FocuStat, CDs, this workshop ... with more to come

focustat

FOCUS DRIVEN STATISTICAL
INFERENCE WITH COMPLEX DATA

Nils Lid Hjort

Department of Mathematics, University of Oslo

Inference With Confidence, May 2015, Oslo

Outline

A FocuStat (2014-2018)

B This workshop & some of its visible themes

C Some less visible themes

D JSPI Special Issue: Confidence Distributions and Related Themes (2016)

A: The FocuStat five-year project

The FocuStat project and work group are partly funded by the Research Council of Norway, from Jan 2014 to Dec 2018. We are one professor + two PostDocs + two PhDs + other associated PhD and Master's level students + links to yet other associated colleagues and projects. Themes include and involve

- ϕ building bridges between parametrics and nonparametrics;
- ϕ combining diverse sources of information;
- ϕ focused model building, selection, averaging;
- ϕ confidence distributions;
- ϕ Bayesian nonparametrics;
- ϕ 'doing things', focused statistics with complex data.

Workshops 2015, 2016, 2017, 2018; annual 'research kitchens'; publishing papers + more (an edited book in 2018?); focus on methodology, but also on real applications; other activities.

Stay tuned – web page, Facebook page.

B: The Inference With Confidence workshop

Check & use the workshop programme, also regarding the practicalities (excursion and Confidence Dinner Tue evening, etc.).

Some of the themes covered:

a. What CDs are: If ψ is a parameter of interest, and

$$[\psi_\alpha, \psi_\beta] = [C^{-1}(\alpha), C^{-1}(\beta)]$$

is a confidence interval of degree $\beta - \alpha$, convert these to a confidence distribution $C(\psi)$ for ψ. Also, the confidence curve

$$\mathrm{cc}(\psi) = |1 - 2\,C(\psi)|$$

is such that $\mathrm{cc}(\psi) = 0.95$ has two solutions, the end points of a 95% confidence interval, etc.

b. Prototypical cases: With data $y_1, \ldots, y_n$ i.i.d. from $\mathrm{N}(\mu, \sigma^2)$:

$$C_1(\mu) = \Phi\Big(\frac{\mu - \bar{y}_{\rm obs}}{\sigma/\sqrt{n}}\Big), \quad C_2(\mu) = F_\nu\Big(\frac{\mu - \bar{y}_{\rm obs}}{\widehat{\sigma}_{\rm obs}/\sqrt{n}}\Big),$$

for σ known and not known, respectively (and $\nu = n - 1$ is the degrees of freedom). Also, from $\widehat{\sigma}^2/\sigma^2 \sim \chi^2_\nu/\nu$,

$$C_3(\sigma) = 1 - \Gamma_\nu(\nu\widehat{\sigma}^2_{\rm obs}/\sigma^2).$$

c. Defining property: $C(\psi, y_{\rm obs})$ is a cdf in ψ for each set of data $y_{\rm obs}$; and $U = C(\psi, Y) \sim \mathrm{unif}$. If $U_n = C_n(\psi, Y_n) \doteq_d \mathrm{unif}$, based on data Y_n of sample size n, then we call $C_n(\psi, y_n)$ an approximate or asymptotic CD.

d. How to construct them:

$$C(\psi) \doteq \Phi((\psi - \widehat{\psi})/\widehat{\kappa})$$

is a first-order correct approximation. Often better, via profile deviance $D_n(\psi) = 2\{\ell_{\rm prof}(\widehat{\psi}) - \ell_{\rm prof}(\psi)\}$:

$$\mathrm{cc}_n(\psi) \doteq \Gamma_1(D_n(\psi)).$$

e. Computations and algorithms (and bags of tricks): With the $C(\psi) = \mathrm{Pr}_\psi\{B \geq b_{\rm obs} \,|\, A = a_{\rm obs}\}$ recipe, how to simulate $B \,|\, (A = a_{\rm obs})$? See talks.

f. How to combine information: This is also related to the interplay between likelihoods and CDs. See various talks, and also the Invited Session on this theme in Rio, ISI 2015.

g. Links from CDs to objective and empirical Bayes: When is

$$\pi_n(\psi \,|\, \mathrm{data}) \propto \int \pi(\psi, \gamma) L_n(\psi, \gamma) \, \mathrm{d}\gamma$$

identical to or close to confidence density $c_n(\psi)$? See talks.

h. What does 'confidence' mean? Interpretation of CDs? They yield posteriors without priors. Different from Bayes? Epistemic vs. aleatory probabilities. See talks.

i. Performance and optimality: What does 'CD_1 is better than CD_2' mean? See talks.

j. CDs may also be used for unobserved variables, not only for parameters, as in prediction confidence distributions for time series, etc. See talks.

k. CDs may be constructed for discrete parameters (population sizes, the number of taxis, etc.). See talks.

l. History: CDs have had their ups and downs and ups (and ups (and ups)), since Fisher 1930, and the fiducial distributions have been generalised and made more generally applicable in various directions. See various talks.

