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

September 2018

Céline Cunen and Nils Lid Hjort

[1]**Department of Mathematics, University of Oslo**

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 $\ell_{\mathrm{conv},j}(\psi_j)$. 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, hard and soft data, 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

$$\widehat{\psi} = \frac{\sum_{j=1}^k y_j/\sigma_j^2}{\sum_{j=1}^k 1/\sigma_j^2} \sim \mathrm{N}\Big(\psi, \Big(\sum_{j=1}^k 1/\sigma_j^2\Big)^{-1}\Big). \tag{1.1}$$

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 $\mathrm{N}(\psi_0, \tau^2)$ distribution. For this type of random effects model, one wishes clear inference strategies for both the overall mean ψ_0 and level of variation τ. We return to this particular problem in Sections 6.1 and 7.2.

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, 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. For our development of II-CC-FF we have attempted to think fundamentally and generally about combination of information problems. Our framework encompasses known meta-analysis methods, but we aim at tackling new and more challenging problems as well. Parts of the meta-analysis literature are quite narrow, with specific methods for specific problems. In that light we hope our more general approach will be useful.

In reasonably general terms, assume there is a parameter ϕ of clear interest, related to parameters $\psi_1, \ldots, \psi_k$, either via a deterministic function $\phi = \phi(\psi_1, \ldots, \psi_k)$ or via some type of random effect distribution, where such a ϕ might be a parameter related to a background distribution of the ψ_j. Suppose further that data source y_j provides information pertaining to ψ_j. For the sake of clear presentation, we let the ψ_j be one-dimensional here; more general cases are considered in Sections 5 and 7. 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 $C_j(\psi_j, y_j)$ 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 $\ell_{\mathrm{conv},j}(\psi_j)$.
- ◇ FF, *Focused Fusion*: In the fixed effect case, the combined confidence log-likelihood function $\ell_{\mathrm{fus}}(\psi_1, \ldots, \psi_k) = \sum_{j=1}^k \ell_{\mathrm{conv},j}(\psi_j)$ is used to reach focused fusion inference for $\phi = \phi(\psi_1, \ldots, \psi_k)$. With random effects, the fusion involves the computation of an integral.

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 Section 5 we provide a standardised version of II-CC-FF, with a generic recipe to follow, but we will see that in many cases one could or should be more careful about the various steps. 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 have to base one's analysis on summary measures, directly or indirectly given via other people's work, reports and publications. The II-CC-FF paradigm is meant to be powerfully applicable in such situations too.

A pertinent question to raise is whether or why there is a need for specific methods for combination of information in the first place; in a suitable sense, all of statistics concerns combination

of information. One might therefore ask why there even exist subfields such as meta-analysis, and specific framework aimed at combination of information such as our own. So isn't meta-analysis just analysis? Two related responses are as follows. (i) Sometimes the full sets of data are not available, with access only to summaries or partial summaries. Issues here are storage, the practicalities of other people's files, privacy concerns, etc. (ii) Sometimes it might be easier, conceptually or practically, to analyse the different sources or studies separately first, and then combine these pieces of summarised information. Also, a statistical prediction is that modern statistics to an increasing degree will be concerned with such issues and challenges, finding and organising bits and pieces of information across different sources, with a need to reach conclusions based on these pieces.

After a motivating illustration, below, 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. In Section 5 we discuss the performance of the combination framework and provide some general guidelines. In Section 6 we explore connections with other approaches, notably well-established meta-analysis methods, and also other methods based on confidence distributions. The three step II-CC-FF machinery is then seen in action through four applications laid out in Section 7.

Motivating illustration

For concreteness we will start off with a meta-analysis example, and present the three steps of the II-CC-FF paradigm at work, but with a minimum amount of details. A non-standard feature of this example is that different studies of the same statistical question have reported different summary measures. Six studies have reported summary statistics based on continuous outcomes, while five other studies reported summaries based on a binary outcome. Also note that the full data of all the studies were not available. The data employed here were first analysed in Whitehead et al. (1999); related problems have been treated in Dominici & Parmigiani (2000) and Liu et al. (2015).

We have eleven randomised trials investigating the use of oxytocic drugs during labour and its potential effect on postpartum blood loss. Each study has two groups of patients, a treatment group receiving oxytocic drug and a control group receiving no drugs of that type. Taking $y_{i,j}$ to be the blood loss for patient i in study j, we may use the simple model $y_{i,j} = \alpha_j + \beta z_{i,j} + \varepsilon_{i,j}$, with the $\varepsilon_{i,j}$ independent and $\mathrm{N}(0, \sigma^2)$, and with $z_{i,j}$ an indicator variable, equal to 0 for patients in the control group and 1 for patients in the treatment group. Here β is the treatment effect and the parameter of main interest.

There are six trials of type A, say, reporting continuous outcomes, and five trials of type B, reporting only binary outcomes relative to a threshold. For the six type A trials, we have the mean and the empirical standard deviation of the blood loss in the two groups of patients. With the simple normal model above, these four summary statistics are sufficient for each trial, and we thus have access to the full log-likelihood $\ell_{A,j}(\beta, \alpha_j, \sigma)$ for each continuous trial j. For the five type B trials, however, we merely have counts of the number of patients in each group having a blood loss of more or less than 500 ml. These numbers constitute a non-sufficient summary; we thus have less information in these studies compared to the continuous ones, and log-likelihood functions not able to inform on β directly, only β/σ. More specifically, based on the normal model above, we

obtain a probit-type log-likelihood for these binary trials, say $\ell_{B,j}(\theta, \gamma_j)$, with $\gamma_j = (500 - \alpha_j)/\sigma$ and $\theta = \beta/\sigma$.

Having made these modelling assumptions, the steps in the II-CC-FF recipe follow straightforwardly. Using the log-likelihood functions described above, we can, by methods described in the next section, construct *confidence curves* for the parameter of interest for each of the studies. From the type A studies we construct $\text{cc}_{A,j}(\beta)$, and from the type B studies $\text{cc}_{B,j}(\theta)$; see the dotted and dashed curves in Figure 1.1. There, we have plugged in the estimated σ from the continuous studies, in order to be able to display the inference from the binary studies on the β scale (rather than on the $\theta = \beta/\sigma$ scale).

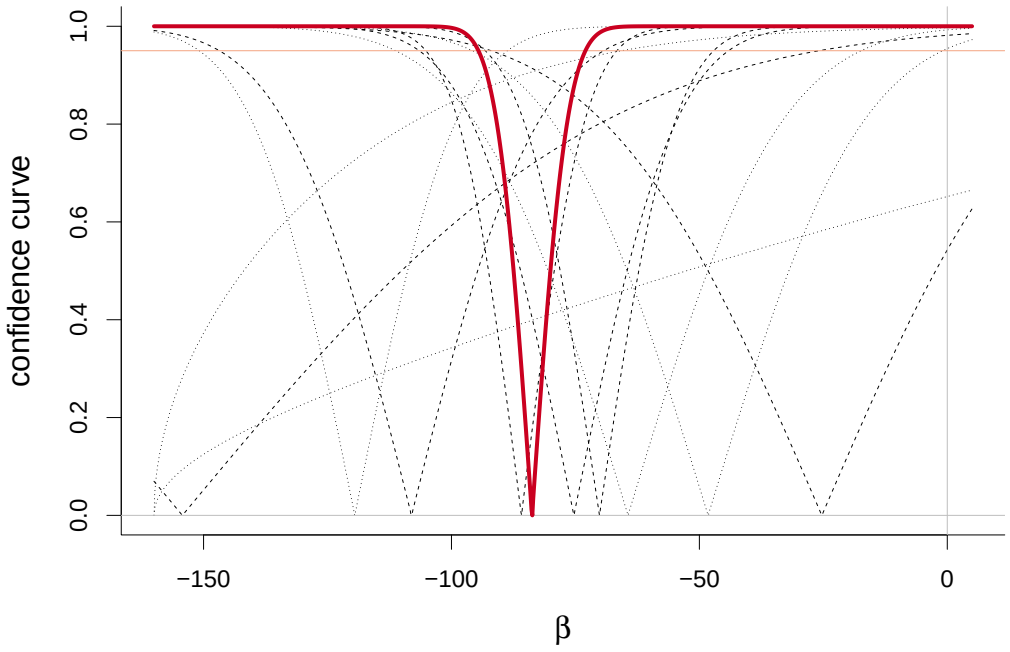

Figure 1.1: Confidence curves for the treatment effect in the 11 trials (dashed lines: continuous studies, dotted lines: binary studies), along with two different combined confidence curve. In red, the confidence curve combining all the 11 studies. The horizontal red line marks the 95% confidence level. The median confidence estimate is −83.7 ml, with 95% interval [−94.4, −73.1].

The CC step is simple in this case, with no extra work required, since the log-likelihood functions were used in the construction of the confidence curves for each study. In other situations we might have to carry out the conversion from confidence statements to log-likelihood functions in ways described in Section 3. Here we have $\ell_{\text{conv},j} = \ell_{A,\text{prof},j}(\beta, \sigma)$ for the continuous studies, and $\ell_{\text{conv},j} = \ell_{B,\text{prof},j}(\beta/\sigma)$ for the binary studies. These profile log-likelihood functions will be explained in the next section. In the FF step we sum the log-likelihood contributions,

$$\text{FF:} \qquad \ell_{\text{fus}}(\beta, \sigma) = \sum_{j=1}^{6} \ell_{A,\text{prof},j}(\beta, \sigma) + \sum_{j=1}^{5} \ell_{B,\text{prof},j}(\beta/\sigma).$$

Further, we profile out σ and obtain the final combined confidence curve by

$$\text{cc}^*(\beta, \text{all data}) = \Gamma_1\big(2\{\max \ell_{\text{fus}}(\beta, \widehat{\sigma}(\beta)) - \ell_{\text{fus}}(\beta, \widehat{\sigma}(\beta))\}\big),$$

with $\Gamma_1(\cdot)$ the c.d.f. of a χ^2_1. In Figure 1.1, the thick red curve is this combined confidence curve.

It is clearly narrower than all the individual curves and placed roughly in the middle of them, as we would expect.

The combined inference clearly indicates that oxytocic drugs reduce postpartum blood loss, which is in agreement with the conclusions in Whitehead et al. (1999). Here we have zoomed in on β as the focus parameter, to pinpoint precisely how much the two groups differ in blood loss. For clinicians it might be of more direct interest to consider the probabilities for having a postpartum blood loss greater than a threshold, like 500 ml, for the two groups, and then focus on the odds ratio, say ρ. Our approach can easily accomodate such an analysis too, with ρ rather than β in the FF step, yielding a figure similar to Figure 1.1, but now for ρ.

