Aspects of Likelihood Inference

Nancy Reid

October 16, 2013
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 $f(y \mid x; \theta)$, $\theta = (\theta_1, \ldots, \theta_d)$

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

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

- $\theta$ could have very large dimension, $d > n$
  typically $y = (y_1, \ldots, y_n)$

- $\theta$ 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 $\theta$ or $f(\cdot)$
- suggested by Fisher as a measure of plausibility

\[
L(\hat{\theta})/L(\theta) \in (1, 3) \quad \text{very plausible;}
\]
\[
L(\hat{\theta})/L(\theta) \in (3, 10) \quad \text{implausible;}
\]
\[
L(\hat{\theta})/L(\theta) \in (10, \infty) \quad \text{very implausible}
\]

- converts a ‘prior’ probability $\pi(\theta)$ to a posterior $\pi(\theta | y)$ via Bayes’ Theorem

- provides a conventional set of summary quantities for inference based on properties of the postulated model
Widely used

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

A. Suyuthi\textsuperscript{a}, B.J. Leira\textsuperscript{a}, K. Riska\textsuperscript{b, c}

\textsuperscript{a} Department of Marine Technology, NTNU, Trondheim, Norway

\textsuperscript{b} Centre of Ships and Offshore Structures (CeSOS), Trondheim, Norway

\textsuperscript{c} Il S. OY. Helsinki, Finland
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
Empirical growth curve estimation considering multiple seasonal compensatory growths of body weights in Japanese Thoroughbred colts and fillies.

(PMID:24085406)

<table>
<thead>
<tr>
<th>Abstract</th>
<th>Citations</th>
<th>BioEntities</th>
<th>Related Articles</th>
<th>External Links</th>
</tr>
</thead>
</table>

Comparative Agricultural Sciences, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan
Journal of Animal Science [2013]

**Type:** Journal Article
... widely used
... widely used

HAVING A MID-LIFE CRISIS? YOU'RE NOT ALONE

A study involving two million people in 72 countries found men and women were less happy in their 40s but that improved in later life.

PROBABILITY OF DEPRESSION BY AGE

PERCENTAGE LIKELIHOOD

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
  \[ \hat{\theta} \sim N(\theta, \frac{1}{n} \chi^2) \]
  \[ \sqrt{n}(\hat{\theta} - \theta) \sim N(0,1) \]
- likelihood function + sample space derivative gives better approximate inference
  \[
  f(\hat{\theta} | a, \theta) = \frac{L(\theta)}{\int L(\theta) \, d\theta} \\
  f(a \, \text{true of } \theta) = \frac{L(\theta; \hat{\theta}, a)}{\int L(\theta; \hat{\theta}, a) \, d\theta}
  \]
Derived quantities

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

- **observed Fisher information**
  \[ j(\hat{\theta}) = -\frac{\partial^2 \ell(\theta)}{\partial \theta^2} \]

- **efficient score function**
  \[ \ell'(\theta) = \frac{\partial \ell(\theta; y)}{\partial \theta} \]
  \[ \ell'(\hat{\theta}) = 0 \text{ assuming enough regularity} \]
  \[ \ell'(\theta; y) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f_{Y_i}(y_i; \theta), \quad y_1, \ldots, y_n \text{ independent} \]
Approximate pivots

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

\[
\hat{\theta} - \theta \sim N(0, j^{-1}(\hat{\theta})) \iff (\hat{\theta} - \theta) \sim N(0, 1)
\]

\[
\begin{align*}
re(\psi; y) &= (\hat{\psi} - \psi)j_p^{1/2}(\hat{\psi}) \sim N(0, 1) \\
r(\psi; y) &= \pm \sqrt{2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\}} \sim N(0, 1) \\
\pi_m(\psi | y) &= N\{\hat{\psi}, j_p^{-1/2}(\hat{\psi})\}
\end{align*}
\]

\[
j_p(\psi) = -\ell''_p(\psi); \text{ profile information}
\]

$\frac{j}{-\frac{\ell''}{\ell''}}$

$z \sim N(0, 1)$

$z^2 \sim X^2_1$

$w \sim X^2_1$

$\pm \sqrt{w} \sim N(0, 1)$
Aspects of Likelihood Inference  Basel, 2013 12

... approximate pivots scalar parameter of interest

\[ r^*(\psi) = r(\hat{\psi}) + \frac{1}{r'(\hat{\psi})} \phi \frac{\phi(\psi)}{\sqrt{r'(\psi)}} \sim N(0,1) \]

(to higher order)

\[ \sim N(0,1) \]

\[ \phi = (\hat{\psi} - \psi) \frac{1}{\sqrt{\psi}} (\hat{\psi}) \]

\[ r = \pm \sqrt{2 \{ - \ell_\theta (\hat{\psi}) - \ell_\theta (\hat{\psi}) \}} \]
approximate pivots

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

$$
y_i \overset{i.i.d.}{\sim} \frac{1}{\beta(\psi)} \left( \frac{\psi}{\lambda} \right)^{\psi_i - 1} \exp\left( - \frac{\psi_i}{\lambda} \right) \sim \text{Gam}(\psi, \lambda)
$$

$$
r(\psi; y) = \pm \sqrt{2 \{ \ell_p(\psi) - \ell_p(\hat{\psi}) \}} \sim N(0, 1)
$$

$$
\theta = (\psi, \lambda) \sim \text{Gam}(\text{slop } \psi, \text{ mean } \lambda)
$$

$$
\ell(4, \chi) = \sum \ell_p f(y_i; \psi, \chi)
$$

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

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

$$
\frac{\partial \ell(\psi, \lambda)}{\partial \chi} = 0
$$
... approximate pivots

- profile log-likelihood \( \ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi) \)
- \( \theta = (\psi, \lambda); \hat{\lambda}_\psi \) constrained maximum likelihood estimator

\[
\begin{align*}
    r_e(\psi; y) &= (\hat{\psi} - \psi) j_p^{1/2}(\hat{\psi}) \\
    r(\psi; y) &= \pm \sqrt{2 \{ \ell_p(\hat{\psi}) - \ell_p(\psi) \}} \\
    \pi_m(\psi | y) &\sim N\{ \hat{\psi}, j_p^{-1/2}(\hat{\psi}) \}
\end{align*}
\]

\[
\begin{align*}
    r^*(\psi; y) &= r(\psi) + \frac{1}{r(\psi)} \log \left\{ \frac{Q_F(\psi)}{\ell_p(\psi)} \right\} \\
    r^*_B(\psi; y) &= r(\psi) + \frac{1}{r(\psi)} \log \left\{ \frac{Q_B(\psi)}{\ell_p(\psi)} \right\}
\end{align*}
\]

\[\text{\textit{\text{Derived from good approx to}} \int \pi_m(\psi | y) \, d\psi}\]
The problem with profiling

- \( \ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi) \) 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 \( \hat{\sigma}^2 = \frac{RSS}{n} \)
  usual estimator \( \frac{RSS}{(n - k)} \) \( k \) the number of regression coefficients
- badly biased if \( k \) large relative to \( n \)
- inconsistent for \( \sigma^2 \) 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
- \( L_p(\psi) \) is a conditional likelihood \( L_c(\psi) \), and estimation of \( \lambda \) has no impact on asymptotic properties
- \( s \) is conditionally sufficient, \( t \) is marginally ancillary, for \( \psi \)
- 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