C: Some CD themes not directly covered

a. 2nd order corrections: $\mathrm{cc}(\psi) = \Gamma_1(D(\psi)/(1+\varepsilon))$; magic formula; expansions; abc bootstrapping; more.

b. Post-selection inference: Most methods are worked out for
(data, given model, focus parameter)
but difficulties arise when the model is selected among candidates via e.g. AIC or FIC.

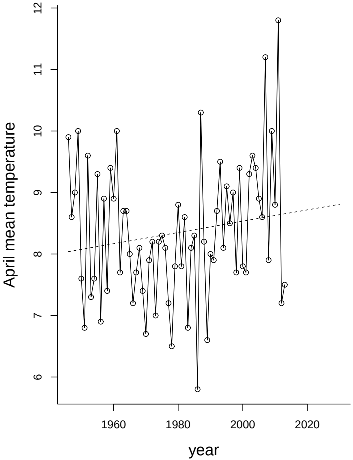
c. CDs associated with nonstandard parameters (and nonstandard estimators with nonstandard distributions), such as

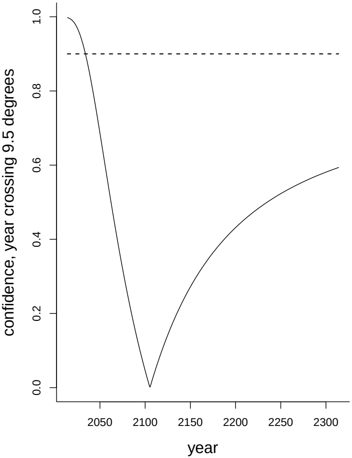
$$C_n(\theta) = 1 - H_n(n^{1/3}(\theta - \widehat{\theta})/\widehat{\kappa})$$

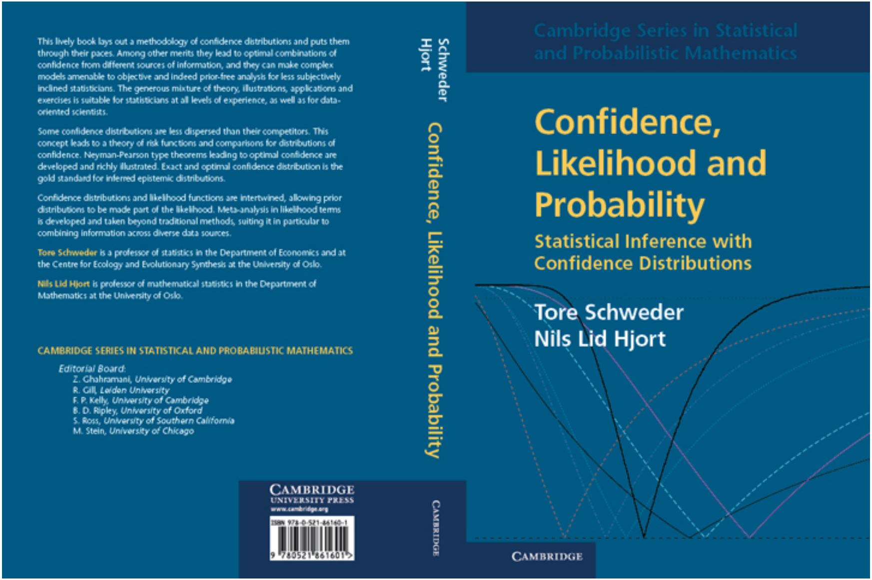
for various situations with cube-root asymptotics. Other nonstandard cases include nonparametric smoothing ($n^{2/5}$ rates), some setups with change-points (n rates), etc.

d. Multidimensional cases are harder (and Fisher went wrong!).

e. CDs for highly nonlinear parameters (where normal approximations are bad). When will mean April temperature cross 9.5 degrees Celsius (if ever)?







Existence proof: CLP, Cambridge University Press, 2015.

D: JSPI Special Issue on this workshop's themes

Special Issue on Confidence Distribution and Related Themes:

- Contributions from this workshop + a few other scholars
- About 15-20 (good!) papers, length 10-15 jspi pages
- Send 'yes!' to Nils and Tore, within May 20, along with tentative title + some lines of abstract
- An author may choose a CD theme different from what was presented here.

- Submission deadline Nov 15, 2015
- Referee work and reports by Mar 1, 2016
- All papers finished by June 15, 2016