Part II: Confidence Distributions & Data Fusion

Nils Lid Hjort

Department of Mathematics, University of Oslo

Geilo Winter School, January 2017

[Note: This is the pdf version of the 2 x 45 minutes Nils Talk II I gave at the Geilo Winter School, January 2017. In my actual presentation I of course did both of (a) saying quite a bit more than is on the page and (b) skidding semi-quickly over chunks of the material, including parts of the mathematics, complete with the usual mixture of hand-waving, glossing over technicalities, and swiping of details under imaginary carpetry. The pdf notes themselves are meant to be decently coherent, though, and may be suitable for study.]

Themes (and background): Confidence distributions; likelihood analysis (with correction tricks); Holy Grail (says Efron): posteriors without priors; optimal inference; GLM (and GLLM); meta-analysis (fusion); empirical likelihood; ..., and applications.

Greater cohesion for statistical inference (making the distance from 'Bayes' to 'Frequentist' a smaller one).

Three revolutions in (parametric) statistical inference: Laplace (1774); Gauss and Laplace (1809–1812); Fisher (1922). There's an ongoing fourth revolution:

- who and how;
- new data, new needs, new methods, new perspectives.

I see CDs (and related tools) as fruitful, promising, not-yet-in-full-bloom methods to play important roles in Revolution Four:

conception, computation, communication of statistical evidence.

Some literature:

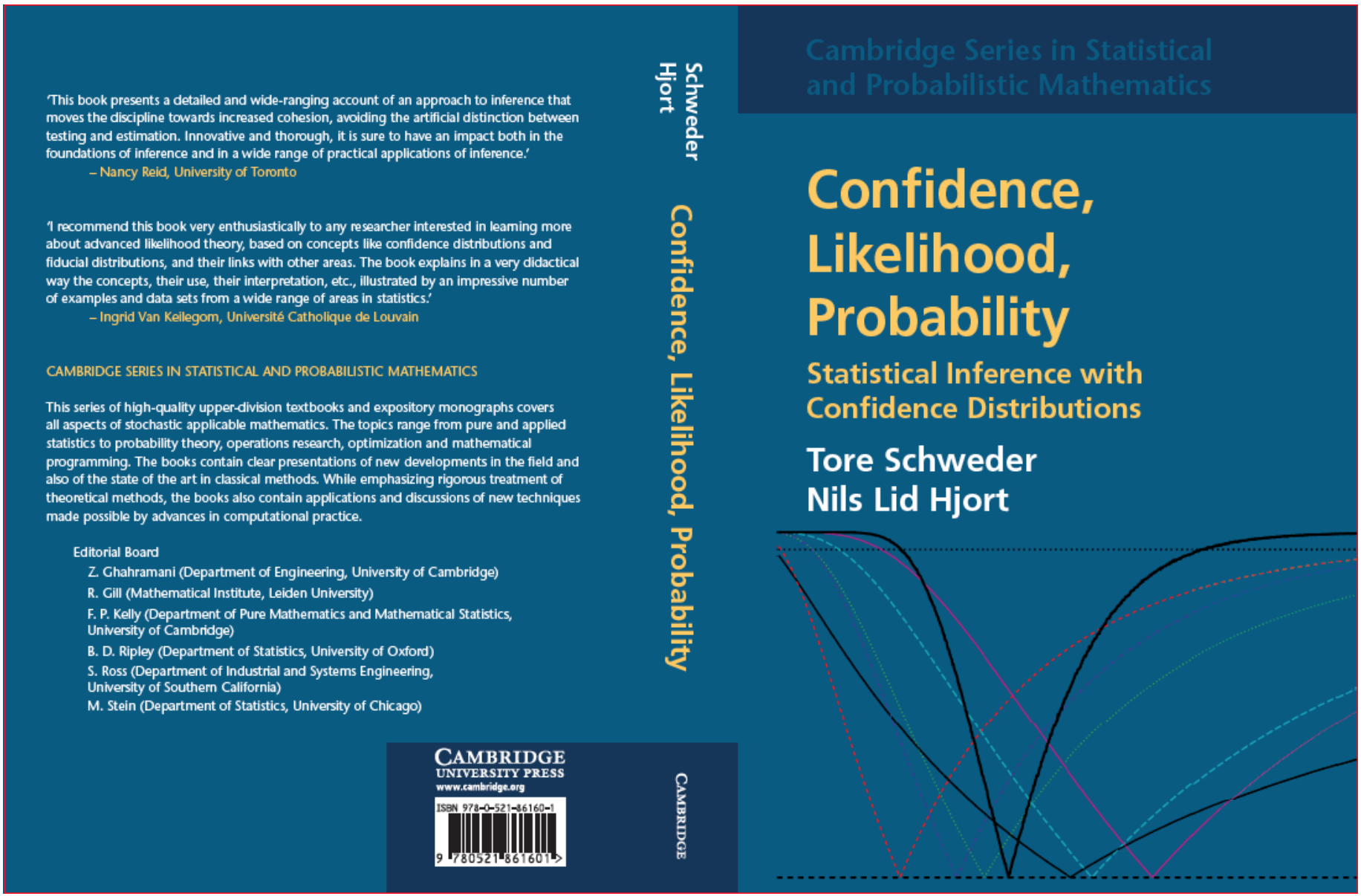
- Fisher, Bartlett, Neyman, many others (1930 to c. 1956);
- Cox, Fraser, Hacking, some others (c. 1958 to c. 1990);
- recent upsurge, many papers, review paper 2013 Xie and Singh (discussants Cox, Efron, Fraser, Parzen, Robert, Schweder and NLH);
- Schweder and NLH (various papers since 1996, plus CUP 2016 book: *Confidence, Likelihood, Probability*);
- JSPI special issue on Confidence Distributions and Related Themes (2017); ...

Some conferences: BFF 1 2014, BFF 2 2015, BFF 3 Rutgers 2016, CDs Oslo 2015, BFF 4 Harvard 2017, BFF 5 Ann Arbor 2018, ...

Plan & outline

1. Holy Grail: distribution for ψ given data, without a prior
2. Confidence distributions, confidence curves, confidence densities
3. Recipes for construction of CDs
4. Partly nonstandard cases (Neyman–Scott, Fieller, boundary parameters, disagreeing with Sims, Nobel Prize 2012)
5. Optimality
6. Better approximative CDs (via various modification tricks)
7. GLMs (and GLLMs)
8. Fusion and meta-analysis: II-CC-FF
9. CDs for prediction
10. Extensions, related themes, questions

CLP, 2016 – see also course website for STK 4180, UiO.



1: The Holy Grail

The Holy Grail of statistics (says Brad Efron): reaching $\pi(\psi \,|\, \mathrm{data})$ without a prior.

It can also be an ordinary & well-working IKEA cup:

Classical framework:

data y;

model $f(y, \theta)$ with $\theta = (\theta_1, \ldots, \theta_p)$;

focus parameter $\psi = \psi(\theta)$.

Aim: probability distribution for ψ given data.

What does Bayes say (in the modern interpretation)?

$$\pi(\theta \,|\, \text{data}) \propto \pi_0(\theta) L(\theta),$$

then integrating out:

$$\pi(\psi \,|\, \text{data}) = \int_{\theta:\ \psi(\theta)=\psi} \pi(\theta \,|\, \text{data})\, \mathrm{d}\theta.$$

This is wondrous – but (at least) two problems: (i) We need $\pi_0(\theta)$ from 'somewhere'; (ii) there are perhaps two types of probabilities at work (so some claim Bayes' theorem can't be applied).

Aim of confidence distributions (CDs): turn data information y, via the model $f(y, \theta)$, into a good, clear (and sometimes optimal) distribution for focus parameter $\psi = \psi(\theta_1, \ldots, \theta_p)$ given data – with no prior and no Bayes theorem.

There are several ideas (related, sometimes equivalent, depending on the framework and conditions) leading to such CDs – from Fisher's fiducial (1930) to pivot transformations to inversions of confidence intervals.

Simplest version (perfect when it works): if we have beautiful confidence intervals of all levels (0 to 1), convert them to a distribution, and derive its confidence density. This is the Holy Grail (or Ikea Cup).

2: Confidence distributions, confidence densities, confidence curves

Example: $y_1, \ldots, y_n$ i.i.d. from exponential θ. With $n = 10$ and $\bar{y}_{\text{obs}} = 0.345$, what can we say about θ?

The log-likelihood is

$$\ell(\theta) = n(\log\theta - \theta\bar{y}),$$

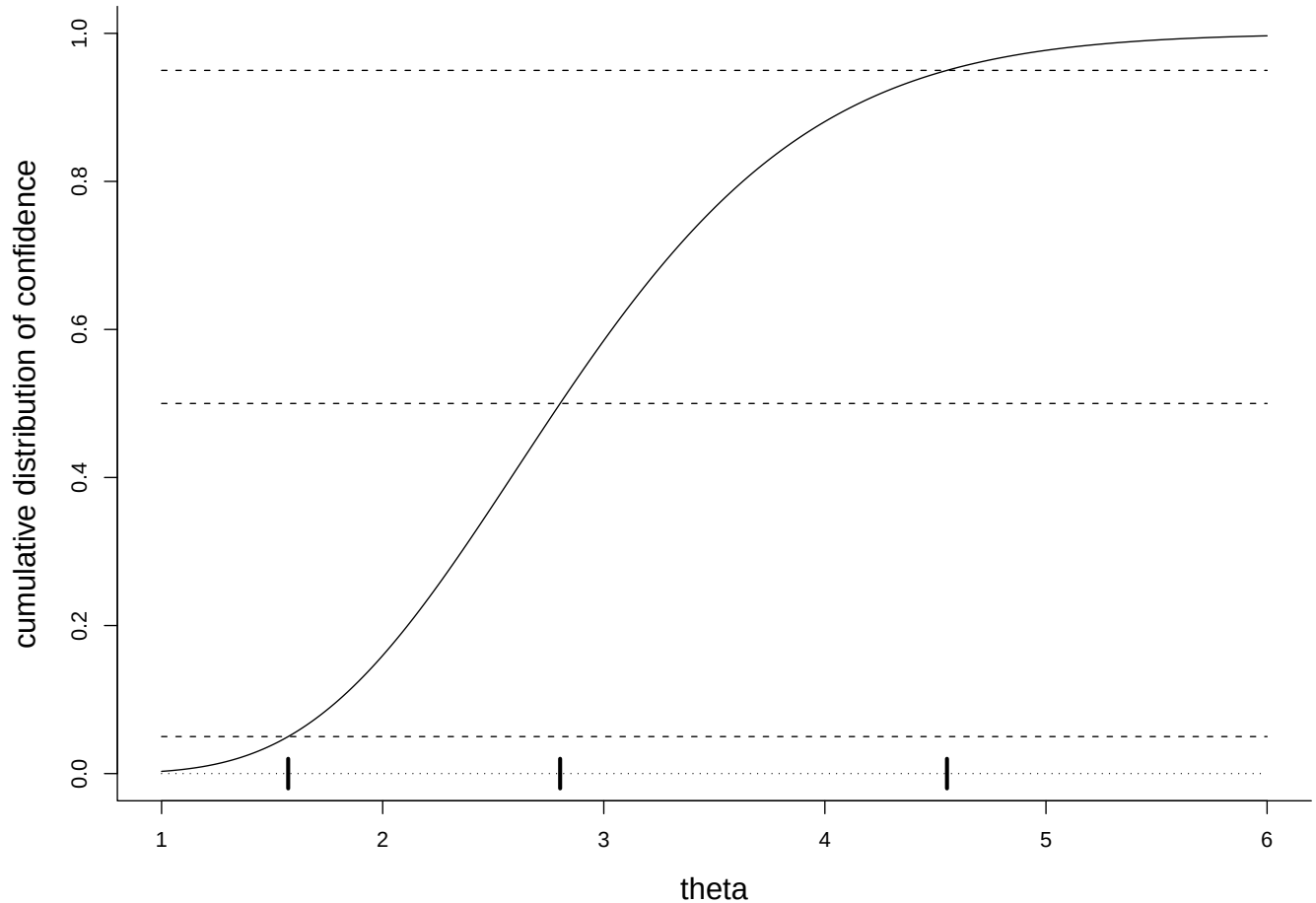
with ML $\widehat{\theta} = 1/\bar{y}$. Also, $\widehat{\theta}/\theta$ is a pivot – distribution G not depending on parameters. We have CD

$$\begin{aligned} C(\theta) &= \Pr_\theta\{\widehat{\theta} \geq \widehat{\theta}_{\text{obs}}\} = \Pr_\theta\{\widehat{\theta}/\theta \geq \widehat{\theta}_{\text{obs}}/\theta\} \\ &= 1 - G(\widehat{\theta}_{\text{obs}}/\theta) = \Gamma_{2n}(2n\theta/\widehat{\theta}_{\text{obs}}). \end{aligned}$$

It's also a post-data p-value function. The confidence density is

$$c(\theta) = \gamma_{2n}(2n\theta/\widehat{\theta}_{\text{obs}})2n/\widehat{\theta}_{\text{obs}}.$$

Interpretation: $[\theta_\alpha, \theta_\beta] = [C^{-1}(\alpha), C^{-1}(\beta)]$ has confidence $\beta - \alpha$: $[C^{-1}(0.33), C^{-1}(0.34)]$ has confidence 0.01, etc.



CD for θ, with ten data points from $\mathrm{Expo}(\theta)$, with $\bar{y} = 0.345$. I've tagged median confidence estimate 2.803 and the 0.05 and 0.95 CD quantiles $[1.573, 4.552]$. Also: $p(\theta_0) = \mathrm{Pr}_{\theta_0}\{\widehat{\theta} \geq \widehat{\theta}_{\mathrm{obs}}\}$ is the confidence in $[0, \theta_0]$.

Suppose $\psi = \psi(\theta)$ is a focus parameter in some model indexed by θ. Three canonical graphical summaries for inference about ψ:

- cumulative confidence function $C_n(\psi)$:
- confidence density $c_n(\psi) = C_n'(\psi)$;
- confidence curve

$$\mathrm{cc}_n(\psi) = |1 - 2C_n(\psi)| = \begin{cases} 1 - 2C_n(\psi) & \text{for } \psi \le \widehat{\psi}_{.50}, \\ 2C_n(\psi) - 1 & \text{for } \psi \ge \widehat{\psi}_{.50}. \end{cases}$$

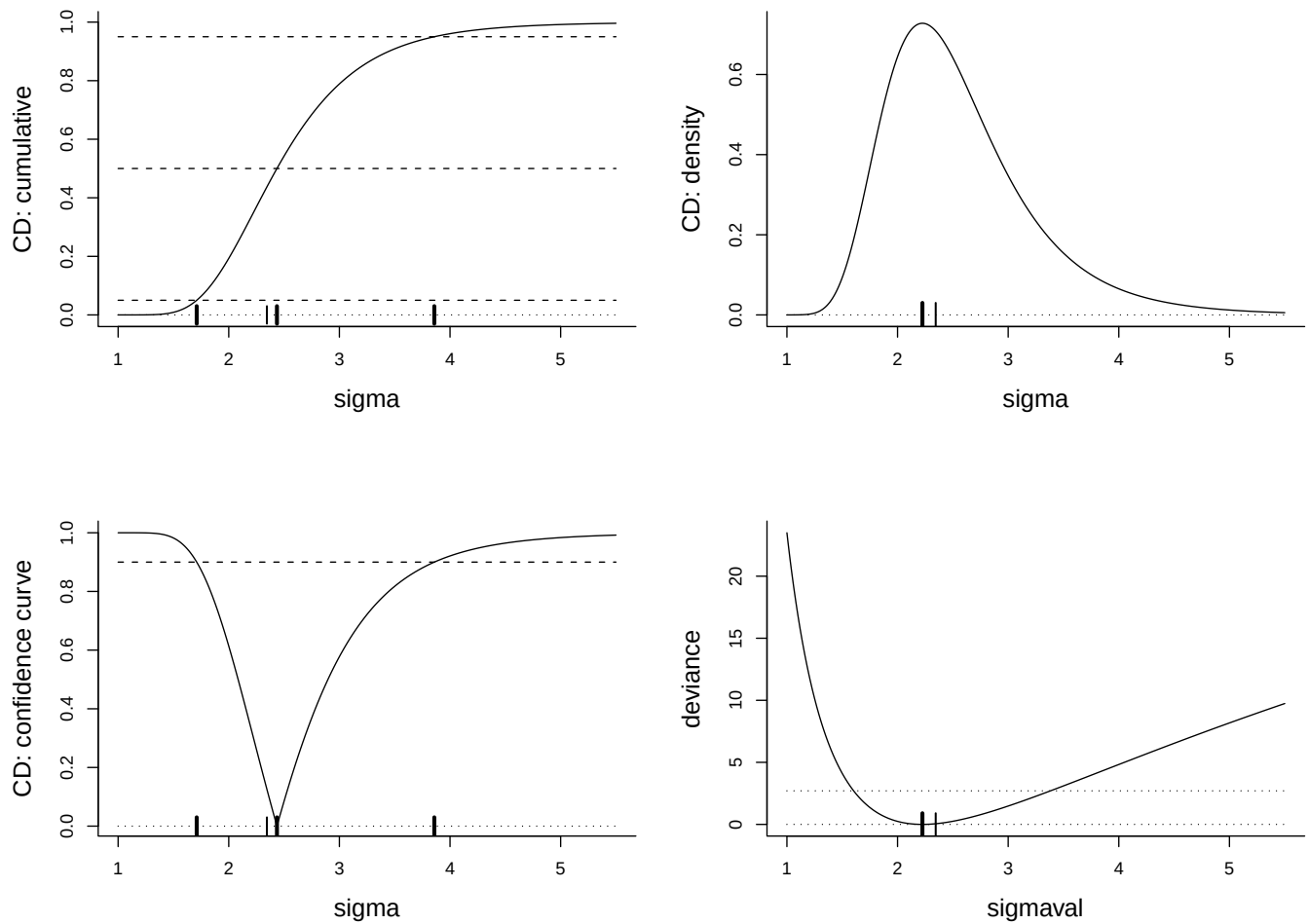
Here $\widehat{\psi}_{.50} = C_n^{-1}(\frac{1}{2})$ is the median confidence estimate. The point is that

$$\mathrm{cc}_n(\psi) = \alpha$$

has two roots giving the level α interval:

$$\text{confidence of } [\psi_{\mathrm{low}}(\alpha), \psi_{\mathrm{up}}(\alpha)] = \alpha.$$

The median confidence estimator is often large-sample equivalent to the ML estimator (but not always).



Graphical inference summaries for normal sd parameter σ, with ten data points having $\widehat{\sigma} = 2.345$: (i) cumulative confidence $C(\sigma)$; (ii) confidence density $c(\sigma)$; (iii) confidence curve $\mathrm{cc}(\sigma)$; (iv) deviance.

3: Recipes for constructing CDs and confidence curves

With data y, from model $f(y, \theta_1, \ldots, \theta_p)$, and focus parameter $\psi = \psi(\theta_1, \ldots, \theta_p)$: $C(\psi, y)$ is a CD for ψ if

- $C(\psi, y)$ is a cdf in ψ, for each dataset y;
- at the true θ_0, with $\psi_0 = \psi(\theta_0)$, we have $\Pr_{\theta_0}\{C(\psi_0, Y) \leq \alpha\} = \alpha$ for each α, i.e. $C(\psi_0, Y) \sim \text{unif}$.

This secures that

$$\Pr_\theta\{\psi \in [C^{-1}(0.05), C^{-1}(0.95)]\} = 0.90,$$

&c.

3A: Pivots. Suppose

$$Z = \text{piv}(\psi, Y)$$

is a pivot, with distribution K not depending on θ. If $\text{piv}(\psi, y)$ is increasing in ψ:

$$C(\psi, y_{\text{obs}}) = K(\text{piv}(\psi, y_{\text{obs}}))$$

is a CD.

Classic Student (1908): $t = \sqrt{n}(\mu - \bar{y})/\widehat{\sigma}$ is a pivot:

$$C(\mu) = F_\nu(\sqrt{n}(\mu - \bar{y}_{\text{obs}})/\widehat{\sigma}_{\text{obs}}).$$

Similarly: $\widehat{\sigma}/\sigma$ is a pivot (with distribution $(\chi^2_\nu/\nu)^{1/2}$), yielding a CD:

$$C(\sigma) = 1 - \Gamma\nu(\nu\widehat{\sigma}^2_{\text{obs}}/\sigma^2).$$

If $Z = \text{piv}(\psi, y)$ is not monotone in ψ: may still construct

$$\text{cc}(\psi, y_{\text{obs}}) = K(\text{piv}(\psi, y_{\text{obs}})),$$

a confidence curve.

