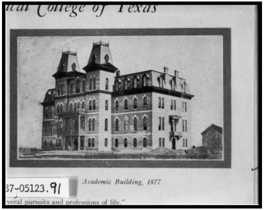
Approximate Likelihoods

Nancy Reid

November 17, 2014



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 $f(y \mid x; \theta)$, $\theta = (\theta_1, \ldots, \theta_d)$
- Data $y = (y_1, \ldots, y_n)$ often independent

- Likelihood function $L(\theta; y) \propto f(y; \theta)$ $(y_1, \ldots, y_n)$
- log-likelihood function $\ell(\theta; y) = \log L(\theta; y)$

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

- θ could have very large dimension, $d > n$

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

Why likelihood?

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

Royall, 1997

$$\begin{array}{ll} L(\hat{\theta})/L(\theta) \in (1,3) & \text{very plausible;} \\ L(\hat{\theta})/L(\theta) \in (3,10) & \text{implausible;} \\ L(\hat{\theta})/L(\theta) \in (10,\infty) & \text{very implausible} \end{array}$$

- converts a 'prior' probability $\pi(\theta)$ to a posterior $\pi(\theta \mid y)$ via Bayes' Theorem

- provides a conventional set of summary quantities: maximum likelihood estimator, score function, ...

... why likelihood?

- likelihood function depends on data
  only through sufficient statistics
- “likelihood map is sufficient” Fraser & Naderi, 2007

- 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
- likelihood function $+$ sample space derivative gives better
  approximate inference

- basis for comparison of models, using AIC or BIC

Complicated likelihoods

generalized linear mixed models

GLM: $y_{ij} \mid u_i \sim \exp\{y_{ij}\eta_{ij} - b(\eta_{ij}) + c(y_{ij})\}$

linear predictor: $\eta_{ij} = x_{ij}^{\mathrm{T}}\beta + z_{ij}^{\mathrm{T}}u_i \qquad j{=}1,\ldots n_i; \quad i{=}1,\ldots m$

random effects: $u_i \sim N_k(0, \Sigma)$

log-likelihood:

$$\begin{aligned}\ell(\beta,\Sigma) &= \sum_{i=1}^{m}\Big(y_i^{\mathrm{T}}X_i\beta - \frac{1}{2}\log|\Sigma| \\ &+ \log\int_{\mathbb{R}^k}\exp\{y_i^{\mathrm{T}}Z_iu_i - 1_i^{\mathrm{T}}b(X_i\beta + Z_iu_i) - \frac{1}{2}u_i^{\mathrm{T}}\Sigma^{-1}u_i\}du_i\Big)\end{aligned}$$

Ormerod & Wand 2012

... complicated likelihoods

Poisson $$f(y_t \mid \alpha_t; \theta) = \exp(y_t \log \mu_t - \mu_t)/y_t!$$

$$\log \mu_t = \beta + \alpha_t$$

autoregression

$$\alpha_t = \phi\alpha_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \quad |\phi| < 1, \quad \theta = (\beta, \phi, \sigma^2)$$

likelihood

$$L(\theta; y_1, \ldots, y_n) = \int \left(\prod_{t=1}^{n} f(y_t \mid \alpha_t; \theta) \right) f(\boldsymbol{\alpha}; \theta) d\boldsymbol{\alpha}$$

$L_{approx}(\theta; y)$ via Laplace with some refinements

Davis & Yau, 2011

... complicated likelihoods

$M/G/1$ queue: exponential arrival times, general service times, single server

observations y_i: times between departures from the queue

unobserved variables V_i: arrival time of customer i

model:

- $V_1 \sim \text{Exp}(\theta_3)$
- $V_i \mid V_{i-1} \sim V_{i-1} + \text{Exp}(\theta_3)$
- $Y_i \mid X_{i-1}, V_i \sim \text{Uniform}\{\theta_1 + max(0, V_i - X_{i-1}),$
  $\theta_2 + max(0, V_i - X_{i-1})\}$ $\quad X_i = \sum_{j=1}^{i} Y_j \quad G = U(\theta_1, \theta_2)$

