Distributional inference by confidence curves

Tore Schweder
University of Oslo

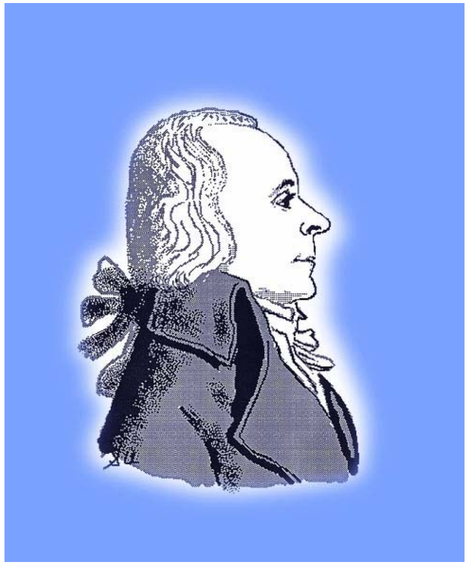
tore.schweder@econ.uio.no

Outline

- Some history
- "Fishers biggest blunder" - the fiducial argument
- Confidence distributions and confidence curves
- Neyman-Pearson lemma
- Confidence and likelihood
- Combining information
- To bias-correct or not to bias-correct
- Box-shaped confidence curves - nested families of confidence bands

- Applications
  - Quantile regression on Norwegian income data
  - Abundance of bowhead whales from Alaskan photo-ID data

History



Laplace 1774-1786

- Inverse probability: Bayesian posteriors from flat priors

$$p(\theta) = \frac{f(x;\theta)}{\int f(x;t)\,dt}$$

R.A. FISHER 1929

R.A. Fisher (1930) *Inverse probability*

"to end 150 years of fog and confusion". "I know of only one case in mathematics .. accepted .. by the most eminent men [Laplace and Gauss...] ...to be fundamentally false and devoid of foundation. Yet that is exactly the position in respect to inverse probability ..error on a question of prime theoretical importance...Inverse probability has, I beleive, survived so long in spite of its unsatisfactory basis, because its critics have until recent times put forward nothing to replace it as a rational theory of learning by experience."

Pivots

- Have the same distribution regardless of the parameter
- Are monotoneous (increasing) in the parameter a.s.

$$piv(\theta, X) : F$$

$$\theta \underset{fidu}{:} F\left(piv(\theta, X_{obs})\right)$$

Example

The chi-square pivot for the empirical variance at *v* degrees of freedom yields the fiducial cdf:

Fiducial density:

$$\frac{\nu s^2}{\sigma^2} : K_\nu$$

$$C(\sigma) = 1 - K_\nu\left(\frac{\nu s_{obs}^2}{\sigma^2}\right)$$

$$c(\sigma) = C'(\sigma) = \frac{2\nu s_{obs}^2}{\sigma^3} k_\nu\left(\frac{\nu s_{obs}^2}{\sigma^2}\right)$$

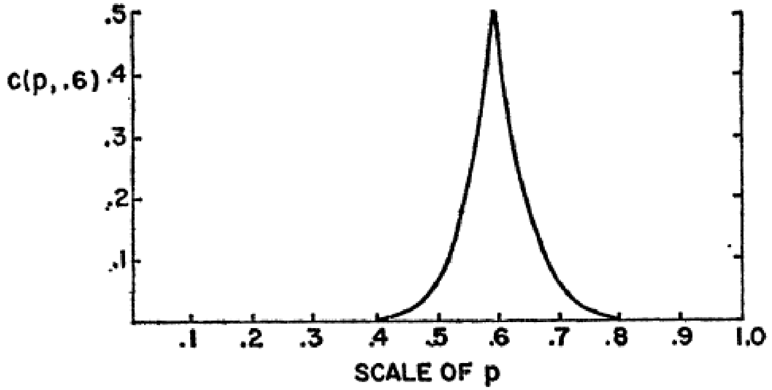
Jerzy Neyman 1930-1941

- Confidence intervals and regions
- Optimality of tests and confidence regions under monotoneous likelihood ratio
- Confidence intervals are obtained from fiducial distributions:

$$s_{obs}\sqrt{\frac{v}{K_v^{-1}\left(1-\frac{\alpha}{2}\right)}}=C^{-1}\left(\frac{\alpha}{2}\right)$$

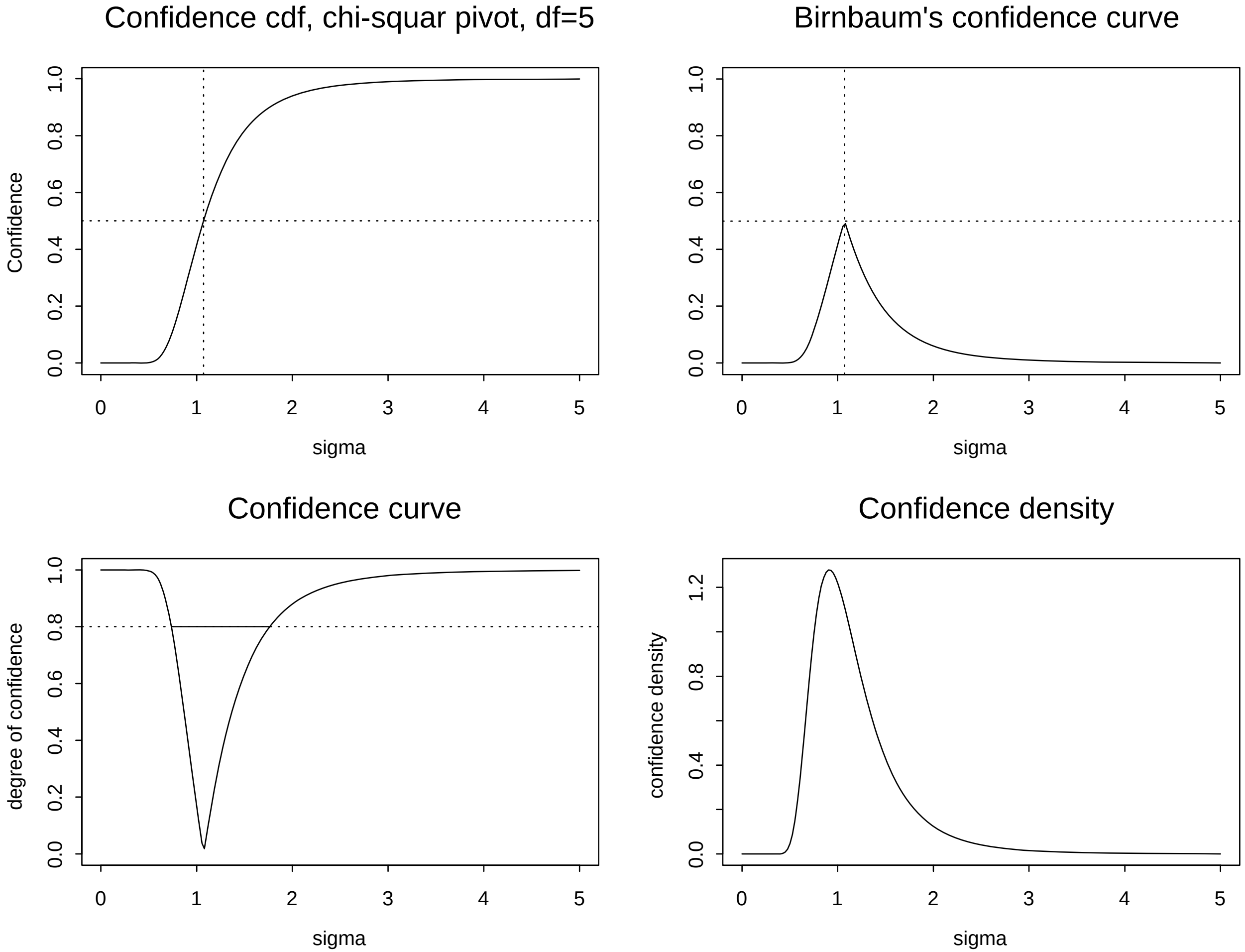
$$s_{obs}\sqrt{\frac{v}{K_v^{-1}\left(\frac{\alpha}{2}\right)}}=C^{-1}\left(1-\frac{\alpha}{2}\right)$$

