Aspects of Likelihood Inference

Nancy Reid

October 16, 2013

JACOBI BERNOULLI,

ARS CONJECTANDI,

OPUS POSTHUMUM.

TRACTATUS

DE SERIEBUS INFINITIS,

DE LUDO PILÆ

RETICULARIS.

BASILEÆ,

Impensis THURNISIORUM, Fratrum.

Models and likelihood

- Model for the probability distribution of y given x
- Density $f(y \mid x)$ with respect to, e.g., Lebesgue measure
- Parameters for the density $f(y \mid x; \theta)$, $\qquad \theta = (\theta_1, \ldots, \theta_d)$

- Likelihood function $L(\theta; y^0) \propto f(y^0; \theta)$

- often $\theta = (\psi, \lambda)$
- θ could have very large dimension, $d > n$

  typically $y = (y_1, \ldots, y_n)$
- θ could have infinite dimension $E(y \mid x) = \theta(x)$ 'smooth', in principle

Why likelihood?

- makes probability modelling central
- emphasizes the inverse problem of reasoning from y^0 to θ or $f(\cdot)$
- suggested by Fisher as a measure of plausibility

Royall, 1994

$L(\hat{\theta})/L(\theta) \in (1, 3)$	very plausible;
$L(\hat{\theta})/L(\theta) \in (3, 10)$	implausible;
$L(\hat{\theta})/L(\theta) \in (10, \infty)$	very implausible

- converts a 'prior' probability $\pi(\theta)$ to a posterior $\pi(\theta \mid y)$ via Bayes' Theorem
- provides a conventional set of summary quantities for inference based on properties of the postulated model

Widely used

Cold Regions Science and Technology

Available online 4 October 2013

In Press, Accepted Manuscript — Note to users

A Generalized Probabilistic Model of Ice Load Peaks on Ship Hulls in Broken-Ice Fields

A. Suyuthi[a], B.J. Leira[a], K. Riska[b, c]

[a] Department of Marine Technology, NTNU, Trondheim, Norway

[b] Centre of Ships and Offshore Structures (CeSOS), Trondheim, Norway

[c] ILS OY, Helsinki, Finland

... widely used

ELSEVIER

Molecular Phylogenetics and Evolution

Available online 3 October 2013

In Press, Uncorrected Proof — Note to users

Diversification of *Scrophularia* (Scrophulariaceae) in the Western Mediterranean and Macaronesia – Phylogenetic relationships, reticulate evolution and biogeographic patterns

Agnes Scheunert, Günther Heubl

Systematic Botany and Mycology, Department Biology I, Ludwig-Maximilians-University, GeoBio Center LMU, Menzinger Strasse 67, 80638 Munich, Germany

... widely used

Empirical growth curve estimation considering multiple seasonal compensatory growths of body weights in Japanese Thoroughbred colts and fillies.

(PMID:24085406)

Abstract | Citations | BioEntities | Related Articles | External Links

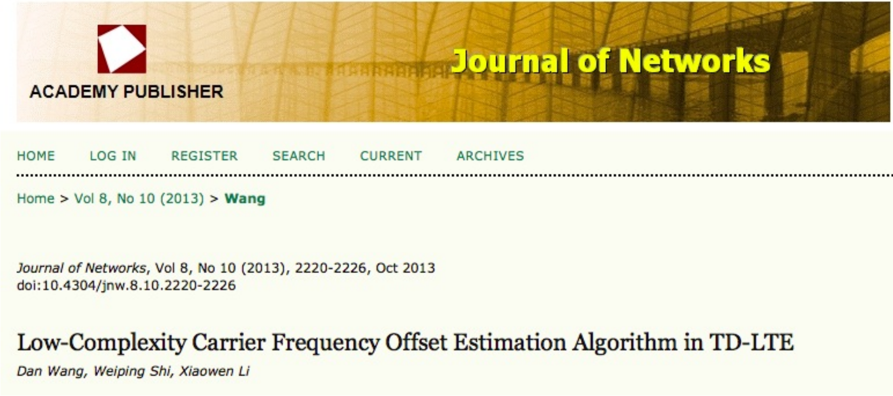
Onoda T, Yamamoto R, Sawmura K, Inoue Y, Murase H, Nambo Y, Tozaki T, Matsui A, Miyake T, Hirai N

Comparative Agricultural Sciences, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan

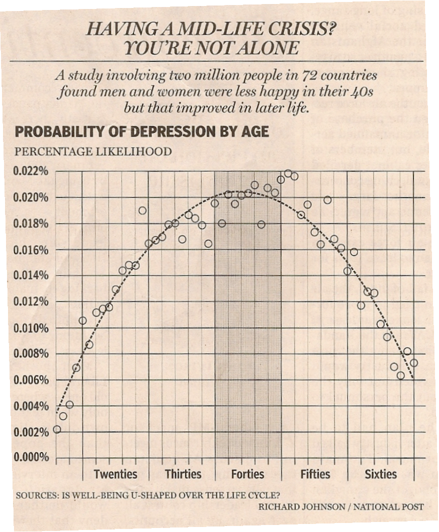
Journal of Animal Science [2013]

Type: Journal Article

... widely used


ACADEMY PUBLISHER
Journal of Networks
HOME LOG IN REGISTER SEARCH CURRENT ARCHIVES
Home > Vol 8, No 10 (2013) > Wang
Journal of Networks, Vol 8, No 10 (2013), 2220-2226, Oct 2013
doi:10.4304/jnw.8.10.2220-2226
Low-Complexity Carrier Frequency Offset Estimation Algorithm in TD-LTE
Dan Wang, Weiping Shi, Xiaowen Li

... widely used



National Post, Toronto, Jan 30 2008

... why likelihood?