Likelihood

$$L(\theta; y) = \int \cdots \int f(v_1 \mid \theta) \prod_{i=1}^{n} f(v_i \mid v_{i-1}, \theta) \prod_{i=1}^{n} f(y_i \mid v_i, x_{i-1}, \theta) dv_1 \cdots dv_n$$

Shestopaloff & Neal, 2013

Fearnhead & Prangle, 2012

... complicated likelihoods

multivariate extremes: example, wind speed at d locations

vector observations: $(X_{1i}, \ldots, X_{di}),\ i = 1, \ldots, n$

component-wise maxima: $Z_1, \ldots, Z_d; Z_j = \max(X_{j1}, \ldots, X_{jn})$

Z_j are transformed (centered and scaled)

joint distribution function:

$$\Pr(Z_1 \leq z_1, \ldots, Z_d \leq z_d) = \exp\{-V(z_1, \ldots, z_d)\}$$

$V(\cdot)$ can be parameterized via Gaussian process models

likelihood : need the joint derivatives of $V(\cdot)$

combinatorial explosion

Davison et al., 2012

... complicated likelihoods

Ising model:

$$f(y;\theta) = \exp\Big(\sum_{(j,k)\in E} \theta_{jk} y_j y_k\Big)\frac{1}{Z(\theta)}$$

observations: $y_i = \pm 1$; binary property of a node i in a graph with K nodes

parameter: θ_{jk} measures strength of interaction between nodes i and j

E is the set of edges between nodes

partition function: $Z(\theta) = \sum_y \exp(\sum_{(j,k)\in E} \theta_{jk} y_j y_k)$

Davison 2000 §6.2

Ravikumar et al. (2010)

Xue et al. (2012)

What’s a poor statistician to do?

- simplify the likelihood
  - composite likelihood
  - variational approximation
  - Laplace approximation to integrals
- change the mode of inference
  - quasi-likelihood
  - indirect inference
- simulate
  - approximate Bayesian computation
  - MCMC

Composite likelihood

- also called pseudo-likelihood
- reduce high-dimensional dependencies by ignoring them

- for example, replace $f(y_{i1}, \ldots, y_{ik}; \theta)$ by

pairwise marginal $\prod_{j<j'} f_2(y_{ij}, y_{ij'}; \theta)$, or

conditional $\prod_{j} f_c(y_{ij} \mid y_{\mathcal{N}(ij)}; \theta)$

- Composite likelihood function

$$CL(\theta; y) \propto \prod_{i=1}^{n} \prod_{j<j'} f_2(y_{ij}, y_{ij'}; \theta)$$

- Composite ML estimates are consistent, asymptotically normal, not fully efficient Besag, 1975; Lindsay, 1988

Example: AR Poisson

Davis & Yau, 2011

- Likelihood

$$L(\theta; y_1, \ldots, y_n) = \int \left(\prod_{t=1}^{n} f(y_t \mid \alpha_t; \theta) \right) f(\boldsymbol{\alpha}; \theta) d\boldsymbol{\alpha}$$

- Composite likelihood

$$CL(\theta; y_1, \ldots, y_n) = \prod_{t=1}^{n-1} \int \int f(y_t \mid \alpha_t; \theta) f(y_{t+1} \mid \alpha_{t+1}; \theta) f(\alpha_t, \alpha_{t+1}; \theta) d\alpha_t d\alpha_{t+1}$$

- consecutive pairs
- Time-series asymptotic regime one vector y of increasing length
- Composite ML estimator still consistent, asymptotically normal, estimable asymptotic variance
- Efficient, relative to a Laplace-type approximation
- Surprises: AR(1), fully efficient; MA(1), poor; ARFIMA(0,d,0), ok

Example: spatial extremes

Davison et al., 2012; & Huser, 2015

$$\Pr(Z_1 \le z_1, \ldots, Z_d \le z_d) = \exp\{-V(z_1, \ldots, z_d; \theta)\}$$

- pairwise composite likelihood used to compare the fits of several competing models
- model choice using "CLIC", an analogue of AIC