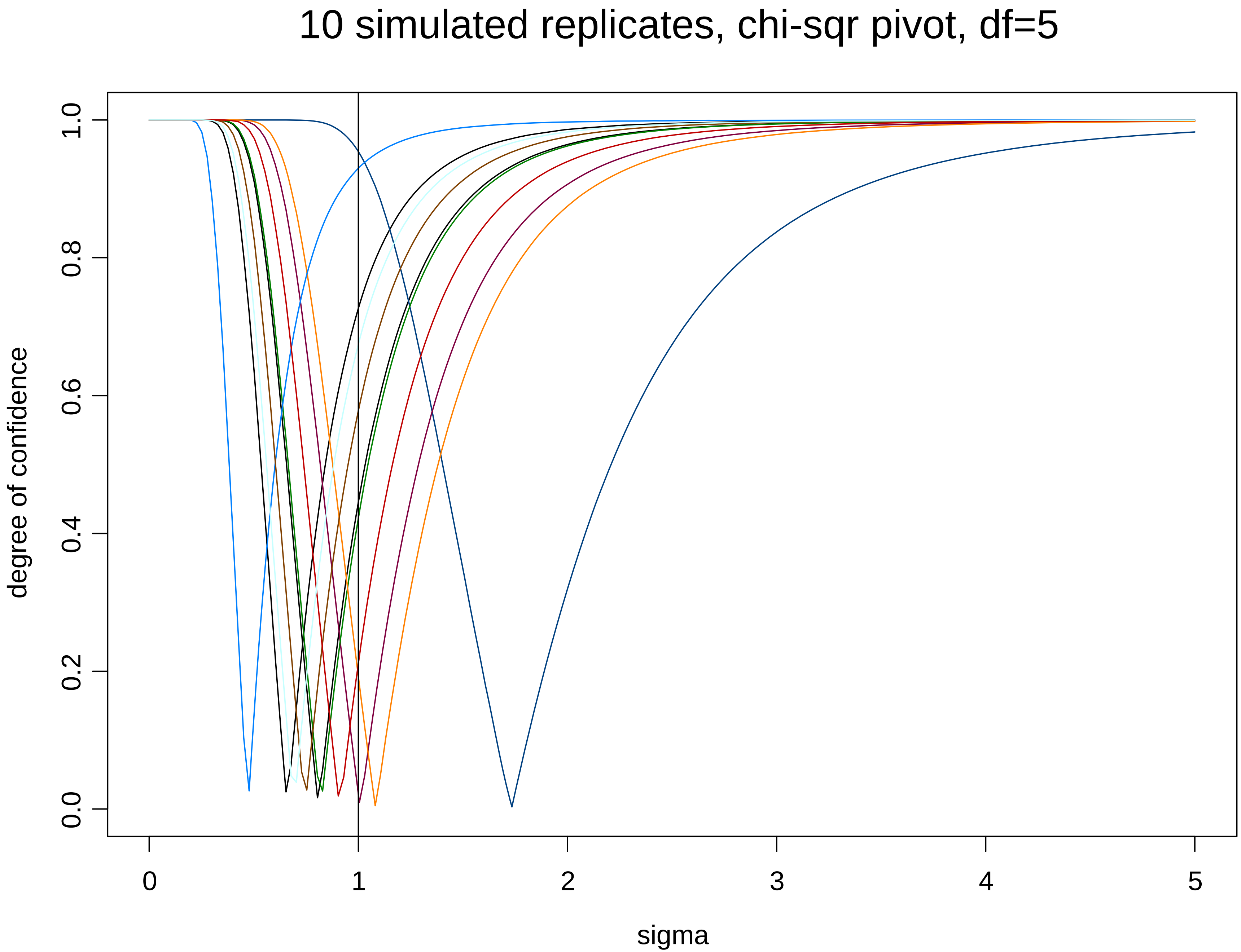
Alan Birnbaum 1961: Confidence curves: an omnibus technique for estimation and testing statistical hypotheses



”incorporating confidence intervals and limits at various limits”

For several parameters ”analogous methods …nested families of confidence regions”.





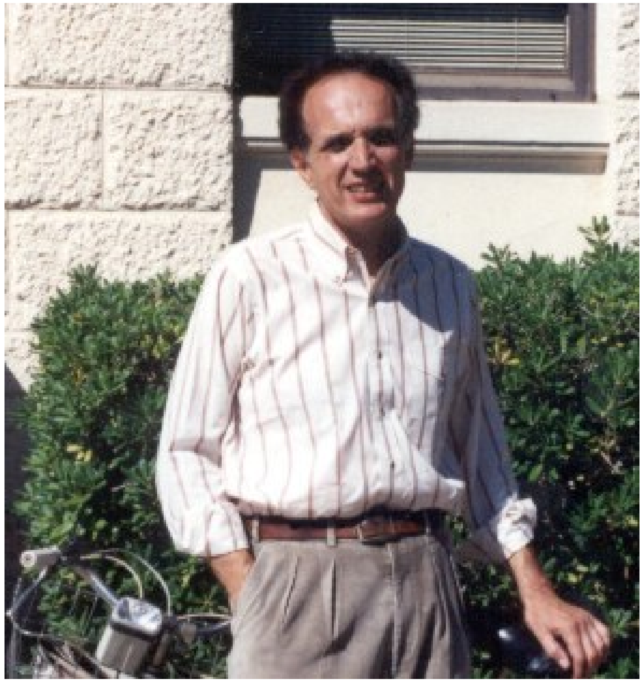
10 simulated replicates, chi-sqr pivot, df=5
degree of confidence
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
sigma

The Bayesian counterrevolution

- Distributional inference.
- Uncertainty presented as posterior distributions.
- Rational updating of information.
- Integrates judgemental and empirical information.
- Computational power and versatility: MCMC.

But

- The posterior distribution (and likelihood inference) might be biased.
- Prior distributions are needed, even when no information is available. Flat prior densities are still “informative”.
- The interpretation of the posterior distribution is unclear when the prior distribution is not a probability distribution.



B. Efron

1987: Better bootstrap confidence intervals

$\theta = h(\gamma)\ \ \hat{\theta} = h(\hat{\gamma})$

$$\frac{\hat{\gamma} - \gamma}{1 + a\gamma} : \ N(b,1)$$

Efron. 1998. R.A. Fisher in the 21st Century

Fiducial probability = “Fisher's biggest blunder”

“I believe that objective Bayes methods will develop for such problems, and that something like fiducial inference will play an important role in this development. Maybe Fisher's biggest blunder will become a big hit in the 21st century!”

“applied statistics seems to need an effective compromise between Bayesian and frequentist ideas, and right now there is no substitute in sight for the Fisherian synthesis.”

Fiducial distribution = *confidence distribution*

The fate of the fiducial argument – can Fisher and Neyman agree?

