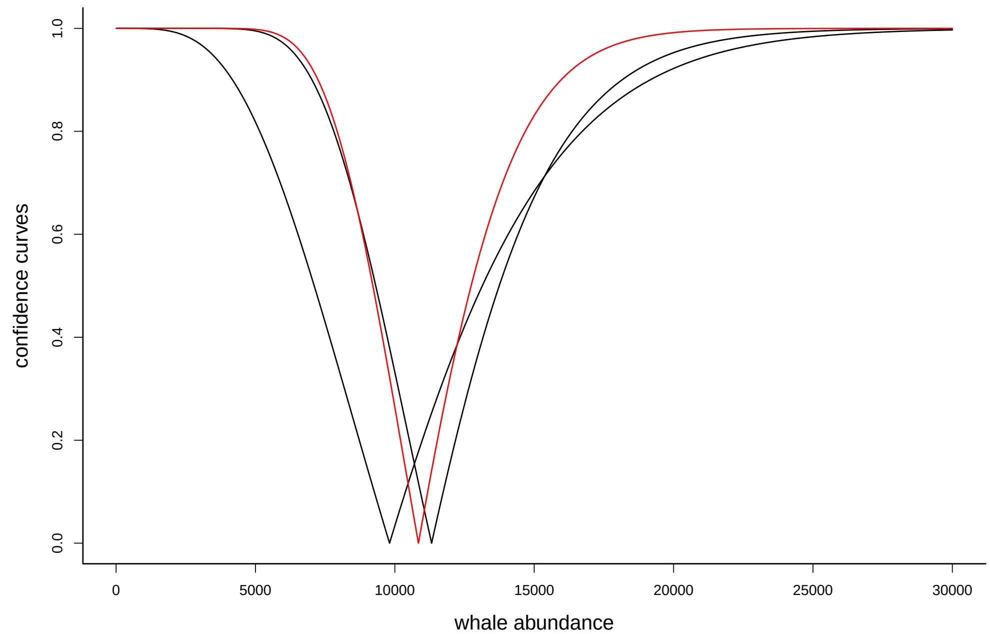

Figure 3: Confidence curves for ψ, the abundance of humpback whales in the North Atlantic. The two black curves represent the inference from each of the two surveys we consider (one in 1995 and another in 2001), while the red curve is the combined confidence curve (under the assumption of a constant abundance). See Section 5.3 and Table 3.

a confidence distribution based on a pivot, say $\mathrm{piv}(y,\psi)$ with density f, the corresponding confidence likelihood is given as $L_c(\psi, y) = f(\mathrm{piv}(y,\psi))\,|\partial \mathrm{piv}(y,\psi)/\partial y|$ (Schweder & Hjort, 2016, Ch. 10). In this case we get

$$\text{CC:} \qquad \ell_{c,j}(\sigma_j) = \tfrac{1}{2}\nu_j\{\log(\widehat\sigma_j^2/\sigma_j^2) - \widehat\sigma_j^2/\sigma_j^2\}.$$

The final FF step requires the profile log-likelihood for ρ, and again the approximate χ_1^2 distribution of the deviance function, as per Section 3,

$$\begin{aligned} \text{FF:} \qquad \ell^*_{c,\mathrm{prof}}(\rho) &= \max\{\ell_{c,1}(\sigma_1) + \ell_{c,2}(\sigma_2) \colon \sigma_2/\sigma_1 = \rho\} = \max_{\sigma_1}\{\ell_{c,1}(\sigma_1) + \ell_{c,2}(\sigma_1\rho)\}, \\ D^*(\rho) &= 2\{\ell_{\mathrm{prof}}(\widehat\rho) - \ell_{\mathrm{prof}}(\rho)\}, \\ \mathrm{cc}^*(\rho) &= \Gamma_1(D^*(\rho)). \end{aligned}$$

The confidence curve for the ratio ρ is displayed as the black full curve in the right panel of Figure 4.

In some cases there may exist some expert knowledge pertaining to at least the focus parameter under study, here $\rho = \sigma_2/\sigma_1$, though not necessarily for the full parameter vector of the combined models, here (σ_1, σ_2). A proper Bayesian analysis requires the statistician to have such a prior for (σ_1, σ_2) – without this ingredient, there is no Bayes theorem leading to a posterior distribution for the model parameters, or indeed for ρ. The II-CC-FF scheme allows however incorporation of such partial prior information, i.e. a prior for ρ without a prior for (σ_1, σ_2). For this illustration we assume that experts provide a Gamma prior with expectation equal to 14.0 and variance 6.0^2. The prior can be represented as a