  $-2\log(\widehat{CL}) + \mathrm{tr}(J^{-1}K)$

- Davison et al. 2012 applied this to annual maximum rainfall at several stations near Zurich
- "fitting max-stable processes to spatial or spatio-temporal block maxima is awkward ... the use of composite likelihoods ... has become widely used" Davison & Huser

... Davison et al, 2012

MODELING OF SPATIAL EXTREMES 175

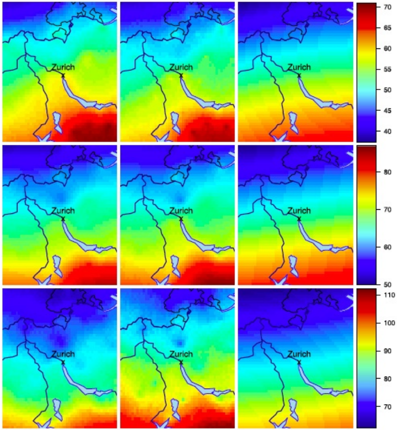
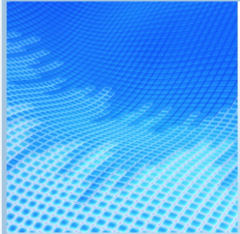


FIG. 3. *Maps of the (predictive) pointwise 25-year return level estimates for rainfall (mm) obtained from the latent variable and max-stable models. The top and bottom rows show the lower and upper bounds of the 95% pointwise credible/confidence intervals. The middle row shows the predictive pointwise posterior mean and pointwise estimates. The left column corresponds to the latent variable model assuming* Gamma(5, 3) *prior on* λ. *The middle column assumes the less informative priors* $\lambda_\eta \sim \mathrm{Gamma}(1, 100)$, $\lambda_\tau \sim \mathrm{Gamma}(1, 10)$ *and* $\lambda_\xi \sim \mathrm{Gamma}(1, 10)$. *The right column corresponds to the extremal t copula model.*

ANNUAL REVIEW OF STATISTICS AND ITS APPLICATION



ABOUT THIS JOURNAL

The *Annual Review of Statistics and Its Application* aims to inform statisticians, and users of statistics more broadly about major methodological advances and the computational tools that allow for their implementation. Statistical methods and ideas are relevant for virtually every intellectual field that involves data and quantitative reasoning, including these related articles from a variety of Annual Reviews journals.

ADD TO FAVORITES

FEATURED REVIEWS

Reviews include developments in the field of statistics, including theoretical statistical underpinnings of new methodology, as well as the developments in specific application domains such as biostatistics and bioinformatics, economics, machine learning, psychology, sociology, and aspects of the physical sciences.

For a list of related articles from other journals, click here.

NEWS

HIDE

Complimentary online access to volume 1 of the *Annual Review of Statistics and Its Application* will be available until January 2015.

JOURNAL CONTENT

JOURNAL HOME
FORTHCOMING
LATEST VOLUME
VOLUME SELECTOR
MOST CITED REVIEWS

MOST RECENT - The most recently available reviews for this journal

select / unselect all reviews

For reviews Select Option

View the full table of contents of the most recently published volume.

Statistical Causality from a Decision-Theoretic Perspective
A. Philip Dawid
Vol. 2
PDF (1526K)

ABSTRACT THUMBNAILS RELATED CONTENT SUPPLEMENTAL MATERIALS

Example: Ising model

Ising model:

$$f(y;\theta) = \exp\Big(\sum_{(j,k)\in E} \theta_{jk} y_j y_k\Big) \frac{1}{Z(\theta)}$$

neighbourhood contributions

$$f(y_j \mid y_{(-j)};\theta) = \frac{\exp(2y_j \sum_{k\neq j} \theta_{jk} y_k)}{\exp(2y_j \sum_{k\neq j} \theta_{jk} y_k) + 1}$$

penalized CL estimation based on sample $y^{(1)}, \dots, y^{(n)}$

$$\max_{\theta} \left\{ \sum_{i=1}^{n} \ell_j(\theta; y^{(i)}) - \sum_{j} \sum_{k} P_{\lambda}(|\theta_{jk}|) \right\}$$