We do not need to have the distribution of a pivot explicitly, as we may simulate.

Normal quantile: $\psi = F^{-1}(0.90) = \mu + 1.282\,\sigma$. Then

$$\begin{aligned} Z &= Z(\psi, \text{data}) \\ &= \frac{\widehat{\psi} - \psi}{c\,\widehat{\sigma}} \\ &= \frac{\widehat{\mu} - \mu + 1.282\,(\widehat{\sigma} - \sigma)}{c\,\widehat{\sigma}} \\ &= \frac{\sigma N/\sqrt{n} + 1.282\,\sigma((\chi^2_\nu/\nu)^{1/2} - 1)}{c\,\sigma(\chi^2_\nu/\nu)^{1/2}} \end{aligned}$$

has a distribution free of parameters, say K. The CD is

$$C(\psi) = 1 - K\Big(\frac{\widehat{\psi} - \psi}{c\,\widehat{\sigma}}\Big).$$

May simulate. Result is fully independent of the c constant.

Example: Pairs (x_i, y_i) from binormal distribution, observed correlation coefficient $\widehat{\rho}$. Then

$$Z = \mathrm{piv}(\rho, \mathrm{data}) = \frac{\rho - \widehat{\rho}}{\widehat{\kappa}}$$

is an approximate pivot, with e.g. $\widehat{\kappa} = 1 - \widehat{\rho}^2$.

$$C(\rho) = K((\rho - \widehat{\rho}_{\mathrm{obs}})/\widehat{\kappa}_{\mathrm{obs}})$$

is an approximate CD (may take K from simulations).

An alternative (and better) pivot:

$$Z' = \mathrm{piv}'(\rho, \mathrm{data}) = c\{h(\rho) - h(\widehat{\rho})\},$$

with

$$h(\rho) = \tfrac{1}{2} \log \frac{1+\rho}{1-\rho}$$

and with any c. This leads to

$$C'(\rho) \doteq \Phi((n-3)^{1/2}\{h(\rho) - h(\widehat{\rho}_{\mathrm{obs}})\}).$$

3B: Via ML and delta method (first order asymptotics).

Under mild regularity conditions:

$$\sqrt{n}(\widehat{\theta} - \theta) \to_d \mathrm{N}_p(0, \Sigma).$$

Delta method for interest parameter $\psi = a(\theta)$:

$$\sqrt{n}(\widehat{\psi} - \psi) \to_d \mathrm{N}(0, \kappa^2) \quad \text{with } \kappa^2 = w^{\mathrm{t}} \Sigma w,$$

where $w = \partial a(\theta)/\partial\theta$. Result:

$$C(\psi) = \Phi(\sqrt{n}(\psi - \widehat{\psi})/\widehat{\kappa})$$

is an approximate (1st order large-sample correct) CD.

3C: Via deviance and Wilks theorem:

The profiled log-likelihood function:

$$\ell_{\text{prof}}(\psi) = \max\{\ell(\theta) \colon a(\theta_1, \ldots, \theta_p) = \psi\}.$$

The deviance:

$$D_n(\psi, y) = 2\{\ell_{\text{prof}}(\widehat{\psi}) - \ell_{\text{prof}}(\psi)\}.$$

The Wilks Theorem: At the true value, and with increasing sample size:

$$D_n(\psi, Y) \to_d \chi^2_1.$$

So

$$\text{cc}(\psi) = \Gamma_1(D_n(\psi, y_{\text{obs}}))$$

is a confidence curve.

Often, the deviance with Wilks provides better approximations than the delta method.

How to prove 1st order asymptotics theorems (including the Wilks theorem)? What are the crucial things going on for likelihood behaviour?

This is the crux:

$$\begin{aligned} A_n(s) &= \ell_n(\theta_0 + s/\sqrt{n}) - \ell_n(\theta_0) \\ &= s^{\mathrm{t}} \ell_n'(\theta_0)\sqrt{n} - \tfrac{1}{2} s^{\mathrm{t}} J_n s + \text{small} \\ &\to_d s^{\mathrm{t}} U - \tfrac{1}{2} s^{\mathrm{t}} J s = A(s), \end{aligned}$$

where $J_n = -n^{-1}\ell_n''(\theta_0) \to_{\mathrm{pr}} J$ and
$U_n = n^{-1/2}\ell_n'(\theta_0) \to_d U \sim \mathrm{N}_p(0, J)$.

Consequence 1: $\mathrm{argmax}(A_n) \to_d \mathrm{argmax}(A)$:

$$\sqrt{n}(\widehat{\theta} - \theta_0) \to_d J^{-1} U \sim \mathrm{N}_p(0, J^{-1}).$$

Consequence 2: $\max A_n \to_d \max A$:

$$\ell_n(\widehat{\theta}) - \ell(\theta_0) \to_d \tfrac{1}{2} U^{\mathrm{t}} J^{-1} U.$$

Nonparametric confidence distributions may also be constructed. The empirical likelihood may be used, via the $-2 \log L_n^*(\theta_0) \to_d \chi_p^2$ property, which implies

$$D_n^*(\psi_0) = -2 \log L_{n,\text{prof}}^*(\psi_0) \to_d \chi_1^2$$

for the empirical deviance.

One particular use of this is via

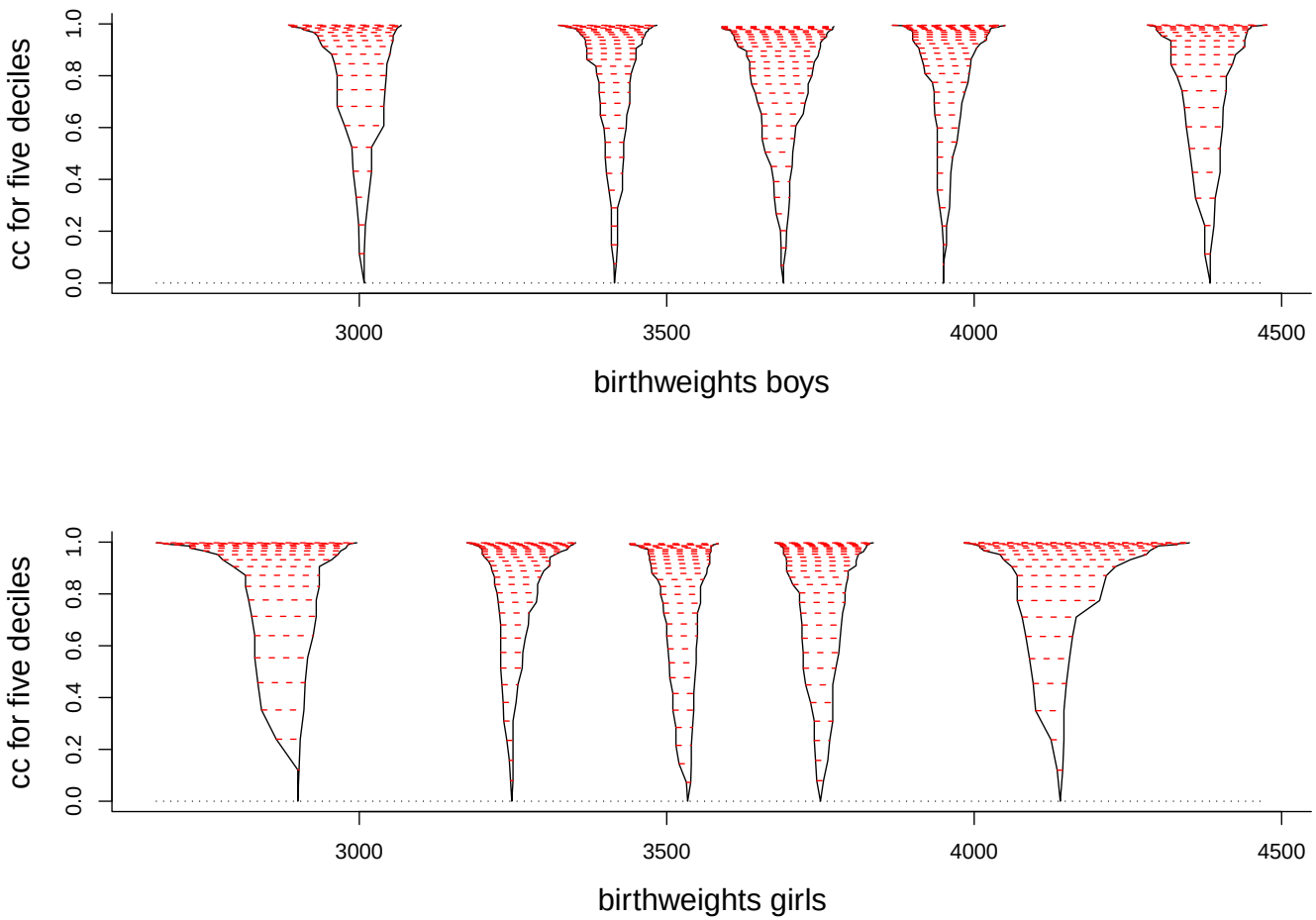
$$\mathrm{E}\, u_j(Y, \theta_0) = 0 \quad \text{for } j = 1, \ldots, p$$

with score functions from a given parametric family. Taking these in the EL yields model robust confidence distributions for each focus parameter: $\mathrm{cc}_n^*(\psi) = \Gamma_1(D_n^*(\psi))$.

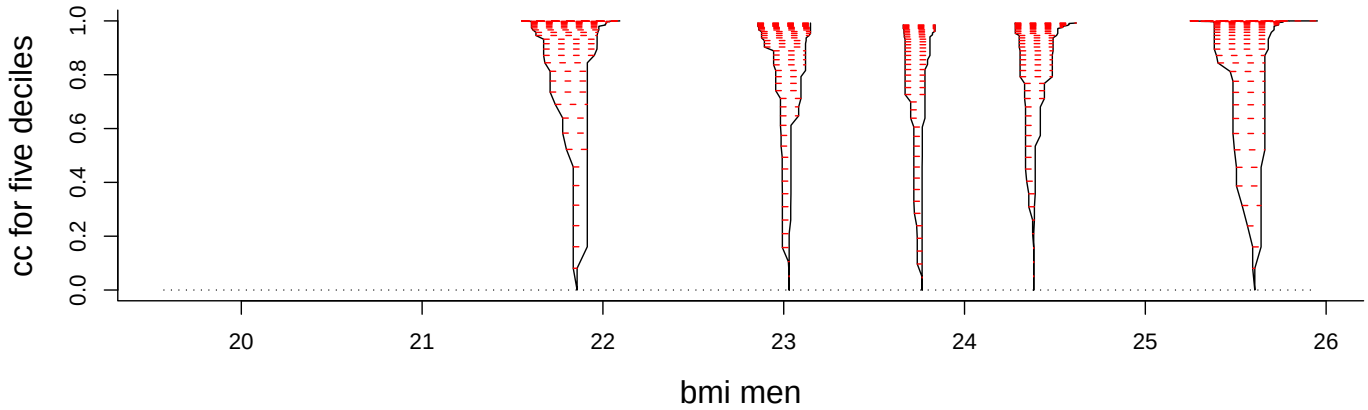
May also construct CDs for the cdf $F(y)$ and quantile function $\mu_p = F^{-1}(p)$. Since $Y_{(i)} = F^{-1}(U_{(i)})$,

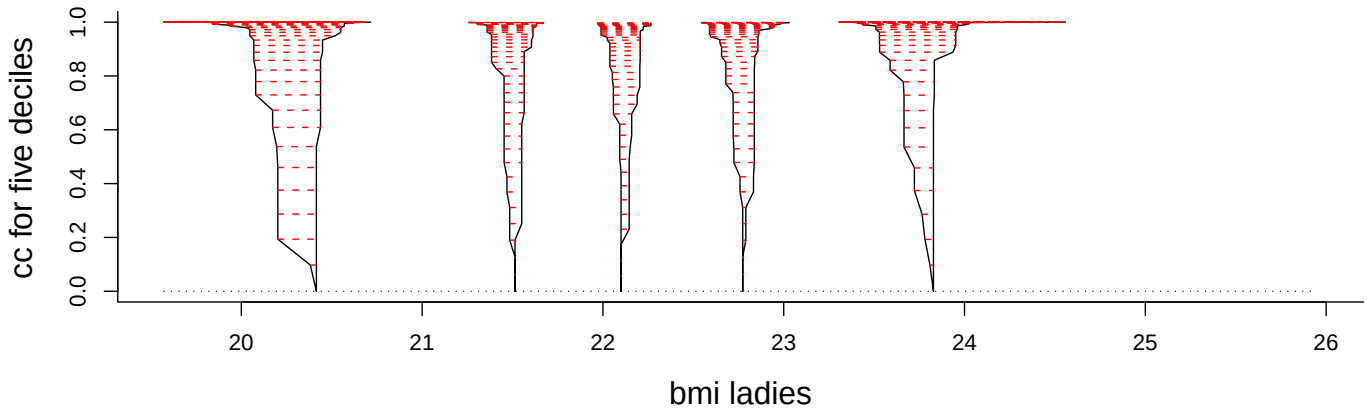
$$s_n(a,b) = \Pr\{Y_{(a)} \le \mu_p \le Y_{(b)}\} = \Pr\{U_{(a)} \le p \le U_{(b)}\}$$
$$= \int_0^p \left\{1 - \mathrm{Be}\left(\frac{p-u}{1-u}, b-a, n-b+1\right)\right\} \mathrm{be}(u, a, n-a+1)\,\mathrm{d}u$$

is the exact coverage probability for $[Y_{(a)}, Y_{(b)}]$, e.g. with $a = [np] - j$ and $b = [np] + j$. We may display them all in a CD plot for quantiles.



Confidence curves $\mathrm{cc}(q)$ for deciles 0.1, 0.3, 0.5, 0.7, 0.9 of birthweight distributions, for boys ($n = 548$) and girls ($n = 480$) born in Oslo 2001–2008.





Confidence curves $\mathrm{cc}(q)$ for deciles 0.1, 0.3, 0.5, 0.7, 0.9 for the BMI of all Olympic speedskaters, 1952 to 2010 (1080 men and 741 ladies). 17% of the male Olympians are overweight (says WHO)!

4: Variations and nonstandard cases

There are cases where maximum likelihood (and Bayes) are in trouble, but where CDs work without problem.

Neyman–Scott situation:

$$Y_{i,1} \sim \mathrm{N}(\mu_i, \sigma^2) \quad \text{and} \quad Y_{i,2} \sim \mathrm{N}(\mu_i, \sigma^2)$$

for $i = 1, \ldots, n$. With $n = 100$ pairs, there are $2n = 200$ observations and $n + 1 = 101$ parameters. Likelihood is hopeless for σ, hence both ML and Bayes are in trouble: $\widehat{\sigma}_{\mathrm{ML}} \to_{\mathrm{pr}} \sigma/\sqrt{2}$. But easy to put up a perfectly good CD.

Length problem: Suppose $Y_i \sim \mathrm{N}(\mu_i, 1)$ for $i = 1, \ldots, p$, with interest in $\psi = \|\mu\|$. The Bayes with flat prior on each component is bad (and becomes worse with increasing p). Easy to construct perfect CD: $\widehat{\psi}^2 \sim \chi^2_p(\psi^2)$, yielding

$$C(\psi) = 1 - \Gamma_p(\widehat{\psi}^2, \psi^2).$$

4A: Fieller problem: ratio of normal means

Assume $\widehat{a} \sim \mathrm{N}(a, 1)$ and $\widehat{b} \sim \mathrm{N}(b, 1)$, interest lies with $\psi = a/b$. Our solution: under a fixed ψ,

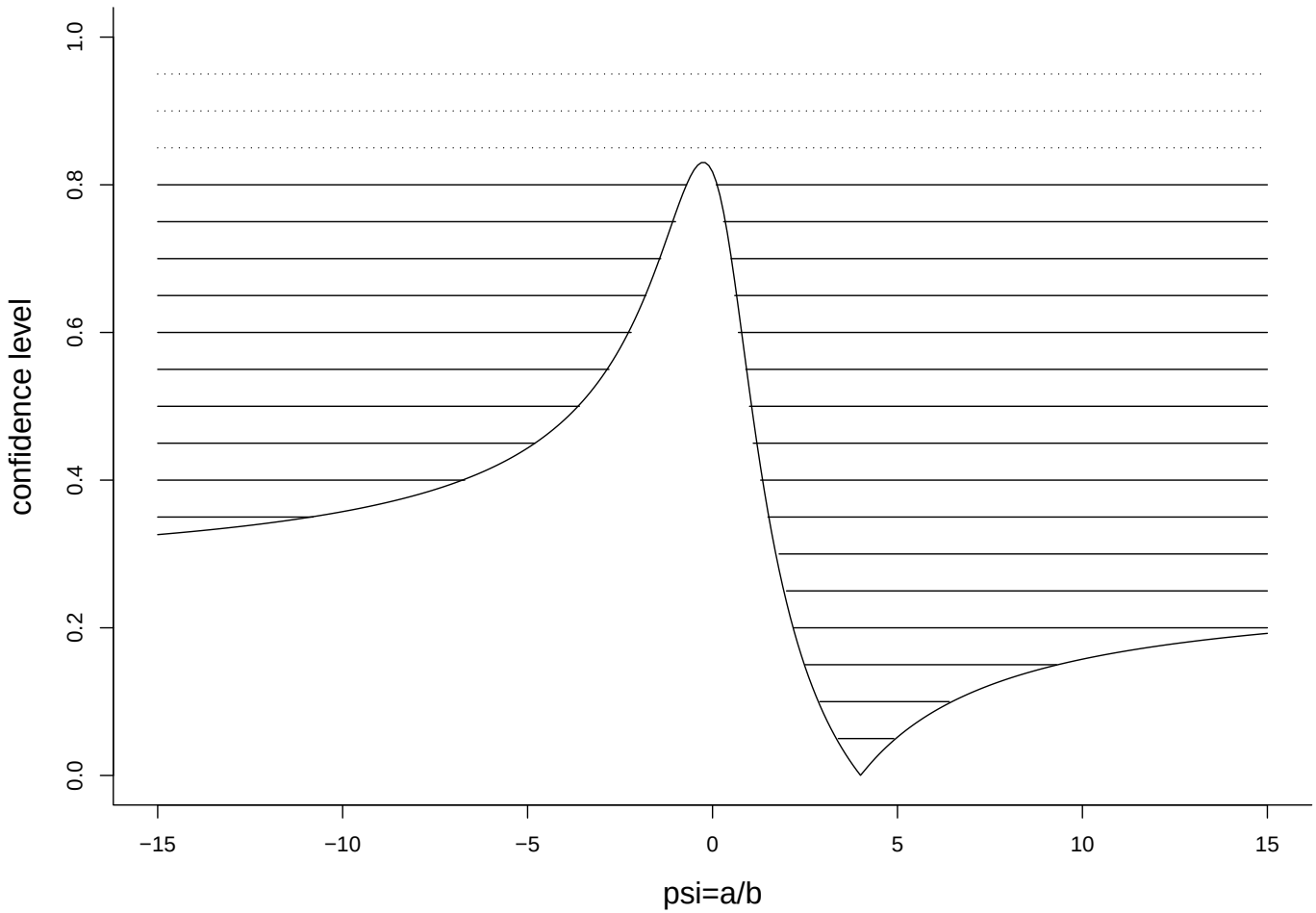
$$D(\psi, \mathrm{data}) = \frac{(\widehat{a} - \psi\widehat{b})^2}{1 + \psi^2} \sim \chi^2_1,$$

so

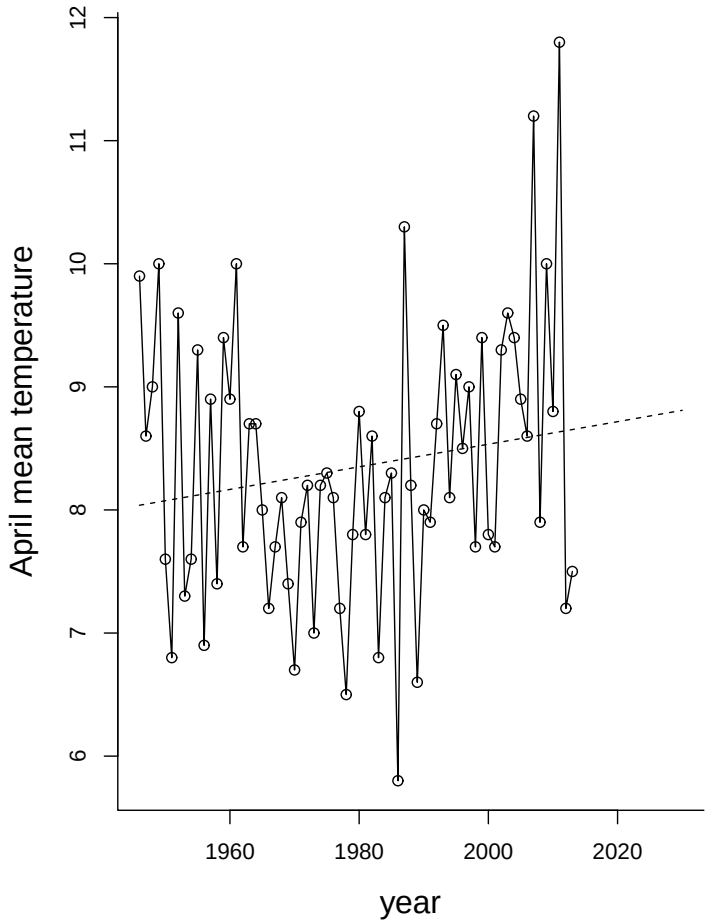
$$\mathrm{cc}(\psi) = \Gamma_1\Big(\frac{(\widehat{a} - \psi\widehat{b})^2}{1 + \psi^2}\Big)$$

