FocuStat, BBB themes, this workshop ... with more to come

Nils Lid Hjort

Department of Mathematics, University of Oslo

Building Bridges, May 2017, Oslo

Outline

A FocuStat (2014-2018)

B Previous Workshops and Research Kitchens, Things Done, (more) Things to be Done

C This workshop & some of its themes

D 2017, 2018, ...

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 focused model building, selection, averaging;
- ϕ confidence distributions;
- ϕ building bridges between parametrics and nonparametrics;
- ϕ combining diverse sources of information;
- 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, blog.

B: Previous Workshops and Kitchens

May 2015: Inference With Confidence, confidence distributions and related themes (and applications); Special Issue of *Journal of Statistical Planning and Inference* coming out this year

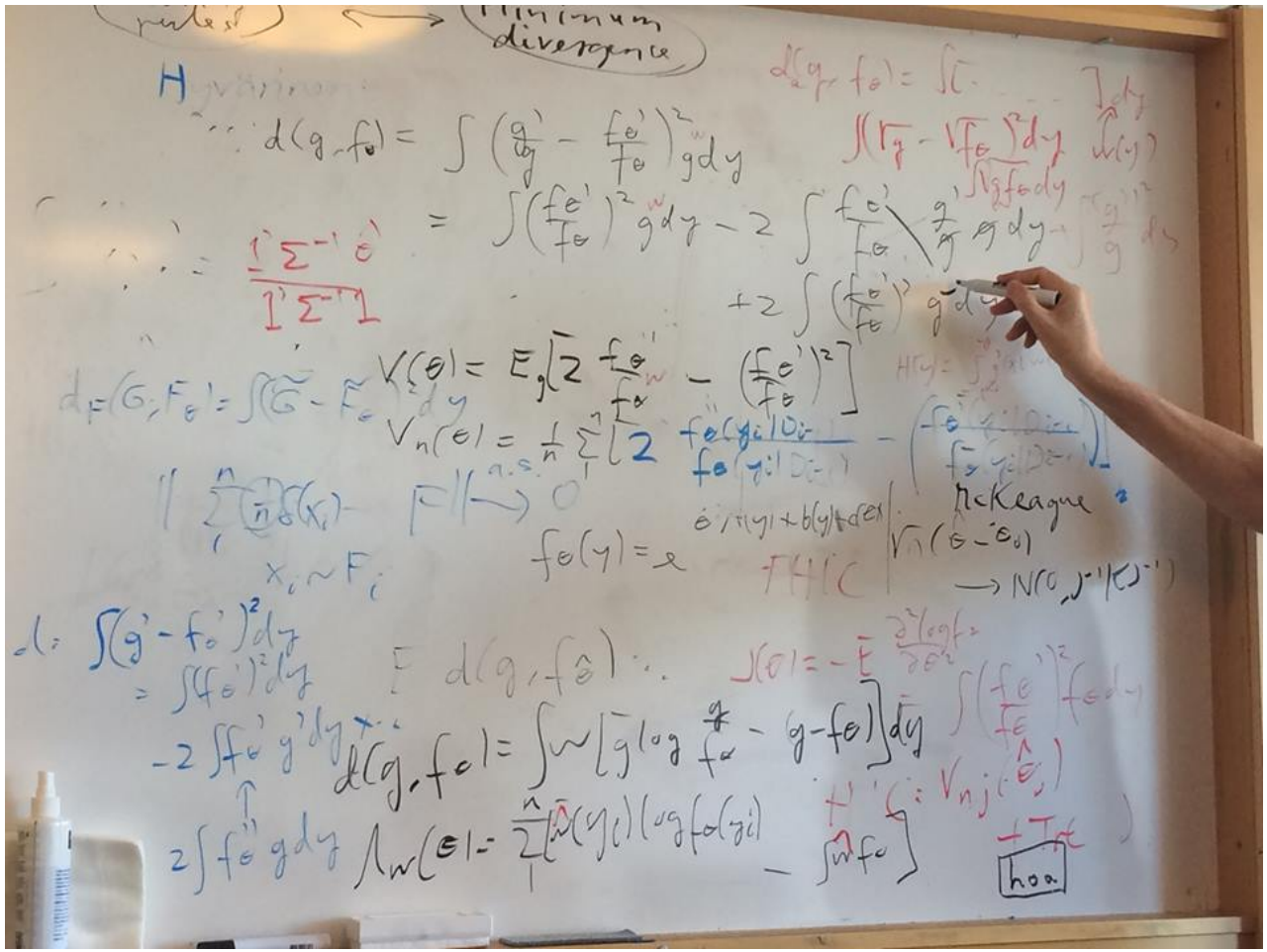
May 2016: FICology, focused model building, model selection, model averaging (with applications involving War & Peace and whaling politics)

May 2017: Building Bridges, bridging parametrics, semi- and nonparametrics

May 2018: perhaps a four-day bigger thing

Research kitchens

Two-three-four colleagues come visiting the group for a few days, just to foster talking and thinking (and perhaps even theorem proving & paper writing).