Xue et al., 2012

Ravikumar et al., 2010

Variational methods

Ormerod & Wand, 2010

- in a Bayesian context, want $f(\beta \mid y)$
  use an approximation $q(\beta)$
- dependence of q on y suppressed

- choose $q(\beta)$ to be
  - simple to calculate
  - close to posterior

- simple to calculate
  - $q(\beta) = \prod q_j(\beta_j)$
  - simple parametric family

- close to posterior: miminize Kullback-Leibler divergence

$$KL(q \mid\mid f_{post}) = \int q(\beta) \log\{q(\beta)/f(\beta \mid y)\} d\beta$$

... variational methods

Titterington, 2006

- close to posterior:

$$\min_q \int q(\beta) \log\{q(\beta)/f(\beta \mid y)\} d\beta = \min_q KL(q \mid\mid f_{post})$$

- equivalent to

best LB for marginal $f(y)$

$$\max_q \int q(\beta) \log\{f(y, \beta)/q(\beta)\} d\beta$$

- in a likelihood context $\log f(y; \theta) = \log \int f(y \mid \beta; \theta) f(\beta) d\beta$

$$= \int q(\beta) \log\{f(y, \beta; \theta)/q(\beta)\} d\beta + KL(q \mid\mid f_{post})$$

-

$$\log f(y; \theta) \geq \int q(\beta) \log\{f(y, \beta; \theta)/q(\beta)\} d\beta$$

here β represent random effects u, or b, or ...

Example: GLMM

Ormerod & Wand, 2012

log-likelihood:

$$\begin{aligned}
\ell(\beta,\Sigma) &= \sum_{i=1}^{m}\Bigg(y_i^{\mathrm{T}} X_i\beta - \frac{1}{2}\log|\Sigma| \\
&+ \log\int_{\mathbb{R}^k} \exp\{y_i^{\mathrm{T}} Z_i u_i - 1_i^{\mathrm{T}} b(X_i\beta + Z_i u_i) - \frac{1}{2} u_i^{\mathrm{T}}\Sigma^{-1}u_i\} du_i \Bigg) \\
&= \sum_{i=1}^{m}\Bigg(y_i^{\mathrm{T}} X_i\beta - \frac{1}{2}\log|\Sigma| \\
&+ \log\int_{\mathbb{R}^k} \exp\{y_i^{\mathrm{T}} Z_i u_i - 1_i^{\mathrm{T}} b(X_i\beta + Z_i u_i) - \frac{1}{2} u_i^{\mathrm{T}}\Sigma^{-1}u_i\} \frac{\phi_{\Lambda_i}(u-\mu_i)}{\phi_{\Lambda_i}(u-\mu_i)} du_i \Bigg)
\end{aligned}$$

variational approx:

$$\begin{aligned}
\ell(\beta,\Sigma) &\geq \sum_{i=1}^{m}\Bigg(y_i^{\mathrm{T}} X_i\beta - \frac{1}{2}\log|\Sigma| \Bigg) \\
&+ \sum_{i=1}^{m} E_{u\sim N(\mu_i,\Lambda_i)} \Bigg(y_i^{\mathrm{T}} Z_i u - 1_i^{\mathrm{T}} b(X_i\beta + Z_i u) - \frac{1}{2} u^{\mathrm{T}}\Sigma^{-1}u - \log\{\phi_{\Lambda_i}(u-\mu_i)\} \Bigg) \\
&\equiv \ell(\beta,\Sigma,\mu,\Lambda) \qquad \text{simplifies to } k \text{ one-dim. integrals}
\end{aligned}$$