“Both Fisher and Neyman would probably have protested against the use of *confidence distributions*” (Efron 1998)

"Fisher was intuitively fully convinced of the importance of "fiducial inference", which he considered the jewel in the crown of the "ideas and nomenclature" for which he was responsible" (Zabel 1992)

"most statisticians, unable to separate the good from the bad in Fisher's arguments, considered the whole fiducial argument Fisher's biggest blunder, or his one great failure, and the whole area fell into disrepute” (Hampel 2002)

The fiducial argument builds a "bridge between aleatory probabilities (the only ones used by Neyman) and epistemic probabilities (the only ones used by Bayesians), by implicitly introducing, as a new type, frequentist epistemic probabilities." (Hampel 2002)

Hampel (2002): The fiducial argument builds a "bridge between aleatory probabilities (the only ones used by Neyman) and epistemic probabilities (the only ones used by Bayesians), by implicitly introducing, as a new type, frequentist epistemic probabilities.“

•*probability* is a good term for aleatory probability
•*confidence* for epistemic "probability“ – the currency of digested information
•*likelihood* – the currency of raw information in data, brings probability and confidence together

•Confidence distributions are *not sigma-additive*! It distributes confidence over intervals or nested families of regions
•Dimensions cannot automatically be reduced by integration.
•*The confidence curve* focus on confidence over regions

Neyman-Pearson lemma (Schweder and Hjort 2002)

A confidence distribution based on a sufficient statistic *S* with monotone likelihood ratio in a scalar parameter is uniformly most powerful:
For any value of the parameter, and
For any spread functional about it
The spread of the confidence distribution based on *S* is stochastically less than that based on another statistic *T*.

$$\Gamma(|t|) \text{Z} \qquad \Gamma(0) = 0$$

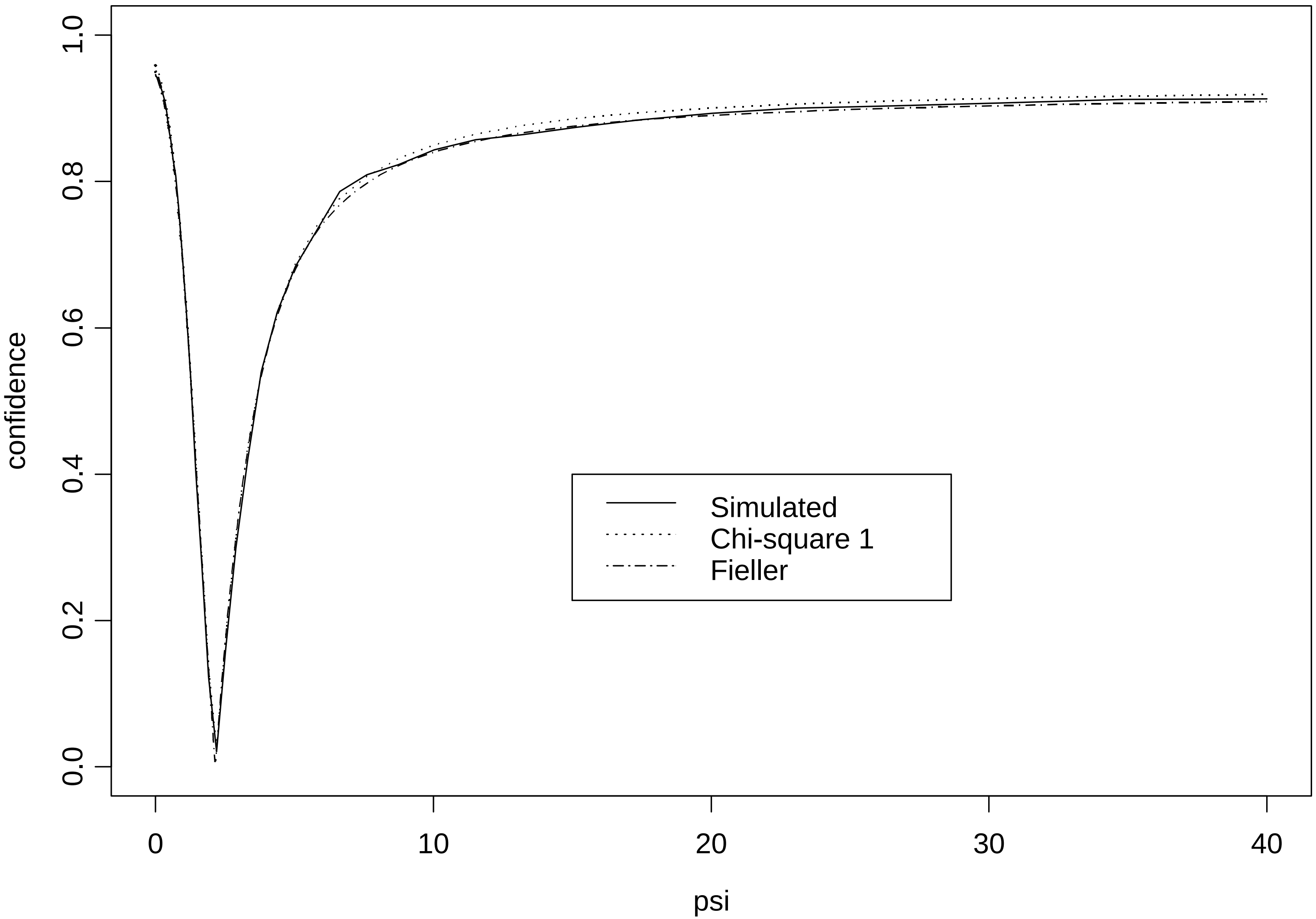
$$\gamma(C) = \int \Gamma(\theta - \theta_0) C(d\theta)$$

