Combining diverse information sources with the II-CC-FF paradigm, with applications in meta-analysis and beyond

FOCUS DRIVEN STATISTICAL INFERENCE WITH COMPLEX DATA

Céline Cunen (joint work with Nils Lid Hjort)

University of Oslo

24/11/2017

First, something about my other projects

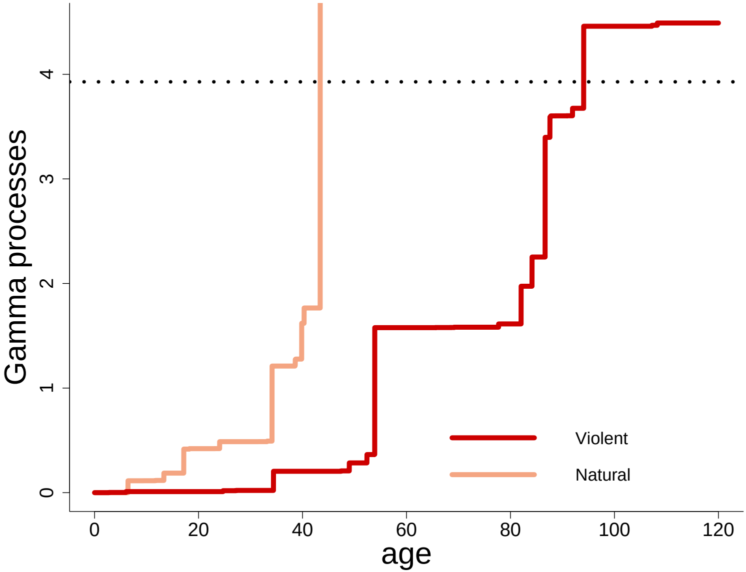
- Cunen, C., Hermansen, G. and Hjort, N.L. (2017). **Confidence distributions for change-points and regime shifts**. *Journal of Statistical Planning and Inference*.

- Model selection for meta-analysis of 2×2 tables using the focused information criteria (FIC) – in progress (with NLH).

- A FIC for linear mixed effect (LME) models:
  - An application: Cunen, C., Walløe, L. and Hjort, N.L. (2017). **Decline in energy storage in Antarctic Minke whales during the JARPA period: Assessment via the Focused Information Criterion (FIC)**. Reports of the Scientific Committee of the International Whaling Commision.
  - A "methods" paper on its way: **Model selection for linear mixed models via the Focused Information Criterion, with an application to whale ecology**.

- Competing risks with gamma process threshold crossing models (with NLH).

Gamma process models for competing risks

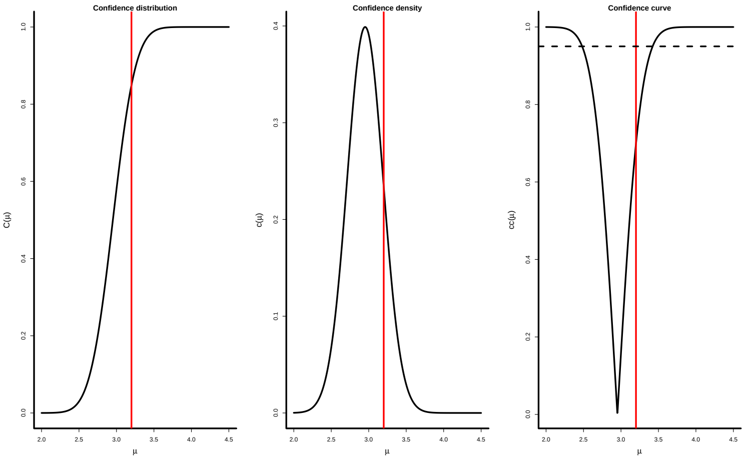


The Problem - Combination of information

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 general framework to provide inference for ϕ in cases like this.
  - Similar to *likelihood synthesis* from Schweder and Hjort (1996)
- Beyond ordinary meta-analysis:
  - not restricted to cases where the sources inform on the *same* parameter – we can deal with complex functions of the parameters from each source: $\phi = \phi(\psi_1, \ldots, \psi_k)$
  - we can deal with cases where we only have summary statistics from some or all of the sources
  - we can handle very diverse sources – for example combining parametric and non-parametric analyses

Confidence distributions (CD)



- ≈ 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 confidence intervals of all levels)

