Confidence distributions: a potential bridge for combining parametric and non-parametric analyses

Céline Cunen (joint work with Nils Lid Hjort)

University of Oslo

22/05/2017
The Problem - Combination of information

We have independent data sources $1, \ldots, k$ providing information about parameters $\psi_1, \ldots, \psi_k$. Our interest is in the overall focus parameter $\phi = \phi(\psi_1, \ldots, \psi_k)$.

- II-CC-FF: a semi-general framework to provide inference for $\phi$ in cases like this

- Beyond ordinary Meta-analysis:
  - not restricted to cases where the sources inform on the same parameter
  - we can deal with cases where we only have summary statistics from some or all of the sources
  - we can handle very diverse sources
- Confidence distributions
- II-CC-FF - general procedure and some illustrations
- Examples combining parametric and non-parametric analyses
Confidence distributions (CD)

- $\approx$ a posterior without having to specify a prior
- a sample-dependent distribution function on the parameter space
- can be used for inference (for example for constructing CIs of all levels)
Requirements for CDs

Definition

A function $C(\theta, Y)$ is called a confidence distribution for a parameter $\theta$ if:

- $C(\theta, Y)$ is a cumulative distribution function on the parameter space
- at the true parameter value $\theta = \theta_0$, $C(\theta_0, Y)$ as a function of the random sample $Y$ follows the uniform distribution $U[0,1]$

- The second requirement ensures that all confidence intervals have the correct coverage.
- More on CDs in *Confidence, Likelihood, Probability.* (Schweder and Hjort, 2016.)
II-CC-FF - overview

Combining information, for inference about a focus parameter $\phi = \phi(\psi_1, \ldots, \psi_k)$:

**II: Independent Inspection:** From data source $y_j$ to estimates and intervals, in the form of a confidence distribution/curve:

$$y_j \implies C_j(\psi_j)$$

**CC: Confidence Conversion:** From the confidence distribution to a confidence log-likelihood,

$$C_j(\psi_j) \implies \ell_{c,j}(\psi_j)$$

**FF: Focused Fusion:** Use the combined confidence log-likelihood $\ell_c = \sum_{j=1}^{k} \ell_{c,j}(\psi_j)$ to construct a CD for the given focus $\phi = \phi(\psi_1, \ldots, \psi_k)$, often via profiling:

$$\ell_c(\psi_1, \ldots, \psi_k) \implies C_{\text{fusion}}(\phi)$$
Illustration I: Classic meta-analysis

Assume all sources inform on the exact same parameter $\psi_1 = \cdots = \psi_k = \psi$, and that each source provide estimators $\hat{\psi}_j$ that are normally distributed $N(\psi, \sigma_j^2)$ with known $\sigma_j$s.

II: Data source $y_j$ leads to $C_j(\psi) = \Phi((\psi - \hat{\psi}_j)/\sigma_j)$.

CC: From $C_j(\psi)$ to $\ell_{c,j}(\psi) = -\frac{1}{2}(\psi - \hat{\psi}_j)^2/\sigma_j^2$.

FF: Summing $\ell_c(\psi) = \sum_{j=1}^k \ell_{c,j}(\psi)$ leads to the classic answer

$$\hat{\psi} = \frac{\sum_{j=1}^k \hat{\psi}_j/\sigma_j^2}{\sum_{j=1}^k 1/\sigma_j^2} \sim N\left(\psi, (\sum_{j=1}^k 1/\sigma_j^2)^{-1}\right).$$
Illustration II: Abundance of Humpback whales

Two studies, from 1995 and 2001, summarised as 2.5%, 50%, 97.5% confidence quantiles – based on very different types of data and very different statistical methods:

<table>
<thead>
<tr>
<th>Year</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>3439</td>
<td>9810</td>
<td>21457</td>
</tr>
<tr>
<td>2001</td>
<td>6651</td>
<td>11319</td>
<td>21214</td>
</tr>
</tbody>
</table>

We assume that the population has not changed and our focus is thus the true abundance, $N = N_1 = N_2$.

II: We need a way to construct confidence distributions from these summaries: $C_j(N) = \Phi((h_j(N) - h_j(\hat{N}_j))/s_j)$ with power transformation $h_j(N) = \text{sgn}(a_j)N^{a_j}$. We need to estimate $(a_j, s_j)$ for both years.
Illustration II: Abundance of Humpback whales

FF: $\ell_c(N) = \ell_{c,1}(N) + \ell_{c,2}(N)$ and Wilks approximation $cc(N) = \Gamma_1(2\{\ell_c(\hat{N}) - \ell_c(N)\})$. 
CC - Confidence conversion

The most difficult step?

\[ C_j(\psi_j) \implies \ell_{c,j}(\psi_j) \]

In some cases we will already have a log-likelihood for \( \psi_j \) from the II-step and then there are no problems.