2 Independent Inspection: confidence distributions

Suppose Y denotes a set of random observations, stemming from a model with parameter θ, typically multidimensional, and with $\psi = \psi(\theta)$. For the ease of presentation, we let ψ_j be a one-dimensional focus parameter for now, but in general combination situations it will typically be multidimensional. 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

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

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

$$\mathrm{cc}(\psi, y_{\rm obs}) = |1 - 2\,C(\psi, y_{\rm obs})|. \tag{2.1}$$

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 7. We also note that the random $\mathrm{cc}(\psi, Y)$ has a uniform distribution, at the true position in the parameter space, since $|1 - 2\,U|$ is uniform when U is. Indeed

$$P_{\theta_0}\{\mathrm{cc}(\psi_0, Y) \le \alpha\} = \alpha, \quad \text{for each } \alpha, \tag{2.2}$$

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 $\mathrm{cc}(\psi, Y)$ may be constructed, with a valid (2.2), 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*, 2018 (Hjort & Schweder, 2018). 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 $\widehat{\psi} \sim \mathrm{N}(\psi, \kappa^2)$, 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, $C(\psi, y) = \Phi((\psi - \widehat{\psi})/\widehat{\kappa})$ 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 (ML), say $\widehat{\psi}_{\mathrm{ml}}$, but other estimators are allowed too in this simple construction. The second is based on the profiled log-likelihood function $\ell_{\mathrm{prof}}(\psi) = \max\{\ell(\theta)\colon \psi(\theta) = \psi\}$, which leads to the deviance function

$$D(\psi) = 2\{\ell_{\mathrm{prof}}(\widehat{\psi}_{\mathrm{ml}}) - \ell_{\mathrm{prof}}(\psi)\} = 2\{\ell_{\mathrm{prof,max}} - \ell_{\mathrm{prof}}(\psi)\}. \tag{2.3}$$

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

$$\mathrm{cc}(\psi, y) = \Gamma_1(D(\psi)), \tag{2.4}$$

with $\Gamma_\nu(\cdot)$ denoting the c.d.f. of a χ^2 with degrees of freedom ν. Typically, the second method (2.4) leads to a better calibrated confidence curve than the the simpler method mentioned first. Further fine-tuning methods are developed, illustrated and discussed in Schweder & Hjort (2016, Chs. 7, 8); see also Section 5.3 below.

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 (2.4) above. It consists in using

$$\ell_{\mathrm{conv}}(\psi) = -\tfrac{1}{2}\Gamma_1^{-1}(\mathrm{cc}(\psi, y)) \tag{3.1}$$

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* $\ell_{\mathrm{conv}}(\psi) = -\frac{1}{2}\{\Phi^{-1}(C(\psi, y))\}^2$. A relevant point here is that one often constructs a confidence curve $\mathrm{cc}(\psi, y)$ directly, not always via (2.1), making (3.1) 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 (3.1) 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 ML 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.

For confidence distributions constructed via a one-dimensional statistics T, one may use *exact conversion* to obtain the confidence log-likelihood. When the statistic has a continuous distribution, the exact conversion of the confidence distribution $C(\psi, T)$, see Schweder & Hjort (2016, Ch. 10), is $\ell_{\rm conv}(\psi) = \log|\partial C(\psi, t)/\partial t|$.

4 Focused Fusion: from full likelihood to focus parameter

Suppose now that the II and CC steps have been successfully carried out, leading to confidence log-likelihood contributions $\ell_{{\rm conv},j}(\psi_j)$ from information sources $j = 1, \ldots, k$. Depending on the application and its context we might then be interested in either a fixed effect approach, with the main focus parameter ϕ is a function of the ψ_j, or a random effect approach, where we introduce an additional layer of heterogeneity through a model for the ψ_j. We will treat the fixed effect case first. The Focused Fusion step will then typically be carried out via profiling of the combined confidence log-likelihood. In Section 4.2 we present the II-CC-FF solution for random effect situations.

4.1 Fixed effects fusion

Assuming the information sources to be independent, the overall confidence log-likelihood function is $\ell_{\rm fus}(\psi_1, \ldots, \psi_k) = \sum_{j=1}^k \ell_{{\rm conv},j}(\psi_j)$. 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:

$$\ell_{\rm fus,prof}(\phi) = \max\{\ell_{\rm fus}(\psi_1, \ldots, \psi_k) \colon \phi(\psi_1, \ldots, \psi_k) = \phi\}.$$

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

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

tends, at the true parameter position and with increasing information volume, to a χ^2_1. Here $\widehat{\phi}$ is the ML, maximising the profiled log-likelihood. Hence

$$\mathrm{cc}^*(\phi, \text{all data}) = \Gamma_1(D^*(\phi))$$

is the outcome of the three step II-CC-FF machine, a confidence curve for the focus parameter. In Section 5 we will come back to some discussion on the meaning of 'increasing information volume' in a combination context. Various fine-tuning techniques may be applied to improve on this first-order approximation method; cf. Schweder & Hjort (2016, Chs. 7, 8) and Section 5.3. In situations where the ψ_j represent the same focus parameter, common across sources, the scheme above simplifies.

4.2 Random effects fusion

In our II-CC-FF setting, we use the term 'random effects' when we wish to introduce an extra layer of heterogeneity in the fusion step. This is more easily presented when assuming that $\psi_1, \ldots, \psi_k$ are scalars. In the random effects case we do not assume that all ψ_j are equal but rather that they come from some underlying distribution. In the most canonical case, this distribution will be governed by some overall mean parameter ψ_0 and some spread parameter τ; specifically we could have $\psi_j \sim \mathrm{N}(\psi_0, \tau^2)$. The parameter of main interest may be either the overall mean, or the spread, or perhaps a quantile, depending on the context.

We propose the following general solution for II-CC-FF with random effects. Suppose the ψ_j are modelled as coming from a background density $f(\psi_j, \kappa)$, say, where the κ could be a centre and a spread parameter, as for (ψ_0, τ) in the normal case. Then, using the confidence log-likelihoods $\ell_{\mathrm{conv},j}(\psi_j)$, we define the fusion log-likelihood for source j to be

$$\ell_{\mathrm{fus},j}(\kappa) = \log\Big[\int \exp\{\ell_{\mathrm{conv},j}(\psi_j)\} f(\psi_j, \kappa)\, \mathrm{d}\psi_j\Big]. \tag{4.1}$$

The likelihood contributions from each source are then summed, $\ell_{\mathrm{fus}}(\kappa) = \sum_{j=1}^k \ell_{\mathrm{fus},j}(\kappa)$. We would usually need to profile again, depending on what we are interested in, say the centre ψ_0 or spread τ for the case of a normal model for the ψ_j. To produce our final confidence curve we will often use the Wilks approximation. This II-CC-FF solution requires the computation of integrals. Sometimes numerical integration routines in R work well enough, other times we will make use of the so-called Template Model Builder package (TMB) and its Laplace approximations in order to compute the integral (Kristensen et al., 2016).

5 General guidelines

The overall objective of the II-CC-FF is to construct a valid confidence curve for each parameter ϕ of particular interest, typically of the form $\phi = \phi(\psi_1, \ldots, \psi_k)$, incorporating the relevant information in all the sources. Below we first present a reasonably standardised version of the II-CC-FF scheme. This framework has limitations and should be used with care, however, which we then discuss in the following subsections. These discussions also highlight various important general issues with methods for combination of information. Our default method is based on profiling, and much of the discussion below relates to that tool, including certain modifications.

5.1 Standard II-CC-FF

Our framework opens up many possibilities for tailored and fine-tuned solutions, where these might exist, as demonstrated in the four applications of Section 7. We are however promoting one versatile version of II-CC-FF, which could be called 'standard', and will work for a range of situations. This version may be developed into an R-package. Here we will describe the steps of this II-CC-FF scheme, when we have the full data available, or sufficient summaries, from all sources. The statistical work starts by finessing the statistical issues into one or more parameters of particular interest, involving relevant parameters $\psi_1, \ldots, \psi_k$ from the k sources. These might be parameter vectors (i.e. need not be one-dimensional), they might differ from source to source, but may also contain common parameters across sources.

- ◇ II, *Independent Inspection*: analyse each source j separately. Assume a parametric model for the observations and put up the likelihood function. Profile out the source-specific parameters, and obtain $\ell_{\text{prof},j}(\psi_j)$.

- ◇ CC, *Confidence Conversion*: in this case we already have the log-likelihood profiles from each source, so the confidence coversion is simple, with $\ell_{\text{conv},j}(\psi_j) = \ell_{\text{prof},j}(\psi_j)$.

- ◇ FF, *Focused Fusion*: here we want to obtain a confidence curve for the parameter of overall interest ϕ. Depending on the situation, (i) if ϕ is assumed to be the same across sources or a function of some source-specific parameters, sum the $\ell_{\text{conv},j}$ and then profile again if necessary; (ii) if some component of the ψ_j are assumed to come from some common distribution, use the random effects solution presented above, and then profile again if needed. We then obtain $\ell_{\text{fus,prof}}(\phi)$, and in both cases we use the Wilks approximation to produce the final, combined confidence curve $\text{cc}^*(\phi, \text{data})$.

5.2 Nuisance parameters

The standard scheme described above is often applicable, but there are situations where it will not work well. As for confidence curves in general, we consider the method to work if the final combined confidence curve $\text{cc}^*(\phi)$ has the right coverage properties, either exactly or approximately (see (2.2)).

Assume we have k sources of information, with n_j observations source j. In combination situations, it is sometimes fruitful to differentiate between two types of nuisance parameters: source-specific and common nuisance parameters. Our default tool for constructing confidence curves, both in the II step and in the FF step, is based on profiling and application of Wilks's theorem. In many situations, we can have two rounds of profiling: first in the II step where we might profile out the source-specific nuisance parameters, and sometimes in the FF step where we might profile out potential common nuisance parameters. The profile log-likelihood is a practical tool, but unfortunately its performance can be poor in some situations, and it is important to be aware of its limitations. We will discuss problematic aspects of the profile likelihood in the next subsection, along with some potential remedies. First, we will briefly describe two different general situations with nuisance parameters.

Often we may find ourselves in the situation where all the nuisance parameters are source-specific. If we have 'large sources', i.e. the sample size n_j of each source is large, we can safely profile in the II step, and we can also safely apply the Wilks theorem in the FF step to produce the final confidence curve, at least in regular models. If some or all the sources are small, however, one should be more careful. Specifically, the profiling might go wrong and one might need certain corrections, as we discuss in the Section 5.3. There are also alternatives to the default profile solution in some cases, see for example Section 5.4 where we describe a somewhat general setting where small-sample exact CDs are available.