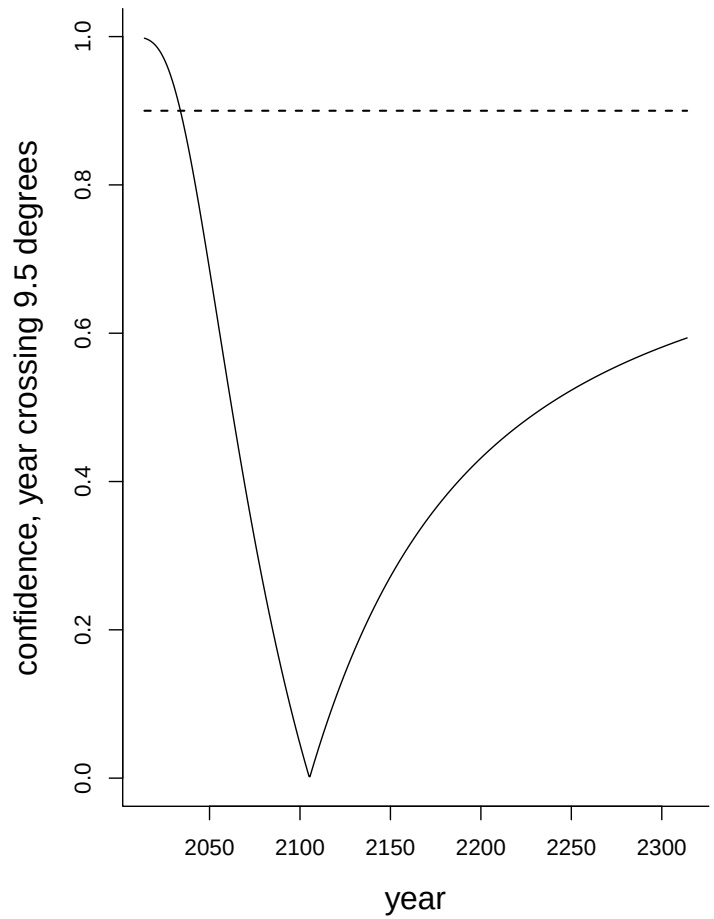
is a perfect and exact confidence curve – even if its shape confuses us (when $|\widehat{b}|$ is small). "If common sense doesn't agree with what comes out of a theory, then there is something wrong either with the theory or with the common sense" (says David Cox, April 2016).

Example: $(\widehat{a}, \widehat{b}) = (1.333, 0.333)$, with $\widehat{\psi} = 4.003$. Confidence regions may be intervals; union of two half-infinite intervals; or the full line.



Inverse regression: $y_i = a + bx_i + \varepsilon_i$, with the ε_i i.i.d. $\mathrm{N}(0, \sigma^2)$. When does $a + bx$ cross a given y_0? Example: CET, temperatures per month in England since 1659. Mean April temperatures since 1946: when will it cross 9.5 degrees? Point estimate $\widehat{x}_0 = 2105$, but 90% interval is $[2028, \infty)$.





4B: Boundary parameters

Some methods and strategies have difficulties with 'boundary parameters', as with variance components.

Prototype situation: $y \sim \mathrm{N}(\theta, 1)$ with $\theta \geq 0$ a priori. My choice: the canonical CD

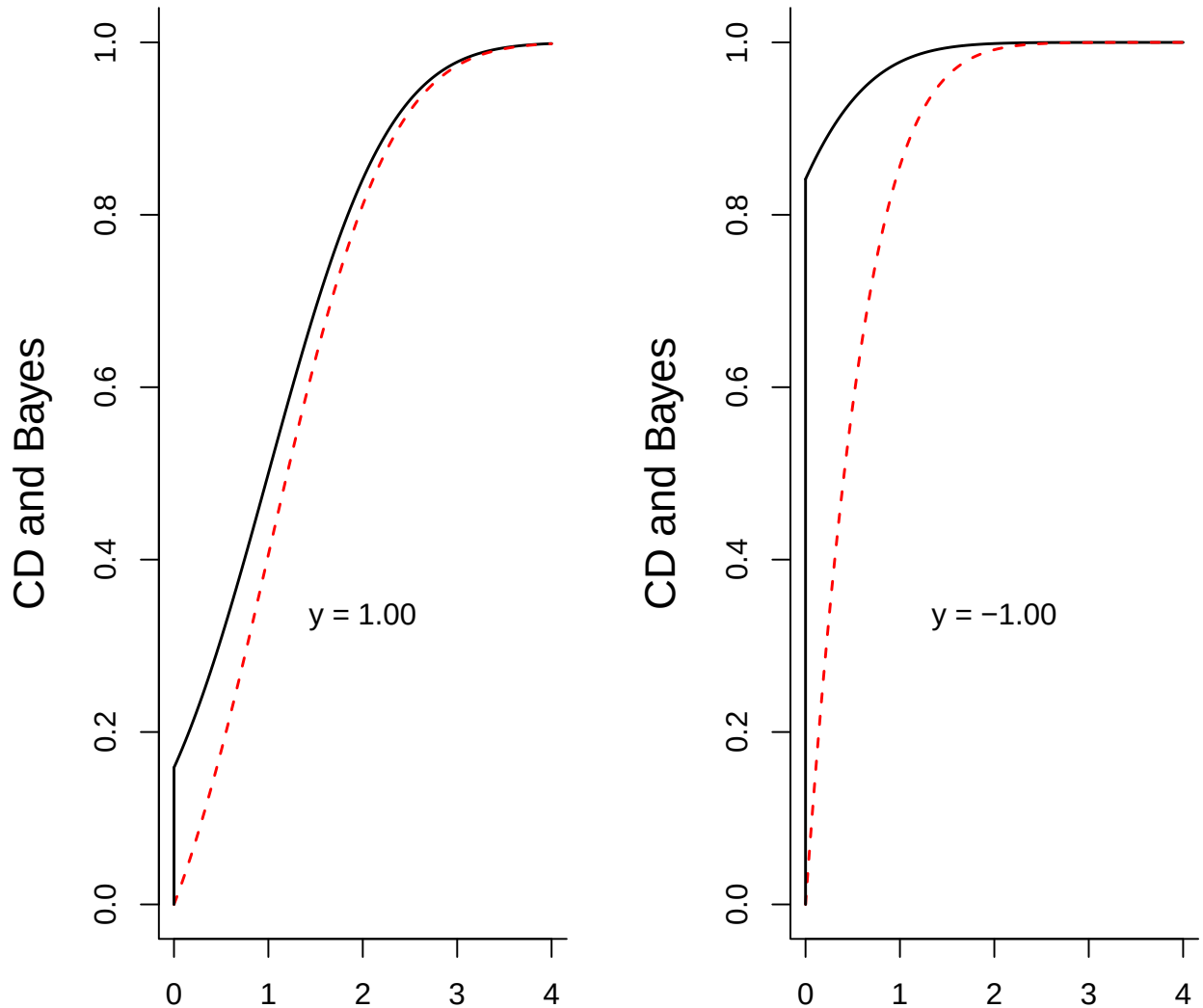
$$C(\theta, y_{\mathrm{obs}}) = \Phi(\theta - y_{\mathrm{obs}}) \quad \text{for } \theta \geq 0.$$

Confidence pointmass $p_0 = \Phi(-y_{\mathrm{obs}})$ at zero. Confidence density is

$$c(\theta, y_{\mathrm{obs}}) = p_0 \delta_0 + (1 - p_0) \frac{\phi(\theta - y_{\mathrm{obs}})}{1 - \Phi(-y_{\mathrm{obs}})}.$$

This CD is the canonical distribution for θ given data (the 'Holy Grail') – but is not equal to the Bayesian posterior distribution, for any fixed prior.

$y \sim \mathrm{N}(\theta, 1)$ with $\theta \geq 0$ a priori: Canonical CD vs. Bayes with flat prior, for cases $y_{\mathrm{obs}} = 1.00$ and $y_{\mathrm{obs}} = -1.00$. Same syndrome as with Schweder & Hjort vs. Sims (Nobel Prize Economics 2012).



4C: C. Sims (2012, Nobel Prize acceptance lecture)

	I	G	C
1929	101.4	146.5	736.3
1930	67.6	161.4	696.8
1931	42.5	168.2	674.9
1932	12.8	162.6	614.4
1933	18.9	157.2	600.8
1934	34.1	177.3	643.7
1935	63.1	182.2	683.0
1936	80.9	212.6	752.5
1937	101.1	203.6	780.4
1938	66.8	219.3	767.8
1939	85.9	238.6	810.7
1940	119.7	245.3	852.7

$$C_t = \text{consumption} = \beta_0 + \beta_1 Y_t + \sigma_C Z_{1,t},$$
$$I_t = \text{investment} = \theta_0 + \theta_1 (C_t - C_{t-1}) + \sigma_I Z_{2,t},$$
$$Y_t = \text{total income} = C_t + I_t + G_t,$$
$$G_t = \text{government spending} = \gamma_0 + \gamma_1 G_{t-1} + \sigma_G Z_{3,t}.$$

Sims' macroeconomic model for (C_t, I_t, G_t) has six regression coefficients plus three standard deviation parameters $\sigma_C, \sigma_I, \sigma_G$. The focus parameter is θ_1 of

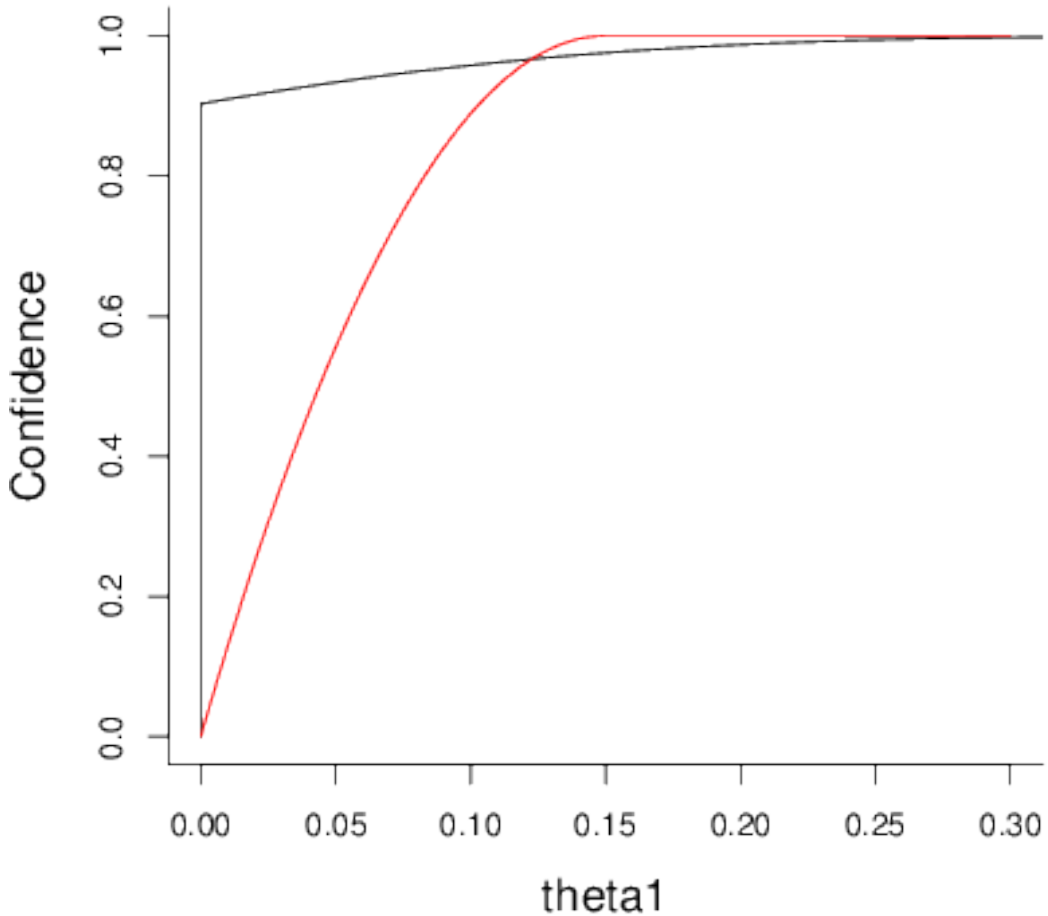
$$I_t = \text{investment} = \theta_0 + \theta_1(C_t - C_{t-1}) + \sigma_I Z_{2,t}.$$

Sims argues that $\theta_1 \geq 0$ a priori, and finds the posterior distribution (via MCMC) using flat priors for all six coefficients plus flat priors for the three $1/\sigma^2$.

We have re-analysed the data. The profile deviance

$$D(\theta_1) = 2\{\ell_{\text{prof}}(\widehat{\theta}_1) - \ell_{\text{prof}}(\theta_1)\}$$

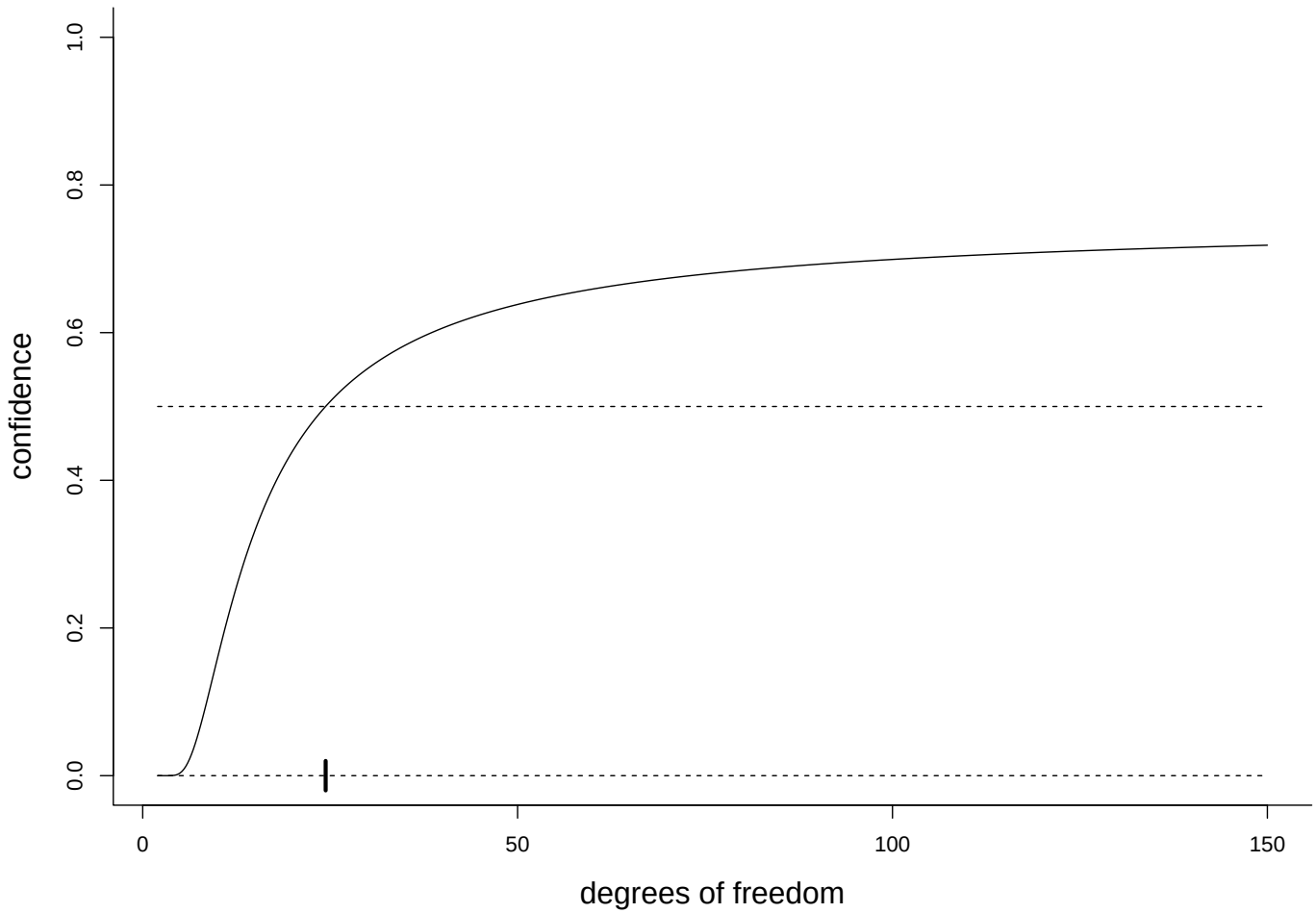
is close to being a pivot (verified by simulation), and bootstrapping yields a confidence distribution markedly different from Sims' Bayesian result.



Posterior cdf for θ_1 reached by Sims (red); our confidence distribution (black), with a high point mass at zero. Pre-war investment was insensitive to government spending!

Example: Three-parameter model for age of mothers

I'm fitting ages for 189 mothers (range 14 to 44) to the model $f(y) = g_\nu((y-\mu)/\sigma)/\sigma$, with g_ν the t-density with ν degrees of freedom. CD for ν has positive pointmass 0.245 at ∞ (i.e. normality).



4D: Are bad-tempered men better at finding good-tempered women than the good-tempered men are?

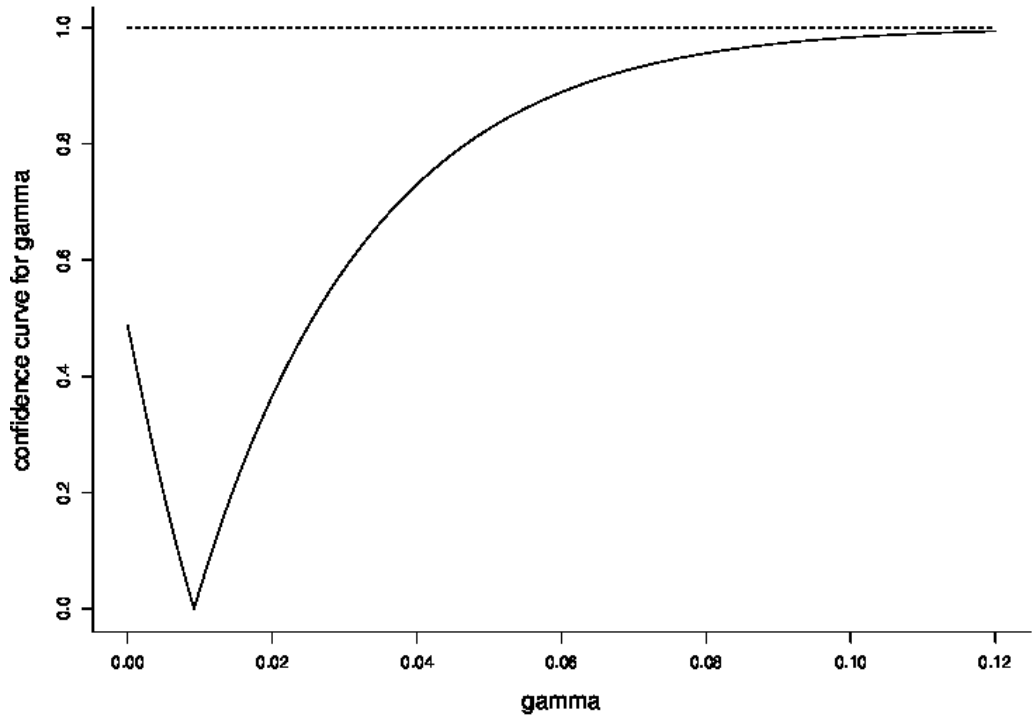
		wife : good	wife : bad
husband :	good	24	27
	bad	34	26

Galton, 1887: "We can hardly, too, help speculating uneasily upon the terms that our own relatives would select as most appropriate to our particular selves."