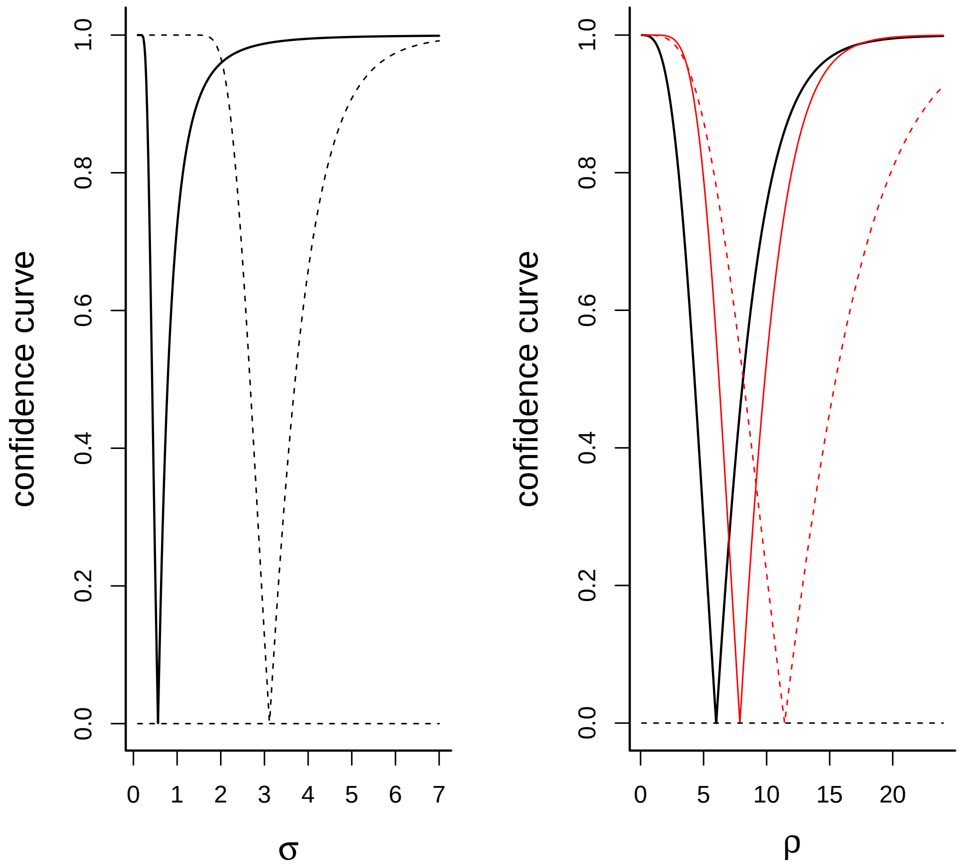

Figure 4: Left panel: confidence curves for σ_1 and σ_2. Right panel: the confidence curve for $\rho = \sigma_2/\sigma_1$ based on the two studies (black curve); the confidence curve based on prior information alone (red dotted curve); and the confidence curve combining the studies and the prior information (red full-drawn curve). See Section 5.4.

confidence curve, supplementing the confidence curve based on the two studies. In order to fuse the prior knowledge and the data we simply add the prior log-likelihood $\ell_B(\rho)$ to the confidence log-likelihoods, in the following way,

$$\text{FF:} \quad \begin{aligned} \ell^*_{\text{prof},B}(\rho) &= \max\{\ell_{c,1}(\sigma_1) + \ell_{c,2}(\sigma_2) + \ell_B(\rho) \colon \sigma_2/\sigma_1 = \rho\} \\ &= \max_{\sigma_1}\{\ell_{c,1}(\sigma_1) + \ell_{c,2}(\sigma_2) \colon \sigma_2/\sigma_1 = \rho\} + \ell_B(\rho) \\ &= \ell^*_{c,\text{prof}}(\rho) + \ell_B(\rho). \end{aligned}$$

We use 'B' as subscript to indicate the in this instance partial and perhaps lazy Bayesian, who does not give a full prior for the model parameters, but contributes a component, namely where it matters the most, about the focus parameter. Of course the log-prior $\ell_B(\rho)$ employed here could have been obtained in the more careful and proper Bayesian way of having started with a full prior for (σ_1, σ_2), and then a transformation, but we do suggest that expert knowledge concerning focus parameters is more often put forward directly, not via the full parameter vector in the fullest model.

Importantly, this extended deviance function does still have an approximate χ^2_1 distribution, by the general approximation arguments involved in the Wilks theorem, unless the log-prior $\ell_B(\rho)$ is sharp and distinctly non-normal. One may conceptually and sometimes practically interpret the log-prior as having resulted from real data in previous experiences, in which case the $\ell_B(\rho)$ would be a genuine profiled log-profile likelihood function from such an information source. Also, as the sample sizes of the studies increase the information from the two studies will dominate the prior and we can safely continue to use the Wilks theorem. As expected, the confidence curve fusing the prior information and the information from the two studies lies between the original confidence curve and the prior confidence curve (see Figure 4). It is also somewhat narrower than both.