Autumn 2014: Ingrid Van Keilegom and Ian McKeague, on empirical likelihood (and cure models).

Autumn 2015: Fabian Krüger, Monica Musio, Thordis Thorarinsdottir, on minimum divergence and scoring rules.

Autumn 2016: Jeff Miller, Tamara Broderick, Peter Grünwald, Peter Müller, Per Mykland, on L^η, многая лета

Winter 2018: Combining information across diverse sources

Later in 2018: Multivariate dependence or Bayesian nonparametrics or CLP packaging or Moby-FIC or cure models or model building via stochastic processes or performance of complex estimators ...

Doing Things!

We're trying to get involved in 'real things', from analyses of datasets or stories that catch our fascination, to bigger stories.

The FocuStat Blog, *some* stories:

- Semifinals influence finals, in Olympic ski sprint
- $\Pr$(gold-medal is shared) in speedskating after four distances
- World's first novel (1460): we solve medieval literary mystery
- Game of Thrones vs. War of the Roses (1455-1487)
- Real Time Real Excitement Plots when watching handball
- Who votes what where?

Some of these stories are occasionally picked up by other media – Céline's GoT analyses; Céline-Gudmund-Nils with Tirant lo Blanch; Emil repeatedly on TV with Trump election; Vinnie's Met hartelijke groente sells well in the Netherlands; ...

FocuStat Blog, stories to come (I hope):

- Sentiment analysis of 100 Paul Simon songs [Sam-Erik]
- Whales, politics, and statisticians [Céline and Nils]
- Tour de France [Gudmund]
- The statistical comprehension level of politicians [Emil]
- Machine learning of French grammar [Céline]
- War and Peace [Gudmund and Nils]
- Personalised medicine [Kristoffer]
- Meteorological precision [Sam-Erik]
- Noorwegen door Nederlands-Koreaanse ogen [Vinnie]

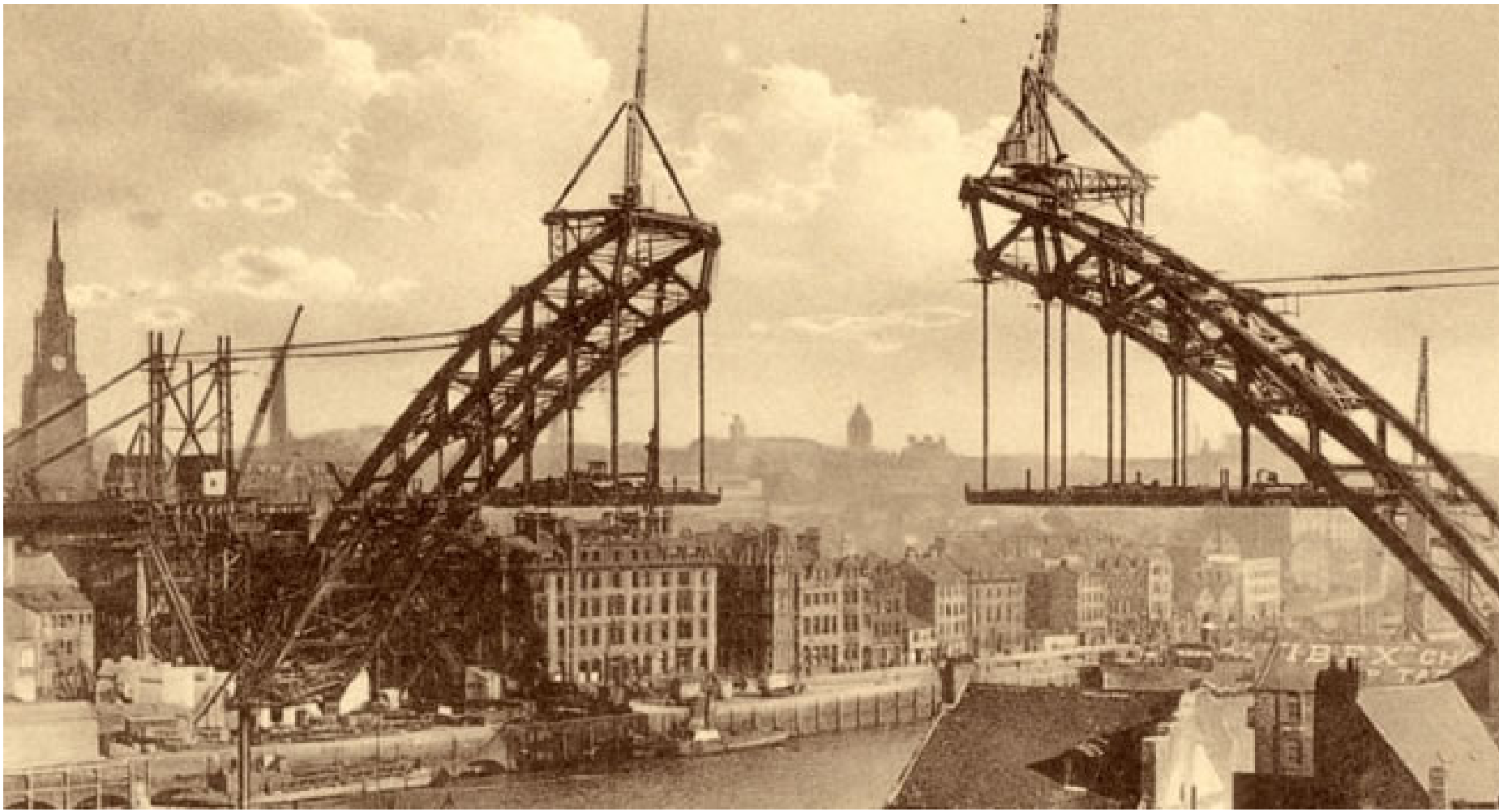
We also give talks and write papers (see our FocuStat website) and complete PhD projects and PostDocs get new jobs etc. ...