Requirements for CDs

Definition

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

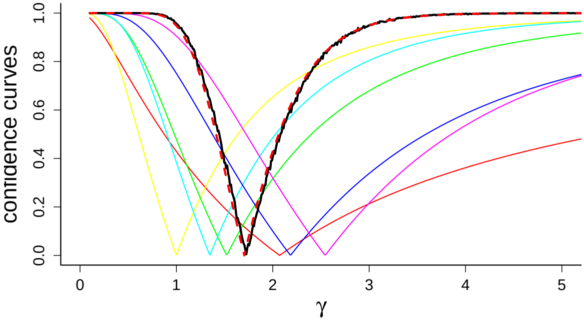
- The second requirement ensures that all confidence intervals have the correct coverage.

- Note that *any* method producing confidence intervals fulfilling these requirements can be used to make CDs (no-matter the underlying paradigm).

- More on CDs in *Confidence, Likelihood, Probability.* (Schweder and Hjort, 2016.)

Outline

- II-CC-FF - general procedure
- Examples illustrating what II-CC-FF can do
  - Classic meta-analysis
  - More complex meta-analysis: Blood loss
  - Random effects: All Blacks
  - Very diverse sources: First word



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_i to estimates and intervals, in the form of a confidence distribution/curve:

$$y_i \Longrightarrow C_i(\psi_i)$$

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

$$C_i(\psi_i) \Longrightarrow \ell_{c,i}(\psi_i)$$

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

$$\ell_f(\psi_1, \ldots, \psi_k) \Longrightarrow C_{fusion}(\phi)$$

Constructing CDs via Wilks' theorem

In many regular cases we have, at the true parameter value ψ_0:

$$2\{\ell_{n,prof}(\hat{\psi}) - \ell_{n,prof}(\psi_0)\} \to_d \chi^2_1,$$

as the sample size n increases. This gives us our favourite approximate confidence curve construction,

$$\mathrm{cc}(\psi) = \Gamma_1(2\{\ell_{n,prof}(\hat{\psi}) - \ell_{n,prof}(\psi)\}).$$

CC - Confidence Conversion

The most difficult step?

$$C_i(\psi_i) \Longrightarrow \ell_{c,i}(\psi_i)$$

In some cases we will already have a log-likelihood for ψ_i 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:

$$\ell_c(\psi) = -\tfrac{1}{2}\Gamma_1^{-1}(\mathrm{cc}(\psi, y)) = -\tfrac{1}{2}\{\Phi^{-1}(C(\psi, y))\}^2.$$

[Note that the confidence log-likelihood is not equal to the log-confidence density ($\log \partial C(\psi)/\partial\psi$). Except under the normal model.]

We can deal with 1: Classic meta-analysis

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

II: Data source y_i leads to $C_i(\psi) = \Phi((\psi - \hat{\psi}_i)/\sigma_i)$.

CC: From $C_i(\psi)$ to $\ell_{c,i}(\psi) = -\frac{1}{2}(\psi - \hat{\psi}_i)^2/\sigma_i^2$.

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

$$\hat{\psi} = \frac{\sum_{i=1}^k \hat{\psi}_i/\sigma_i^2}{\sum_{i=1}^k 1/\sigma_i^2} \sim \mathrm{N}\left(\psi, (\sum_{i=1}^k 1/\sigma_i^2)^{-1}\right).$$

We can deal with 2: more complicated meta-analysis

We do not have access to the full dataset (only summaries), and the studies differ in their reported outcomes: some studies report continuous outcomes, others report counts of a binary outcome.

Example from Whitehead et al. (1999): Blood loss during labor. Does treatment with oxytocic drugs help reduce blood loss?
Total of 11 studies, 6 studies report summary statistics of the continuous outcome (the actual blood loss in ml):

	Treatment	n	Mean	SD
Study 1	Control	510	325.75	288.61
	Treatment	490	255.60	213.74
Study 2	...	...	...	...

5 studies report counts of a binary outcome (yes = blood loss greater than 500 ml):

	Treatment	Yes	No	Total
Study 7	Control	152	697	849
	Treatment	50	796	846
Study 8	...	...	...	...