We can also have situations with common nuisance parameters. These common parameters can be of different kinds, and here we will particularly concern ourselves with nuisance parameters arising from the random effect distribution in the FF step. For example, if we have $\psi_j \sim \text{N}(\psi_0, \tau^2)$ and our focus parameter is ψ_0, then τ is a common nuisance parameter of that type. Then, we need to be careful with the profiling in the FF step: if the number of sources k is large we can safely profile and use the Wilks theorem in regular models. If k is small we may need to resort to

some of the corrections described in Section 5.3, or seek exact (but case-by-case) solutions. Often we may have both source-specific parameters and common nuisance parameters arising from the random effect distribution in the FF step. In that case, we ideally need to have a large number of large sources in order to produce (close to) valid CDs with the default profiling-based method. Again corrections may be needed, in both the II and FF steps. Note that in these cases, if k is too small, large sources will not necessarily help. Conversely, if the n_j are too small, a large k will not in general be able to remedy the mistakes coming from profiling in the II step.

5.3 Corrections to the log-likelihood profile

In situations with nuisance parameters using the profile log-likelihood can lead to "inefficient and even inconsistent estimates" (McCullagh & Tibshirani, 1990). There is a large literature on this topic, concerning second-order approximations and corrections or modifications of the profile likelihood. Cox & Reid (1993) state that there are two related reasons for modifying the profile log-likelihood: coming closer to the χ^2_1 distribution, and avoiding 'failure' due to nuisance parameters.

The different corrections appearing in the literature have varying performance and complexity; see for instance Barndorff-Nielsen (1986), Cox & Reid (1993), Diciccio & Efron (1992), Stern (1997), DiCiccio et al. (1996), Schweder & Hjort (2016, Ch. 7). There is also a whole subfield of integrated likelihood methods with partly similar aims, see Berger, Liseo & Wolpert (1999). A thorough investigation of all these methods is outside the scope of this article, and we will therefore only present one rather simple, somewhat limited solution. Alternative methods might work better, or at least in a more general setting, but these are often more complicated to compute.

In Cox & Reid (1987), the authors present what we will term the simple Cox–Reid correction. This is possibly the easiest correction to compute among those suggested above. It can be considered a special case of the correction in the general modified profile likelihood of Barndorff-Nielsen, but the simple Cox–Reid correction is limited to situations with orthogonal parameters (i.e. that the off-diagonal terms in the expected information matrix are equal to zero). Assume we have a scalar parameter of interest ψ and some vector of nuisance parameters λ. As usual, the profile log-likelihood for ψ is defined as $\ell_{\text{prof}}(\psi) = \ell(\psi, \widehat{\lambda}_\psi)$, where $\widehat{\lambda}_\psi$ is the ML estimate of λ for each fixed ψ value. The simple Cox–Reid correction gives the following modification of the profile log-likelihood,

$$\ell_{\text{cprof}}(\psi) = \ell_{\text{prof}}(\psi) - \tfrac{1}{2}\log\{\det J_{\lambda\lambda}(\psi, \widehat{\lambda}_\psi)\} \tag{5.1}$$

where $J_{\lambda\lambda}(\psi, \widehat{\lambda}_\psi)$ is the observed information for the λ components. The simple Cox–Reid correction can be used both in the II and FF steps, for models with orthogonal parameters. We will start by illustrating its use in the II step.

5.3.1 Using the Cox–Reid correction in the II step

In the II step the correction is particularly necessary when the sample sizes n_j within each source are small. The Neyman–Scott problem is an extreme example of such a situation. It can be presented of as a meta-analysis problem. We have a large number k of studies, but each study has only two observations (so $n_1 = \cdots = n_k = n = 2$). From each source j we observe $y_{i,j} \sim \text{N}(\mu_j, \sigma^2)$ where $i = 1, 2$ and $j = 1, \ldots, k$. Each sources has a specific mean parameter, but the variance, which is the parameter of main interest, is common across sources. This problem is popular in the

literature concerning corrections to the profile likelihood (see e.g. Schweder & Hjort (2016, Ch. 7)), since there exists a simple and exact solution, which serves as a gold standard against which to compare different corrections. We will compare this gold standard solution, as found in Schweder & Hjort (2016, Ch. 4), against the 'standard' II-CC-FF solution and the corrected II-CC-FF solution using the simple Cox–Reid correction.

The pivot $\widehat{\sigma}^2/\sigma^2$ can be seen to have a $\chi^2_k/(2k)$ distribution, which gives the exact confidence distribution,

$$C_{\text{gold}}(\sigma) = 1 - \Gamma_k(2k\widehat{\sigma}^2/\sigma^2),$$

where $\Gamma_k(\cdot)$ is the c.d.f. of the χ^2_k distribution and $\widehat{\sigma}^2 = \sum_{j=1}^k S_j^2/(2k) = \sum_{j=1}^k \frac{1}{2}(y_{1,j} - y_{2,j})^2/(2k)$ is the ML estimate (we note that this is a famous case where the ML estimator is inconsistent). The gold standard CD may be turned into a confidence curve using (2.1) and is displayed in black in Figure 5.1. For the sake of comparisons, we can write out the confidence likelihood implied by this confidence distribution,

$$\ell_{\text{conv,gold}}(\sigma) = -k\log\sigma - \tfrac{1}{2}(1/\sigma^2)\sum_{j=1}^k S_j^2. \tag{5.2}$$

With the standard II-CC-FF solution we start with the II step where we deal with each source separately: we profile out μ_j and get $\ell_{\text{prof},j}(\sigma) = -2\log\sigma - \frac{1}{2}(1/\sigma^2)S_j^2$. All the sources inform on exactly the same focus parameter and we can just sum the log-likelihood contributions in the fusion step,

$$\ell_{\text{prof}}(\sigma) = -2k\log\sigma - \tfrac{1}{2}(1/\sigma^2)\sum_{j=1}^k S_j^2. \tag{5.3}$$

Comparing this to the confidence log-likelihood for the gold standard in (5.2) we see that they differ by an extra '2' in the first term, which causes the inconsistency of the ML estimator. We may nevertheless construct our confidence curve in the general II-CC-FF manner, $\text{cc}_1^*(\sigma, \text{data}) = \Gamma_1\big(2\{\ell_{\text{prof}}(\widehat{\sigma}) - \ell_{\text{prof}}(\sigma)\}\big)$.

For this model, the simple Cox–Reid correction term for each source is $\log\sigma$. The corrected profile log-likelihood for each source is then $\ell_{\text{prof},j}(\sigma) + \log\sigma$, and for the full data we obtain a corrected profile log-likelihood identical to (5.2). Following the II-CC-FF recipe we construct the confidence curve with the Wilks approximation, $\text{cc}_2^*(\sigma, \text{data}) = \Gamma_1\{2(\ell_{\text{conv,gold}}(\widehat{\sigma}) - \ell_{\text{conv,gold}}(\sigma))\}$.

Figure 5.1 gives the three confidence curves in a specific example with $k = 20$ sources. The standard II-CC-FF solution in red is clearly far from the exact black curve. With k increasing, it converges to the wrong value, $\sigma/\sqrt{2}$. The blue curve, on the other hand, corresponding to the II-CC-FF solution with Cox–Reid correction, is close to the exact curve, even though it is constructed using the Wilks approximation. When k increases, for instance to 50, the blue and black curves are virtually identical.

5.3.2 Using the Cox–Reid correction in the FF step

Corrections of the type discussed here may be necessary in the FF step when there are common nuisance parameters arising from the random effect distribution in the Fusion step. In particular, we propose that this correction should readily be applied when then random effect distribution is assumed to be normal. Here, the correction would be notable with small k. First, we will present an example of the Cox–Reid correction in a classic model for random effect meta-analysis. We may

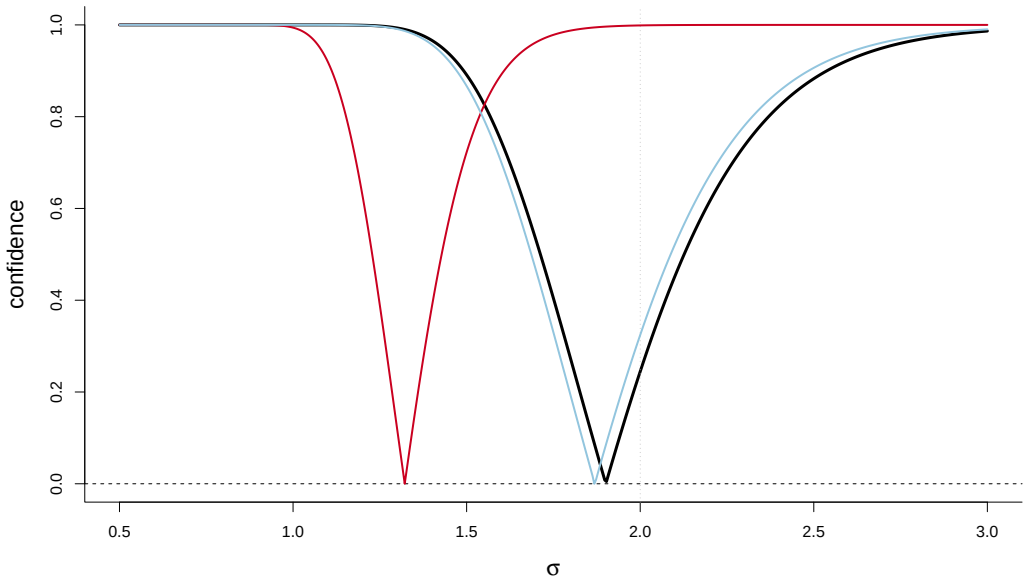

Figure 5.1: Neyman–Scott example with $k = 20$ sources/studies, the parameter of main interest $\sigma = 2$, and the source specific means drawn from a uniform between -3 and 3. The exact confidence curve in black, the standard uncorrected II-CC-FF solution in red, and the corrected II-CC-FF solution in blue.

call that model normal-normal, since both the model for the observations and the model for the random effects is normal. Next, we will discuss a situation where the observations are non-normal.