... variational approximations

Ormerod & Wand, 2012

- $$\ell(\beta, \Sigma) \geq \ell(\beta, \Sigma, \mu, \Lambda)$$
- variational estimate:
$$\ell(\tilde{\beta}, \tilde{\Sigma}, \tilde{\mu}, \tilde{\Lambda}) = \arg\max_{\beta,\Sigma,\mu,\Lambda} \ell(\tilde{\beta}, \tilde{\Sigma}, \tilde{\mu}, \tilde{\Lambda})$$
- inference for $\tilde{\beta}, \tilde{\Sigma}$? consistency? asymptotic normality?
  Hall, Ormerod, Wand, 2011; Hall et al. 2011
- emphasis on algorithms and model selection
  e.g. Tan & Nott, 2013, 2014
- VL: approx $L(\theta; y)$ by a simpler function of θ, e.g. $\prod q_j(\theta)$
- CL: approx $f(y; \theta)$ by a simpler function of y, e.g. $\prod f(y_j; \theta)$

Some Links between Variational Approximation and Composite Likelihoods?

S. Robin

UMR 518 AgroParisTech / INRA Applied Math & Comput. Sc.

MSTGA, Paris, November 22-23, 2012

http://carlit.toulouse.inra.fr/AIGM/pub/Reunion_nov2012/MSTGA-1211-Robin.pdf

Laplace approximation

$$\ell(\theta; y) = \log \int f(y \mid b; \theta) g(b) db = \log \int \exp\{Q(b, y, \theta)\} db, \text{say}$$

$$\ell_{Lap}(\theta; y) = Q(\tilde{b}, y, \theta) - \frac{1}{2} \log |Q''(\tilde{b}, y, \theta)| + c$$

using Taylor series expansion of $Q(\cdot, y, \theta)$ about $\tilde{b}$

simplification of the Laplace approximation leads to PQL:

$$\ell_{PQL}(\theta, b; y) = \log f(y \mid b; \theta) - \frac{1}{2} b^{\mathrm{T}} \Sigma^{-1} b$$

Breslow & Clayton, 1993

to be jointly maximized over b and θ and parameters in Σ

PQL can be viewed as linearizing $E(y)$ and then using results for linear mixed models Molenberghs & Verbeke, 2006

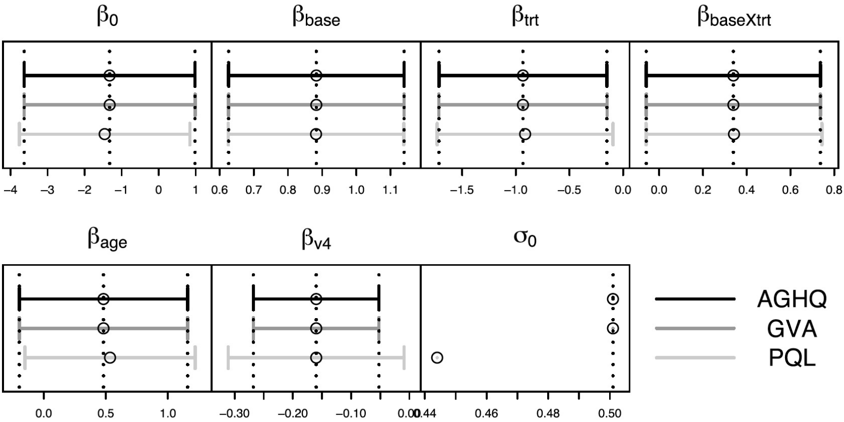


Figure 2. Point estimates and approximate 95% confidence intervals based on each of AGHQ, GVA, and PQL for the *Epilepsy* data random intercept model. The vertical dotted lines correspond to the AGHQ values.

implemented in `lme4` as `glmer`, in `MASS` as `glmmPQL`

Ormerod & Wand, 2012

Extensions of Laplace approximations

- expansions valid with $p = o(n^{1/3})$ Shun & McCullagh, 1995
- expansions for mixed linear models to higher order
  Raudenbush et al., 2000
- use REML for variance parameters Nelder & Lee, 1996
- integrated nested Laplace approximation Rue et al., 2009
  - model $f(y_i \mid \theta_i)$; prior $\pi(\theta \mid \vartheta)$ parameters and hyper-par
  - posterior $\pi(\theta, \vartheta \mid y) \propto \pi(\theta \mid \vartheta)\pi(\vartheta) \prod f(y_i \mid \theta_i)$
  - marginal posterior

$$\pi(\theta_i \mid y) = \int \underbrace{\pi(\theta_i \mid \vartheta, y)}_{\text{Laplace}} \underbrace{\pi(\vartheta \mid y)}_{\text{Laplace}} d\vartheta$$