On The Road

ϕ Gudmund is behind Oslo Data Science Meetup, a grand success.
ϕ We wish to do War and Peace statistics, with PRIO.
ϕ Céline and Nils fought hard in the Scientific Committee of the International Whaling Commission (and have invented Moby-FIC).

C: Building Bridges

We are (approximately) here: Something (para) on the left; something (nonpara) on the right; how to build, or combine, or select, or to borrow, or to meld?



What is nonparametrics?

This portmanteau word means (and has meant) different thing to different statisticians (over time):

- tests with few assumptions (rank tests etc.);
- certain tests being distribution-free in the limit;
- analyses based on empirical distributions (Kolmogorov etc.);
- density estimation, nonparametric regression, smoothing
- remove constraints, if not needed;
- formulate assumptions in non-finite-parametric ways ('convex', 'monotone', 'low interactions')
- create a nonparametric envelope around a parametric structure
- finding structure in data, without clear models
- machine learning

... and we can all fill in more.

What is the what?

Well, what is it^c? Theorem: If the world is frequentist or Bayes, and parametric or nonparametric, then

$$\text{IV} = (\text{I} \cup \text{II} \cup \text{III})^c.$$

	frequentist	Bayes
parametric	I	II
nonparametric	III	IV

I: Smallish finite models, estimation and inference for aspects of θ.

II: Smallish finite models, estimation and posterior inference, via prior $\pi(\theta)$ (this was all of Bayes inference, from c. 1774 to c. 1973).

III: Bigger models, density estimation, nonparametric regression, confidence bands, etc.

IV: Priors and posteriors for random functions, bigger structures, hierarchies of hierarchies, ...

Some bridging themes

a. Bayesian nonparametrics (Sonia, Igor): e.g. nonparametric envelope around parametric model, as in

$$\bar{F}(y) = H(F(y,\theta)), \text{random } H \text{ centred at uniform.}$$

b. Add on ε neighbourhood around parametric structure,

$$G = (1-\varepsilon)F_\theta + \varepsilon H,$$

with H left unspecified.

c. Many density estimation schemes combine para and nonpara (Ingrid, Dag, Bård, Håkon):

$$\begin{aligned}
\widehat{f}(x) &= f(x,\widehat{\theta})\widehat{r}(x); \quad \text{[Hjort–Glad]} \\
\widehat{f}(x) &= f(x,\widehat{\theta}(x)); \quad \text{[Hjort–Jones]} \\
\widehat{f}(x) &= (\widehat{f}_{\rm para}(x)\widehat{f}_{\rm nonpara}(x))^{1/2}/k; \\
\widehat{f}(x) &= \widehat{w}_0 f_{\rm para}(x) + \widehat{w}_1 f_{\rm nonpara}(x); \quad \text{[Olkin]}
\end{aligned}$$

etc., with parallels for para and nonpara regression.

d. Growing models, dimension of parameter vector grows with sample size (Riccardo, Christian R.?). Example (Sam-Erik):

$$f(y) = \exp\{\theta_1 T_1(y) + \cdots + \theta_p T_p(y) - c_p(\theta_1, \ldots, \theta_p)\}.$$