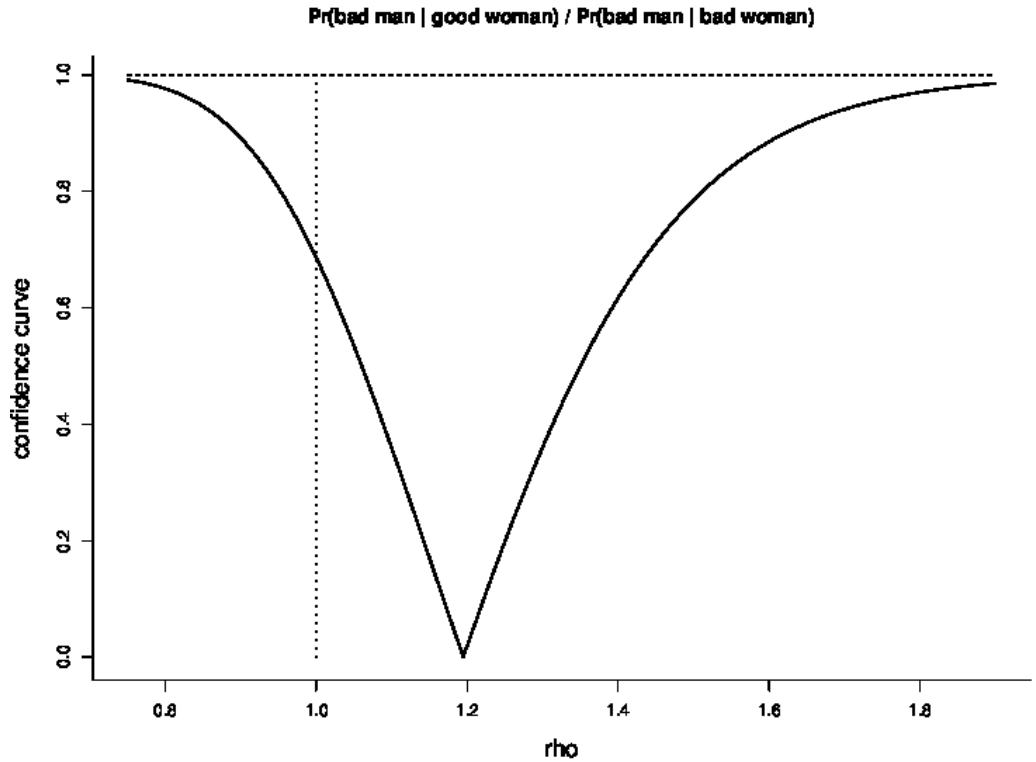
Confidence inference for

$$\gamma = \sum_{i,j} \frac{(p_{i,j} - a_i b_j)^2}{a_i b_j}$$

with a_i and b_j the marginals; $Z_n = n\widehat{\gamma}$ is the Pearson statistic.



Confidence curve $\mathrm{cc}(\gamma)$ for $\gamma = \sum_{i,j}(p_{i,j} - a_i b_j)^2/(a_i b_j)$ parameter with Galton's data on good- and bad-tempered husbands and wives. This is more informative than saying "I've checked the Pearson test, and do not reject H_0 of independence".



Can similarly provide a CD for each focus parameter. Here:
Confidence curve $\mathrm{cc}(\rho)$ for
$\rho = \Pr(\text{bad man} \,|\, \text{good woman})/\Pr(\text{bad man} \,|\, \text{bad woman})$.

4E: Other types of models

Above: I've essentially discussed i.i.d. and regression type models. But the confidence concepts are quite general, and likelihood tools generalise to various other models – Markov chains; time series (unless the memory is too strong); survival analysis models.

As long as

$$n^{-1/2}\ell_n'(\theta_0) \to_d \mathrm{N}_p(0, J) \quad \text{and} \quad -n^{-1}\ell_n''(\theta_0) \to_{\mathrm{pr}} J,$$

we're (very much) in business. For different model setups, these 'things are going to turn out as hoped for' criteria may be checked.

Survival analysis models: data take the form of triples

- t_i: time to event (censored or not);
- x_i: vector of covariates;
- δ_i: 1 of observed, 0 if censored

for patients or objects $i = 1, \ldots, n$.

Survival analysis models

Model for hazard rates:

$$h_i(s) = h_i(s, \theta) \quad \text{for } i = 1, \ldots, n,$$

with cumulative hazard rates $H_i(t, \theta) = \int_0^t h_i(s, \theta)\, \mathrm{d}s$. The log-likelihood becomes

$$\begin{aligned} \ell_n(\theta) &= \sum_{i=1}^n \int_0^\tau \{\log h_i(s, \theta)\, \mathrm{d}N_i(s) - Y_i(s) h_i(s, \theta)\, \mathrm{d}s\} \\ &= \sum_{i=1}^n \{h_i(t_i, \theta)\delta_i - H_i(t_i, \delta)\}. \end{aligned}$$

Here

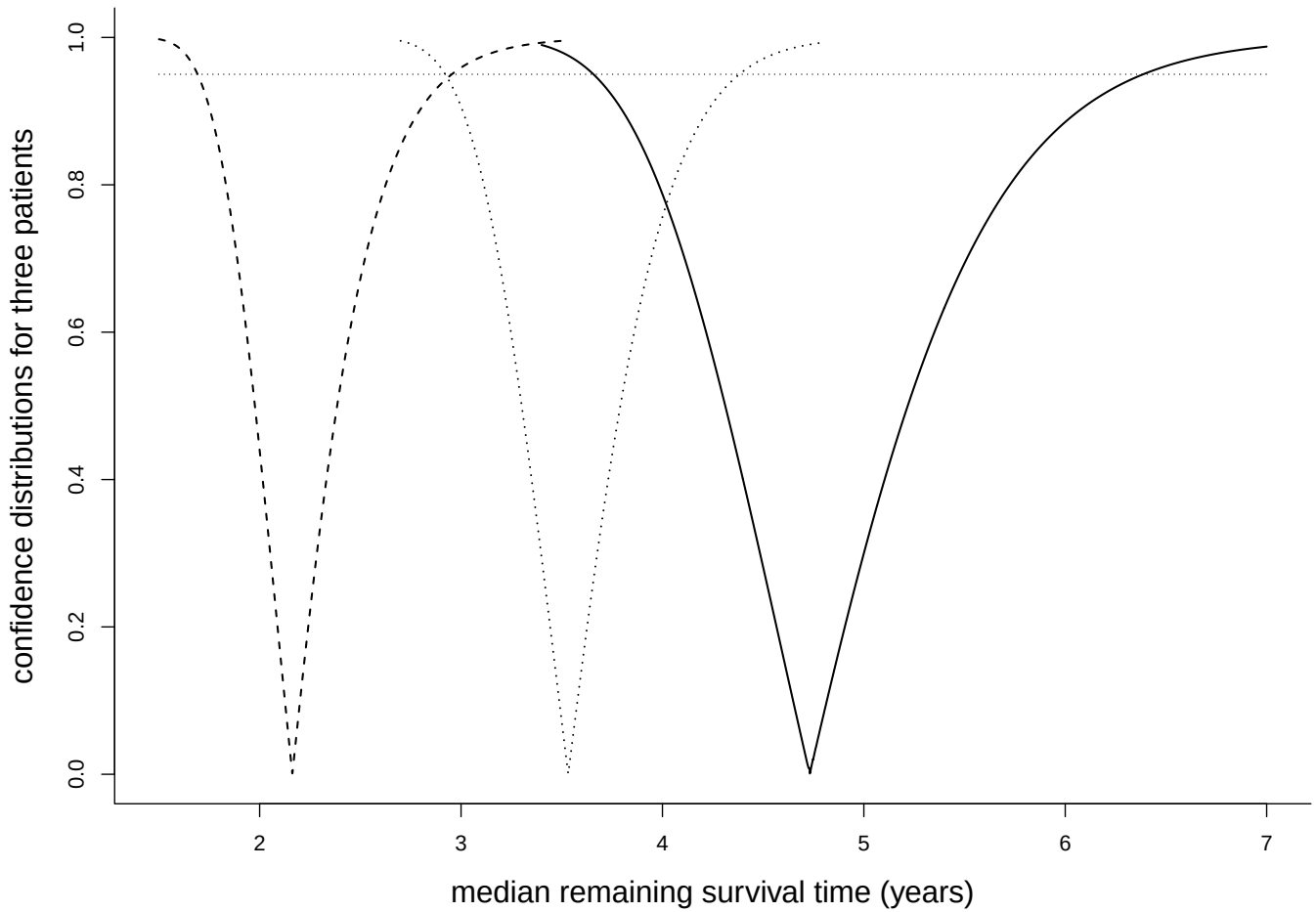
$$\begin{aligned} Y_i(s) &= I\{\text{no. } i \text{ is still at risk at time } s\}, \\ \mathrm{d}N_i(s) &= I\{\text{no. } i \text{ dies in } [s, s + \mathrm{d}s]\}. \end{aligned}$$

The 'things are running well' conditions are met: ML asymptotics and Wilks work, so we can construct CDs and confidence curves.

Example: Median remaining survival time after operation (here: for three types of patients). Carcinoma of the oropharynx, data (t_i, x_i, δ_i) for 193 patients,

$$h_i(s) = \gamma s^{\gamma-1} \exp(\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_4 x_{i,4}),$$

$$m(t_0, x) = \left(t_0^{\gamma} + \frac{\log 2}{\exp(x^{\mathrm{t}}\beta)} \right)^{1/\gamma}.$$



Inference for σ in a normal sample $y_i \sim \mathrm{N}(\mu, \sigma^2)$: may use

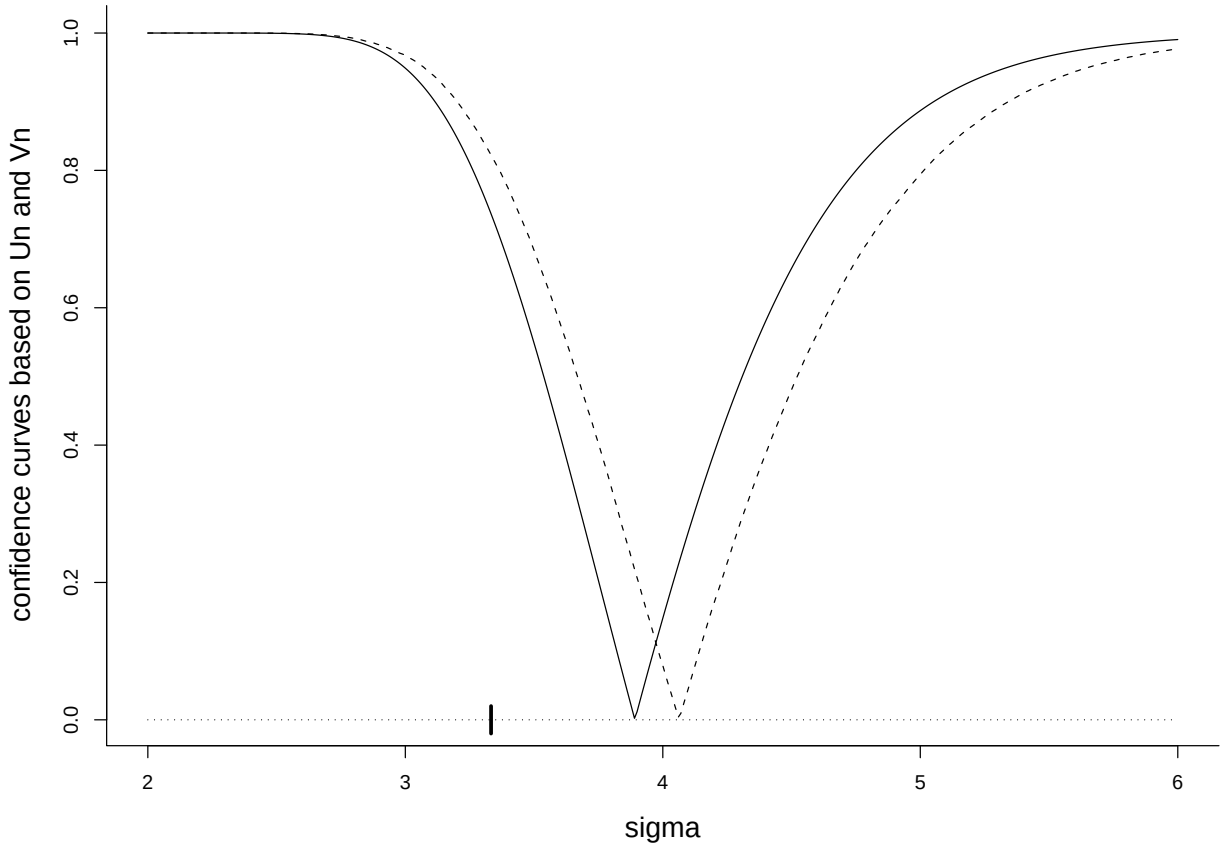
$$U_n = n^{-1} \sum_{i=1}^{n} |y_i - \bar{y}| \quad \text{and} \quad V_n = n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

Here U_n/σ and V_n/σ^2 are pivots, with distributions G_n and H_n, so may use

$$C_{n,u}(\sigma) = 1 - G_n(U_{n,\mathrm{obs}}/\sigma) \quad \text{and} \quad C_{n,v}(\sigma) = 1 - H_n(V_{n,\mathrm{obs}}/\sigma^2).$$

But which is best?

Sir Arthur Eddington (1914) preferred U_n; Sir Ronald Fisher (1922) felt the need to disagree (and introduced sufficiency to prove he was right).



Normal σ, with $n = 25$, using U_n and V_n: We need notions of 'some CDs are better than others', and we need reduction principles (invariance, sufficiency, other).

5A: Invariance

When meeting the same type of data, and with the same focus parameter, we should use the same CD each time. – Suppose model says $y \sim f(y, \theta)$, and that for various g,

$$g(y) \sim f(y, \bar{g}(\theta)).$$

Consequence:

$$C(\psi(\theta), y) = C(\psi(\bar{g}(\theta)), g(y)) \quad \text{for all } g \in \mathcal{G}.$$

Example A: $X \sim \text{Expo}(\gamma a)$ and $Y \sim \text{Expo}(a)$, focus on γ. Then need

$$C(\gamma, cX, cY) = C(\gamma, X, Y) \quad \text{for all } c > 0,$$

which means an invariant CD must depend on $Z = Y/X$ alone. Rest is then easy:

$$C^*(\gamma) = 1 - F_{2,2}(z/\gamma) = \frac{\gamma}{\gamma + z} \quad \text{and} \quad c^*(\gamma) = \frac{z}{(\gamma + z)^2}.$$

Example B: $Y = (Y_1, \ldots, Y_n)$ with means zero and $\mathrm{Var}\, Y = \sigma^2 I_n$. Then $Y' = PY$ has same type of structure, for each orthogonal P. Hence an invariant CD must depend on Y via $\|Y\|$ only. – If in addition Y is normal, then there is only one remaining invariant CD:

$$C(\sigma) = 1 - \Gamma_n\Big(\sum_{i=1}^{n} y_i^2/\sigma^2\Big).$$

Example C: Regression model

$$y_i = a + bx_i + \sigma\varepsilon_i \quad \text{for } i = 1, \ldots, n,$$

with the $\varepsilon_i \sim \mathrm{N}(0,1)$: The classical

$$C(b) = F_\nu\Big(M^{1/2}\frac{b - \widehat{b}}{\widehat{\sigma}}\Big)$$

is best invariant CD (among those based on sufficiency), where $M = \sum_{i=1}^{n}(x_i - \bar{x})^2$.

5B: Risk functions for CDs

Risk functions for competing CDs: at position θ_0 in parameter space, with $\psi_0 = \psi(\theta_0)$:

$$\text{risk}(C, \theta_0) = \text{E}_{\theta_0} \int \Gamma(\psi - \psi_0) \, \text{d}C(\psi, Y).$$

Here $\Gamma(z)$ is any convex function with $\Gamma(0) = 0$, like $\Gamma(z) = |z|$.

Can be computed and interpreted as

$$\text{risk}(C, \theta_0) = \frac{1}{B} \sum_{j=1}^{B} \Gamma(\psi_{\text{CD},j} - \psi_0),$$

with two levels of variability: (i) Y_j is drawn from $f(y, \theta_0)$, leading to $C(\psi, Y_j)$; (ii) $\psi_{\text{CD},j}$ is drawn from this CD.

With $\Gamma(z) = z^2$:

$$\begin{aligned}\mathrm{risk}(C, \theta_0) &= \mathrm{E}_*(\psi_{\mathrm{CD}} - \psi_0)^2 \\ &= \mathrm{E}\left(\mathrm{Var}\, \psi_{\mathrm{CD}} \mid Y\right) + \mathrm{E}\left\{\left(\mathrm{E}\, \psi_{\mathrm{CD}} \mid Y\right) - \psi_0\right\}^2.\end{aligned}$$

1st order large-sample approximation: Suppose $\widehat{\psi} \approx_d \mathrm{N}(\psi, \kappa^2/n)$. Then

$$\mathrm{risk}(C, \theta_0) \doteq \kappa^2/n + \kappa^2/n.$$

May also work with more complex (and context-driven) loss and risk functions, i.e. with specially designed $\Gamma(z)$.

Neyman–Pearson for confidence power in the mean: With S a sufficient statistic, suppose

$$L(\psi_2, S)/L(\psi_1, S) \text{ is increasing in } S \text{ for } \psi_2 > \psi_1.$$

May then prove that the CD based on S is uniformly most powerful in terms of confidence risk.

Optimal confidence for exponential families: Suppose

$$f(y, \psi, \lambda) = \exp\{\psi A(y) + \lambda_1 B_1(y) + \cdots + \lambda_p B_p(y) - k(\psi, \lambda_1, \ldots, \lambda_p)\}.$$

Then the uniformly most powerful confidence distribution for ψ is

$$C^*(\psi) = \Pr_\psi\{A \geq A_{\text{obs}} \,|\, B_1 = B_{1,\text{obs}}, \ldots, B_p = B_{p,\text{obs}}\}.$$

This theorem covers a long list of cases inside e.g. generalised linear models.