The most canonical type of random effect meta-analysis, which we term the basic random effect model, starts with k independent estimators $y_1, \ldots, y_k$ aiming at the parameters $\psi_1, \ldots, \psi_k$, with $y_j \,|\, \psi_j \sim \mathrm{N}(\psi_j, \sigma_j^2)$, and $\psi_j \sim \mathrm{N}(\psi_0, \tau^2)$. We can investigate the standard II-CC-FF solution, along with the simple Cox–Reid correction. When the σ_j are assumed known, the integral in (4.1) has an explicit solution and the full combined profile likelihood from the FF step becomes

$$\ell_{\mathrm{fus,cprof}}(\psi_0) = \sum_{j=1}^{k} \left\{ -\tfrac{1}{2} \log(\widehat{\tau}_\psi^2 + \sigma_i^2) - \tfrac{1}{2} \frac{(y_i - \psi_0)^2}{\widehat{\tau}_\psi^2 + \sigma_i^2} \right\} \\ -\tfrac{1}{2} \log \left[\sum_{j=1}^{k} \left\{ \left| -\tfrac{1}{2} \frac{1}{(\widehat{\tau}_\psi^2 + \sigma_i^2)^2} + \frac{(y_i - \psi_0)^2}{(\widehat{\tau}_\psi^2 + \sigma_i^2)^3} \right| \right\} \right].$$

The first part of the formula is the ordinary profile log-likelihood, the second part the simple Cox–Reid correction. Brief simulation studies indicate that confidence curves produced with this corrected profile log-likelihood and the Wilks approximation have good coverage properties. The correction is especially important when there are few studies and the heterogeneity between them is large.

In the normal-normal set-up above, the parameters are orthogonal, but we suggest that one could use the simple Cox–Reid correction even if the model for the observations in each source is non-normal. Suppose we are in a setting like in (4.1), and assume that the random effect distribution is normal, with the parameter of main interest being the overall mean ψ_0. Inside each source we can have any (regular) model. In a simple normal model, the Cox–Reid correction when

profiling out the variance τ^2 would be equal to $-\log\{\widehat{\tau}^2(\psi)\}$. We propose to routinely use the following corrected likelihood profile construction in this type of random effect setting,

$$\ell_{\rm fus,cprof}(\psi_0) = \sum_{j=1}^k \Big\{ \log\Big[\int \exp\{\ell_{{\rm conv},j}(\psi_j)\} \frac{1}{\widehat{\tau}(\psi_0)} \varphi\left(\frac{\psi_j - \psi_0}{\widehat{\tau}(\psi_0)} \right) {\rm d}\psi_j \Big] \Big\} + \log\{\widehat{\tau}(\psi_0)^2\}, \quad (5.4)$$

where $\widehat{\tau}(\psi_0)$ is the ML estimate of τ for each fixed ψ_0 value and φ is the standard normal density. The orthogonality of ψ_0 and τ in the full (integrated) distribution does not necessarily hold, but it would hold if the sources were large, since then $\ell_{{\rm conv},j}(\psi_j)$ above will approach a normal likelihood. We therefore consider formula (5.4) to be an approximate correction for the situation when k is small, but the n_j are sufficiently large. This idea appears to work well, for instance in the situation of meta-analysis a2×2 tables which is mentioned in Section 6.1.

5.4 Confidence power and optimal methods

Profiling the log-likelihood is a general method for elimination of nuisance parameters, and as we have seen it may have unsatisfying performance in some situations. For some parameters in exponential families, there is an alternative method which is much more powerful, producing *optimal confidence distributions*. In our II-CC-FF setting, this method might come into play both in the II step and the FF step, see the application in Section 7.1.

For ease of presentation, we present the optimal confidence method in the case where all the k sources inform on a common focus parameter $\psi = \psi_1 = \cdots = \psi_k$. This constitutes a situation where the method is used in the final FF step. Suppose again that ψ is the focus parameter, and that we have m nuisance parameters $\gamma_1, \ldots, \gamma_m$, which may be both source-specific or common to all k sources. Suppose also that the log-likelihood function at work, based on information sources $y_1, \ldots, y_k$, can be written in the form

$$\ell(\psi, \gamma_1, \ldots, \gamma_m) = \psi A + \gamma_1 B_1 + \cdots + \gamma_m B_m - d(\psi, \gamma_1, \ldots, \gamma_m) + h(y_1, \ldots, y_k), \quad (5.5)$$

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

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

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

To illuminate the exact meaning of 'most powerful' in this setting, one needs to consider the theory for loss and risk functions for confidence distribution developed in Schweder & Hjort (2016, Ch. 5). Confidence power is measured via the risk function

$$r(C, \psi, \gamma) = {\rm E}_{\psi,\gamma} \int \Gamma(\psi_{\rm cd} - \psi) \, {\rm d}C(\psi_{\rm cd}, Y), \quad (5.6)$$

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.

6 Connections with other approaches

II-CC-FF provides a general framework for combination problems. These problems have been studied in the statistics field for a long time, and II-CC-FF naturally has many connections to that literature. In this section we will review related methods and offer some comparisons with II-CC-FF. We start by describing parts of the meta-analysis literature, before treating two groups of CD-based methods for combination of information.

6.1 Meta-analysis methods

As mentioned in the small start example (1.1), some common meta-analysis methods flow more or less directly from the II-CC-FF framework. In addition, the framework also invites more general, principled and non-standard solutions. The meta-analysis literature is vast, and we have only investigated the connections between II-CC-FF and a couple of widely encountered meta-analysis methods. We will start with a short discussion of the basic random effect model, before we go on to discuss the famous case of meta-analysis of 2×2 tables. While II-CC-FF certainly can handle these common cases, the framework naturally opens the door to non-standard analyses, for example using other distributions than the normal (see Section 7.1) or situations where different studies have reported different summaries (as with our start illustration in Section 1).

The basic random effect model was discussed in Section 5.3. Again, we have k independent estimators $y_1, \ldots, y_k$ aiming at the parameters $\psi_1, \ldots, \psi_k$, with $y_j \,|\, \psi_j \sim \mathrm{N}(\psi_j, \sigma_j^2)$ and $\psi_j \sim \mathrm{N}(\psi_0, \tau^2)$. Usually the source-specific standard deviations σ_j are assumed known, and the most commonly used method for inference on the overall mean ψ is the DerSimonian-Laird method. Several simulation studies have revealed that this method can have poor performance and often produces too narrow confidence intervals, see Partlett & Riley (2017) and references therein. There are several possible solutions that follow naturally from the general CD-methodology and II-CC-FF, see Schweder & Hjort (2016, Ch. 13). Specifically, we investigated the standard II-CC-FF solution, along with the simple Cox–Reid correction in Section 5.3. Likelihood-based method for the basic random effect model, even exploring higher order corrections, have been investigated earlier, see for instance Hardy & Thompson (1996) and Noma (2011). For a more general likelihood approach see O'Rourke (2008).

If the sources are small, the assumption of known σ_j does not hold and II-CC-FF can provide more sophisticated solutions. In the II step, we have exact CDs for each ψ_j based on the Student's t distribution, which we can convert to a confidence log-likelihood by exact conversion in the CC step. For the FF step, we use the general random effect method from (4.1), either with numerical integration or using the TMB package. Corrections in both the II and FF step may be considered.

In meta-analyses of 2×2 tables, each study is usually modelled with a pair of binomially distributed variables, one for the control group and another for the treatment group; $Y_{0,j} \sim \mathrm{binom}(m_{0,j}, p_{0,j})$ and $Y_{1,j} \sim \mathrm{binom}(m_{1,j}, p_{1,j})$, with $p_{0,j} = \exp(\theta_j)/(1 + \exp(\theta_j))$ and $p_{1,j} = \exp(\theta_j + \psi_j)/(1 + \exp(\theta_j + \psi_j))$. Each source has a specific nuisance parameter θ_j, governing the event probability in the control group, and ψ_j the log odds ratio. We will first treat the fixed effect case where the log odds ratios are assumed common across all sources, $\psi_1 = \cdots = \psi_k = \psi$, before we come to the random effect case in the next paragraph. The information available in each source depends on the size of the binomial sample sizes $m_{0,j}$ and $m_{1,j}$ (and on the event probabilities). If the number of studies increases while the size of each study stays constant, it is known that the

ML estimator is inconsistent (Breslow, 1981). We are therefore in a Neyman–Scott type situation and using the standard II-CC-FF will not be good when the sources are small (especially if the event probabilities are low). Also, the simple Cox–Reid correction to the profile in each source is not immediately available because ψ_j and θ_j are not orthogonal. However, there exists an optimal CD for the common ψ based on the theory from Section 5.4,

$$C_{\text{opt}}(\psi, \text{data}) = P_\psi(B_k > b \,|\, z_1, \ldots, z_k) + \tfrac{1}{2} P_\psi(B_k = b \,|\, z_1, \ldots, z_k). \tag{6.1}$$

Here, $z_j = y_{0,j} + y_{1,j}$ and $B_k = \sum_{j=1}^k Y_{1,j}$. The CD is obtained by simulating the distribution of B_k given $Z_1, \ldots, Z_k$. Note also that we similarly have an optimal CD for ψ_j within each source,

$$C_{\text{opt},j}(\psi_j, y_{0,j}, y_{1,j}) = P_\psi(Y_{1,j} > y_{1,j} \,|\, z_j) + \tfrac{1}{2} P_\psi(Y_{1,j} = y_{1,j} \,|\, z_j). \tag{6.2}$$

This CD is simple to compute as $Y_{1,j} \,|\, Z_j$ has an eccentric hypergeometric distribution. Starting from (6.2) for each source in the II step, we can obtain an approximation to the optimal solution in (6.1) which is faster to compute and also lends itself to a natural random effect extension, as we will see. In the CC step, we use exact conversion to obtain the confidence log-likelihoods $\ell_{\text{conv},j}(\psi_j) = \log g_j(y_{1,j}, \psi_j)$, where

$$g_j(y_{1,j}, \psi_j) = \frac{\binom{m_{0,j}}{z_j - y_{1,j}} \binom{m_{1,j}}{y_{1,j}} \exp(\psi_j y_{1,j})}{\sum_{u=0}^{z_j} \binom{m_{0,j}}{z_j - u} \binom{m_{1,j}}{u} \exp(\psi u)} \quad \text{for } y_{1,j} = 0, 1, \ldots, \min(z_j, m_{1,j}) \tag{6.3}$$

is the density function of the eccentric hypergeometric distribution. We sum these confidence log-likelihoods to get $\ell_{\text{fus}}(\psi) = \sum_{j=1}^k \ell_{\text{conv},j}(\psi_j)$, find the ML estimate $\widehat{\psi}$ and the deviance, and use the Wilks approximation:

$$\text{cc}^*(\psi, \text{data}) = \Gamma_1(2\{\ell_{\text{fus}}(\widehat{\psi}) - \ell_{\text{fus}}(\psi)\}). \tag{6.4}$$

Even though there is some level of approximation in this solution, it tends to work very well even for small k.