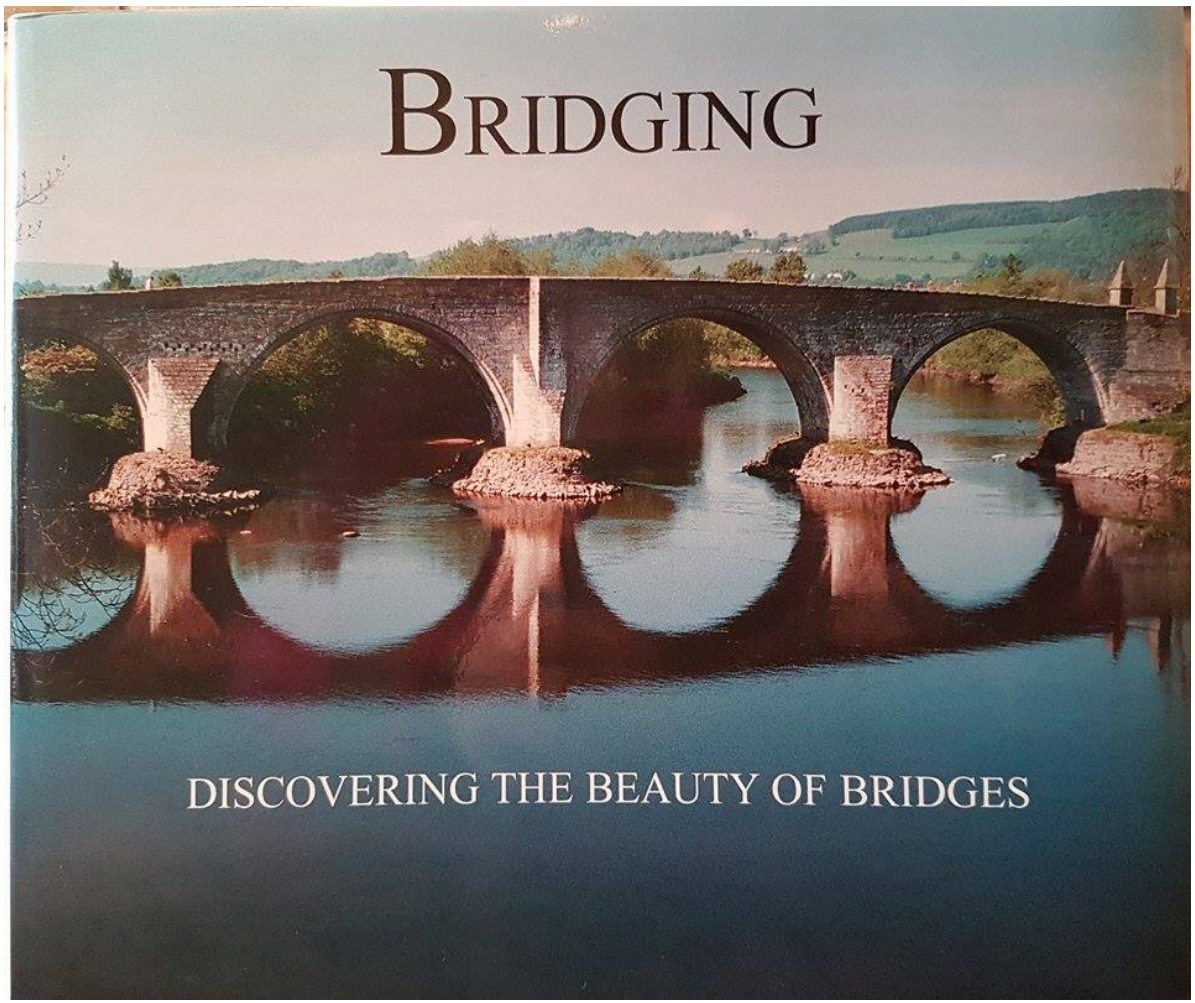
e. Combining para and nonpara information, e.g. via confidence curves and translations to log-likelihoods (Céline).

f. Regularising models and likelihoods via nonparametric controls (Gudmund, Nils).

g. Letting data choose between competing parametric models *and* nonparametrics (Martin, Vinnie, Sam-Erik); averaging over the best.

E: FocuStat 2016, 2017, 2018, ...

FocuStat Workshop May 2017: Building Bridges, themes connecting parametrics, semiparametrics, nonparametrics.



FocuStat Conference (May) 2018: could be a bit bigger than Workshops 2015, 2016, 2017 (and perhaps an edited book).