Example: Poisson pairs. Suppose $X_j \sim \text{Pois}(\lambda_j)$, $Y_j \sim \text{Pois}(\lambda_j \gamma)$ for $j = 1, \ldots, k$. The log-likelihood is

$$\ell = \sum_{j=1}^{k} y_j \log \gamma + \sum_{j=1}^{k} (x_j + y_j) \log \lambda_j - \sum_{j=1}^{k} \lambda_j (1 + \gamma).$$

With $S = \sum_{j=1}^{k} y_j$ and $z_j = x_j + y_j$: the optimal CD for γ is

$$\begin{aligned} C^*(\gamma) = \text{Pr}_\gamma\{S > S_{\text{obs}} \,|\, Z_1 = z_{1,\text{obs}}, \ldots, Z_k = z_{k,\text{obs}}\} \\ + \tfrac{1}{2} \text{Pr}_\gamma\{S > S_{\text{obs}} \,|\, Z_1 = z_{1,\text{obs}}, \ldots, Z_k = z_{k,\text{obs}}\} \end{aligned}$$

which we compute using

$$y_j \,|\, z_j \sim \text{Bin}(z_j, \gamma/(1+\gamma)) \quad \text{for } j = 1, \ldots, k,$$

so

$$S \,|\, (z_{1,\text{obs}}, \ldots, z_{k,\text{obs}}) \sim \text{Bin}\Big(\sum_{j=1}^{k} z_j, \frac{\gamma}{1+\gamma}\Big).$$

Example: Odds ratio. Suppose

$$Y_0 \sim \text{Bin}(m_0, p_0) \quad \text{and} \quad Y_1 \sim \text{Bin}(m_1, p_1),$$

with

$$p_0 = \frac{\exp(\theta)}{1+\exp(\theta)} \quad \text{and} \quad p_1 = \frac{\exp(\theta+\psi)}{1+\exp(\theta+\psi)}.$$

The odds ratio is

$$\rho = \frac{p_1/(1-p_1)}{p_0/(1-p_0)} = \exp(\psi).$$

From the joint likelihood $L = f(y_0, \theta) f(y_1, \theta+\psi)$ comes

$$\ell = y_1 \log \rho + z\theta - m_0 \log\{1+\exp(1+\theta)\} - m_1 \log\{1+\exp(\theta+\psi)\}$$

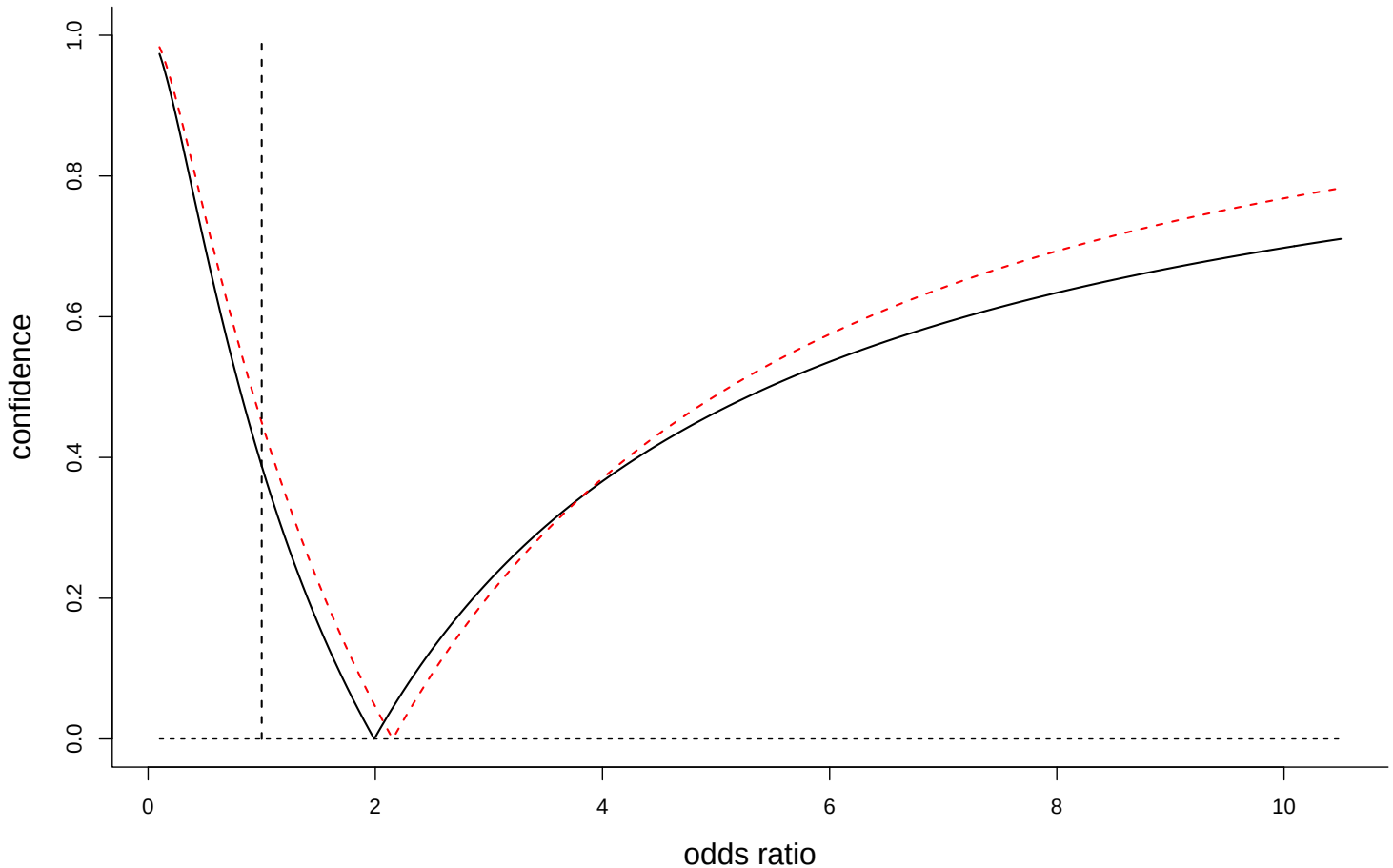
with $z = y_0 + y_1$. The optimal CD for the odds ratio:

$$C^*(\rho) = \Pr_\rho\{Y_1 > y_{1,\text{obs}} \,|\, Z = z_{\text{obs}}\} + \tfrac{1}{2} \Pr_\rho\{Y_1 = y_{1,\text{obs}} \,|\, Z = z_{\text{obs}}\}.$$

We find

$$g(y_1 \mid z) = \binom{m_0}{z - y_1} \binom{m_1}{y_1} \rho^{y_1} \Big/ \sum_{y_1'=0}^{z} \binom{m_0}{z - y_1'} \binom{m_1}{y_1} \rho^{y_1'}$$

for $y_1 = 0, 1, \ldots, \min(z, m_1)$. Illustration: Optimal and 'standard' for $(m_0, m_1) = (15, 15)$, $(y_0, y_1) = (1, 2)$.



Example: Strauss model. A basic model for a random point pattern x on $[0,1]^2$ is

$$f(x,\rho) = c(\rho)\rho^{M(x)} \quad \text{where } M(x) = \sum_{i<j} I\{\|x_j - x_i\| \le r\}.$$

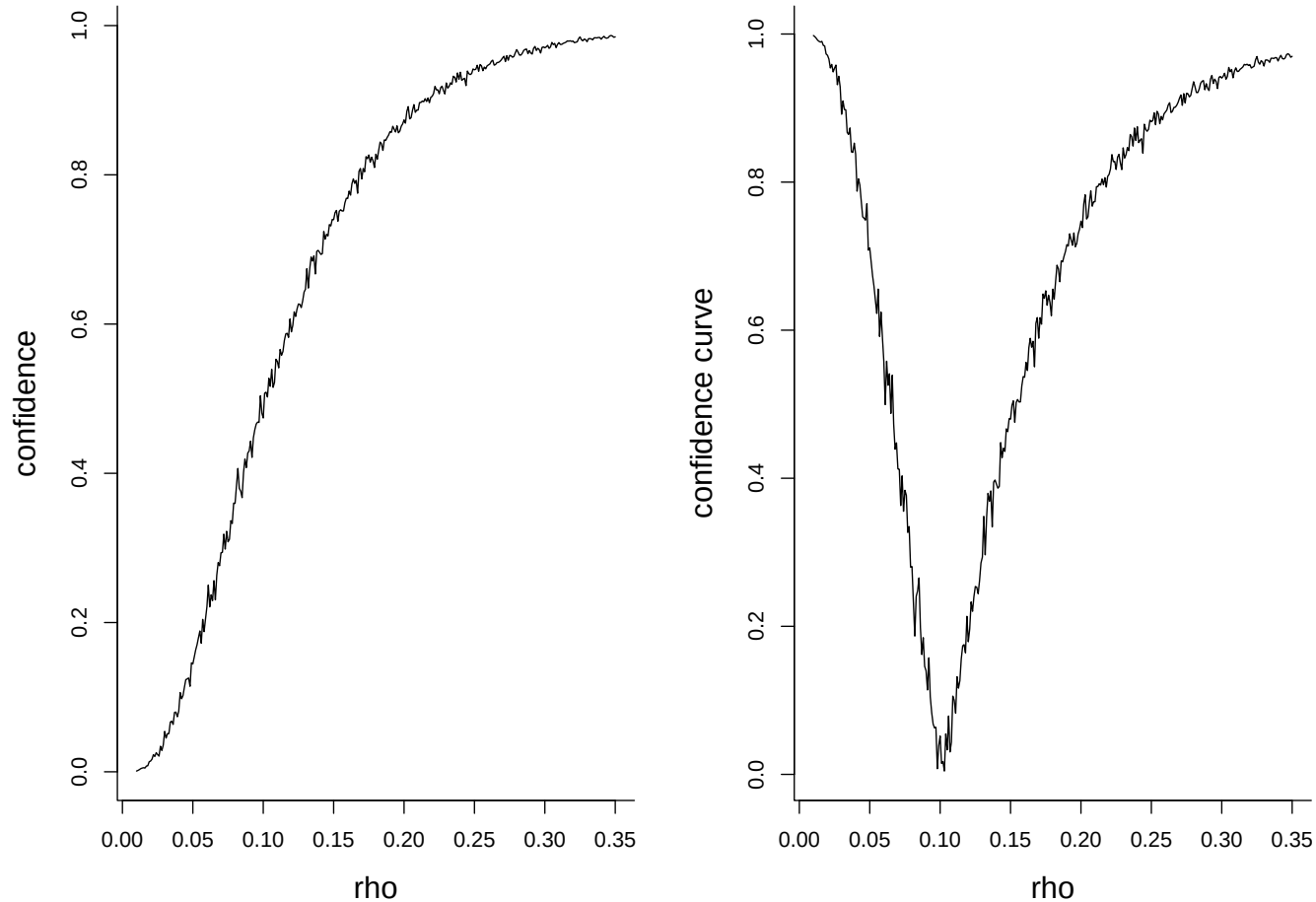
Suppose $n = 25$ with $M = 4$ pairs of points having distance less than $r = 0.15$. Moyeed and Baddeley (1991) wrote a full paper to compute $\widehat{\rho}_{\mathrm{ML}} = 0.099$.

The optimal confidence distribution is

$$C^*(\rho) = \Pr_\rho\{M(X) > 4\} + \tfrac{1}{2}\Pr_\rho\{M(X) = 4\}.$$

I've used a primitive MCMC to simulate 10^5 Strauss realisations for each ρ in a grid (see figure).

The same works for bigger point process models of the exponential type, for the Ising model, the Potts and Gibbs type models, etc.



For $n = 25$ points from the Strauss model with $M = 4$ pairs inside radius $r = 0.15$: optimal confidence distribution (left) and optimal confidence curve (right).

6: Better approximative CDs (via modifications and tricks)

1st order large-sample approximations: with $\sqrt{n}(\widehat{\psi} - \psi) \to_d \mathrm{N}(0, \kappa^2)$,

$$C_1(\psi) = \Phi(\sqrt{n}(\psi - \widehat{\psi})/\widehat{\kappa})$$

is asymptotically correct. Also, via Wilks theorem:

$$C_2(\psi) = \Gamma_1(D(\psi)).$$

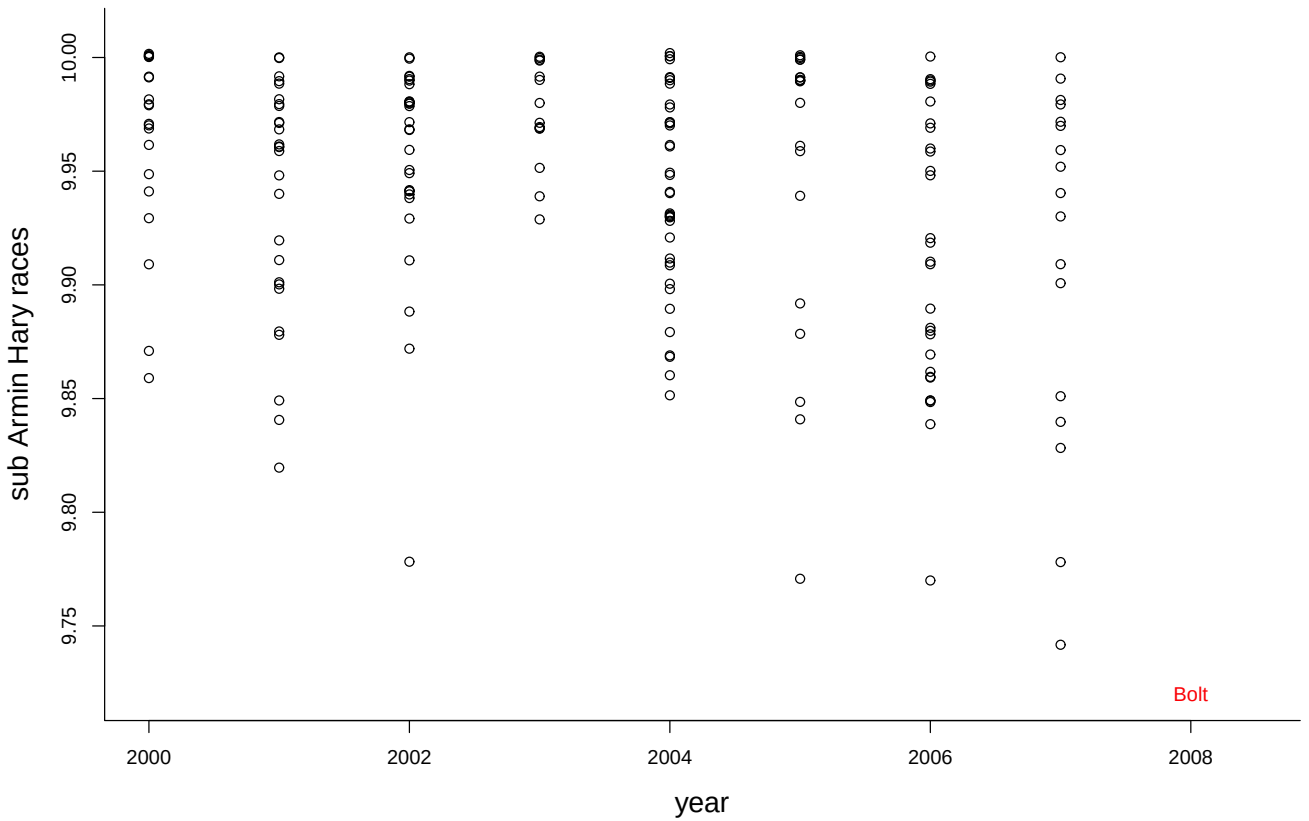
Modifications & tricks: Bartletting often helps. Suppose $\mathrm{E}_\theta\, D(\psi, Y) = 1 + \varepsilon$ (typically, $\varepsilon \doteq \varepsilon_0/n$). Then $D(\psi, Y)/(1 + \widehat{\varepsilon})$ is closer to χ^2_1:

$$C_3(\psi) = \Gamma_1\Big(\frac{D_n(\psi, y_{\mathrm{obs}})}{1 + \widehat{\varepsilon}}\Big).$$

6A: Bolt from Heaven

On 31 May 2008, Usain Bolt did 9.72. How surprised were we?

I spent the following night tracking down (and then analysing) the 195 sub-10.00-races recorded from seasons 2000, 2001, ..., 2007.



6B: The extreme value distribution

With r_i 100 m running times, I work with

$$y_i = 10.005 - r_i$$

and care only about sub-Armin-Hary races, i.e. $y_i > 0$.

Theorems of Fisher, Tippet, Gnedenko imply (under some conditions) that these very fast races follow the c.d.f.

$$G(y, a, \sigma) = 1 - (1 - ay/\sigma)^{1/a}$$

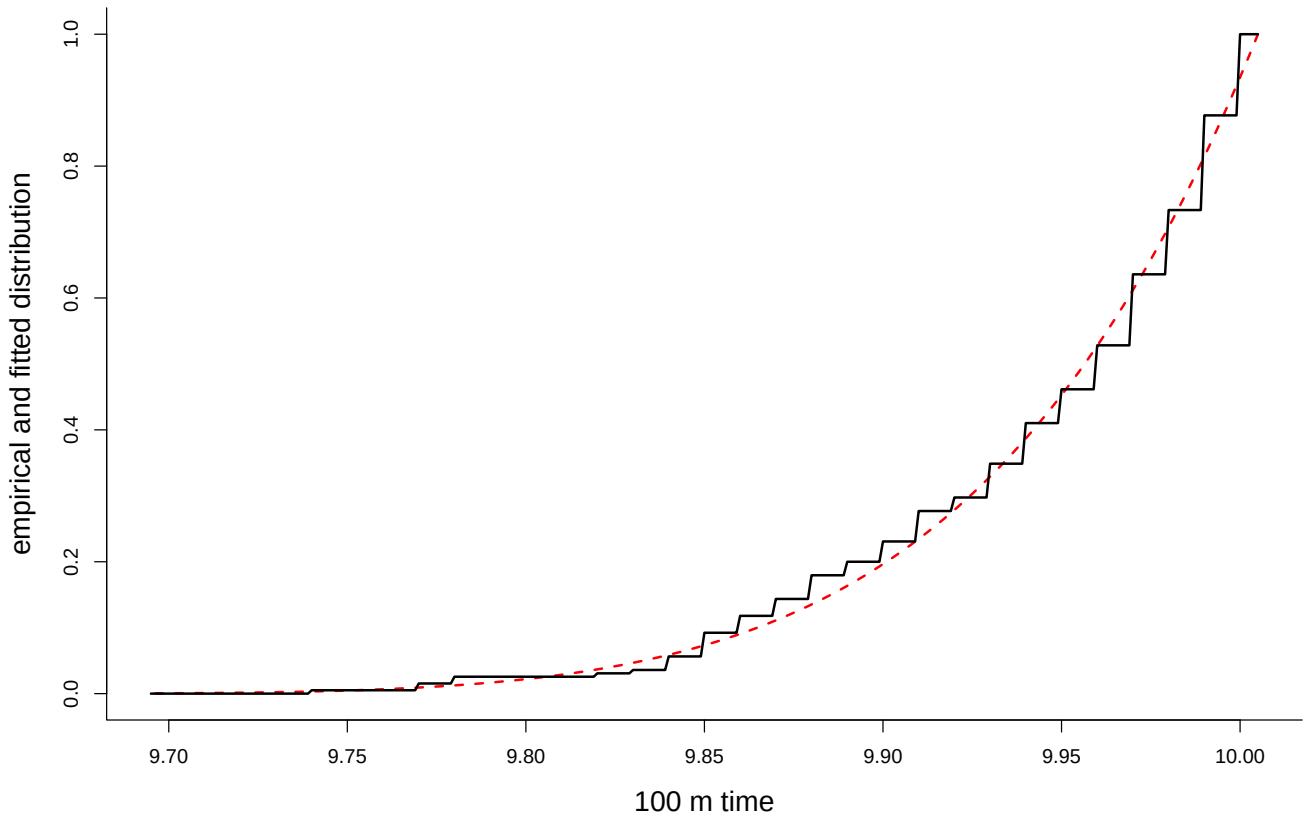
for some (a, σ). For the $n = 195$ (bona fide) races of 2000–2007, can hence form the log-likelihood function

$$\ell_n(a, \sigma) = \sum_{i=1}^{n} \{-\log \sigma + (1/a - 1) \log(1 - ay_i/\sigma)\}.$$

This leads to maximum likelihood estimates (and estimates of their precision).

The two-parameter extreme value model fits data very well:

$$\widehat{a} = 0.1821 \ (\mathrm{sd} = 0.0701), \widehat{\sigma} = 0.0745 \ (\mathrm{sd} = 0.0074).$$



6C: Distribution of best race in a season: a formula

For a season with N top races (below Hary threshold), consider

$$p = p(a, \sigma) = \Pr\{\max(Y_1', \ldots, Y_N') \geq w\}.$$

With $N \sim \mathrm{Pois}(\lambda)$, we find

$$p = p(a, \sigma) = 1 - \exp\{-\lambda(1 - aw/\sigma)^{1/a}\}.$$

I use $\lambda = 195/8 = 24.375$, rate of top races per year. For each threshold w we may estimate p and its approximate standard error.

With $w = 10.005 - 9.72 = 0.285$, for 31 May 2008, I find $\widehat{p} = 0.035$:

3.5% probability of seeing a 9.72 or better in the course of 2008, as judged from 1 January 2008.

6D: Full confidence distribution for $p = p(a, \sigma)$

Traditional approach: ML is approximately normal, so with delta method $0.035 \pm 1.96\,\widehat{\mathrm{sd}}$ would be a 0.95 interval, etc. This doesn't work well here, in spite of $n = 195$, and in spite of $(\widehat{a}, \widehat{\sigma})$ being approximately binormal: $p = p(a, \sigma)$ is not close to linear in this part of the parameter space.