From this approximate fixed effect approach we find a natural extension to random effects. Assuming that the log-odd ratios from the different sources come from a common normal distribution, we have the following log-likelihood contribution from each source to the overall parameters,

$$\ell_{\text{fus},j}(\psi, \tau) = \log\Big\{ \int g_j(y_{1,j}, \psi_j) \frac{1}{\tau} \varphi\left(\frac{\psi_j - \psi}{\tau} \right) \, \mathrm{d}\psi_j \Big\}, \tag{6.5}$$

where $g_j(y_{1,j}, \psi_j)$ is the density function of the eccentric hypergeometric distribution, pointed to above. We sum the contributions from each source, profile out τ, and use the Wilks approximation. If the number of sources is small we add the approximate Cox–Reid correction $\log \widehat{\tau}^2(\psi)$ from (5.4) to the profile log-likelihood. In applications, we computed the integral using the TMB package. This approach seems promising with good coverage properties in simulations. The uncorrected version is similar to the hypergeometric–normal model in Stijnen et al. (2010), but we find the correction to be important when k is not very large.

6.2 Other combination methods based on CDs

There is a steadily growing literature on combination of information with confidence distributions. Here we will briefly discuss methods by Singh et al. (2005) and Liu et al. (2015). These CD

approaches are sometimes collected under the same umbrella, called Fusion Learning (Cheng et al., 2017).

We start by discussing the approach of Singh et al. (2005), valid when all confidence components relate to a common focus parameter. Suppose 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 et al. (2011), Xie & Singh (2013), Liu et al. (2014), 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 $\Phi(\cdot)$ again 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

$$\bar{C}(\psi, y) = \Phi\Big(\sum_{j=1}^k w_j \Phi^{-1}(C_j(\psi, y_j))\Big) \tag{6.6}$$

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 (1.1), the individual confidence distributions take the form $C_j(\psi, y_j) = \Phi((\psi - y_j)/\sigma_j)$, and the general (6.6) 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 (6.6).

Recipe (6.6) 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) are of this type. In such cases recipe (6.6) 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

$$\bar{C}(\psi, y) = H\Big(\sum_{j=1}^k \widehat{w}_j \Phi^{-1}(C_j(\psi, y_j))\Big), \tag{6.7}$$

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 (6.7) leading back to (6.6). In yet other words, method (6.6) remains correct to the first-order large-sample degree, even though more careful versions of (6.7) would tend to work better for smaller samples.

The approach described above yields approximative solutions for the basic normal-normal random effect model, partly helped by the fact that the unconditional density in that case has an explicit normal form, $y_j \sim \mathrm{N}(\psi_0, \sigma_j^2 + \tau^2)$. It is not clear how the method in Xie et al. (2011) can incorporate more general random effect models, however.

Under the 'Fusion learning' umbrella there are other methods. The method in Liu et al. (2015) may be termed a 'confidence density method' and can be considered as a special case of II-CC-FF, as we will see. The method is proposed for a fixed effect setting, but where the studies may differ in reported outcomes, in measured covariates, or have source-specific nuisance parameters. Thus, some of the studies may only contain indirect information about the parameter of interest. Let θ be the full parameter vector for all the studies and $\gamma_j = M_j(\theta)$ the parameters in study j, with M_j denoting a known mapping function. Liu et al. (2015) summarise the information in each source with multivariate normal CDs, $C_j(\gamma_j, y_j)$, transform these to confidence densities $c_j(\gamma_j, yj) = \partial C_j(\gamma_j, yj)/\partial\gamma_j$, which are then multiplied into a combined confidence density, which informs on the full θ. The authors stress that the approach is general in the sense that it can be used with a wide range of parametric models for the sources. This generality is achieved because the authors assume that the number of observations in each source increases to infinity.

The normal CDs for each study only requires the estimated parameter vector $\widehat{\gamma}_j$ and estimated covariance matrix $\widehat{\Sigma}_j$ for $\widehat{\gamma}_j$, and the authors therefore highlight that the approach only needs summary statistics rather than the full data. Also, they prove that their approach is asymptotically equally efficient as a traditional likelihood approach using the full data.

For location parameters in normal models the confidence density and confidence likelihood are proportional. The approach in Liu et al. (2015) can therefore be considered a special case of II-CC-FF. Sometimes the confidence density might be easier to obtain than the exact confidence log-likelihood, and could be used also in connection with II-CC-FF. However, one might need to be careful, as this approach could introduce mistakes. The confidence density is equal to $\partial C(\psi, T)/\partial\psi$, while the exact confidence likelihood takes the derivative with respect to T, as we saw in Section 3. The difference between the confidence density and confidence likelihood will be the most pronounced when the sample sizes are small, with the difference going away with increasing sample sizes.

7 Applications

Below we illustrate the capacity for the II-CC-FF paradigm to solve problems in rather different application settings. The first application concerns a meta-analysis of 2×2 tables, where we consider both fixed and random effect approaches. Besides demonstrating the use of II-CC-FF in a typical meta-analysis setting, we also aim at investigating the effect of the CC step, particularly the difference between exact conversion and the approximate chi-squared inversion method. In the second application we analyse an interesting archaeological dataset. Here we use the so-called basic random effect model which was discussed in Sections 5.3 and 6.1, but atypically our parameter of main interest is the spread parameter τ. For this parameter there exists an exact confidence distribution in the FF step.

The annual growth rate of humpback whales is the focus of the third application. There, we illustrate how to construct confidence curves based on non-sufficient summary statistics; we only have access to a point-estimate and a highly non-symmetric confidence interval. In this example we

also demonstrate how partial prior information can be incorporated into our II-CC-FF framework. Finally, the last application illustrates the combination of 'hard' with 'soft' data. Here 'hard' designates data sources of high quality which inform directly on the focus parameter. 'Soft' data, on the other hand, may be of lower quality, with more noise and biases, or simply containing less direct information on the focus parameter. Such large, noisy datasets are increasingly available in a number of fields, for example from webscraping or text-mining, but lead to challenges when attempting to fuse the sources. We illustrate the combination of 'hard' and'soft' data with a question from the field of peace research; is there evidence for The long peace, and in that case, when did it start?

7.1 Meta-analysis of 2×2 tables: do corticosteroids increase the risk of gastrointestinal bleeding?

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) provide optimal confidence distributions 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 (5.5), 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.

Table 7.1: Corticosteroids and gastrointestinal bleeding: Number of bleeding ulcer events in two groups of patients, one receiving corticosteroids and the other not receiving them, in five independent studies; see Section 7.1 and Figure 7.1.

m_1	m_0	y_1	y_0
49	50	5	0
101	99	0	1
41	40	1	2
63	63	1	0
198	202	1	0

Narum, Westergren & Klemp (2014) provide a medical dataset with five studies investigating gastrointestinal bleeding for a certain subgroup of patients, those that are in ambulatory care (they have other data for hospitalised patients). The treatment group received corticosteroids and the control group did not, and the number of bleeding events were recorded for each group. We will use this example to illustrate the use of II-CC-FF in a typical but somewhat difficult meta-analysis setting. The difficulty here comes from the low event probability, which translates to several studies with zero bleeding events in one of the two groups; in particular, this implies that the ML estimators are far away from the usually assumed approximate normality. The different studies also exhibit very different treatment effects, as we will see. This example will also serve as an illustration of the effect of the confidence conversion step.

As discussed in Section 6, each study is usually modelled with a pair of binomially distributed variables,

$$Y_{0,j} \sim \text{binom}(m_{0,j}, p_{0,j}) \quad \text{and} \quad Y_{1,j} \sim \text{binom}(m_{1,j}, p_{1,j}),$$

with subscript '1' indicating treatment and '0' control, with $p_{0,j} = \exp(\theta_j)/\{1+\exp(\theta_j)\}$ and $p_{1,j} = \exp(\theta_j+\psi_j)/\{1+\exp(\theta_j+\psi_j)\}$. The focus parameter is the treatment effect; here we will consider $\gamma_j = \exp(\psi_j)$, the odds ratios. First, we will treat the fixed effect case, where the odds ratios are assumed common across all sources, $\gamma_1 = \cdots = \gamma_k = \gamma$; afterwards we come to the random effect case where the γ_j are assumed to come from a common background distribution.

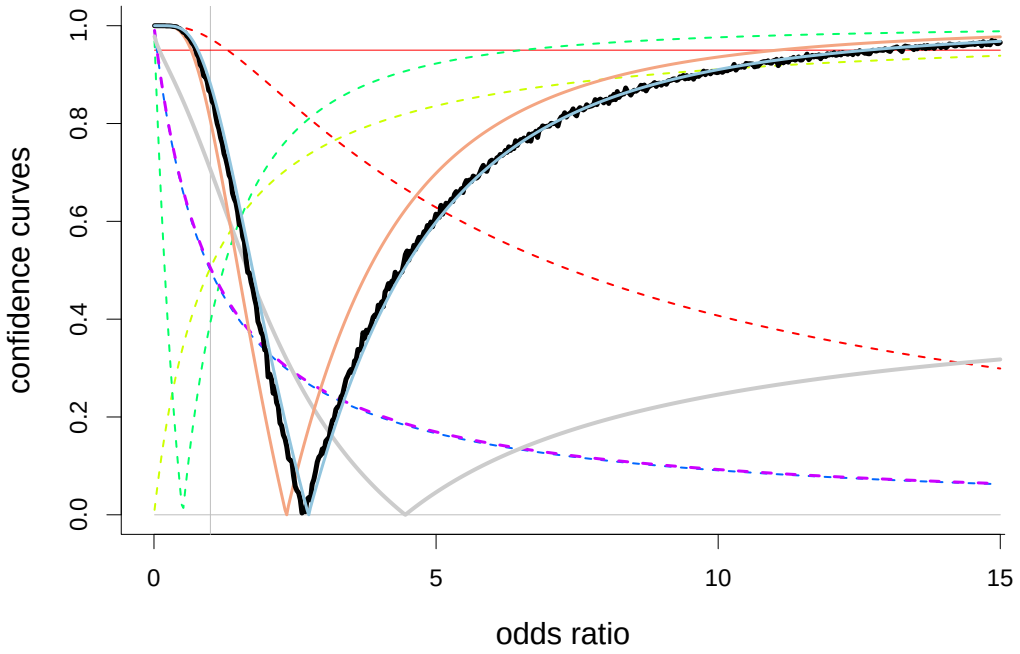

Figure 7.1: The coloured dashed curves are the confidence curves for the odds ratio from each of the five studies. The thick black curve is the optimal combined confidence curve, assuming fixed effects, while the grey curve corresponds to a random effect model. The light blue curve is $cc_1^*(\gamma)$ and the orange curve is $cc_2^*(\gamma)$. The horizontal red line marks the 95% confidence level and the vertical grey line corresponds to an odds ratio equal to one. See Section 7.1 and Table 7.1.