2: Blood loss

Model: $y_{ij} = \alpha_i + \beta z_{ij} + \epsilon_{ij} \qquad \epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$

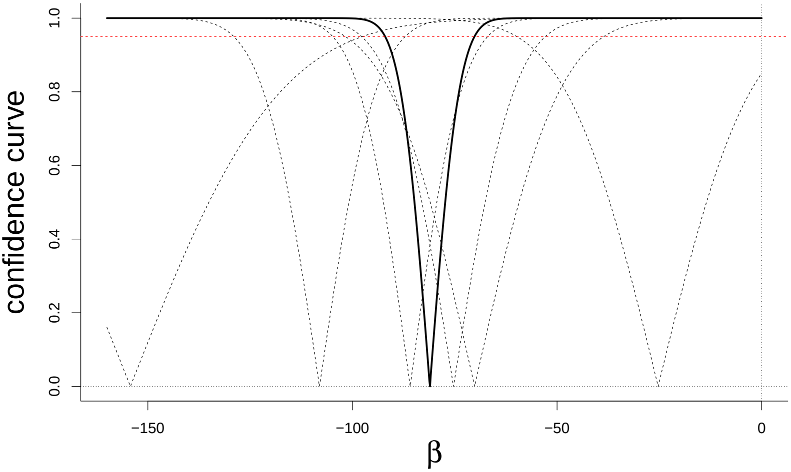
II and CC: If study i has a continuous outcome we have $\ell_{c,i}(\alpha_i, \beta, \sigma) = -(n_1 + n_2)\log(\sigma) - \{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + n_1(\bar{y}_1 - \alpha_i)^2 + n_2(\bar{y}_2 - \alpha_i - \beta)^2\}/(2\sigma^2)$.

If study i has a binary outcome we have $\ell_{b,i}(\alpha_i, \beta, \sigma) = x_{01}\log\{\Phi((500 - \alpha_i)/\sigma)\} + x_{11}\log\{1 - \Phi((500 - \alpha_i)/\sigma)\} + x_{02}\log\{\Phi((500 - \alpha_i - \beta)/\sigma)\} + x_{12}\log\{1 - \Phi((500 - \alpha_i - \beta)/\sigma)\}$.

FF: Summing $\ell_f(\alpha_1, ..., \alpha_k, \beta, \sigma) = \sum_{i=1}^{k_c} \ell_{c,i}(\alpha_i, \beta, \sigma) + \sum_{i=1}^{k_b} \ell_{b,i}(\alpha_i, \beta, \sigma)$ and then profiling $\ell_{f,prof}(\beta) = \ell_f(\hat{\alpha}_1(\beta), ..., \hat{\alpha}_k(\beta), \beta, \hat{\sigma}(\beta))$ and we get a combined confidence curve for β by using the Wilks' approximation

$$\mathrm{cc}(\beta) = \Gamma_1\{2(\ell_{f,prof}(\hat{\beta}) - \ell_{f,prof}(\beta))\}.$$

2: Blood loss



Oxytocic drugs is seen to decrease blood loss.

We can deal with 3: Random effects!

3: All Blacks

We have measures of "passage times" in 10 Rugby games ($\approx$ studies). There are 5 games before a certain change of rules and 5 after.

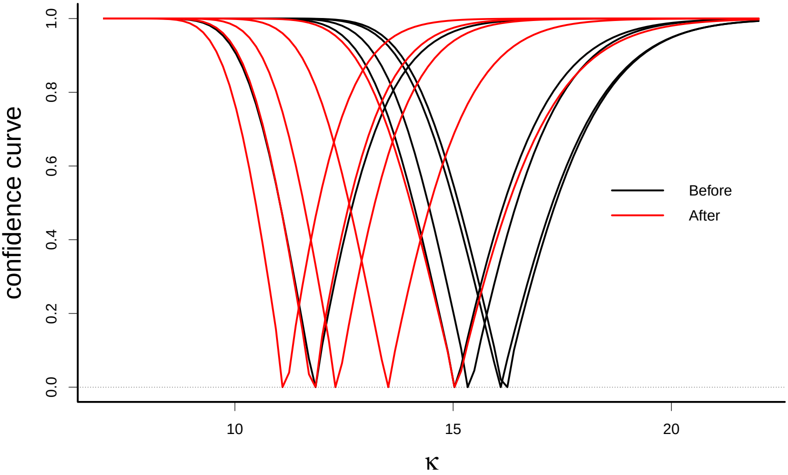
Model: $y_{ij} \sim \mathrm{Gamma}(a_i, b_i)$

Say we are interested in the standard deviation of passage times $\kappa_i = \sqrt{a_i}/b_i$.