$$\gamma(C^S) \underset{ST}{\overset{\theta_0}{\leq}} \gamma(C^T)$$

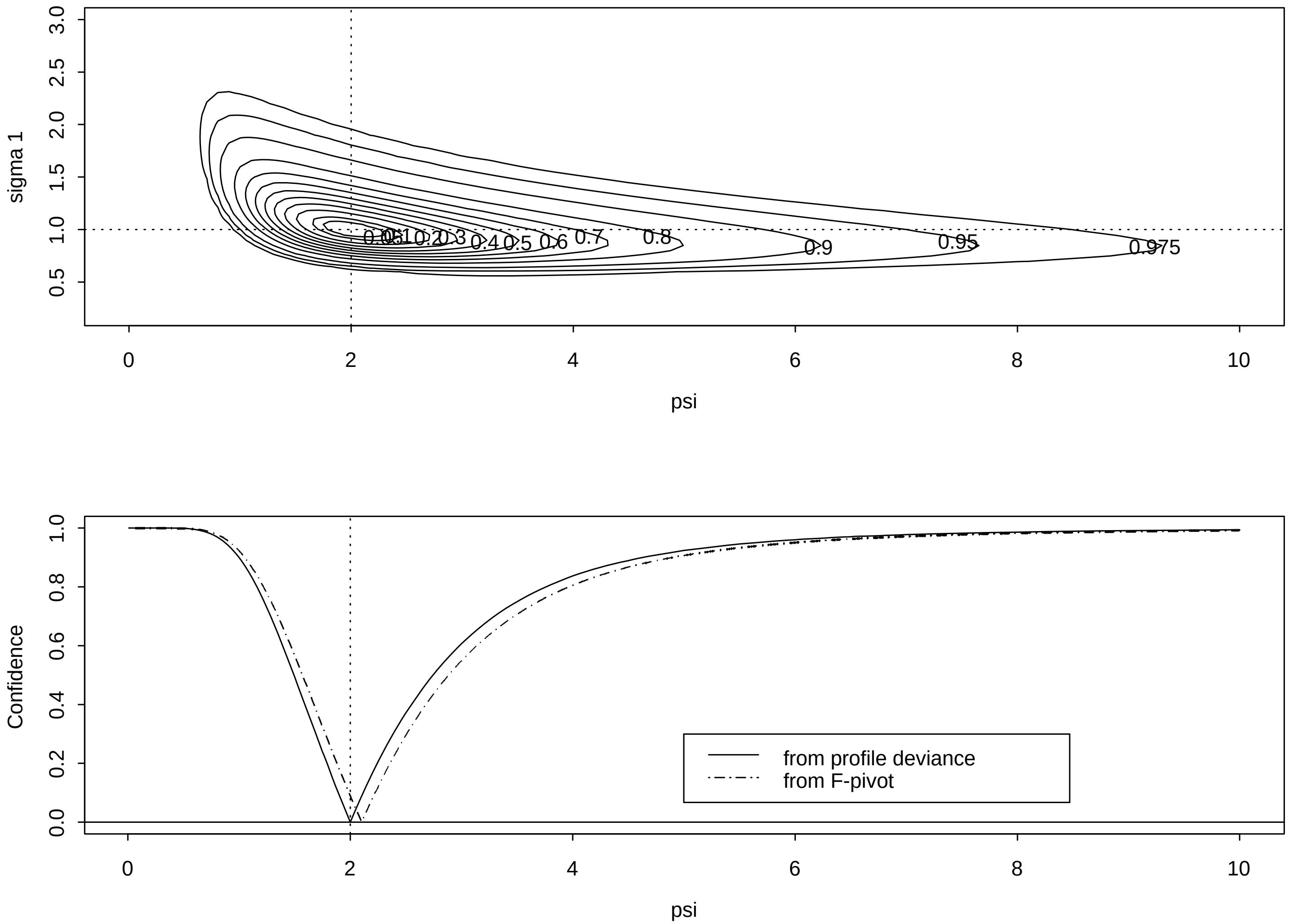
Confidence and likelihood

Deviance	$D(\theta; X) = -2\log\left(L(\theta; X) / L(\hat{\theta}; X)\right)$
Null distribution cdf	$D(\theta; X) : \; F_{\theta}$
Confidence from deviance	$N(\theta; X) = F_{\theta}\left(D(\theta; X)\right)$
Profile deviance	$D_p(\psi; X) = -2\log\left(\max_{\chi} L((\psi, \chi); X) / L(\hat{\theta}; X)\right)$
Null distribution	$D_p^{\chi}(\psi; X) : \; F_{\psi,\chi}$
Confidence from profile deviance or other pseudo likelihood	$N^{\chi}(\psi; X) = F_{\psi,\chi}\left(D_p^{\chi}(\psi; X)\right)$ $N(\psi; X) = F_{\psi,\hat{\chi}(\psi)}\left(D_p^{\hat{\chi}(\psi)}(\psi; X)\right)$

Example: Ratio of two regression parameters. Dotted line is chi-square calibration of the profile deviance. The Fieller method is exact. Only ca 87% finite support!



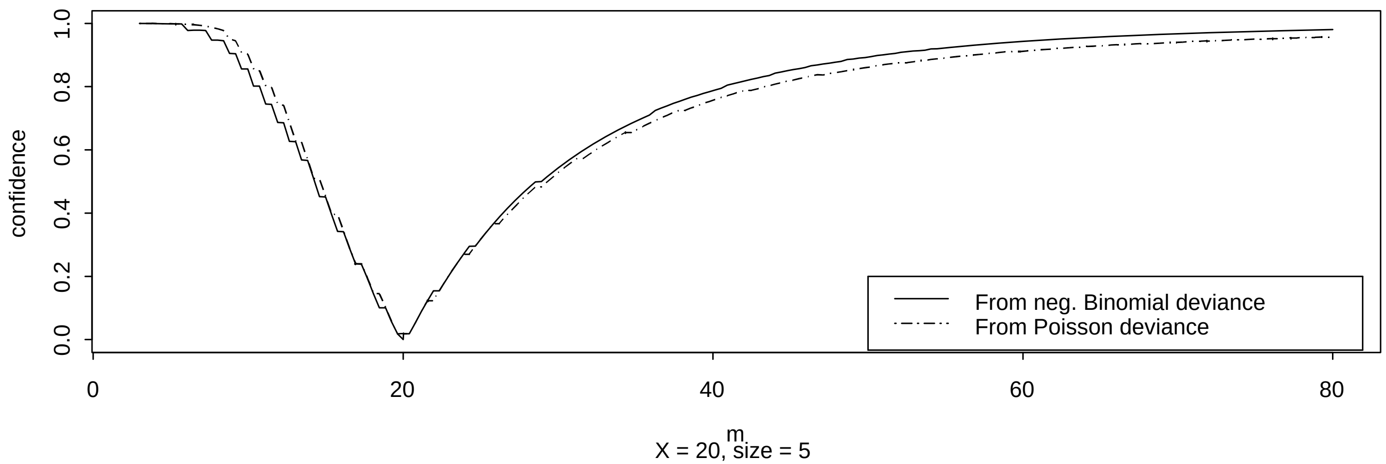
Example: Two normal samples, df's 9 and 4 $\psi = \frac{\sigma_1}{\sigma_2}$



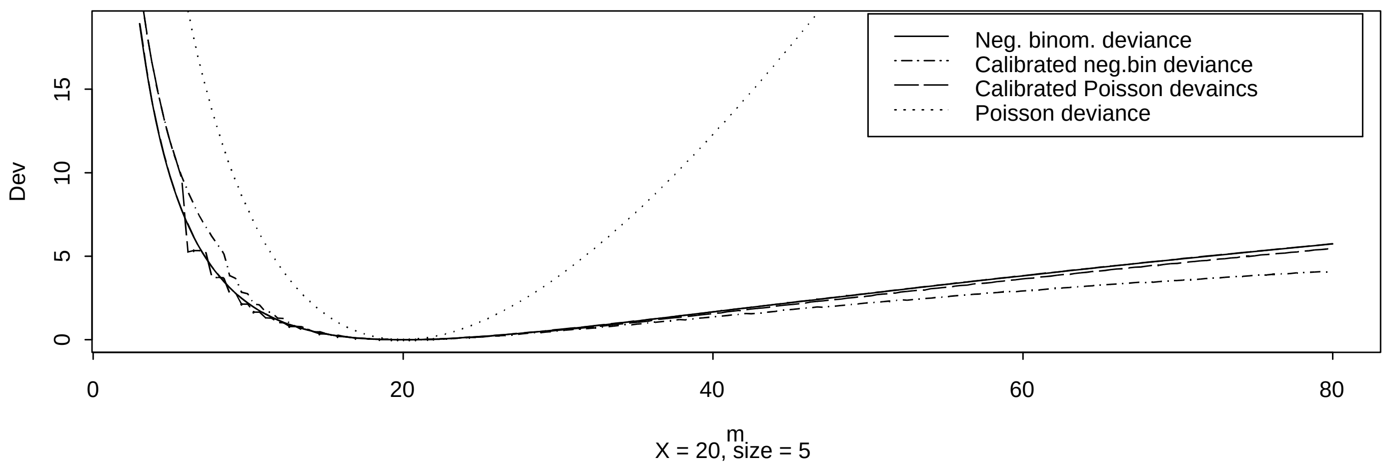
Pseudo deviance calibrated to confidence curve. Confidence curve calibrated to approximate deviance.

$$D_{pseudo}(\theta; X) : \; F_\theta$$
$$N(\theta; X) = F_\theta\left(D_{pseudo}(\theta; X)\right)$$
$$D_{appr}(\theta; X) = K_\nu^{-1}\left(N(\theta; X)\right)$$

Confidence curve, Negative binomial



Deviance functions, negative binomial model



Combining information

J independent sources.