- likelihood function depends on data only through sufficient statistics
- "likelihood map is sufficient" Fraser & Naderi, 2006

- gives exact inference in transformation models
- "likelihood function as pivotal" Hinkley, 1980

- provides summary statistics with known limiting distribution
- leading to approximate pivotal functions, based on normal distribution $\hat{\theta} \sim N(\theta, i^{-1}(\theta))$ $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0,\cdot)$
- likelihood function $+$ sample space derivative gives better approximate inference

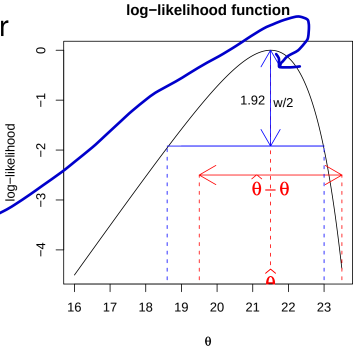
$$f(\hat{\theta} \mid a; \theta) = L(\theta) \Big/ \int L(\theta)\, d\theta$$

$f(a)$ free of θ

$$= L(\theta; \hat{\theta}, a) \Big/ \int L(\theta; \hat{\theta}, a)\, d\theta$$

Derived quantities

$\theta \in \mathbb{R}$



- maximum likelihood estimator
  $\hat{\theta} = \arg\sup_\theta \log L(\theta; y)$
  $= \arg\sup_\theta \ell(\theta; y)$

- observed Fisher information
  $j(\hat{\theta}) = -\partial^2 \ell(\theta)/\partial\theta^2 \big|_{\hat{\theta}}$

- efficient score function
  $\ell'(\theta) = \partial\ell(\theta; y)/\partial\theta$

  $\ell'(\hat{\theta}) = 0$ assuming enough regularity

- $\ell'(\theta; y) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f_{Y_i}(y_i; \theta)$, $\quad y_1, \ldots, y_n$ independent

Approximate pivots

scalar parameter of interest

$\theta = (\psi, \lambda)$

- profile log-likelihood $\ell_p(\psi) = \ell(\psi, \hat\lambda_\psi)$
- $\theta = (\psi, \lambda)$; $\hat\lambda_\psi$ constrained maximum likelihood estimator

$$\hat\theta - \theta \;\dot\sim\; N(0, j^{-1}(\hat\theta)) \Leftarrow (\hat\theta - \theta) j^{1/2}(\hat\theta) \;\dot\sim\; N(0,1)$$

$$\begin{aligned} r_e(\psi; y) = (\hat\psi - \psi) j_p^{1/2}(\hat\psi) \quad &\dot\sim \quad N(0,1) \\ r(\psi; y) = \pm\surd[2\{\ell_p(\hat\psi) - \ell_p(\psi)\}] \quad &\dot\sim \quad N(0,1) \\ \pi_m(\psi \mid y) \quad &\dot\sim \quad N\{\hat\psi, j_p^{-1/2}(\hat\psi)\} \end{aligned}$$

$j_p(\psi) = -\ell_p''(\psi)$; profile information

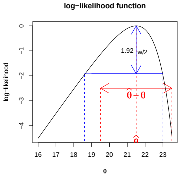
$$j = -\frac{\partial^2 \ell}{\partial\theta\,\partial\theta^T}$$

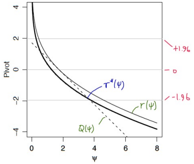
$Z \sim N(0,1)$

$Z^2 \sim \chi^2_1$

$w \sim \chi^2_1$

$\pm\sqrt{w} \sim N(0,1)$



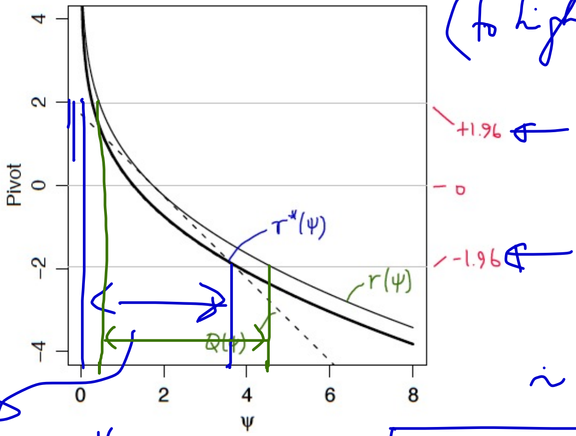


... approximate pivots

scalar parameter of interest

$$r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log \frac{q(\psi)}{r(\psi)} \sim N(0,1)$$