Better (both for approximation quality and for representing uncertainty): the confidence distributions of Schweder and Hjort (Confidence, Likelihood, Probability, 2015). I compute the confidence curve

$$\mathrm{cc}(p) = \Gamma_1(D_n(p))$$

via the profiled deviance:

$$D_n(p) = 2\{\ell_n(\widehat{a}, \widehat{\sigma}) - \ell_{n,\mathrm{prof}}(p)\}.$$

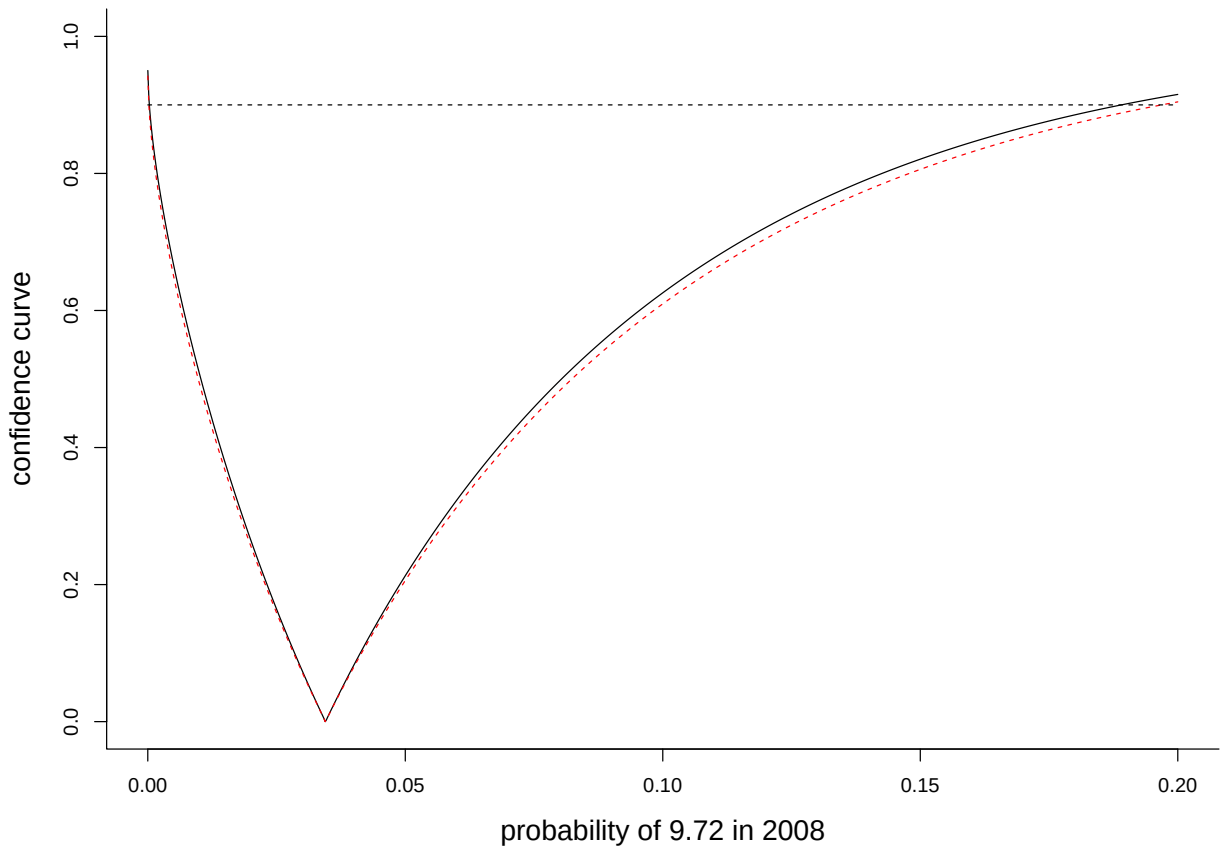
The $\mathrm{cc}(p)$ curve gives both the estimate (3.5%) and each confidence interval. These are highly skewed; the 90% interval is [0%, 18.9%], etc.

Even better with a mean correction to make distribution of $D_n(p)/(1+\varepsilon_n)$ closer to χ^2_1. For the case of

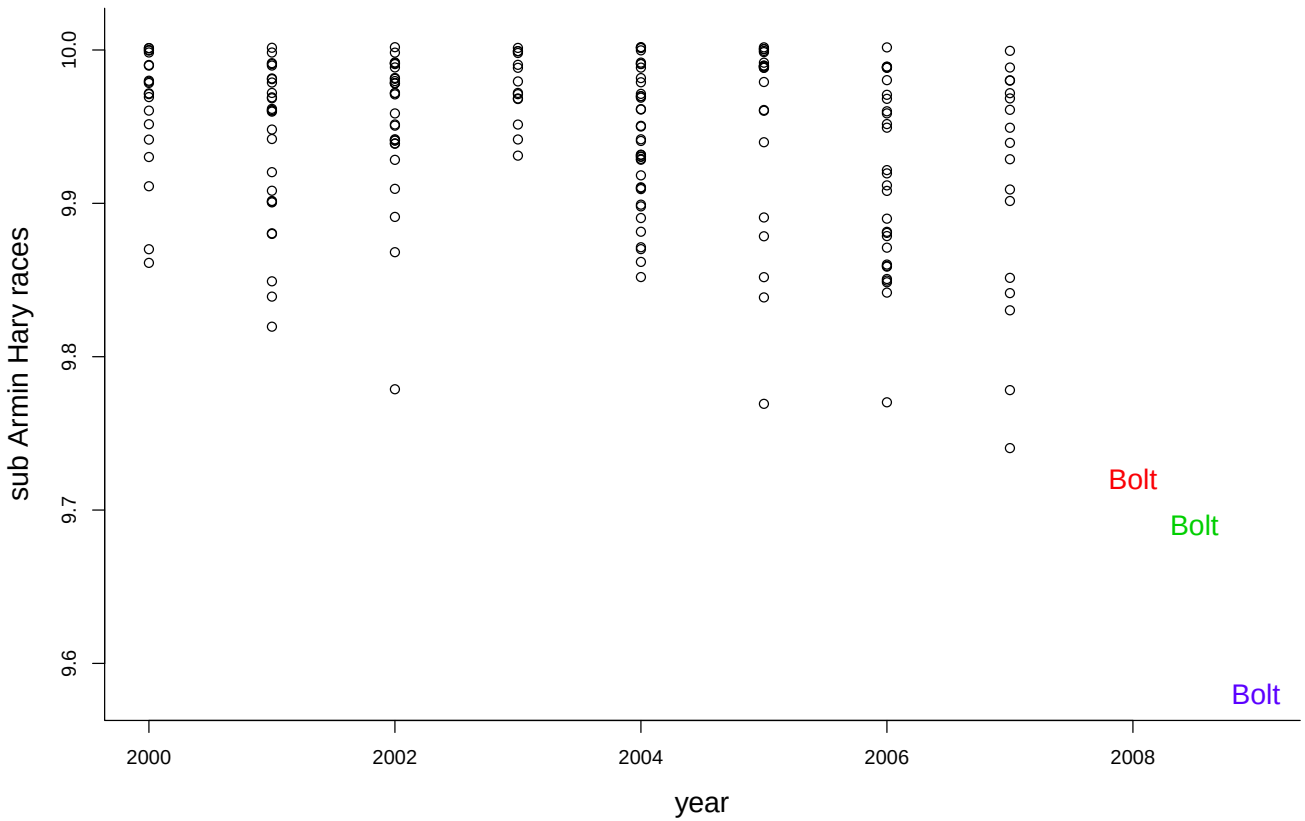
$$p = \Pr\{\text{seeing a 9.72 or better in 2008}\}$$

(as happened on May 31),

$$\mathrm{cc}^*(p) = \Gamma_1(D_n(p)/1.07).$$



Shock barometer (set up and calibrated as of January 2008):
Gay 9.77 (June 2008): shock = 79.7
Bolt 9.72 (May 2008): shock = 96.5
Bolt 9.69 (August 2008): shock = 99.3
Bolt 9.58 (August 2009): shock = beamonesque 100.0



6E: t-bootstrapping

We wish a CD for ψ. We work with

$$t = \frac{h(\psi) - h(\widehat{\psi})}{\widehat{\kappa}} \sim G(\cdot, \theta).$$

At the estimated position $\widehat{\theta}$ in parameter space:

$$t^* = \frac{h(\widehat{\psi}) - h(\widehat{\psi}^*)}{\widehat{\kappa}^*} \sim K = G(\cdot, \widehat{\theta}).$$

We estimate K via bootstrapping, and produce

$$C(\psi) = K\Big(\frac{h(\psi) - h(\widehat{\psi}_{\rm obs})}{\widehat{\kappa}_{\rm obs}}\Big) = G\Big(\frac{h(\psi) - h(\widehat{\psi}_{\rm obs})}{\widehat{\kappa}_{\rm obs}}, \widehat{\theta}\Big).$$

This often works very well. In case t is an exact pivot, the method is exact.

6F: abc bootstrapping

Suppose that on *some* transformed scale $\psi \to \gamma = h(\psi)$, $\widehat{\psi} \to \widehat{\gamma} = h(\widehat{\psi})$,

$$\frac{\gamma - \widehat{\gamma}}{1 + a\gamma} - b \sim \mathrm{N}(0, 1),$$

for certain (typically small) a (acceleration) and b (bias).

With $\widehat{G}(x) = \Pr_*\{\widehat{\psi}^* \le x\}$ the bootstrap distribution,

$$C_{\mathrm{abc}}(\psi) = \Phi\Big(\frac{\Phi^{-1}(\widehat{G}(\psi)) - b}{1 + a\Phi^{-1}(\widehat{G}(\psi)) - b} - b\Big)$$

works as an acceleration and bias corrected version.

It works very well in various test cases (where one knows the h transformation; the abc method does not use this knowledge).

7: Exponential model, GLMs (and GLLMs)

Exponential model class:

$$f(y) = \exp\{h_1(\theta)\,T_1(y) + \cdots + h_p(\theta)\,T_p(y) + m(y) - k(\theta)\}.$$

Long list of standard models are of this type (normal, binomial, Poisson, gamma, beta, geometric, multinormal, ...).

Canonical parametrisation: $\eta_1 = h_1(\theta), \ldots, \eta_p = h_p(\theta)$. Then, with data $y_1, \ldots, y_n$:

$$\ell_n(\eta) = n\{\eta_1\,\bar{T}_1 + \cdots + \eta_p\,\bar{T}_p - k_0(\eta)\}.$$

So $(\bar{T}_1, \ldots, \bar{T}_p)$ is sufficient, and theory is strong and clean.

There are optimal CDs for each η_j (and for linear combination of these).

Example: Optimal CD for a and b in Beta distribution

With

$$f = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1} \quad \text{for } y \in (0,1),$$

$$\ell_n = (a-1)\sum_{i=1}^{n} \log y_i + (b-1)\sum_{i=1}^{n} \log(1-y_i) + n\{\log\Gamma(a+b) - \log\Gamma(a) - \log\Gamma(b)\}.$$

For a certain dataset, I observe

$$U_n = \sum_{i=1}^{n} \log y_i = -1336.002 \quad \text{and} \quad V_n = \sum_{i=1}^{n} \log(1-y_i) = -255.978.$$

The power optimal CDs can be computed with (clever) simulation:

$$C^*(a) = \Pr_a\{U_n \geq -1336.002 \,|\, V_n = -255.978\},$$
$$C^*(b) = \Pr_b\{V_n \geq -255.978 \,|\, U_n = -1336.002\}.$$

Regression data (x_i, y_i), with

$$f(y_i, \theta_i, \phi) = \exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\},$$

and a link function

$$g(\mathrm{E}(Y_i \mid x_i)) = x_i^{\mathrm{t}}\beta \quad \text{for } i = 1, \ldots, n.$$

Long list of well-known special cases – linear (and non-linear) normal regression; Poisson regression; logistic and probit regression; gamma regression models; etc.

Theory works particularly well when the natural parameter is linear in covariates – then optimal CDs for each regression coefficient, etc.

Example: Gamma regression

Classic dataset (from 1965): y_i is survival time (in years) after leukaemia diagnosis, x_i is white blood count at time of diagnosis. Good model (better than in Cox and Snell etc.):

$$Y_i \sim \mathrm{Gamma}(\beta_0 + \beta_1(x_i - \bar{x}), \nu).$$

Sufficient statistics:

$$U = n^{-1}\sum_{i=1}^{n} Y_i, \quad V_0 = n^{-1}\sum_{i=1}^{n} \log Y_i, \quad V_1 = n^{-1}\sum_{i=1}^{n} \log(x_i - \bar{x})\, Y_i,$$

with values $1.201, -0.589, -0.693$. May then compute optimal CD for β_1 via (clever) simulation (I've used methods of Lindqvist and Taraldsen, 2006, 2007):

$$C^*(\beta_1) = \Pr_{\beta_1}\{V_1 \geq -0.693 \,|\, U = 1.201, V_0 = -0.589\}.$$

(Context, figure, discussion: CLP Example 8.6.)

7C: Extending models by exponential tilting

Consider a 'start model' $f_0(y, \theta)$ with $(\theta_1, \ldots, \theta_p)$. May then extend this to

$$f(y, \theta, \gamma) = f_0(y, \theta) \exp\{\gamma^{\mathrm{t}} T(y) - k(\theta, \gamma)\},$$

with

$$k(\theta, \gamma) = \log\Big[\int f_0(y, \theta) \exp\{\gamma^{\mathrm{t}} T(y)\} \, \mathrm{d}y\Big].$$

May carry out ML analysis etc. for the full (θ, γ), and in particular provide optimal CDs for the extension parameter $\gamma_1, \ldots, \gamma_q$. This may be used for goodness-of-fit purposes etc.

Example: extended and tilted Beta distribution $f(y, a, b, \gamma_1, \gamma_2)$, of the type

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1} \exp\{\gamma_1 y + \gamma_2 y^2 - k(a, b, \gamma_1, \gamma_2)\}.$$

7D: Generalised linear-linear models (GLLM)

GLMs: express covariance influence on one of the model parameters, like

$$y_i \sim \mathrm{N}(\exp(x_i^{\mathrm{t}}\beta), \sigma^2).$$

But we may easily extend this to covariance influence on two parameters:

$$y_i \sim \mathrm{N}(x_i^{\mathrm{t}}\beta, \exp(x_i^{\mathrm{t}}\gamma)).$$

Similarly:

$$y_i \sim \mathrm{Gamma}(a_i, b_i) \quad \text{with } a_i = \exp(x_i^{\mathrm{t}}\beta), b_i = \exp(x_i^{\mathrm{t}}\gamma).$$

This might be supplemented with AIC or FIC analyses; also, CD checks on all 'extra parameters'.

Discussion and applications: CLP Section 8.8.

8: Fusion and meta-analysis (and II-CC-FF)

Suppose there are information sources $1, \ldots, k$, each of (direct or indirect) relevance to focus parameter ψ. Often: information source j relates to ψ_j, and ψ is a function of $\psi_1, \ldots, \psi_k$.

The II-CC-FF paradigm:

- Independent Inspection of different available information sources, leading to $C_1(\psi_1), \ldots, C_k(\psi_k)$;
- Confidence Conversion from CDs to (profiled) log-likelihoods, translating to $\ell_1(\psi_1), \ldots, \ell_k(\psi_k)$;
- Focused Fusion based on $\sum_{j=1}^{k} \ell_j(\psi_j)$, with further profiling etc.

Classic meta-analysis with a common mean: $y_j \sim \mathrm{N}(\theta, \sigma_j^2)$ with common θ. Then the above gives

$$\ell = -\tfrac{1}{2} \sum_{j=1}^{k} \frac{(\theta - y_j)^2}{\sigma_j^2}$$

and

$$C^*(\psi) = \Phi\Big(\frac{\theta - \theta^*}{\sigma^*}\Big),$$

where

$$\theta^* = \frac{\sum_{j=1}^k y_j/\sigma_j^2}{\sum_{j=1}^k 1/\sigma_j^2} \quad \text{and} \quad \frac{1}{(\sigma^*)^2} = \sum_{j=1}^{k} \frac{1}{\sigma_j^2}.$$

Fusion with non-common means:

$$y_j \,|\, \theta_j \sim \mathrm{N}(\theta_j, \sigma_j^2) \quad \text{and} \quad \theta_j \sim \mathrm{N}(\theta_0, \tau^2).$$

Then $y_j \sim \mathrm{N}(\theta_0, \sigma_j^2 + \tau^2)$. Interest in both grand mean θ_0 and in spread parameter τ.

For given τ,

$$\widehat{\theta}(\tau) = \sum_{j=1}^{k} \frac{y_j}{\sigma_j^2 + \tau^2} \Big/ \sum_{j=1}^{k} \frac{1}{\sigma_j^2 + \tau^2}$$

is best. Also,

$$Q(\tau) = \sum_{j=1}^{k} \frac{\{y_j - \widehat{\theta}(\tau)\}^2}{\sigma_j^2 + \tau^2} \sim \chi^2_{k-1}.$$

Exact CD for τ:

$$C(\tau) = 1 - \Gamma_{k-1}(Q(\tau)),$$

with pointmass $C(0) = 1 - \Gamma_{k-1}(Q(0))$ at zero.

Fusion 1: Effective population size ratio for cod

A certain population of cod is studied. Of interest is both actual population size N and effective population size N_e (the size of a hypothetical stable population, with the same genetic variability as the full population, and where each individual has a binomially distributed number of reproducing offspring). The biological focus parameter in this study is $\phi = N_e/N$.

Steps II-CC for N: A CD for N, with confidence log-likelihood: A certain analysis leads to confidence log-likelihood

$$\ell_c(N) = -\tfrac{1}{2}(N - 1847)^2/534^2.$$

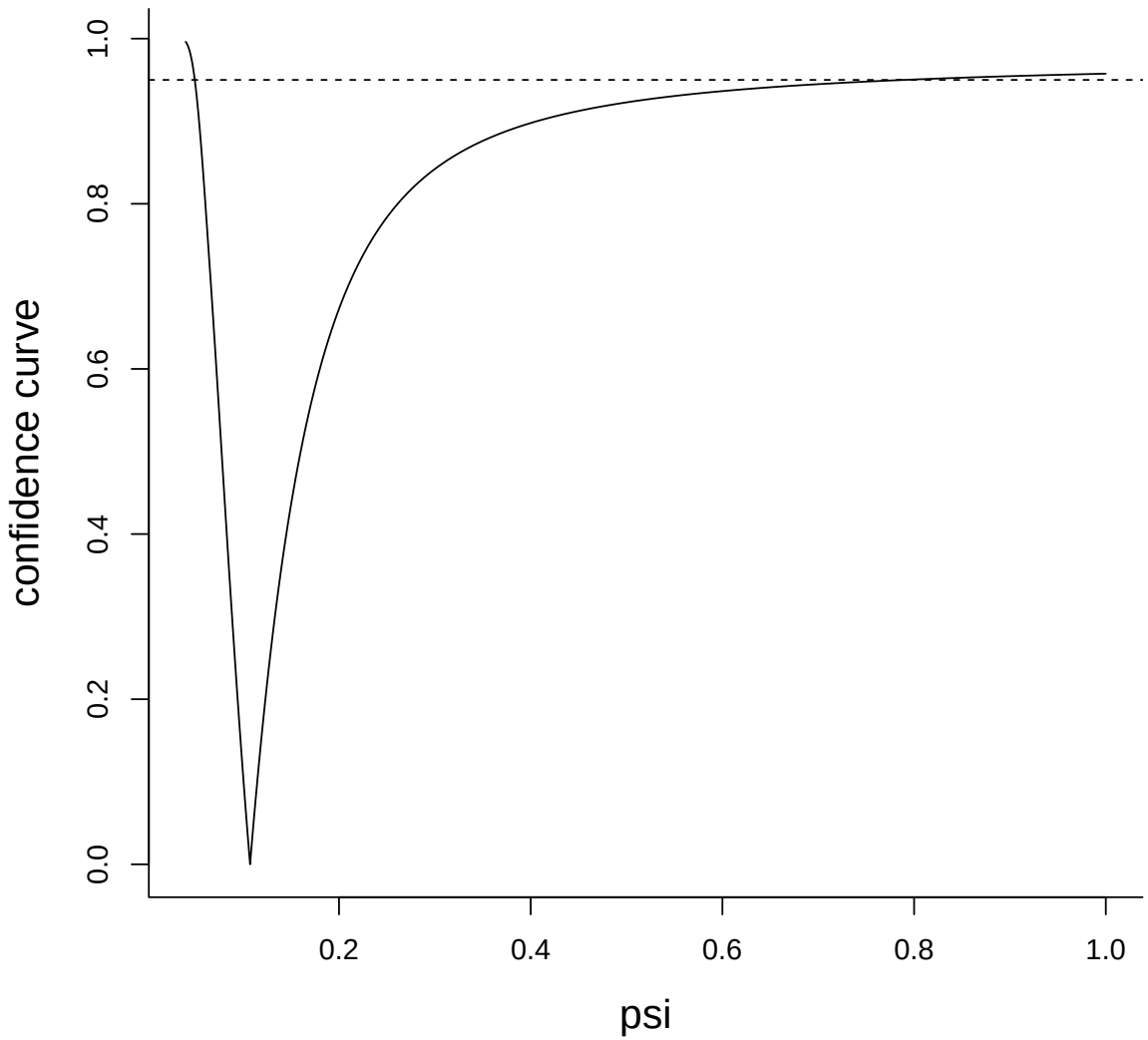
Steps II-CC for N_e: A CD for N_e, with confidence log-likelihood: This is harder, via genetic analyses, etc., but yields confidence log-likelihood

$$\ell_{c,e}(N_e) = -\tfrac{1}{2}(N_e^b - 198^b)/s^2$$

for certain estimated transformation parameters (b, s).