Quasi-likelihood

- simplify the model
- $$E(y_i; \theta) = \mu_i(\theta); \qquad \text{Var}(y_i; \theta) = \phi\nu_i(\theta)$$
- consistent with generalized linear models
- example: over-dispersed Poisson responses
- PQL uses this construction, but with random effects

  Molenberghs & Verbeke, Ch. 14
- why does it work?
- score equations are the same as for a 'real' likelihood

  hence unbiased
- derivative of score function equal to variance function

  special to GLMs

Indirect inference

- composite likelihood estimators are consistent

under conditions ...

- because $\log CL(\theta; y) = \sum_{i=1}^{n} \sum_{j<j'} \log f(y_j, y_{j'}; \theta)$
- derivative w.r.t. θ has expected value 0

- what happens if an estimating equation $g(y; \theta)$ is biased?
- $g(y_1, \dots, y_n; \tilde{\theta}_n) = 0; \qquad \tilde{\theta}_n \to \theta^*$ $\qquad E g(Y; \theta^*) = 0$

- $\theta^* = \tilde{k}(\theta)$; invertible? $\theta = k(\theta^*)$ $\qquad \tilde{k}^{-1} \equiv k$

- new estimator $\hat{\theta}_n = k(\tilde{\theta}_n)$
- $k(\cdot)$ is a bridge function, connecting wrong value of θ to the right one $\qquad$ Yi & R, 2010; Jiang & Turnbull, 2004

... indirect inference

Smith, 2008

- model of interest

$$y_t = G_t(y_{t-1}, x_t, \epsilon_t; \theta), \quad \theta \in \mathbb{R}^d$$

- likelihood is not-computable, but can simulate from the model

- simple (wrong) model

$$y_t \sim f(y_t \mid y_{t-1}, x_t; \theta^*), \quad \theta^* \in \mathbb{R}^p$$

- find the MLE in the simple model, $\hat{\theta}^* = \hat{\theta}^*(y_1, \ldots, y_n)$, say

- use simulated samples from model of interest to find the 'best' β
- 'best' θ gives data that reproduces $\hat{\theta}^*$

Shalizi, 2013

... indirect inference

Smith, 2008

- simulate samples y_t^m, $\quad m = 1, \ldots, M$ at some value θ

- compute $\hat{\theta}^*(\theta)$ from the simulated data

$$\hat{\theta}^*(\theta) = \arg\max_{\theta^*} \sum_m \sum_t \log f(y_t^m \mid y_{t-1}^m, x_t; \theta^*)$$

- choose θ so that $\hat{\theta}^*(\theta)$ is as close as possible to $\hat{\theta}^*$

- if $p = d$ simply invert the 'bridge function'
- usually $p > d$
  - $\hat{\theta}_1 = \arg\min_\theta \{\hat{\theta}^*(\theta) - \hat{\theta}\}^{\mathrm{T}} W \{\hat{\theta}^*(\theta) - \hat{\theta}\}$
  - $\hat{\beta}_2 = \arg\min_\theta (\sum_t \log f(y_t \mid y_{t-1}, x_t, \hat{\theta}^*(\theta)) - \sum_t \log f(y_t \mid y_{t-1}, x_t, \hat{\theta}))$
- estimates of θ are consistent, asymptotically normal, but not efficient

Approximate Bayesian Computation

Marin et al., 2010

- simulate θ' from $\pi(\theta)$
- simulate data z from $f(\cdot; \theta')$
- if $z = y$ then θ' is an observation from posterior $\pi(\cdot \mid y)$
- actually $s(z) = s(y)$ for some set of statistics
- actually $\rho\{s(z), s(y)\} < \epsilon$ for some distance function $\rho(\cdot)$

Fearnhead & Prangle, 2011

- many variations, using different MCMC methods to select candidate values θ'