It is relatively straightforward to construct confidence curves for each κ_i (by profiling and Wilks' approximation).

II: $\mathrm{cc}_i(\kappa_i) = \Gamma_1\{2(\ell_{i,prof}(\hat{\kappa}_i) - \ell_{i,prof}(\kappa_i))\}$

3: All Blacks



There seems to be smaller standard deviations in passage times (smaller κs) after the rule change.

But there is substantial spread in the mean κs between different games - do we need random effects?

3: All Blacks

We assume :

- Before rule change $\kappa_1, ...\kappa_5 \sim \mathrm{N}(\kappa_B, \tau_B^2)$ and
- After rule change $\kappa_6, ...\kappa_{10} \sim \mathrm{N}(\kappa_A, \tau_A^2)$.

We are interested in making confidence curves for κ_B, κ_A, and the ratio between them $\delta = \kappa_B/\kappa_A$.

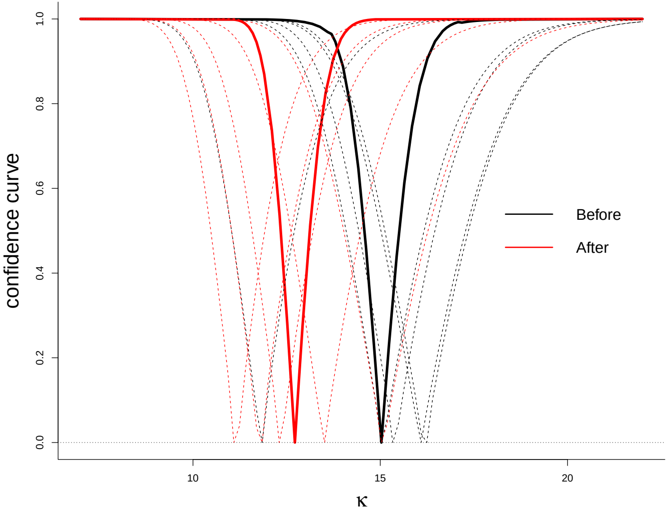
CC: We already have the $\ell_{i,prof}(\kappa_i)$ from the II-step, but now we need $\ell_i(\kappa_B, \tau_B)$ (and similarly for the parameters after the rule change):

$$\ell_i(\kappa_B, \tau_B) = \log[\int \exp\{\ell_{i,prof}(\kappa_i) - \ell_{i,prof}(\hat{\kappa}_i)\} \frac{1}{\tau_B} \phi\left(\frac{\kappa_i - \kappa_B}{\tau_B}\right) d\kappa_i]$$

(We integrate numerically or use Laplace approximation)

3: All Blacks

FF: Summing $\ell_f(\kappa_B, \tau_B) = \sum_{i=1}^{k_B} \ell_i(\kappa_B, \tau_B)$, profile to get $\ell_{f,prof}(\kappa_B)$ and Wilks' approximation to get $\mathrm{cc}(\kappa_B)$. Similarly for $\mathrm{cc}(\kappa_A)$



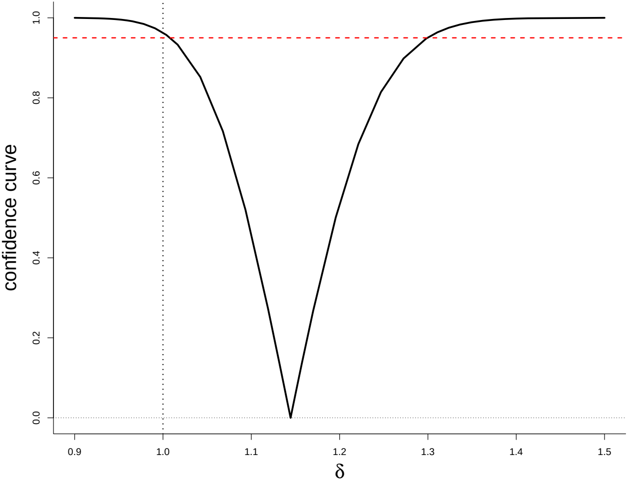
3: All Blacks - computation

For fast computation of the Laplace approximation and the subsequent profiling, we can make use of the TMB package (Template Model Builder, related to the ADMB). https://github.com/kaskr/adcomp/wiki

One simply(?) needs to write down an objective function/likelihood in C++. Then one specifies which parameters to profile out and which to integrate over.