Confidence curves $N_j(\theta)$

Deviance functions $D_j(\theta)$

Likelihood synthesis

K is the chi-square cdf with sum df

$$D(\theta) = \sum D_j(\theta)$$

$$N(\theta) \approx K_\nu\left(D(\theta)\right) \quad \nu = \sum \nu_j$$

Confidence synthesis

Singh, Xie and Strawderman (2005).

H double exponential cdf,

G the convolution cdf of *J* double exponentials.

Scalar parameter.

$$N(\theta) = G\left(\sum H^{-1}C_j(\theta)\right)$$

$$H^{-1}C_j(\theta) = \pm\log\left(1 - N_j(\theta)\right)$$

To bias-correct, or not to bias-correct?

A scalar parameter is median bias-corrected by *b*

$$P_{\theta}\left(b\left(\hat{\theta}\right) \leq \theta\right) = \frac{1}{2}$$

The bias-corrected deviance is calibrated to a confidence curve

$$D\left(b(\theta)\right) \sim_{\theta} G_{\theta}$$

$$N_{bc}(\theta) = G_{\theta}\left(D\left(b(\theta)\right)\right)$$

For a sample from the Efron-family, the calibrated bias-corrected deviance is to second order the tail-symmetric confidence curve from the pivot

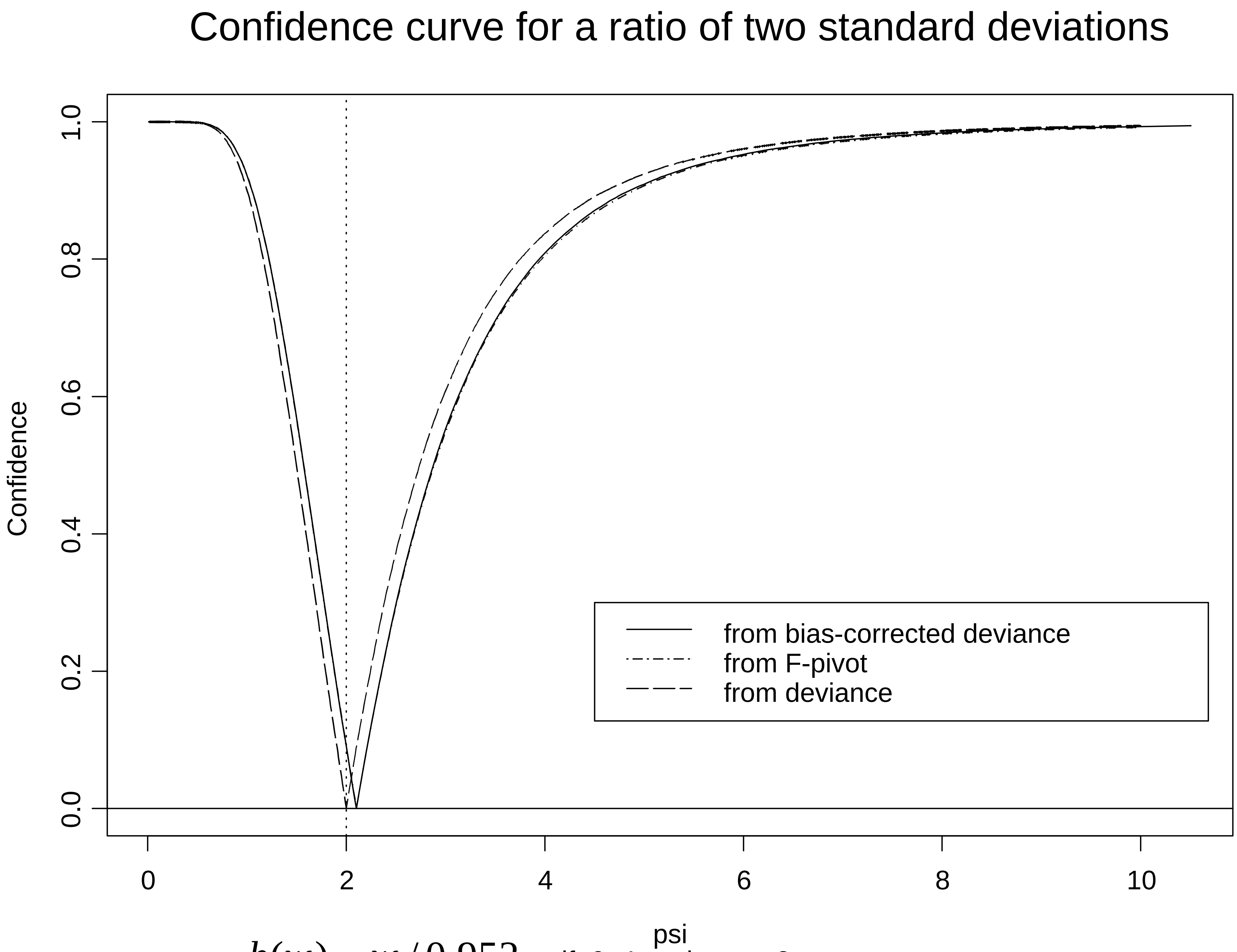
$$\theta = h(\gamma) \quad \hat{\theta} = h(\hat{\gamma})$$

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

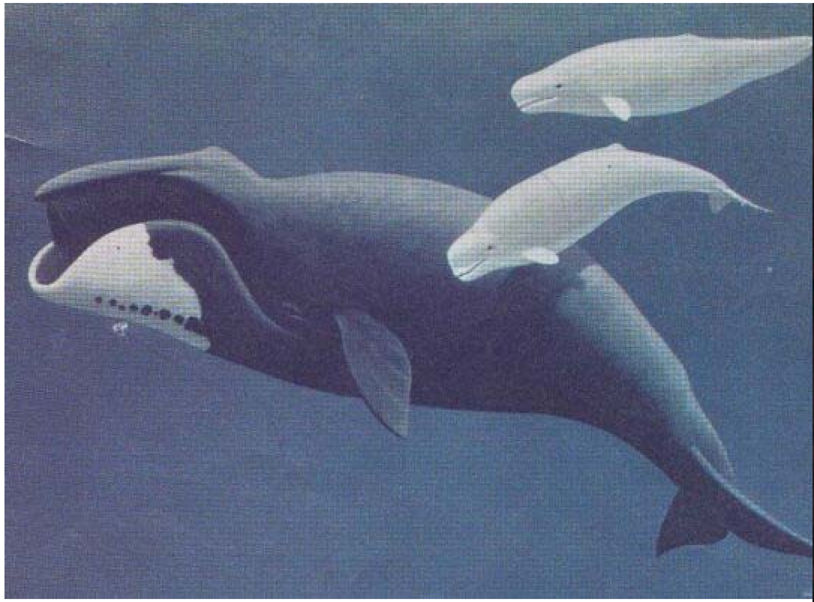
The confidence curve from the deviance has maximum power at infinitessimal levels of confidence