(to higher order)



$\sim N(0,1)$

$$q = (\hat{\psi} - \psi) j_p^{1/2}(\hat{\psi})$$

$\sim N(0,1)$

$$r = \pm \sqrt{2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\}}$$

... approximate pivots   scalar parameter of interest

- profile log-likelihood $\ell_{\mathrm{p}}(\psi) = \ell(\psi, \hat\lambda_\psi)$
- $\theta = (\psi, \lambda)$; $\hat\lambda_\psi$ constrained maximum likelihood estimator

$$r_e(\psi; y) = (\hat\psi - \psi) j_{\mathrm{p}}^{1/2}(\hat\psi) \ \dot\sim\ N(0,1)$$

$$r(\psi; y) = \pm\sqrt{[2\{\ell_{\mathrm{p}}(\hat\psi) - \ell_{\mathrm{p}}(\psi)\}]} \ \dot\sim\ N(0,1)$$

$$\pi_m(\psi \mid y) \ \dot\sim\ N\{\hat\psi, j_{\mathrm{p}}^{-1/2}(\hat\psi)\}$$

$$r^*(\psi; y) = r(\psi) + \frac{1}{r(\psi)} \log\left\{\frac{Q_F(\psi)}{r(\psi)}\right\} \ \dot\sim\ N(0,1)$$

$$r_B^*(\psi; y) = r(\psi) + \frac{1}{r(\psi)} \log\left\{\frac{Q_B(\psi)}{r(\psi)}\right\} \ \dot\sim\ N(0,1)$$

y_i iid $\frac{1}{\Gamma(\psi)} \left(\frac{\psi}{\lambda}\right)^{\psi} y_i^{\psi-1} e^{-\psi y_i/\lambda}$

$\theta = (\psi, \lambda)$ Gamma(shape ψ, mean λ)

$$\ell(\psi, \lambda) = \sum \log f(y_i; \psi, \lambda)$$

$$= -n \log \Gamma(\psi) + \dots + \psi \sum \log y_i - \frac{\psi}{\lambda} y_+$$

$$\frac{\partial \ell(\psi, \lambda)}{\partial \lambda} = 0$$

... approximate pivots scalar parameter of interest

- profile log-likelihood $\ell_{\rm p}(\psi) = \ell(\psi, \hat\lambda_\psi)$
- $\theta = (\psi, \lambda)$; $\hat\lambda_\psi$ constrained maximum likelihood estimator

$$r_e(\psi; y) = (\hat\psi - \psi) j_{\rm p}^{1/2}(\hat\psi) \quad \dot\sim \quad N(0,1) \left\{1 + O_p\left(\frac{1}{\sqrt{n}}\right)\right\}$$

$$r(\psi; y) = \pm\sqrt{[2\{\ell_{\rm p}(\hat\psi) - \ell_{\rm p}(\psi)\}]} \quad \dot\sim \quad N(0,1) \quad ''$$

$$\pi_{\rm m}(\psi \mid y) \quad \dot\sim \quad N\{\hat\psi, j_{\rm p}^{-1/2}(\hat\psi)\} \quad ''$$

$$r^*(\psi; y) = r(\psi) + \frac{1}{r(\psi)} \log\left\{\frac{Q_F(\psi)}{r(\psi)}\right\} \quad \dot\sim \quad N(0,1) \left\{1 + O_p\frac{1}{(\sqrt{n})^3}\right\}$$

$$r_B^*(\psi_0; y) = r(\psi_0) + \frac{1}{r(\psi_0)} \log\left\{\frac{Q_B(\psi_0)}{r(\psi_0)}\right\} \quad \dot\sim \quad N(0,1) \quad ''$$

derived from good approx to $\int_{\psi_0}^{\infty} \pi_m(\psi \mid y)\, d\psi$

The problem with profiling

- $\ell_{\mathsf{p}}(\psi) = \ell(\psi, \hat{\lambda}_\psi)$ used as a 'regular' likelihood, with the usual asymptotics
- neglects errors in the estimation of the nuisance parameter
- can be very large when there are many nuisance parameters

- example: normal theory linear regression $\hat{\sigma}^2 = RSS/n$ usual estimator $RSS/(n - k)$ k the number of regression coefficients
- badly biased if k large relative to n
- inconsistent for σ^2 if $k \to \infty$ with n fixed
- example fitting of smooth functions with large numbers of spline coefficients

Conditional and marginal likelihoods

$$f(y; \psi, \lambda) \propto f_1(s \mid t; \psi) f_2(t; \lambda)$$

- $L(\psi, \lambda) \propto L_c(\psi) L_m(\lambda)$, where L_1 and L_2 are genuine likelihoods, i.e. proportional to genuine density functions
- $L_{\text{p}}(\psi)$ is a conditional likelihood $L_c(\psi)$, and estimation of λ has no impact on asymptotic properties
- s is conditionally sufficient , t is marginally ancillary, for ψ

- hardly ever get so lucky
- but might expect something like this to hold approximately, which it does, and this is implemented in r_F^* formula
  automatically   Brazzale, Davison, R 2007