In the first step, II, we use (6.2), providing the optimal CD for the odds ratio from each of the five studies. These confidence curves are shown in Figure 7.1 as dashed coloured lines. We see that the studies indicate wildly different odds ratios; two studies have a point estimate smaller than one, while three studies have point estimates larger than one, suggesting that corticosteroids increase the risk of bleeding. These three studies actually have a ML point estimate equal to infinity, due to the control group having zero events. Note however that two of these studies have extremely wide confidence intervals, in fact spanning the entire positive line for most confidence levels.

For the CC step we will investigate two possibilities, (1) using exact conversion, and (2) using the more automatic and approximate normal conversion (chi-squared inversion method, as per Section 3), which corresponds to a situation where we would have access to the individual confidence curves, but did not know how they were constructed. The two alternatives lead to

$$\text{CC:} \qquad \ell_{\text{conv},1,j}(\gamma) = \log g_j(y_{1,j},\gamma), \qquad \ell_{\text{conv},2,j}(\gamma) = -\tfrac{1}{2}\Gamma_1^{-1}(\text{cc}_j(\gamma,\text{data}_j)) \quad \text{for } j=1,\ldots,k,$$

where $g_j(\cdot)$ is the density function of the eccentric hypergeometric distribution in (6.3), $\text{cc}_j(\gamma,\text{data}_j)$ is the confidence curve from study j, and Γ_1^{-1} is the quantile function of the χ_1^2 distribution. For both choices of confidence likelihood the FF step is the same, we sum the log-likelihoods, find the

combined deviance function, and apply the Wilks's theorem, as per Section 4:

$$\text{FF:} \qquad \ell_{\rm fus,1}(\gamma) = \sum_{j=1}^{6} \ell_{{\rm conv},1,j}(\gamma), \qquad {\rm cc}_1^*(\gamma, {\rm data}) = \Gamma_1(2\{\ell_{\rm fus,1}(\widehat{\gamma}) - \ell_{\rm fus,1}(\gamma)\}),$$

$$\text{or } \ell_{\rm fus,2}(\gamma, {\rm data}) = \sum_{j=1}^{6} \ell_{{\rm conv},2,j}(\gamma), \qquad {\rm cc}_2^*(\gamma) = \Gamma_1(2\{\ell_{\rm fus,2}(\widehat{\gamma}) - \ell_{\rm fus,2}(\gamma)\}).$$

We will compare these two confidence curves with the confidence curve for γ which is given in (6.1) and is based on simulating the distribution of a sum of eccentric hypergeometrically distributed variables. This curve is optimal in the sense of Section 5.4, but can be quite heavy to compute. As seen in Figure 7.1, the curve resulting from the exact conversion, $cc_1^*(\gamma, {\rm data})$, perfectly matches the optimal confidence curve for γ, even though we use the Wilks approximation. The curve resulting from normal conversion does not match the optimal curve completely, indicating that the normal conversion introduces some errors in this case. In other meta-analyses of 2×2 tables, we have seen that the normal conversion works very well, and the success of that approximation depends on the size of the tables, the event probabilities and crucially on whether we have some studies with zero events in one or both groups.

We can use a random effect approach as well, assuming that the log odds ratios ψ_j come from a normal distribution and using (6.5) and the TMB package for computation of the integral. The simple Cox–Reid correction does not work well here because each tables has very little information in them (and $g(\cdot)$ is therefore far from a normal density). We find considerable heterogeneity between the studies, with $\widehat{\tau} = 2.16$ being the estimated standard deviation of the ψ_j. This results in the grey confidence curve in Figure 7.1, which is considerably wider than the other combined confidence curves. At any rate none of the curves indicates a significant effect on the 95% level, so there is little evidence for corticosteroids causing gastrointestinal bleeding for patients in ambulatory care.

7.2 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 $y = (y_1, y_2, y_3, y_4)^{\rm t}$, 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

$$Y_{j,i} \sim {\rm N}_4(\xi_j, \Sigma_j) \quad \text{for } i = 1, \ldots, 30,$$

for each of the five epochs j. 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

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

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 $a^{\rm t}Y$ of the four skull measurements, normalised to have coefficient vector length $\|a\| = 1$. This parameter is one of several natural measures of the 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 $\psi_1, \ldots, \psi_5 \sim {\rm N}(\psi_0, \tau^2)$. 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 7.2: Skulls: For each of the five time epochs, the table gives the estimate $\widehat{\psi}$ and its estimated standard deviation $\widehat{\sigma}$. See Section 7.2 and Figure 7.2.

epoch	$\widehat{\psi}$	$\widehat{\sigma}$
-4000	2.652	0.561
-3300	2.117	0.444
-1850	1.564	0.331
-200	2.914	0.620
150	1.764	0.373

Table 7.2 gives point estimates

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

for the five time epochs, along with estimated standard deviations σ_j 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_j approximately known. Figure 7.2 displays point estimates with 0.90 confidence intervals (left panel), for the five epochs. The log-likelihood for these five estimates, under the implied ${\rm N}(\psi_0, \sigma_j^2 + \tau^2)$ model, writing k for the number of data sources involved, is

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

The ensuing profiled log-likelihood is

$$\ell_{\rm prof}(\tau) = -\tfrac{1}{2} \sum_{j=1}^k \Big[\log(\sigma_j^2 + \tau^2) + \frac{\{\widehat{\psi}_j - \widetilde{\psi}_0(\tau)\}^2}{\sigma_j^2 + \tau^2} \Big], \text{ with } \widetilde{\psi}_0(\tau) = \frac{\sum_{j=1}^k \widehat{\psi}_j/(\sigma_j^2 + \tau^2)}{\sum_{j=1}^k 1/(\sigma_j^2 + \tau^2)}. \quad (7.1)$$

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

$$Q(\tau) = \sum_{t=1}^k \frac{\{\widehat{\psi}_j - \widetilde{\psi}_0(\tau)\}^2}{\sigma_j^2 + \tau^2} \quad \text{and} \quad C(\tau, \text{data}) = 1 - \Gamma_{k-1}(Q(\tau)),$$

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 has a confidence point mass $C(0, \text{data}) = 0.222$ at zero, and is shown in the right panel of Figure 7.2. The confidence point-mass 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 0.389.

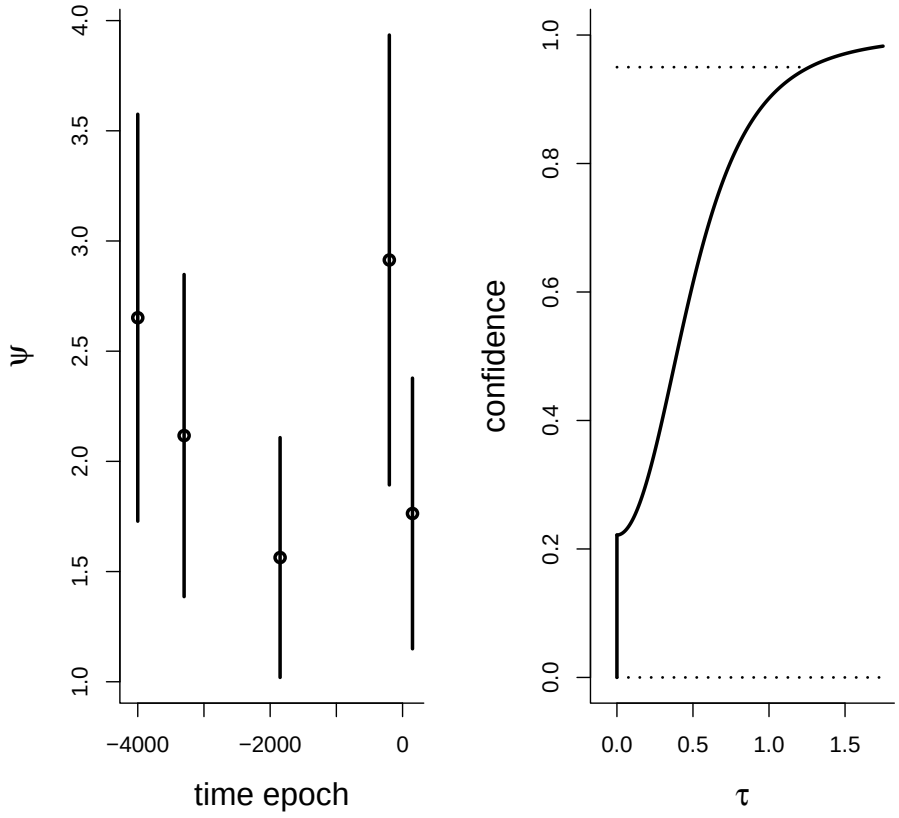

Figure 7.2: Left panel: Point estimates $\widehat{\psi}_j$ with 90% confidence intervals, for the skull stretch parameter ψ, across five time epochs (see Table 7.2). Right panel: Confidence distribution for the variability parameter τ.

In this specific example we had access to a tailored and exact recipe, which we might compare with the performance of the standard II-CC-FF method of Section 5.1. The standard II-CC-FF makes use of the log-likelihood profile in (7.1) and the Wilks theorem. Short investigations reveal that the standard II-CC-FF recipe is not working well in this case. This is partly due to the relatively small number of groups $k = 5$, but primarily to problems related to τ lying close to the boundary of its parameter space (the ML estimate is 0.061). Corrections related to boundary problems, as suggested in Schweder & Hjort (2016, Ch. 4), can be considered.

In other applications of this type of extended meta-analysis machinery the centre value ψ_0 of the background distribution of the ψ_j might be of high importance, and methods in Schweder & Hjort (2016, Ch. 13) may be used to produce an accurate $\text{cc}(\psi_0, \text{data})$. For the skulls analysis the primary question is whether the ψ_j parameter, or other similar parameters associated with the Σ_j matrices, have changed over the course of four thousand years, and the precise value of ψ_0 is of

secondary importance. We report, though, that its point estimate is 2.067, with an accurate 90% interval stretching from 1.713 to 2.522 (see left panel of Figure 7.2).

7.3 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 includes bootstrapping. The available information is as presented in Table 7.3; 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 7.3: 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 7.3 and Figure 7.3.

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

For this illustration we are interested in the underlying true abundances underlying these two studies. Let ψ_1 be the population size in 1995 and ψ_2 be the size in 2001. Our main interest may lie in the annual growth rate underlying these two population sizes. We define $\rho = (\psi_2 - \psi_1)/(6\psi_1)$, a simple (and in some sense approximate) definition of annual growth rate.