3: All Blacks

The ratio $\delta = \kappa_B/\kappa_A$. Sum, profile and Wilks':
$\ell_{ff}(\kappa_B, \delta) = \ell_{f,prof,B}(\kappa_B) + \ell_{f,prof,A}(\kappa_B/\delta)$, then
$\ell_{ff}(\delta) = \ell_{ff}(\hat{\kappa}_B(\delta), \delta)$.



We can deal with 4: Very diverse sources

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).

Focus: When do girls start to speak? And when do boys start to speak?

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

Question: can the large (low-quality) study improve our analysis of the small study?

Data from: Schneider, Yurovsky & Frank (2015). Large-scale investigations of variability in children's first words. In CogSci2015 Proceedings.

4: Age at first word

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

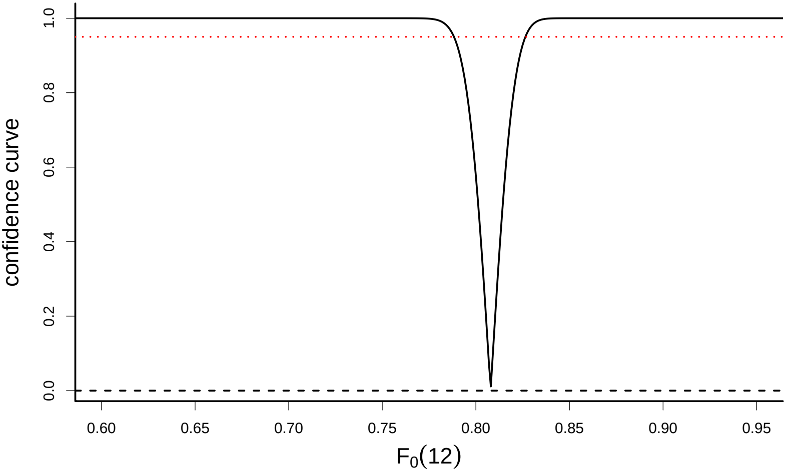
$$S(t_0|x0) = e^{-H_0(t_0)e^{x_0^t\beta}} = S_0(t_0)^{e^{x_0^t\beta}} = (1-F_0(t_0))^{e^{x_0^t\beta}} \quad \text{with } t_0 = 12.$$

- Large study: will give information about F_0 at t_0 - $\Longrightarrow \mathrm{cc}_1(F_0(t_0))$. Non-parametric!
- Small study: will give information about $\beta \Longrightarrow \mathrm{cc}_2(\beta)$. Cox model - Semi-parametric!

with II-CC-FF we can combine these and obtain a cc for $S(t_0|x0)$ with $t_0 = 12$.

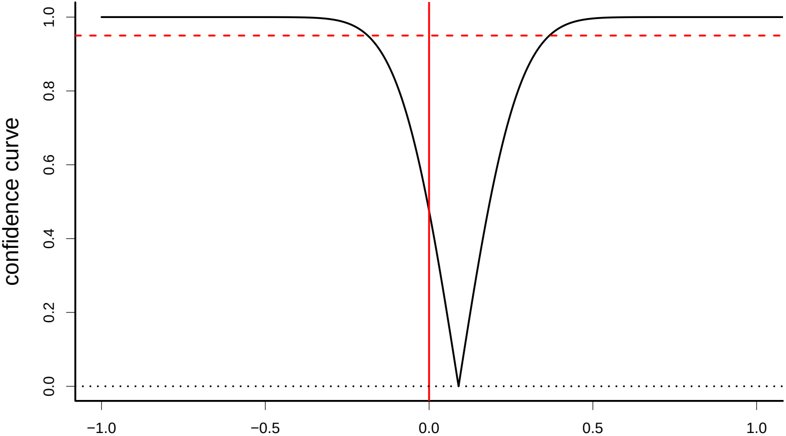
4: Age at first word

Obtaining a confidence curve for the “baseline” F_0 (at $t_0 = 12$).
An exact CD based on the binomial distribution.



4: Age at first word

Obtaining a confidence curve for the coefficient β (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 (here we only need the summary statistics: estimate and standard error).