6. Discussion

We have proposed, developed and demonstrated the broad usefulness of the II-CC-FF paradigm for combining information across diverse sources. We intend to develop certain types of applications further, in situations where some of the three analysis steps are harder, and shall report on this in future publications.

We note that versions of the II-CC-FF scheme, though not necessarily in three full steps, have been used earlier, in various guises. Generally speaking, a fair portion of meta-analyses can be cast in II-CC-FF terms, as indicated in our introduction, and in Schweder & Hjort (2016, Ch. 13). The challenge is to exploit and employ the general line of thinking in more complicated situations. An early demonstration is in the field of whale abundance assessment, in Schweder & Hjort (1996), where other methods, including those pertaining to Bayesian melding, were shown to have pitfalls; see also Schweder & Hjort (2002).

There is manoeuvring room for variations, fine-tuning and improvement inside the II-CC-FF scheme, as is also clear from the four applications discussed above. Since several substeps must be decided on by the context of the given problem, we do not yet foresee a general automated II-CC-FF package, though making certain specialised versions could be contemplated. This will also be considered in work to come.

Acknowledgements. The work reported on here has been partially funded via the Norwegian Research Council's five-year project *FocuStat: Focused Statistical Inference With Complex Data* (2014–2018), led by Hjort. The authors are grateful to Tore Schweder for always fruitful discussions related to issues and methods worked with in this paper.

References

CLAESKENS, G. & HJORT, N. L. (2008). *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.

CUNEN, C. M. L. & HJORT, N. L. (2015). Optimal inference via confidence distributions for two-by-two tables modelled as poisson pairs: Fixed and random effects. In *Proceedings 60th World Statistics Congress, 26-31 July 2015, Rio de Janeiro*, vol. I. Amsterdam: International Statistical Institute, pp. 3581–3586.

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

EFRON, B. & HASTIE, T. (2016). *Computer Age Statistical Inference*. Cambridge: Cambridge University Press.

GOULD, S. J. (1981). *The Mismeasure of Man*. New York: W.W. Norton & Company.

LIU, D., LIU, R. Y. & XIE, M.-G. (2014). Exact meta-analysis approach for discrete data and its application to 2×2 tables with rare events. *Journal of the American Statistical Association* **109**, 1450–1465.

LIU, D., LIU, R. Y. & XIE, M.-G. (2015). Multivariate meta-analysis of heterogeneous studies using only summary statistics: Efficiency and robustness. *Journal of the American Statistical Association* **110**, 326–340.

NORMAND, S.-L. T. (1999). Tutorial in biostatistics meta-analysis: Formulating, evaluating, combining, and reporting. *Statistics in Medicine* **18**, 321–359.

PAXTON, C. G. M., BURT, M. L., HEDLEY, S. L., VÍKINGSSON, G. A., GUNNLAUGSSON, T. & DESPORTES, G. (2009). Density surface fitting to estimate the abundance of Humpback whales based on the NASS-95 and NASS-2001 aerial and shipboard surveys. *NAMMCO Scientific Publications* **7**, 143–160.

SCHWEDER, T. & HJORT, N. L. (1996). Bayesian synthesis or likelihood synthesis – what does Borel's paradox say? *Reports of the Internaional Whaling Commision* **46**, 475–479.

SCHWEDER, T. & HJORT, N. L. (2002). Likelihood and confidence. *Scandinavian Journal of Statistics* **29**, 309–322.

SCHWEDER, T. & HJORT, N. L. (2013a). Discussion contribution to the Xie and Singh paper. *International Statistical Review* **81**, 56–68.

SCHWEDER, T. & HJORT, N. L. (2013b). Integrating confidence intervals, likelihoods and confidence distributions. In *Proceedings 59th World Statistics Congress, 25-30 August 2013, Hong Kong*, vol. I. Amsterdam: International Statistical Institute, pp. 277–282.

SCHWEDER, T. & HJORT, N. L. (2016). *Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions*. Cambridge: Cambridge University Press.

SIMPSON, R. J. S. & PEARSON, K. (1904). Report on certain enteric fever inoculation statistics. *The British Medical Journal* **3**, 1243–1246.

SINGH, K., XIE, M.-G. & STRAWDERMAN, W. E. (2005). Combining information from independent sources through confidence distributions. *Annals of Statistics* **33**, 159–183.

THOMSON, A. & RANDALL-MACIVER, R. (1905). *Ancient Races of the Thebaid: Being an Anthropometrical Study of the Inhabitants of Upper Egypt from the Earliest Prehistoric Times to the Mohammedan Conquest, Based Upon the Examination of Over 1,500 Crania*. Oxford: Oxford University Press.

XIE, M.-G. & SINGH, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review [with discussion and a rejoinder]. *International Statistical Review* **81**, 3–39.