Step FF for the ratio: A CD for $\phi = N_e/N$. This is achieved via log-likelihood profiling and median-Bartletting,

$$\ell_{\text{prof}}(\phi) = \max\{\ell_c(N) + \ell_{c,e}(N_e) \colon N_e/N = \phi\}.$$



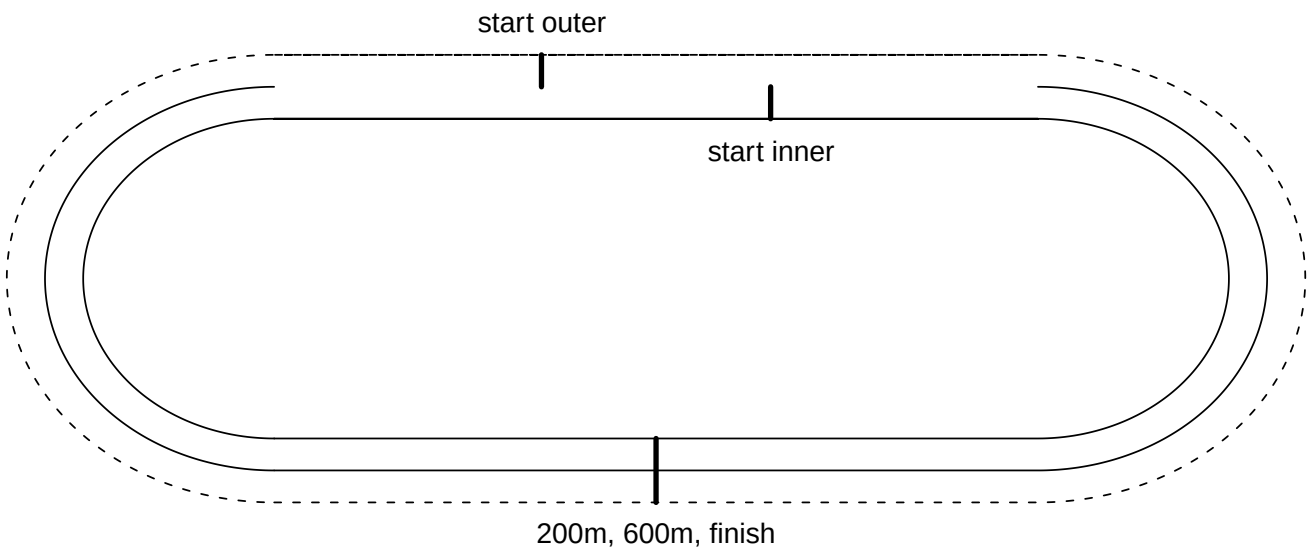
Fusion 2: The Olympic unfairness of the 1000 m

Olympic speedskaters run the 1000 m in less than 70 seconds (speed more than 50 km/h). They skate two and a half laps, in pairs, with a draw determining inner/outer. Acceleration matters ($mv^2/r_1 > mv^2/r_2$ with $r_1 = 25$ m and $r_2 = 29$ m), and so does fatigue at end of race.

Start in inner lane: three inners, two outers.

Start in outer lane: two inners, three outers.

I shall estimate the Olympic unfairness parameter d, the difference between outer and inner, for top skaters.



In the Olympics: only one race. In the annual World Sprint Championships: they race 500 m and 1000 m both Saturday and Sunday, and they switch start lanes.

The six best men, from Calgary, January 2012, Saturday and Sunday, with 'i' and 'o' start lanes, and passing times:

			200 m	600 m	1000 m		200 m	600 m	1000 m
1	S. Groothuis	i	16.61	41.48	1:07.50	o	16.50	41.10	1:06.96
2	Kyou-Hyuk Lee	i	16.19	41.12	1:08.01	o	16.31	40.94	1:07.99
3	T.-B. Mo	o	16.57	41.67	1:07.99	i	16.27	41.54	1:07.99
4	M. Poutala	i	16.48	41.50	1:08.20	o	16.47	41.55	1:08.34
5	S. Davis	o	16.80	41.52	1:07.25	i	17.02	41.72	1:07.11
6	D. Lobkov	i	16.31	41.29	1:08.10	o	16.35	41.26	1:08.40

I need a model for (Sat, Sun) results (Y_1, Y_2), utilising passing times $u_{i,1}, v_{i,1}$ for Sat race and $u_{i,2}, v_{i,2}$ for Sun race, along with

$$z_{i,1} = \begin{cases} -1 & \text{if no. } i \text{ starts in inner on Saturday,} \\ 1 & \text{if no. } i \text{ starts in outer on Saturday,} \end{cases}$$

$$z_{i,2} = \begin{cases} -1 & \text{if no. } i \text{ starts in inner on Sunday,} \\ 1 & \text{if no. } i \text{ starts in outer on Sunday.} \end{cases}$$

to get hold of d.

My model for (Sat, Sun) results, for skater i:

$$Y_{i,1} = a_1 + bu_{i,1} + cv_{i,1} + \tfrac{1}{2}dz_{i,1} + \delta_i + \varepsilon_{i,1},$$
$$Y_{i,2} = a_2 + bu_{i,2} + cv_{i,2} + \tfrac{1}{2}dz_{i,2} + \delta_i + \varepsilon_{i,2}.$$

Here $u_{i,1}, u_{i,2}$ are 200 m passing time, $v_{i,1}, v_{i,2}$ are 600 m passing time; δ_i follows the skater, with $\delta_i \sim \mathrm{N}(0, \kappa^2)$ across skaters; and $\varepsilon_{i,1}, \varepsilon_{i,2}$ are independent $\mathrm{N}(0, \sigma^2)$. The inter-skater correlation is $\rho = \kappa^2/(\sigma^2 + \kappa^2)$.

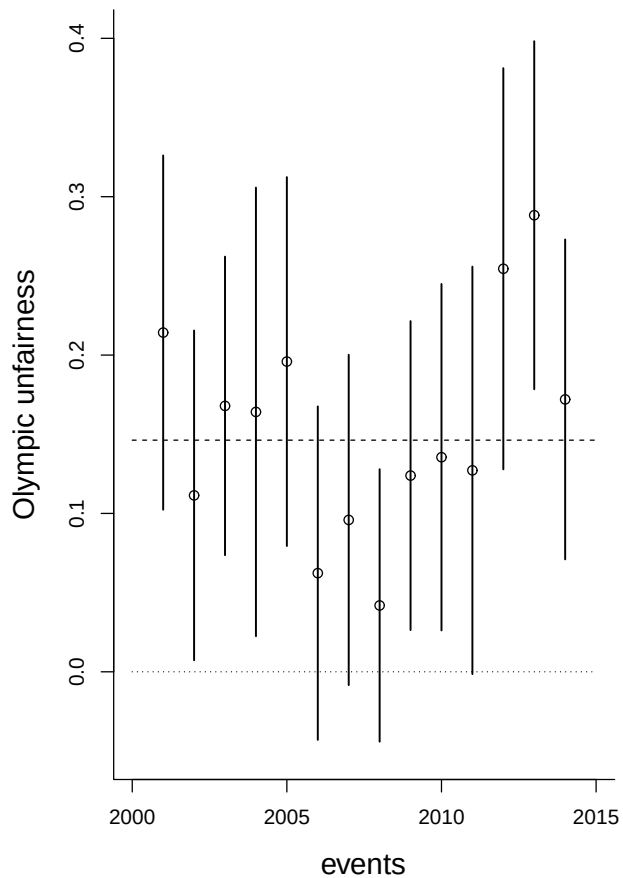
Crucially, outer lane start means adding $\frac{1}{2}d$, inner lane start means adding $-\frac{1}{2}d$, so d is overall difference due to start lane. Fairness means d should be very close to zero.

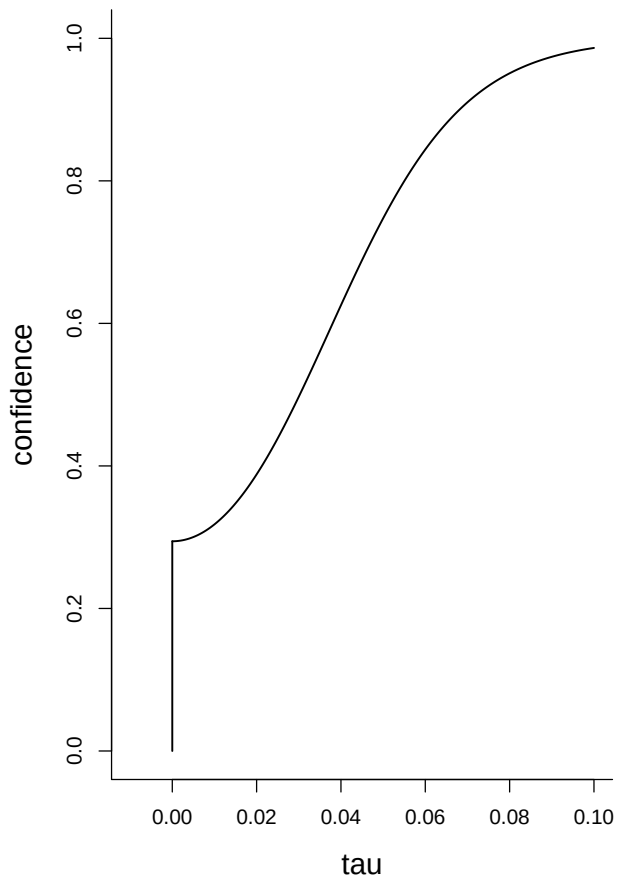
The model has seven parameters, and I need full analysis of dataset from each World Sprint Championships event to get hold of a CD for the focus parameter d.

From full analysis of World Sprint events 2014, ..., 2001 (seven parameters in each model), I get hold of

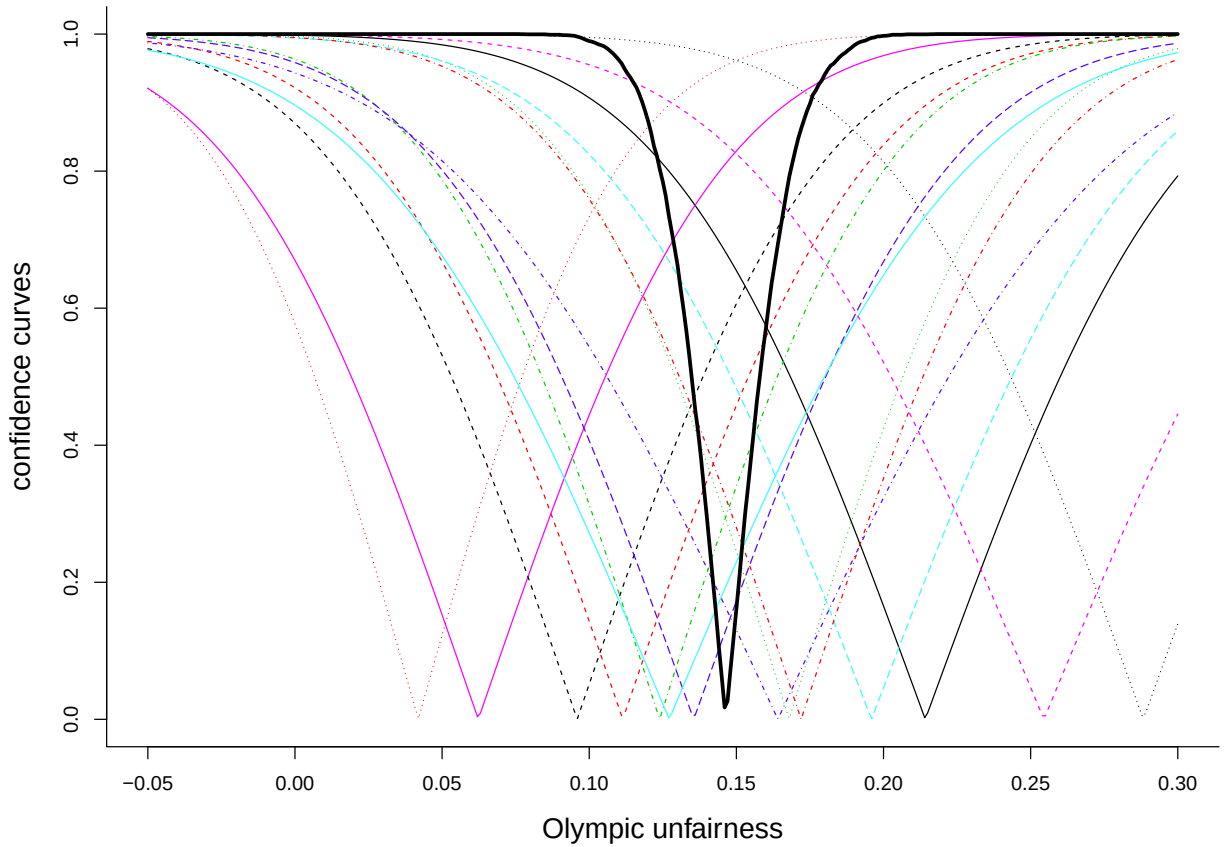
$$\widehat{d}_j \sim \mathrm{N}(d_j, \sigma_j^2),$$

and I then use $d_j \sim \mathrm{N}(d_0, \tau^2)$. Full CD analyses are then available for d_0 and for τ.





Confidence curves $\mathrm{cc}(d_j)$ for the fourteen unfairness parameters, over 2014 to 2001. The overall estimate 0.14 seconds (advantage inner-starter) is very significant, and big enough to make medals change necks.



Conclusion: The skaters need to run twice. (I've told the ISU.)

Fusion 3: Rosiglitazone drug and paired binomials

The drug is for type 2 diabetes mellitus and is controversial – annual sales exceed two billion dollars, but it is the subject of 13,000 lawsuits, due to alleged association with heart attacks. Meta-analyses have been carried out based on 48 two-by-two tables,

$$Y_{i,0} \sim \text{Bin}(m_{i,0}, p_{i,0}) \quad \text{and} \quad Y_{i,1} \sim \text{Bin}(m_{i,1}, p_{i,1}).$$

The model used in several papers takes

$$\theta_i = \log \frac{p_{i,0}}{1-p_{i,0}} \quad \text{and} \quad \theta_i + \psi = \log \frac{p_{i,1}}{1-p_{i,1}}$$

with ψ the main focus parameter. Analysis is troubled by 'null cells' and 'null tables' – should these be included?, what information do these provide?

The log-likelihood for table i is

$$\ell_i(\theta_i, \psi) = y_{i,1}\psi + z_i\theta_i - m_{i,0}\log\{1+\exp(\theta_i)\} - m_{i,1}\log\{1+\exp(\theta_i+\psi)\},$$

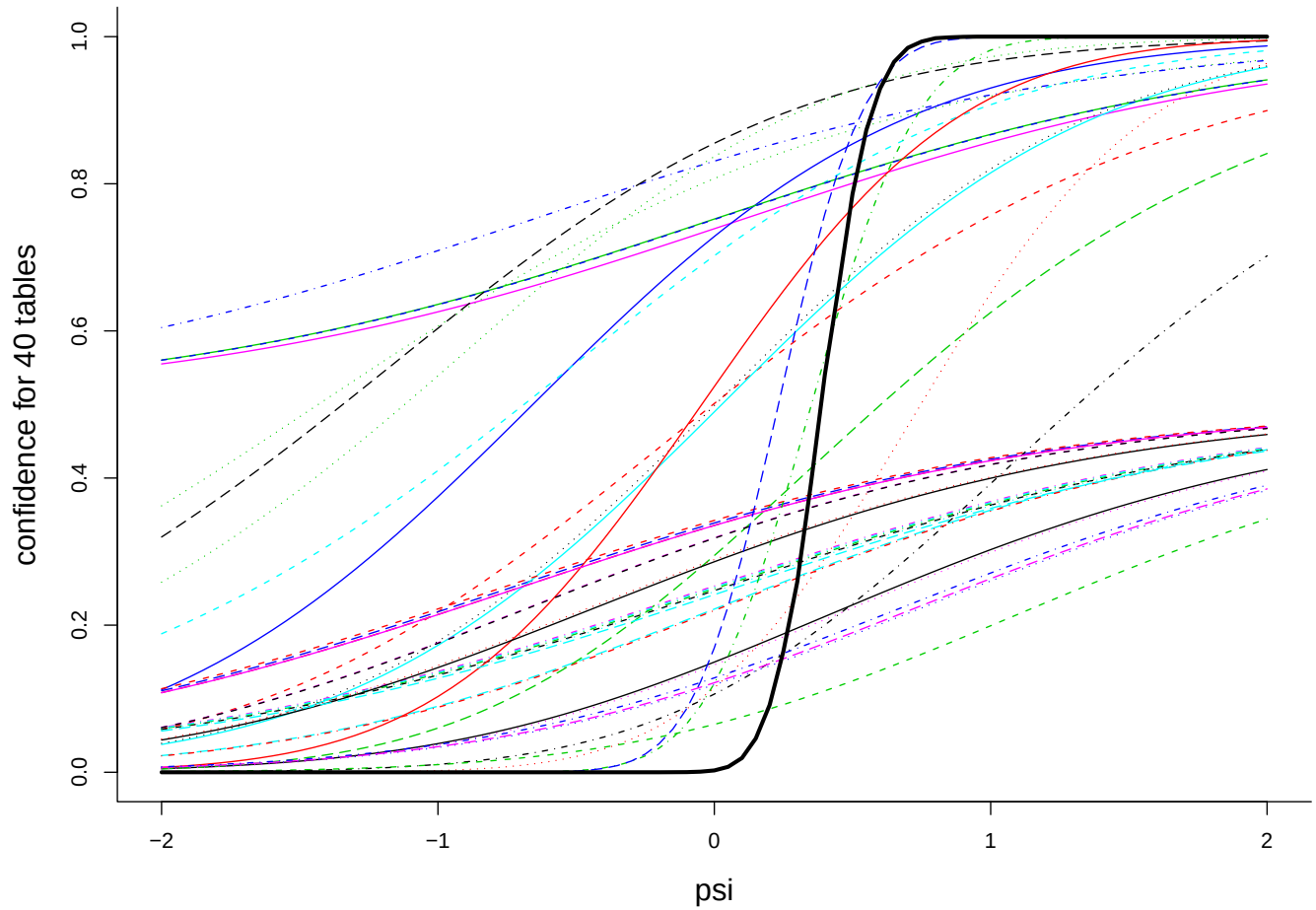
where $z_i = y_{i,0} + y_{i,1}$. The power theorem yields optimal CD for each table, needing

$$f(y_{i,1} \mid z_i) = \frac{\binom{m_{i,0}}{z_i - y_{i,1}}\binom{m_{i,1}}{y_{i,1}}\exp(\psi y_{i,1})}{\sum_{y_1'=0}^{z_i}\binom{m_{i,0}}{z_i - y_1'}\binom{m_{i,1}}{y_1'}\exp(\psi y_1')}.$$

It also yields the optimal overall CD for ψ utilising all 48 tables, via

$$\ell(\theta_1, \ldots, \theta_{48}, \psi) = \Big(\sum_{i=1}^{48} y_{i,1}\Big)\psi + \sum_{i=1}^{48} z_i\theta_i - k(\theta_1, \ldots, \theta_{48}, \psi).$$

The conditional distribution of $\sum_{i=1}^{48} y_{i,1}$ given $z_1, \ldots, z_{48}$ may be simulated. Bayes is in trouble here.



Optimal CDs for log-odds difference ψ for each of the 48 separate studies, along with the optimal overall CD (fat curve).