The first step, *Independent Inspection*, requires us to construct confidence distributions for ψ_1 and ψ_2 from 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

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

with a power transformation $h(\psi, a) = \text{sgn}(a)\psi^a$; 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,

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

where $[\psi_L, \psi_R]$ 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 the left panel of Figure 7.3. In this case the confidence log-likelihoods in the *Confidence Conversion* step are easily obtained. For year j,

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

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

$$\begin{aligned} \text{FF:} \quad \ell_{\text{fus,prof}}(\rho) &= \max\{\ell_{\text{conv},1}(\psi_1) + \ell_{\text{conv},2}(\psi_2) \colon (\psi_2 - \psi_1)/(6\psi_1) = \rho\}, \\ \text{cc}^*(\rho) &= \Gamma_1(2\{\ell_{\text{fus}}(\widehat{\rho}) - \ell_{\text{fus}}(\rho)\}). \end{aligned}$$

Here we obtain the blue curve in the right panel of Figure 7.3, with $\widehat{\rho} = 0.026$ and a 95% confidence interval $[-0.094, 0.454]$.

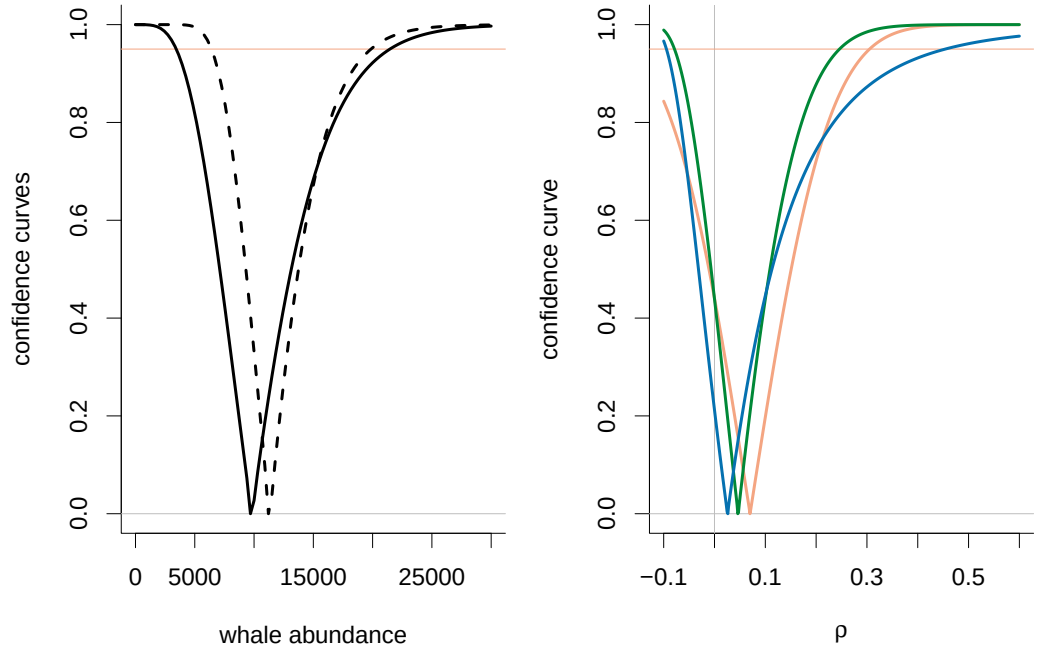

Figure 7.3: Left panel: confidence curves for ψ_1 and ψ_2, the abundance of humpback whales in the North Atlantic in 1995 (fully drawn line) and 2001 (dashed line). Right panel: the confidence curve for $\rho = (\psi_2 - \psi_1)/(6\psi_1)$ based on the two surveys (blue curve); the confidence curve based on prior information alone (orange curve); and the confidence curve combining the studies and the prior information (green curve). See Section 7.3 and Table 7.3.

In some cases there may exist some expert knowledge pertaining to at least the focus parameter under study, here the annual growth rate ρ, though not necessarily for the full parameter vector of the combined models, here (ψ_1, ψ_2) the two population sizes. A proper Bayesian analysis requires the statistician to have such a prior for (ψ_1, ψ_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 (ψ_1, ψ_2). For this illustration we assume that whale biologists provide a normal prior with expectation equal to 0.07 and variance 0.12^2. This prior may come from knowledge of other humpback whale populations or simulation-based life-history models (see for example Zerbini et al. (2010), giving a similar point estimate as we have used).

The prior can be represented as a 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,

$$\begin{aligned}\text{FF:}\qquad \ell_{\text{fus,prof},B}(\rho) &= \max\{\ell_{\text{conv},1}(\psi_1)+\ell_{\text{conv},2}(\psi_2)+\ell_B(\rho)\colon (\psi_2-\psi_1)/(6\psi_1)=\rho\}\\ &= \max_{\sigma_1}\{\ell_{\text{conv},1}(\psi_1)+\ell_{\text{conv},2}(\psi_2)\colon (\psi_2-\psi_1)/(6\psi_1)=\rho\}+\ell_B(\rho)\\ &= \ell_{\text{fus,prof}}(\rho)+\ell_B(\rho).\end{aligned}$$

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 (ψ_1,ψ_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 the right panel of Figure 7.3). It is also somewhat narrower than both.

7.4 Combining hard and soft data: battle deaths and Ngrams

In this last illustration, we will examine the use of the II-CC-FF framework in a highly non-standard setting, where one wishes to combine hard data, sources that inform directly on the focus parameter, with softer data sources, which only contain indirect or noisier information about the focus parameter. This kind of combination has wide potential in various fields where 'soft' data could be based on webscraping, using twitter accounts or other social media, but raises specific issues and challenges.

The question we investigate here is the extent of statistical evidence for The long peace, the period of relative peace and stability following the second world war (and still lasting, presumably). Specifically, do we find evidence of a change-point τ when analysing the sequence of battle deaths in interstate wars between 1823 and today? This question has been investigated in Cunen, Hjort & Nygård (2018b) using the Correlates of War (CoW) datatset (Sarkees & Wayman, 2010). The authors found evidence of an abrupt change in the battle death distribution at some point after the second world war, from a distribution with a high median battle death to a distribution with a lower median (and also a less heavy tail). Here, we want to extend the analysis in Cunen, Hjort & Nygård (2018b) and investigate whether there might be benefits in combining the battle death data with other sources assumed to be informing on τ.

Some political scientists consider the afore-mentioned decrease in battle deaths to reflect a moral and political shift within a large portion of the world's population. At some point in the 20th century, it is argued, the perception of war changed, from being seen as something natural and

inevitable, sometimes even positive, to being perceived as highly negative, evil and unacceptable; cf. Pinker (2011, Ch. 5). This change in norms has likely manifested itself in various ways, including cultural, artistic and political expressions, for example through text. We will therefore collect sequences representing the usage of certain relevant words, and then attempt to combine the change-point inference from such an Ngram analysis (suggested to us by Steven Pinker, personal communication), with the change-point inference from the battle deaths data.

We might provide a more thorough analysis of this question in future publications, but for the sake of this illustration we limit ourselves to analysing one word: 'anti-war'. We collected the rate of usage of 'anti-war' for each year between 1823 and 2003 from the Google Books Ngram viewer (Michel et al., 2010), see Figure 7.4. The rate of usage is the number of times that word appears in each year divided by the total number of words in the Google Books corpus from each year. For a more thorough analysis we would build a score based on several such Ngrams, or even a joint model for several Ngrams, but those efforts are outside the scope of this illustration. Naturally, the whole analysis rests upon a strong assumption: that the change-point parameter underlying the sequence of battle-deaths and the (potential) change-point parameter underlying the Ngram are somehow the same parameter. We thus assume that changes in the battle death distribution and in the 'anti-war' distribution are two different manifestations of the same underlying process.

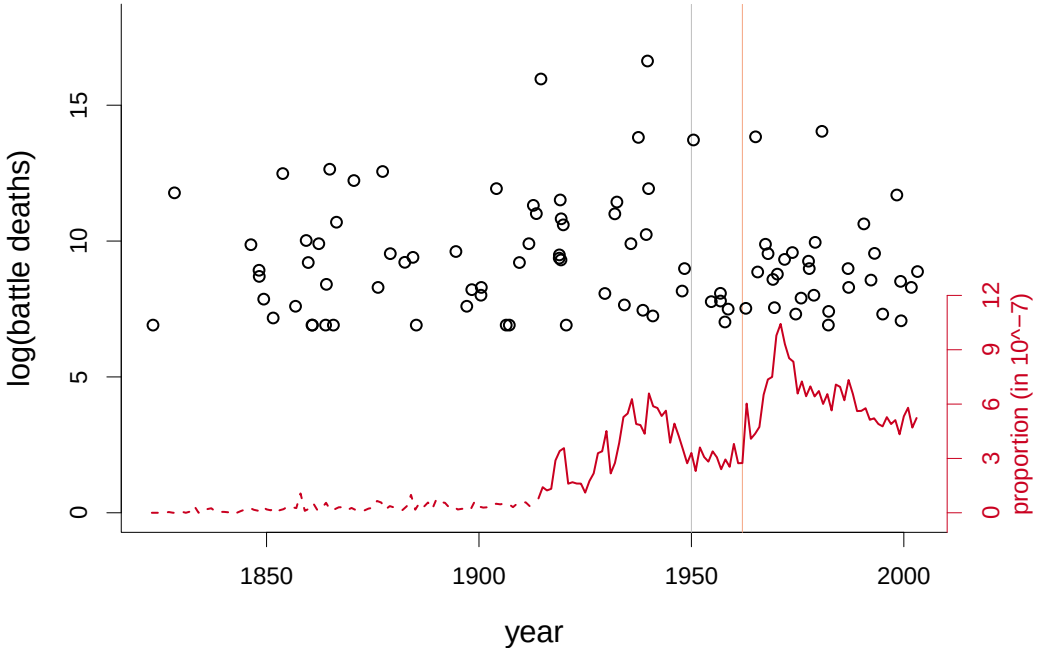

Figure 7.4: The points represent the battle deaths, on log scale, for 95 wars between 1823 and 2003. Note that the CoW dataset only includes wars with at least 1000 battle deaths. The vertical grey line gives the point estimate for the change-point based on the battle deaths data. The red line shows the Ngram for 'anti-war', i.e. the number of times that word appears in each year divided by the total number of words in the corpus from each year. The counts between 1823 and 1913 were not used in the change-point analysis (and hence dashed). The vertical red line gives the point estimate for the change-point based on the Ngram.