$$D(\theta) \sim_{\theta} F_{\theta}$$

$$N(\theta) = F_{\theta}\left(D(\theta)\right)$$



$b(\psi) = \psi / 0.952$ df=9, 4; estimate =2

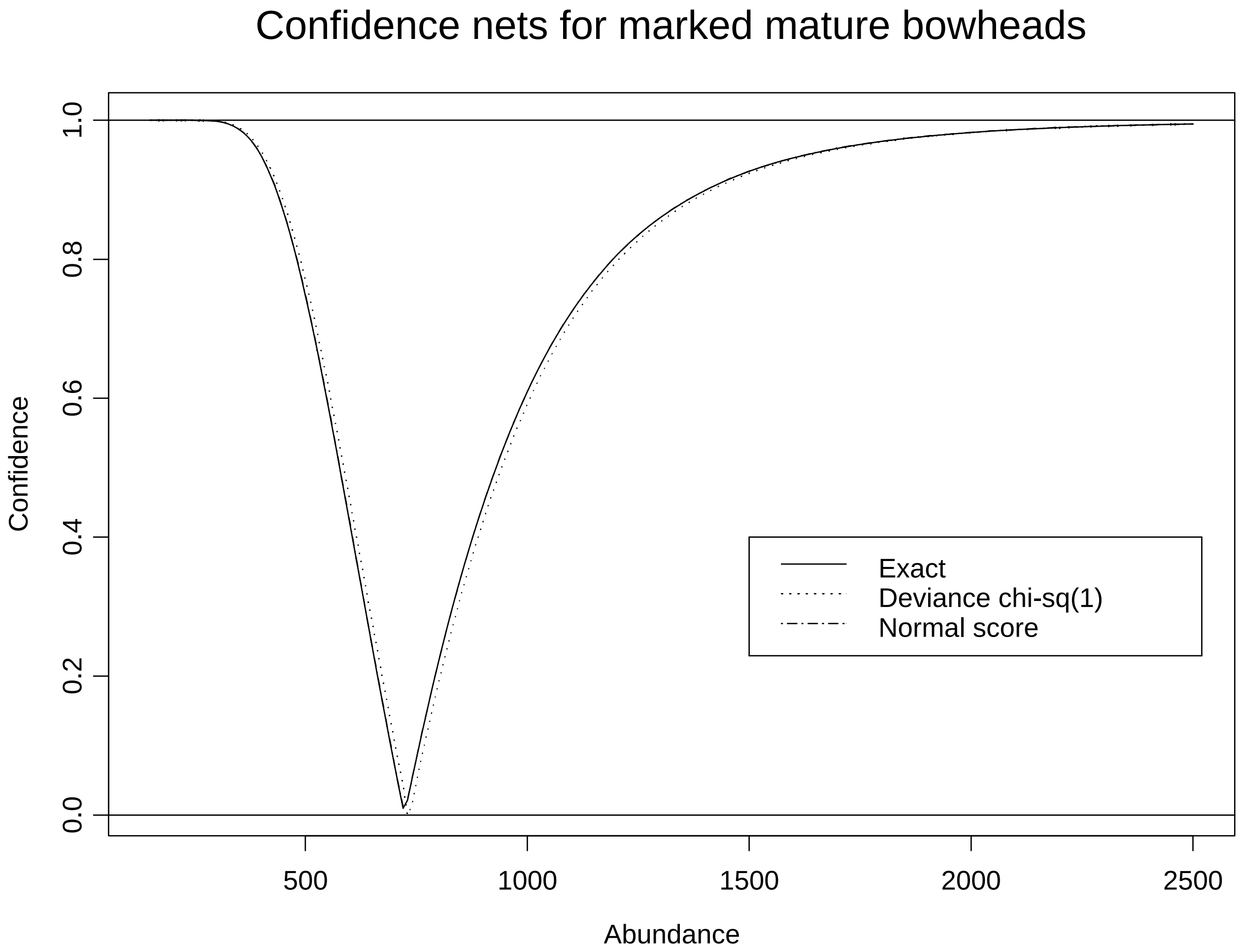


Application: Four photo surveys of bowhead whales off Barrow, Alaska.

Schweder (2003)

# of marked n, recaptured r, and unmarked bowheads u, and # of unique marked X. Spring and Fall 1985-86.					
Survey	S 85	F85	S85	F85	*X*
immatures	15, 0, 191	32, 6, 353	9, 0, 67	11, 0, 99	62
matures	44, 7, 128	20, 3, 66	49, 2, 75	7, 1, 39	113

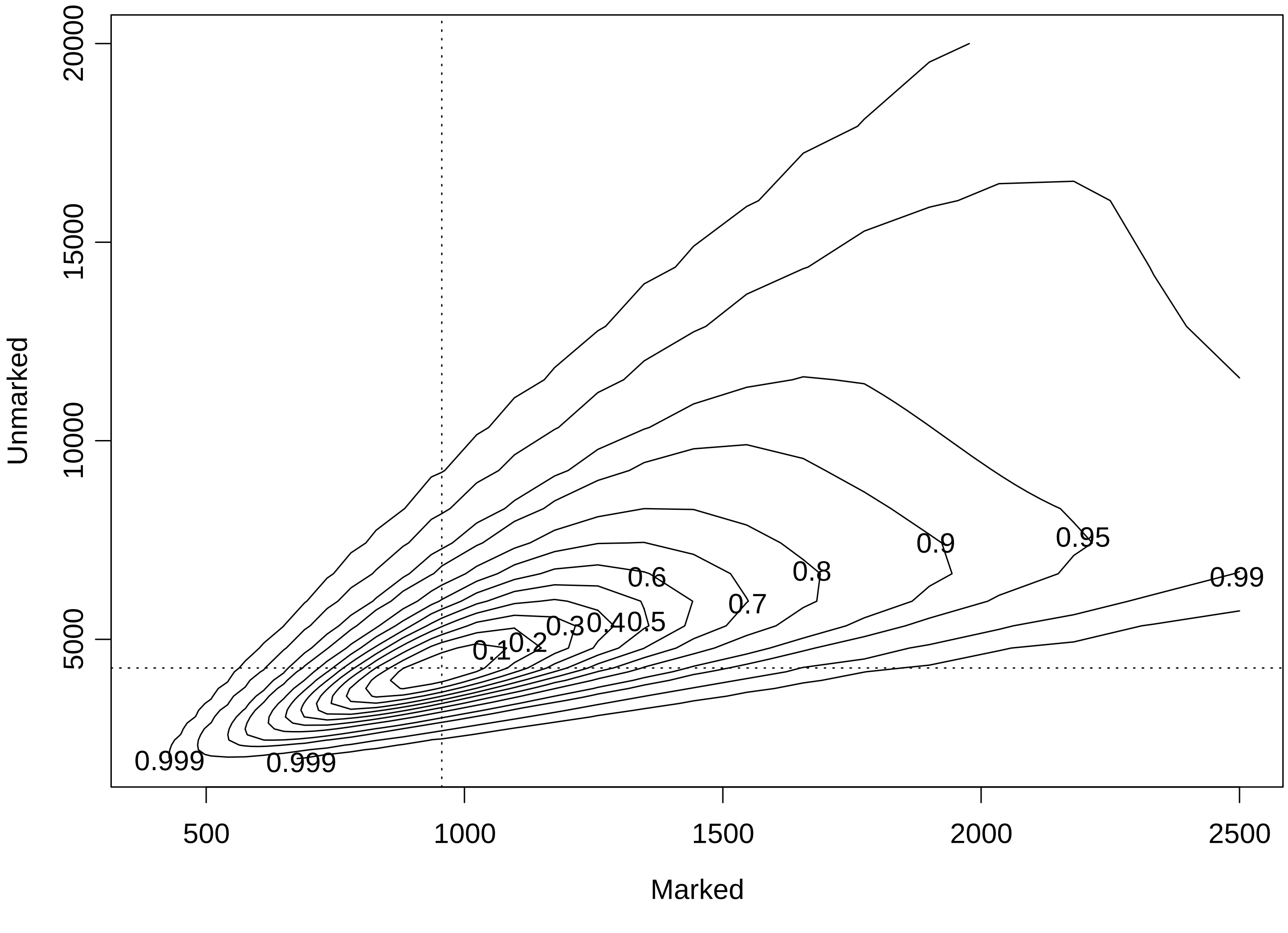
Darroch's conditionallikelihood for number of unique captures in multiple capture-recapture surveys. Marked mature bowheads only



$$D(\theta) = P\left(X > x_{obs} \mid \left\{n_j\right\}\right) + \frac{1}{2} P\left(X = x_{obs} \mid \left\{n_j\right\}\right)$$

$$N_{chi-sq}(\theta) = K_1\left(D(\theta)\right)$$

Confidence curve for number of marked and unmarked , 1986. Profiled deviance, simulated null distribution. Schweder (2003).