In other cases, the confidence curves from the II-step are not constructed via likelihoods.

Then we need to do something else (and be more careful). A simple and general method - the normal conversion:

\[ \ell_c(\psi) = -\frac{1}{2} \Gamma_1^{-1}(cc(\psi, y)) = -\frac{1}{2} \left\{ \Phi^{-1}(C(\psi, y)) \right\}^2. \]
What can we do?

- Deal with summary statistics
- Deal with complex functions of the parameters from each source: \( \phi = \phi(\psi_1, \ldots, \psi_k) \)
- Deal with very diverse sources - for example combining parametric and non-parametric analyses
Example I - difference between medians

With simulated data. We have two sources:

- A large source (250 observations): here we compute an exact, non-parametric cc for the median $\mu_1$ by using properties of order statistics
- A small source (20 observations): here we assume normally distributed data and obtain a cc for $\mu_2$

We are interested in $\psi = \mu_2 - \mu_1$.

One way to compute a non-parametric cc for the median:

- Order your $m$ observations $(Y_{(1)}, \ldots, Y_{(m)})$
- Consider all $m/2$ intervals of the type $[Y_a, Y_{m-a}]$
- Compute the exact coverage of each interval by

$$r_m(a, m - a) = \int_0^{1/2} \left\{ 1 - \text{Be} \left( \frac{1}{2} - \frac{u}{1-u}, m, a + 1 \right) \right\} \text{be}(u, a, m - a + 1) du$$

- The confidence curve is $\text{cc}(\mu_1) = \min \{ r_m(a, m - a) : \mu_1 \in [Y_a, Y_{m-a}] \}$
Example I - difference between medians

**CC:** normal conversion \( \ell_{c,1}(\mu_1) = -\frac{1}{2} \Gamma_1^{-1}(cc(\mu_1, y)) \)

**FF:**
- Summing: \( \ell_c(\mu_1, \mu_2) = \ell_{c,1}(\mu_1) + \ell_{c,2}(\mu_2) \)
- Profiling: \( \ell^*(\psi) = \max\{\ell_c(\mu_1, \mu_2) : \mu_2 - \mu_2 = \psi\} \)
- Wilks approximation: \( cc_{fusion}(\psi) = \Gamma_1(2\{\ell^*(\hat{\psi}) - \ell^*(\psi)\}) \)
Example II - age at first word

We have two sources:

- **A large study:** 1640 parents report the age (in months) at which their child said its first word. Ranges from 1 (!) to 25.

- **A small study:** 51 parents report the age (in months) at which their child said its first word. Here we have some covariate information: gender (of the child).

**Possible hypothesis:** Girls speak earlier than boys?

**Model:** proportional hazards model (no censoring here - but we could have dealt with that too!)

Focus: probability that a child with covariate information $x_0$ does not speak at the age of 12 months

$$S(t_0|x_0) = e^{-H_0(t_0)e^{x_0t\beta}} = S_0(t_0)e^{x_0t\beta} = (1-F_0(t_0))e^{x_0t\beta} \text{ with } t_0 = 12.$$ 

- **Large study**: will give information about $F_0$ at $t_0$ - $\Longrightarrow cc_1(F_0(t_0))$. Non-parametric!

- **Small study**: will give information about $\beta \Longrightarrow cc_2(\beta)$. Cox model - Semi-parametric!

with II-CC-FF we can combine these and obtain a $cc$ for $S(t_0|x_0)$ with $t_0 = 12$. 
Example 2 - age at first word

Obtaining a confidence curve for the “baseline” $F_0(t_0)$.
An exact CD based on the binomial distribution.
Example 2 - age at first word

Obtaining a confidence curve for the coefficient $\beta$ (taking care to define gender as 1/-1, so that the value 0 corresponds to the overall mean)

Approximate CD based on the normal distribution.
Example 2 - age at first word

CC: we use normal conversion. FF: profiling

\[ \ell^*_c(S(t_0|x_0)) = \max\{ \ell_c(F_0(t_0), \beta) : (1 - F_0(t_0))e^{x_0^t\beta} = S(t_0|x_0) \} \]

and then Wilks approximation.
Example 2 - age at first word

Comparing with results from small source only.

Confidence curve

S(12)

Girl
Boy
Example 2 - age at first word

Comparing with results from merging small and large source.