... approximate Bayesian computation

$M/G/1$ queue: exponential arrival times, general service times, single server

observations y_i: times between departures from the queue

unobserved variables V_i: arrival time of customer i

model:

- $V_1 \sim \text{Exp}(\theta_3)$
- $V_i \mid V_{i-1} \sim V_{i-1} + \text{Exp}(\theta_3)$
- $Y_i \mid X_{i-1}, V_i \sim \text{Uniform}\{\theta_1 + max(0, V_i - X_{i-1}),$
  $\theta_2 + max(0, V_i - X_{i-1})\} \qquad X_i = \sum_{j=1}^{i} Y_j$
- service time $\sim U(\theta_1, \theta_2)$

ABC: use quantiles of departure times as summary statistics

Indirect Inference: use $\bar{y}$, $y_{(1)}$, $\hat{\theta}_2$ from steady-state model

Table 7. Mean quadratic losses for various analyses of 50 $M/G/1$ data sets†

Method	θ_1	θ_2	θ_3
Comparison	1.1	2.2	0.0013
Comparison + regression	*0.020*	1.1	*0.0013*
Semi-automatic ABC	*0.022*	1.0	*0.0013*
Semi-automatic predictors	0.024	1.2	0.0017
Indirect inference	0.18	*0.42*	0.0033

†Losses within 10% of the smallest values for that parameter are italicized.

Fearnhead & Prangle, 2011

ABC and Indirect Inference

Cox & Kartsonaki, 2012

- both methods need a set of parameter values from which to simulate: θ' or θ
- both methods need a set of auxiliary functions of the data $s(y)$ or $\hat{\theta}^*(y)$

- in indirect inference, $\hat{\theta}^*$ is the 'bridge' to the parameters of real interest, θ

- C & K use orthogonal designs based on Hadamard matrices to chose θ'
- and calculate summary statistics focussed on individual components of θ

- MCMC estimation of log-likelihood function

Geyer & Thompson, 1992

cond. comp. likelihood poor for Ising model   Okabayashi et al., 2011

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician* **24**, 179–195.
Breslow, N.E. & Clayton, D. G. (1993). Approximate inference in generalizsed linear mixed models. *J. Am. Statist. Assoc.* **88**, 9–25.
Cox, D.R. & Kartsonaki, C. (2012). The fitting of complex parametric models. *Biometrika* **99**, 741–747.
Davis, R. & Yau, C.Y. (2011). Comments on pairwise likelihood in time series. *Statistica Sinica* **21**, 255–277.
Davison, A.C. (2000). *Statistical Models*. CUP Cambridge.
Davison, A.C., (2012). Statistical modeling of spatial extremes. *Statistical Science* **27**, 161–186.
Davison, A.C. & Huser, R.(2015). Statistics of Extremes *Annual Reviews* **2**, to appear.
Fearnhead, P. & Prangle, (2012). Approximate lieklihood methods for estimating local recombination rates *J. R. Statist. Soc.* B **64**, 657–680.
Fraser, D.A.S. & Naderi, A. (2007). Minimal sufficient staststics ...*Int. J. Statist. Sci.* **6** 55–61.
Geyer, C. & Thompson, E.A. (1992). Constrained MC maximum likelihood... *J. R. Statist. Soc.* B **54**, 657–699.
Hall, P., Ormerod, J.T. & Wand, M.P. (2011). Theory of Gaussian variational approximation... *Stat. Sinica* **21**, 369–389.
Hall, P., Pham, T., Wand, M.P. & Wang S.S.J. (2011). Asymptotic normality and valid inference for Gaussian variational approximation. *Ann. Statist.* **39** 2502–2532.
Hinkley, D.V. (1980). Likelihood as approximate pivotal. *Biometrika* **67**, 287–292.
Jiang, W. & Turnbull, B. (2004). The indirect methods ... *Statistical Science* **19**, 239–263 .
Lindsay, B. (1988). Composite likelihood methods. *Contemp. Math.* **80**, 220–239.
Marin, J.-M. et al. (2010). Approximate Bayesian computational methods. *Stat. & Computing* **22**, 1167–1180.
Molenberghs, G. & Verbeke, G. (2006). *Discrete Longitudinal Data* Springer, New York.
Okabayashi, X. Johnson, L.& Geyer, C.J. (2011). Extending pseudo-likelihood *Statistica Sinica* **21**, 331–347.
Ormerod, & Wand, M. (2010). Explaining variational approximations. *Am. Stat.* **64**, 140–153.
Ormerod, & Wand, M. (2012). Gaussian variational approximate inference... *J Comp Graph Statist***21**, 2–17.
Raudenbush, S., Yang, M.-L. & Yosef, M. (2000). *J. Comp. Graph. Statist.* **9**, 141–157
Ravikumar et al. (2010). High-dimensional Ising model selection... *Ann. Statist.* **38**, 1287–1319.
Robin, S. (2011). online talk
Royall, R.J. (1997). *Statistical Evidence...*. Chapman & Hall, London.
Rue, H. , (2009). Approximation Bayesian inference ... *J. R. Statist. Soc.* B **71**, 319–392.
Shalizi, C. (2013). Notebooks. indirect inference
Shestaopaloff, A. & Neal, R. (2013). On Bayesian inference for the M/G/1 queue... online report.
Shun, Z. & McCullagh, P. (1995). Laplace approximation ... *J. R. Statist. Soc.* B *57*, 749–760.
Smith, A.A. (2008). Indirect inference. in *New Palgrave Dictionary of Economics* 2nd ed.
Tan, & Nott, (2013). Variational inference for generalized linear mixed models... *Statist. Sci.* **28**, 168–188.
Tan & Nott (2014). Variational approximation for mixtures of linear models. *J. Comp. Graph. Statist.* **23**, 564–585.
Titterington, D.M. (2006). Bayesian methods for neural networks ... *Statistical Science* **19**, 128–139.
Xue, L., Zou, H. & Cai, T. (2012). Nonconcave penalized composite conditional likelihood... *Ann. Statist.* **40**, 1403–1429.