Box-shaped confidence curves for curve-parameters

families of simultaneous confidence bands Schweder (2007)

A curve-parameter is of high dimension *T*.

A box-shaped confidence region is a simultaneous confidence band.

Beran (1988) adjusts the degree of confidence for a point-wise confidence band to make it a simultaneous confidence band.

When the point-wise confidence curve is based on bootstrapping and Efron’s abc-method, the point-wise confidence curve is adjusted to a box-shaped confidence curve by the bootstrap distribution of the maximum point-wise curve.

$$\theta = \left(\theta_1, \ldots, \theta_T\right)$$

$$\theta_t^* \sim H_t$$

$$\max_t |1 - 2H_t\left(\theta_t^*\right)| \sim K$$

$$N_{box}\left(\theta\right) = \left\{K\left(N_{abc}^t\left(\theta_t\right)\right)\right\}$$

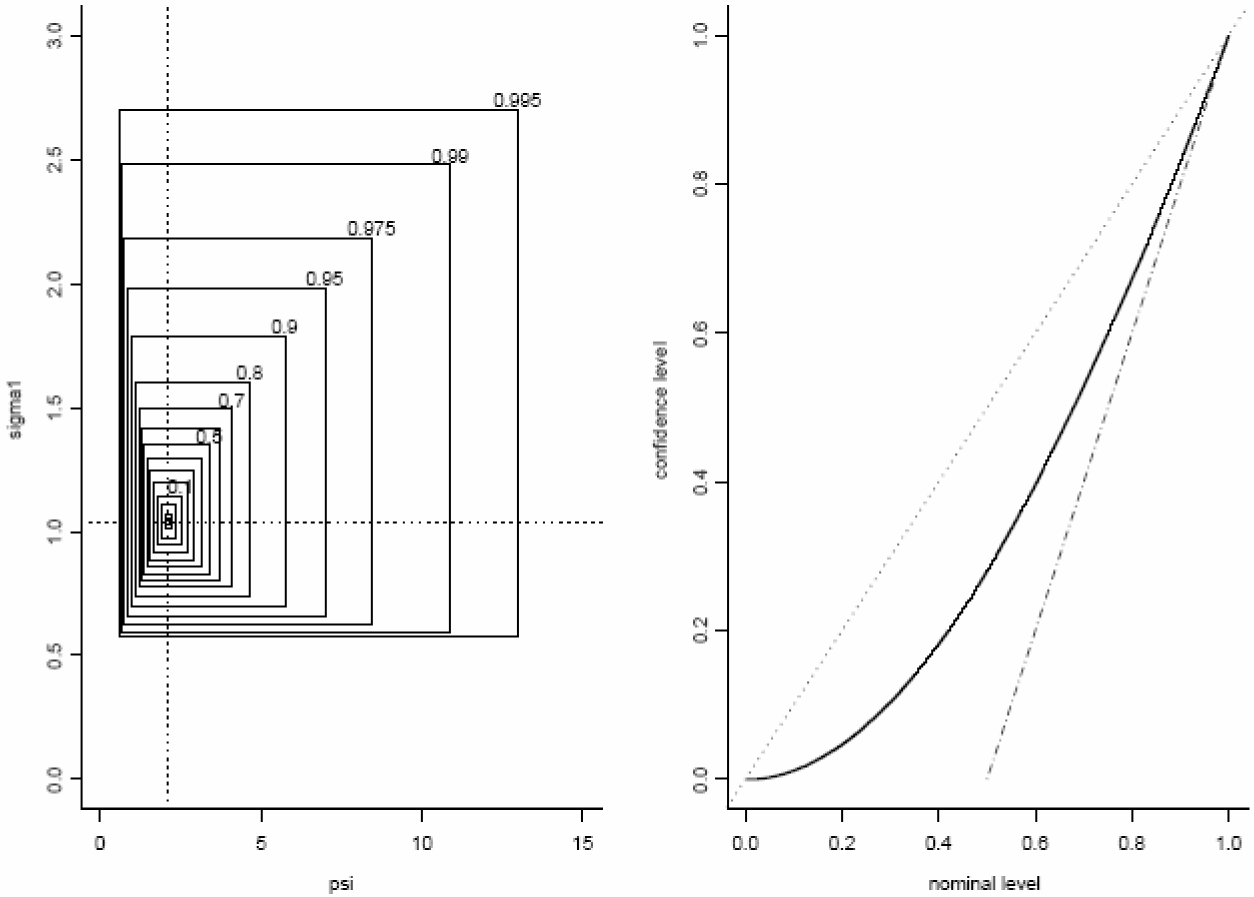


Fig. 4. Level sets of a product confidence net for $\psi = \sigma_2/\sigma_1$ and σ_1 (left panel), and K (solid line, right panel) with the diagonal and also the simple Bonferroni adjustment function (dashed line). $\nu_1 = 9$ and $\nu_2 = 4$. 100,000 simulations.

Application: Income and wealth survey, Norway 2002. 22496 Males. 95% Quantile regression of capital income on other income (wage), controlled for age. 1000 bootstrap replicates.