We modelled the battle deaths with a fat-tailed inverse Burr distribution and used the methods in Cunen, Hermansen & Hjort (2018a) to compute a full-confidence curve for the change-point

parameter τ. More details on the method can be found in Cunen, Hermansen & Hjort (2018a) and Cunen, Hjort & Nygård (2018b), but the important part is that it is based on the profile log-likelihood $\ell_{B,\text{prof}}(\tau)$, and then using simulations in order to compute the distributions of the deviance at each potential change-point (which is far from χ^2_1 for a change-point parameter). The confidence curve for a change-point is unusual looking since it often will provide disjoint confidence sets; see the red curve based on the battle deaths in Figure 7.5. The curve reveals a point estimate for the change-point in 1950, but with considerable uncertainty; the years 1939 and 1965 are also considered likely candidates for the change.

We model the 'anti-war' Ngram with a simple normal model with an autoregressive correlation structure of order 1. We allow the change-point to influence both the expectation and variance parameters of the model, but it turns out that it is primarily the expectation that changes across the change-point (it increases). The correlation between consecutive years is high (0.80). From Figure 7.4 it is clear that there are at least one very clear change-point in the sequence of usage rates: from 1823 to 1914 'anti-war' is hardly used at all (dashed line in the figure), and then the use increases. This increase might reflect a genuine increase in usage, or simply that the Google Books corpus is less complete for older texts. At any rate, we will assume that the change-point around 1914 (from no use to some use) is not the one we are interested in, but rather that the change in norms we are searching for must be reflected in a potential later change-point (from some use to more use). We will therefore only use the Ngram data for the years after 1914; one must bear in mind that this entails that the Ngram can only influence the change-point inference for the latter part of the full sequence of war years.

Using the autoregressive model and the method from Cunen, Hermansen & Hjort (2018a), we obtain another log-likelihood profile $\ell_{N,\text{prof}}(\tau)$ and also the full confidence curve based on the Ngram information (in blue in the left panel of 7.5). This curve has a point estimate at 1962, but with considerable confidence for the change rather taking place in 1927 or in 1971.

In the fusion step, the most straightforward solution is simply to sum the two log-likelihood profiles, calculate the deviance, and run the simulations to find the distribution of the deviance at each potential change-point (in the way as described in Cunen, Hermansen & Hjort (2018a)). This raises the question on whether it is appropriate to treat the two sources of information equally, however. There could be good reasons to consider the battle death data to be more directly informative for τ than the 'anti-war' data. These arguments invite a combined confidence log-likelihood of the form

$$\text{FF}: \qquad \ell_{\text{fus}}(\tau) = \ell_{B,\text{prof}}(\tau) + \omega \ell_{N,\text{prof}}(\tau), \quad \text{with } \omega \in [0,1].$$

The down-weighting parameter ω should reflect the degree of relevance of the soft data source. In most situations, there will be no information available that could help us estimate ω, and the parameter must therefore be chosen by the analyst; also, the effect of the choice should be communicated clearly and openly. In the right panel of Figure 7.5 we display the confidence curve with $\omega = 0.2$ (in light violet), along with the curve without down-weighting ($\omega = 1$, in dark violet). Both combined curves indicate a point-estimate of 1965; this was not the point-estimate in any of the two sources, but that year had a high confidence in both. The combined curve with no downweighting gives the appearance of higher precision than the light violet curve, but this might be misleading if we do not trust the 'soft data' fully. Then we might prefer the combined curve with $\omega = 0.2$, which is more similar to the original curve based on the battle death information only.

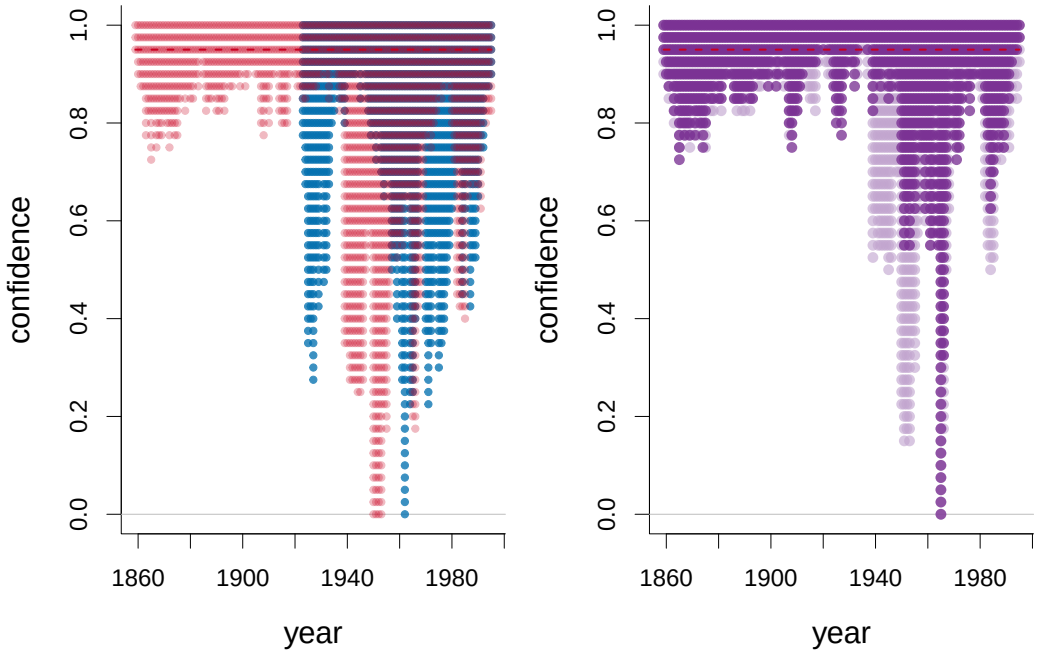

Figure 7.5: Left panel: in red the confidence curve based on the battle death data (point estimate 1950), in blue the one based on the Ngram (point estimate 1962). Right panel: combined confidence curves, dark violet with no down-weighting, light violet with down-weighting of the Ngram information. Both these two curves give a point estimate equal to 1965.

We must emphasise that we do not recommend the use of this subjective down-weighting in most combination settings. In usual settings, the 'degree of informativeness' of each source is already sufficiently well represented by the likelihood component from that source. However, down-weighting can be considered in situations like the present one, with combination of soft and hard data, where there might be stark differences in quality or relevance between the sources.

In this application we have illustrated a situation where one information source was considered to be of higher quality and relevance than the other; a combination of hard and soft data. Note that such combination attempts often require the users to make strong assumptions, for example that very different sources inform on the exact same parameter. Low-quality, large-data sources are expected to play an increasingly important role in statistics in years to come (especially via scraping of the internet). The combination of such data sources with more high-quality sources raises various issues, and we will end with a note of caution. In a best case scenario, the analyst manages to be benefit from a large, low-quality source and can obtain more precise statements than those from the smaller, high-quality sources alone. In the worst case scenario, the analyst is contaminating good data with irrelevant noise, and does not learn anything of value.

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 comments from Steven Pinker, and to Tore Schweder for always fruitful discussions related to issues and methods worked with in this paper.

References

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–322.

Berger, J. O., Liseo, B. & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science* **14**, 1–28.

Breslow, N. (1981). Odds ratio estimators when the data are sparse. *Biometrika* **68**, 73–84.

Cheng, J. Q., Liu, R. Y. & Xie, M.-g. (2017). Fusion learning. *Wiley StatsRef: Statistics Reference Online* .

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

Cox, D. R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society, Series B* **49**, 1–39.

Cox, D. R. & Reid, N. (1993). A note on the calculation of adjusted profile likelihood. *Journal of the Royal Statistical Society, Series B* **55**, 467–471.

Cunen, C., Hermansen, G. & Hjort, N. L. (2018a). Confidence distributions for change-points and regime-shifts. *Journal of Statistical Planning and Inference* **195**, 14–34.

Cunen, C. & 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.

Cunen, C., Hjort, N. L. & Nygård, H. (2018b). Statistical sightings of better angels: Analysing the distribution of battle deaths in interstate conflict over time. *Submitted for publication* .

Diciccio, T. & Efron, B. (1992). More accurate confidence intervals in exponential families. *Biometrika* **79**, 231–245.

DiCiccio, T. J., Martin, M. A., Stern, S. E. & Young, G. A. (1996). Information bias and adjusted profile likelihoods. *Journal of the Royal Statistical Society, Series B* **58**, 189–203.

Dominici, F. & Parmigiani, G. (2000). Combining studies with continuous and dichotomous responses: A latent-variables approach. In *Meta-analysis in Medicine and Health Policy*. CRC Press, pp. 99–118.

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.

Hardy, R. J. & Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* **15**, 619–629.

Hjort, N. L. & Schweder, T. (2018). Confidence distributions and related themes [introduction to the special issue, by the guest editors]. *Journal of Statistical Planning and Inference* **195**, 1–13.

Kristensen, K., Nielsen, A., Berg, C., Skaug, H. & Bell, B. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software* **70**, 1–21.

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.

McCullagh, P. & Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society. Series B* **52**, 325–344.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, Matthew K Brockman, W., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A. & Aiden, E. L. (2010). Quantitative analysis of culture using millions of digitized books. *Science* **331**.

Narum, S., Westergren, T. & Klemp, M. (2014). Corticosteroids and risk of gastrointestinal bleeding: a systematic review and meta-analysis. *BMJ Open* **4**.

Noma, H. (2011). Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Statistics in Medicine* **30**, 3304–3312.

O'Rourke, K. (2008). *The Combining of Information: Investigating and Synthesizing What is Possibly Common in Clinical Observations or Studies Via Likelihood.* Ph.D. thesis, University of Oxford.

Partlett, C. & Riley, R. D. (2017). Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. *Statistics in Medicine* **36**, 301–317.

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.

Pinker, S. (2011). *The Better Angels of Our Nature: Why Violence Has Declined.* Toronto: Viking Books.

Sarkees, M. R. & Wayman, F. W. (2010). *Resort to War: a Data Guide to Inter-state, Extra-state, Intra-state, and Non-state Wars, 1816-2007.* Washington, DC: CQ Press.

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.

Stern, S. E. (1997). A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *Journal of the Royal Statistical Society, Series B* **59**, 653–665.

Stijnen, T., Hamza, T. H. & Özdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine* **29**, 3046–3067.

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.

Whitehead, A., Bailey, A. J. & Elbourne, D. (1999). Combining summaries of binary outcomes with those of continuous outcomes in a meta-analysis. *Journal of Biopharmaceutical Statistics* **9**, 1–16.

Xie, M., Singh, K. & Strawderman, W. E. (2011). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association* **106**, 320–333.

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.

Zerbini, A. N., Clapham, P. J. & Wade, P. R. (2010). Assessing plausible rates of population growth in humpback whales from life-history data. *Marine Biology* **157**, 1225–1236.