http://www.fields.utoronto.ca/programs/scientific/14-15/bigdata/

FIELDS

THEMATIC PROGRAM ON STATISTICAL INFERENCE, LEARNING, AND MODELS FOR BIG DATA

JANUARY - JUNE, 2015

PROGRAM

JANUARY 12 – 23, 2015

Opening Conference and Boot Camp

Organizing Committee: Nancy Reid (Chair), Sallie Keller, Lisa Lix, Bin Yu

JANUARY 26 – 30, 2015

Workshop on Big Data and Statistical Machine Learning

Organizing committee: Ruslan Salakhutdinov (Chair), Dale Schuurmans, Yoshua Bengio, Hugh Chipman, Bin Yu

FEBRUARY 9 – 13 , 2015

Workshop on Optimization and Matrix Methods in Big Data

Organizing Committee: Stephen Vavasis (Chair), Anima Anandkumar, Petros Drineas, Michael Friedlander , Nancy Reid, Martin Wainwright

FEBRUARY 23 – 27, 2015

Workshop on Visualization for Big Data: Strategies and Principles

Organizing Committee: Nancy Reid (Chair), Susan Holmes, Snehelata Huzurbazar, Hadley Wickham, Leland Wilkinson

MARCH 23 – 27, 2015

Workshop on Big Data in Health Policy

Organizing Committee: Lisa Lix (Chair), Constantine Gatsonis , Sharon-Lise Normand

APRIL 13 – 17, 2015

Workshop on Big Data for Social Policy

This thematic program emphasizes both applied and theoretical aspects of statistical inference, learning and models in big data. The opening conference will serve as an introduction to the program, concentrating on overview lectures and background preparation. Workshops throughout the program will highlight cross-cutting themes, such as learning and visualization, as well as focus themes for applications in the social, physical and life sciences. It is expected that all activities will be webcast using the FieldsLive system to permit wide participation. Allied activities planned include workshops at PIMS in April and May and CRM in May and August.

ORGANIZING COMMITTEE

Yoshua Bengio (Montréal)
Hugh Chipman (Acadia)
Sallie Keller (Virginia Tech)