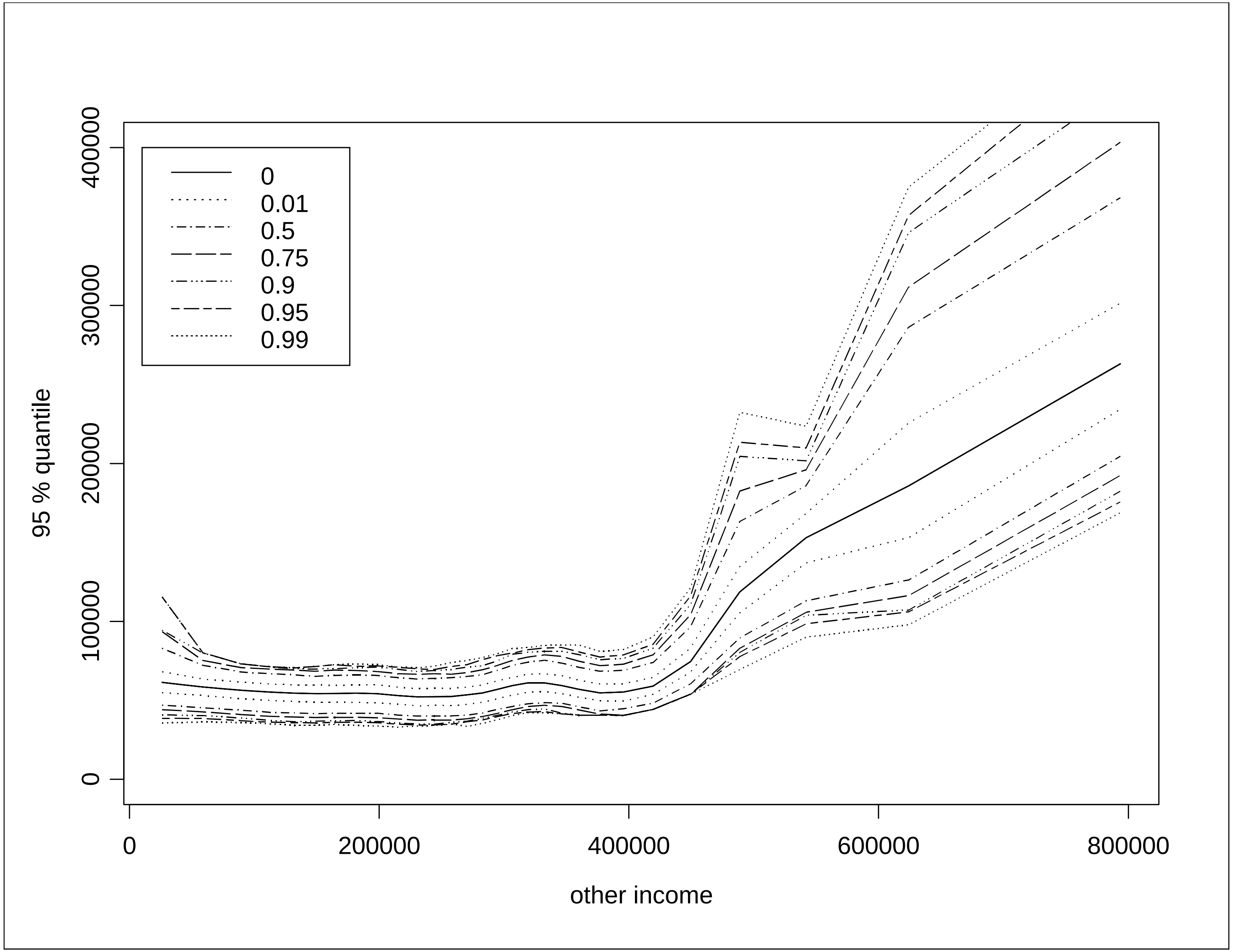


Table 1. Nominal pointwise confidence needed to obtain the abc net and the simple Bonferroni net at given simultaneous confidence. The default smoothing by gam in Splus is denoted by default df.

	.50	.75	.90	.95	.99
$abc : K^{-1}(\alpha)\ df = 10$	.908	.964	.988	.994	.998
$abc : K^{-1}(\alpha)$ default df	.816	.926	.976	.990	.998
Bonferroni : $1 - (1 - \alpha)/29$	.983	.991	.997	.998	1.00

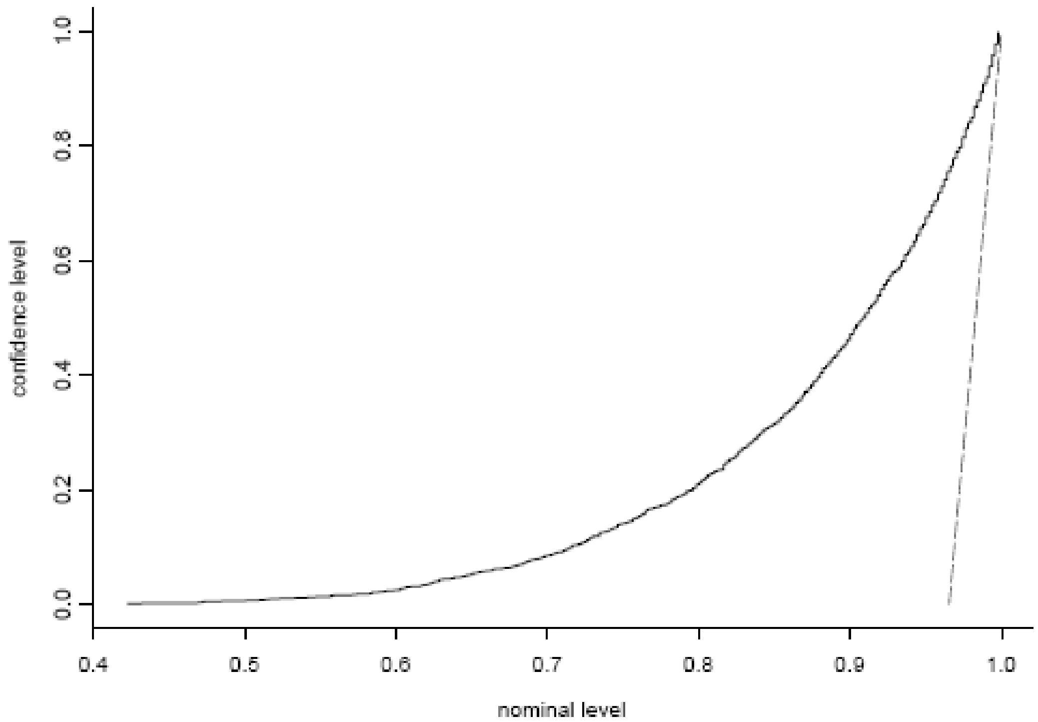


Fig. 7. Adjustment function K for the abc product net (solid curve) and the simple Bonferroni adjustment (dashed line).

Summary

Confidence distributions and confidence curves

- provides distributional inference on par with Bayesian posterior distributions, not based on priors
- allows coherent learning
- apply to prediction
- distributes confidence over regions, are not sigma-additive
- might not be proper (the Fieller problem), can be median bias-corrected
- reduce dimension by profiling rather than integration
- keeps aleatory and epistemic probability apart, but provides a bridge between the two: probability vs. confidence
- are transformation invariant
- provides optimal inference in simple models

References

- Beran, R. 1988. Balanced simultaneous confidence sets. *J. Amer. Statis. Assoc*. 83: 679-686.
- Birnbaum, A. 1961. Confidence curves: an omnibus technique for estimation and testing statistical hypotheses. *J. Amer. Statis. Assoc*. 56: 246-249.
- Efron, B. 1987. Better bootstrap confidence intervals, (with discussion). J. Amer. Statist. Assoc. 82, 171–200.
- Efron, B. 1998. R.A. Fisher in the 21st century (with discussion). Statistical Science, 13:95-122.
- Fisher, R.A. 1930. Inverse probability. *Proc. Cambridge Phil. Society*. 26: 528-535.
- Hampel, F. 2003. The proper fiducial argument. ftp://ftp.stat.math.ethz.ch/Research-Reports/114.pdf.
- Neyman, J. 1941. Fiducial argument and the theory of confidence intervals. *Biometrika* 32: 128-150.
- Schweder, T. 2003. Abundance estimation from photo-identification data: confidence distributions and reduced likelihood for bowhead whales off Alaska. *Biometrics* 59: 976-985.
- Schweder, T. 2007. Confidence nets for curves. In *Advances in Statistical Modeling and Inference Essays in Honor of Kjell A. Doksum* (ed V. Nair). World Scientific.
- Schweder, T. and Hjort, N.L. 2002. Confidence and likelihood. *Scandinavian Journal of Statistics*. 29: 309-332.
- Singh, K., M. Xie and W.E. Strawderman. 2005. Combining information through confidence distributions *Annals of Statistics* 33:159-183.
- Zabell S. L. 1992.. R. A. Fisher and the fiducial argument. *Statistical Science* 7:369--387.