Alternative Poisson model:

$$Y_{i,0} \sim \text{Pois}(e_{i,0}\lambda_{i,0}) \quad \text{and} \quad Y_{i,1} \sim \text{Pois}(e_{i,1}\lambda_{i,1})$$

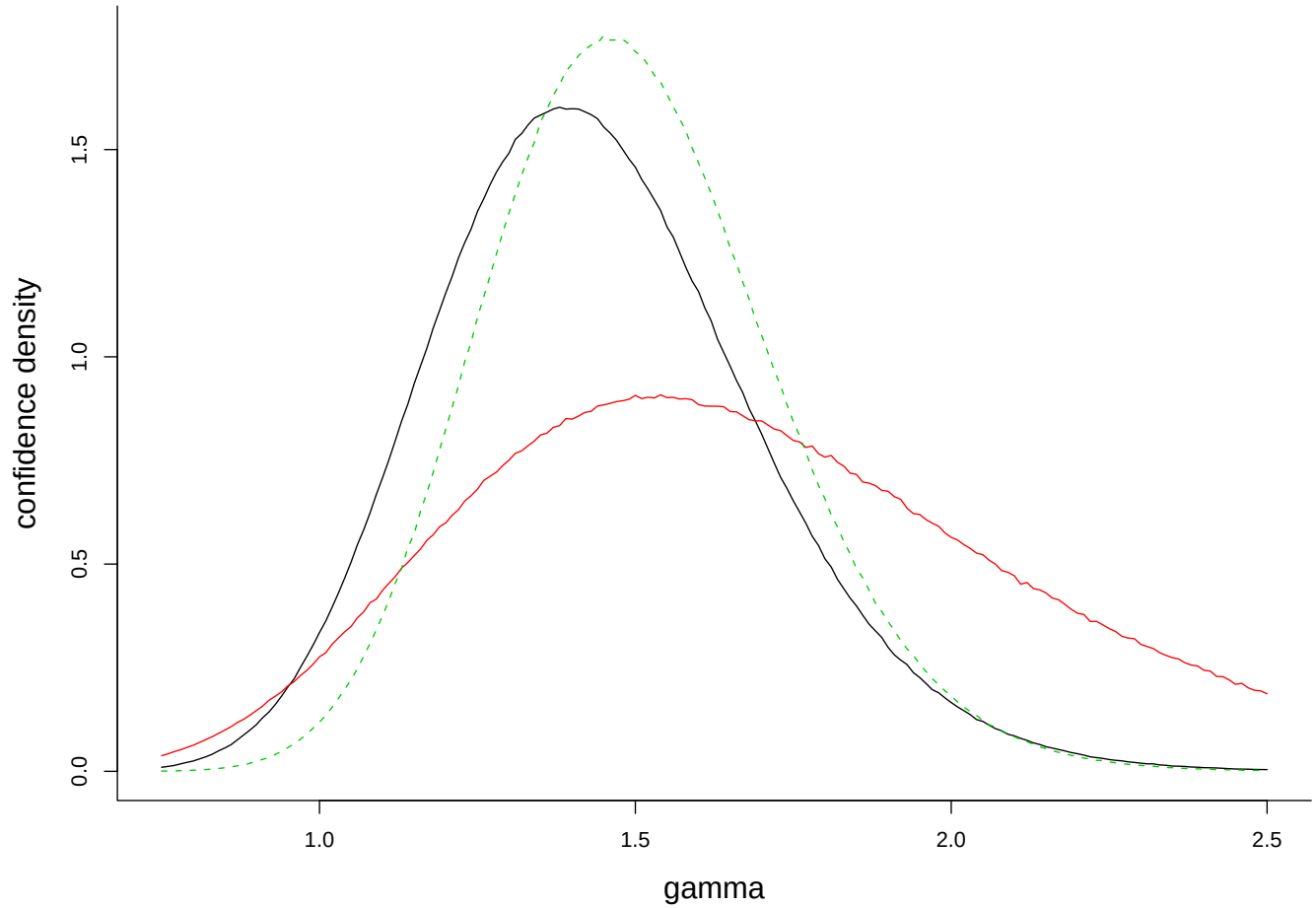
with $\lambda_{i,1} = \lambda_{i,0}\gamma$.

May use the power theorem to provide optimal overall CD for the focus parameter γ, as

$$\ell(\gamma, \lambda_{1,0}, \ldots, \lambda_{48,0}) = \sum_{i=1}^{48} y_{i,1} \log \gamma + \sum_{i=1}^{48} z_i \log \lambda_{0,i} - k(\gamma, \lambda_{1,0}, \ldots, \lambda_{48,0}).$$

The conditional distribution is a sum over 48 binomials with different parameters, and is evaluated via simulation, for each γ.

The figure displays optimal confidence curves $\text{cc}(\gamma)$ for (i) m.i. deaths, (ii) c.v.d. deaths, (iii) m.i. + c.v.d. deaths.



Optimal confidence curves $\mathrm{cc}(\gamma)$ for the risk proportionality parameter γ for m.i. only (ML = 1.421), for c.v.d. only (ML = 1.659), and for m.i. + c.v.d. (ML = 1.482).

Fusion 4: Whale abundance

These are published results of a humpack whale population size (via complicated data collections, models, analyses):

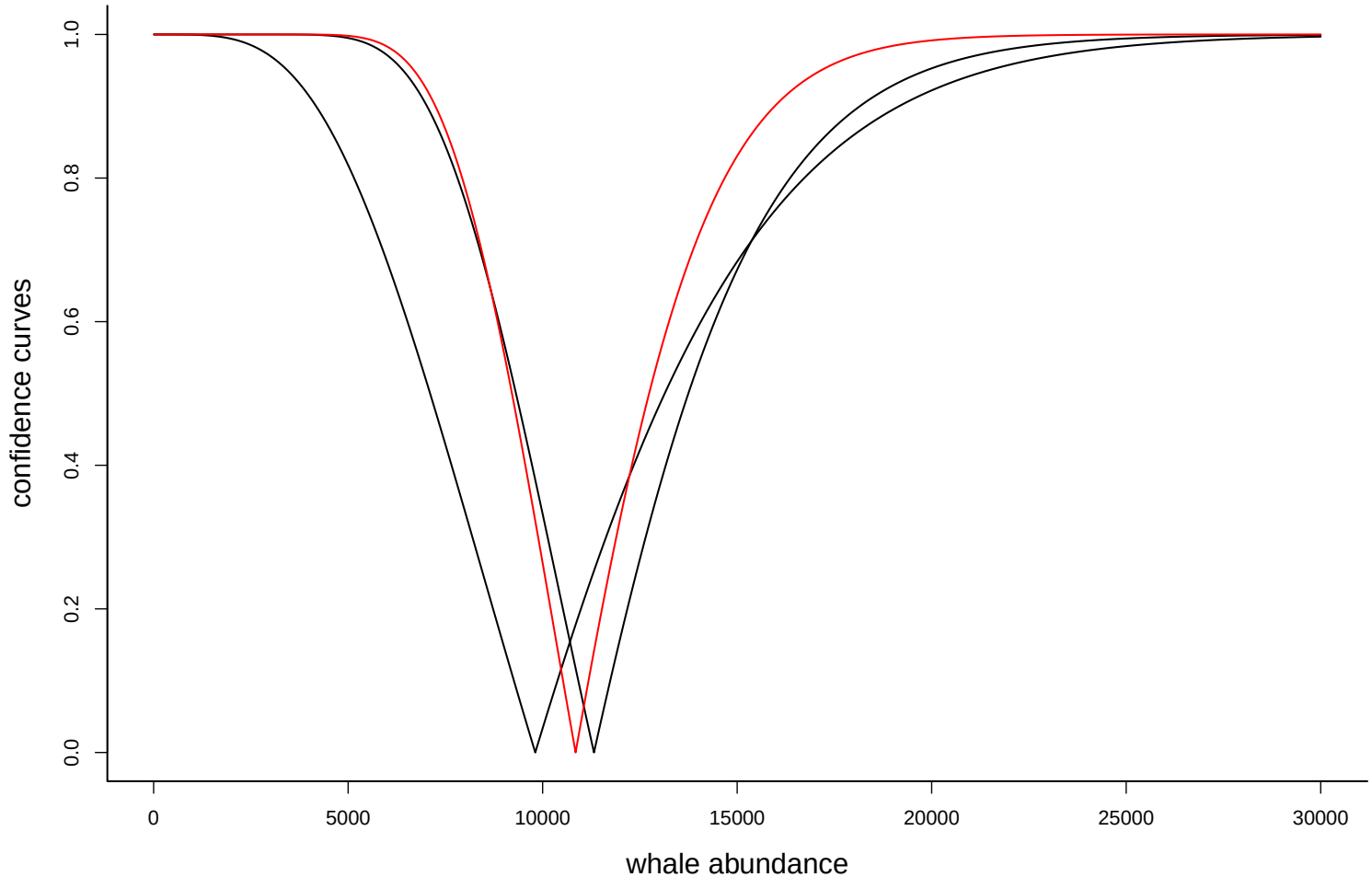
	2.5%	50%	97.5%
1995	3439	9810	21457
2001	6651	11319	21214

Suppose that this is all available information (no raw data; we just read the summaries of the two papers).

- How can we translate this to confidence curves for abundance, for 1995 and 2001?
- How can we do data fusion, arriving at one cc?

(Here we're *not* assuming normality.)

Via certain transformation tricks (etc.), Cunen and Hjort (2016):



9: CDs for prediction

What is the distribution of a y_{new} (not yet observed)?
(i) $C_{\text{pred}}(y_{\text{new}}, y)$ should be a cdf in y_{new} for each dataset y;
(ii) $C_{\text{pred}}(Y_{\text{new}}, Y) \sim \text{unif}$.

Example: the next normal. With $y_1, \ldots, y_n$ i.i.d. $\text{N}(\mu, \sigma^2)$, we have $y_{\text{new}} - \bar{y} \sim \text{N}(0, \sigma^2(1 + 1/n))$. With σ known:

$$t = \frac{y_{\text{new}} - \bar{y}}{\sigma(1 + 1/n)^{1/2}} \sim \text{N}(0, 1),$$

and

$$C_{\text{pred}}(y_{\text{new}}, \text{data}) = \Phi\Big(\frac{y_{\text{new}} - \bar{y}}{\sigma(1 + 1/n)^{1/2}}\Big).$$

With σ unknown:

$$C_{\text{pred}}(y_{\text{new}}, \text{data}) = F_\nu\Big(\frac{y_{\text{new}} - \bar{y}}{\widehat{\sigma}(1 + 1/n)^{1/2}}\Big).$$

Regression: Suppose $y_i \sim \mathrm{N}(x_i^{\mathrm{t}}\beta, \sigma^2)$ for $i = 1, \ldots, n$. What is the value of a new y_{new}, with covariates x_{new}? Here

$$\widehat{\beta} = (X^{\mathrm{t}}X)^{-1}X^{\mathrm{t}}y \sim \mathrm{N}_p(\beta, \sigma^2\Sigma^{-1}),$$

with $\Sigma = X^{\mathrm{t}}X = \sum_{i=1}^{n} x_i x_i^{\mathrm{t}}$.

For a new x_{new}:

$$\frac{y_{\mathrm{new}} - x_{\mathrm{new}}^{\mathrm{t}}\widehat{\beta}}{\widehat{\sigma}(1 + x_{\mathrm{new}}^{\mathrm{t}}\Sigma^{-1}x_{\mathrm{new}})^{1/2}} \sim t_\nu,$$

so

$$C_{\mathrm{pred}}(y_{\mathrm{new}}, \mathrm{data}) = F_\nu\Big(\frac{y_{\mathrm{new}} - x_{\mathrm{new}}^{\mathrm{t}}\widehat{\beta}}{\widehat{\sigma}(1 + x_{\mathrm{new}}^{\mathrm{t}}\Sigma^{-1}x_{\mathrm{new}})^{1/2}}\Big).$$

More challenging (but more important): time series, spatial models, kriging, etc.

From approximate CD predictive to a better CD predictive:

$$F(Y_{\text{new}}, \widehat{\theta}) \to_d F(Y_{\text{new}}, \theta_0) \sim \text{unif},$$

so for large n, plug-in $C_{\text{pred},0}(y_{\text{new}}, y) = F(y, \widehat{\theta})$ is ok.

But it can be improved upon:

$$F(Y_{\text{new}}, \widehat{\theta}) \sim G(\cdot, \theta) \text{ (not yet fully uniform)},$$

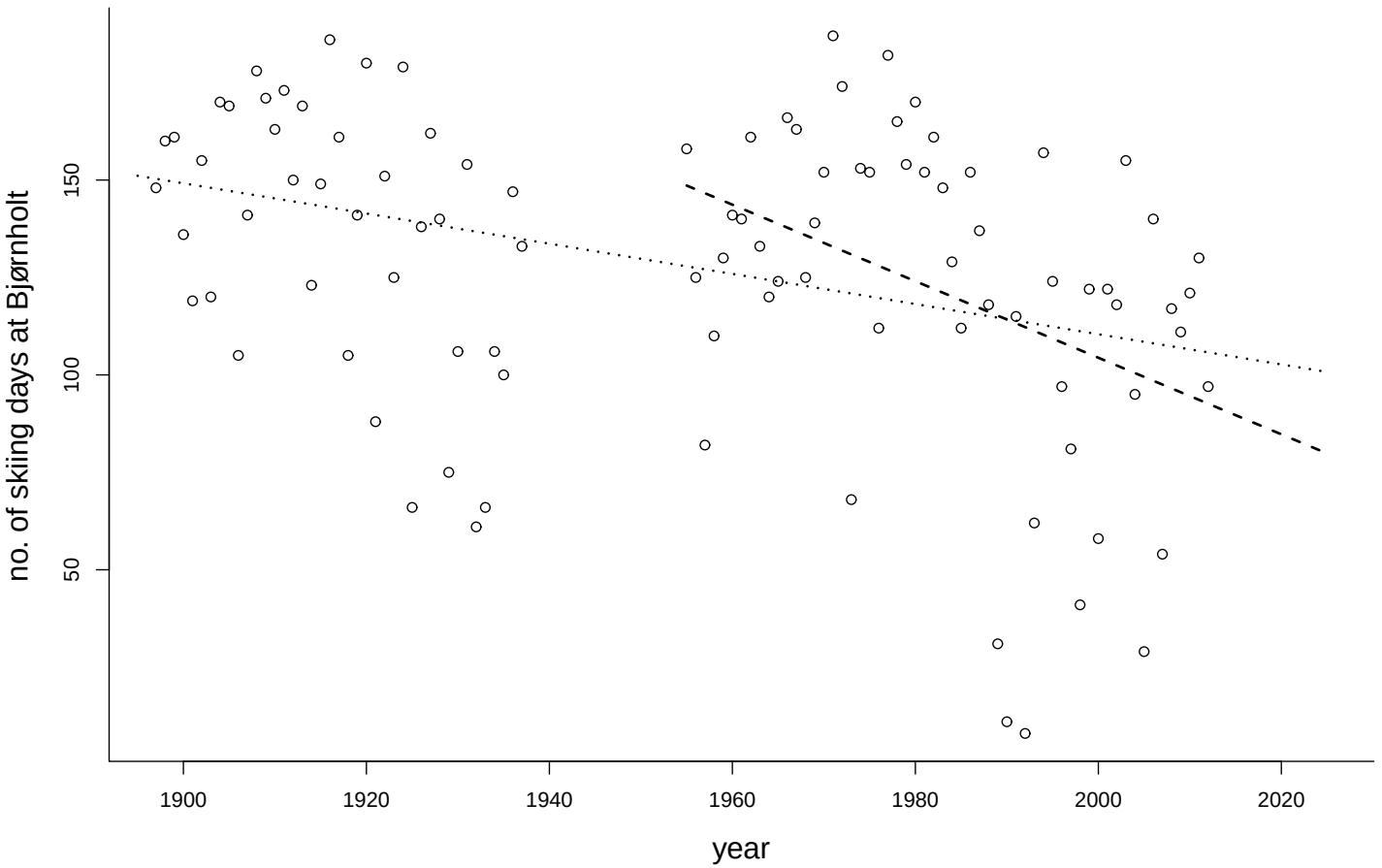
so

$$G(F(Y_{\text{new}}, \widehat{\theta}), \theta) \sim \text{unif (exact)}.$$

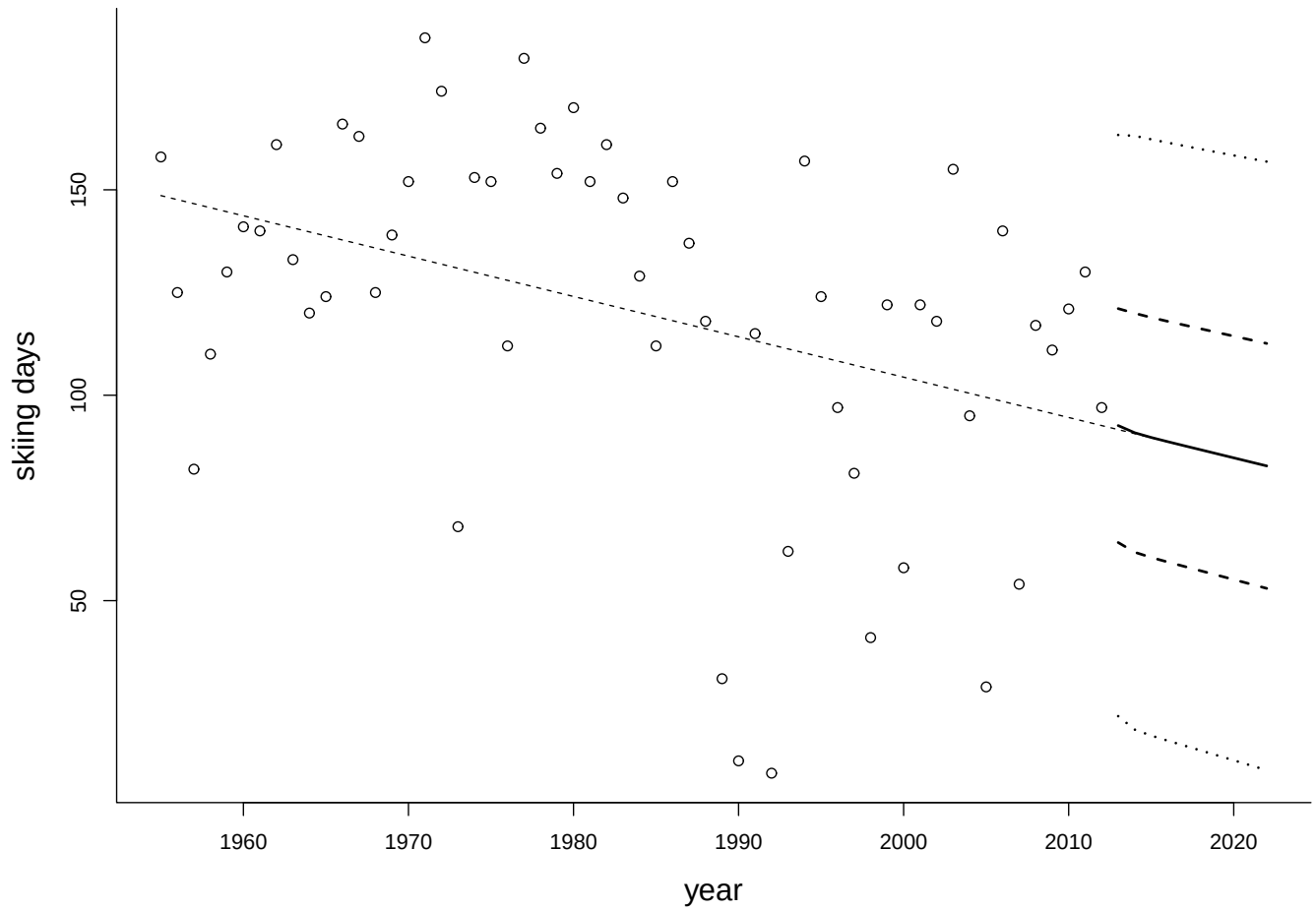
We estimate, using a 'second round', via simulation or bootstrapping:

$$\begin{aligned} C_{\text{pred},1}(y_{\text{new}}, y) &= G(F(y_{\text{new}}, \widehat{\theta}), \widehat{\theta}) \\ &= \frac{1}{B} \sum_{j=1}^{B} I\{F(y^*_{j,\text{new}}, \widehat{\theta}^*_j) \leq F(y_{\text{new}}, \widehat{\theta}_{\text{obs}})\}. \end{aligned}$$

Skiing days at Bjørnholt



Quo vadimus? The number of skiings days per year at Bjørnholt.
When did the gradient change? What's y_{2030}?



Predictive confidence distributions for the number of skiing days at Bjørnholt, 2014 to 2022, based on data up to 2013. Displayed are 0.05, 0.25, 0.50, 0.75, 0.95 quantiles.

10: Extensions and related themes

Efron (1998): "Maybe Fisher's biggest blunder will become a big hit in the 21st century!"

- confidence curves for change-points and regime shifts (Cunen, Hermansen, Hjort, JSPI 2017)
- use of models when they're not (necessarily) correct
- robust estimation (Walker and Hjort, 2017)
- post-model-selection
- bigger models
- epistemic vs. aleatory vs. subjective probability
- factoring in Bayes in II-CC-FF:

$$\log \pi_0(\psi) + \sum_{j=1}^{k} \ell_j(\psi_j)$$

(requiring prior only for ψ, not for full $(\theta_1, \ldots, \theta_p)$)