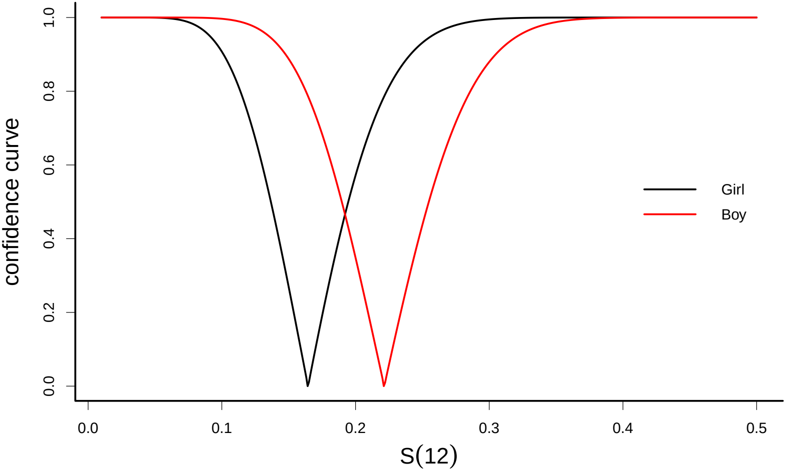


4: Age at first word

FF: summing $\ell_f(F_0, \beta) = \ell_1(F_0) + \ell_2(\beta)$ and profiling

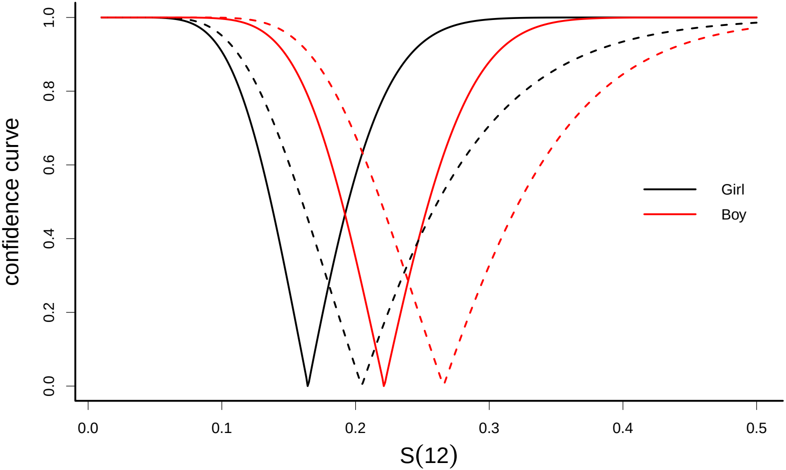
$$\ell_{f,prof}(S(t_0|x_0)) = \max\{\ell_f(F_0, \beta) : (1 - F_0(t_0))^{e^{x_0^t \beta}} = S(t_0|x_0)\}$$

and then Wilks' approximation.



4: Age at first word

Comparing with results from small source only.



Concluding remarks

What can II-CC-FF 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 (hard and soft data; improving inference from small, carefully designed studies by using large, "less-informative" datasets; ...)

Some remaining challenges:

- Neyman-Scott type problems
- Corrections for improving the Wilks' approximation
- Real applications!

References

- Cunen and Hjort (2016). Combining information across diverse sources: The II-CC-FF paradigm. *Proceedings from the Joint Statistical Meeting 2016, the American Statistical Association*.
- Liu, Liu and Xie (2015): Multivariate meta-analysis of heterogeneous studies using only summary statistics: efficiency and robustness. *JASA*.
- Schneider & Frank (2015). Large-scale investigations of variability in children's first words. *In CogSci2015 Proceedings*.
- Schweder & Hjort (1996). Bayesian synthesis or likelihood synthesis – what does Borel's paradox say? *Reports of the International Whaling Commision, 46*.
- Schweder & Hjort (2013). Integrating confidence intervals, likelihoods and confidence distributions. *Proceedings 59th World Statistics Congress*.
- Schweder & Hjort (2016). *Confidence, Likelihood, Probability*. Cambridge University Press.
- Whitehead, Bailey and Elbourne (1999). Combining summaries of binary outcomes with those of continuous outcomes in a meta-analysis. *Journal of Biopharmaceutical Statistics*.