1. Introduction

Goal: $p$-value for scalar parameter

Solution: correction to likelihood ratio statistic via maximum likelihood or score type statistic

$$p - \text{value} = \Phi(r) + \phi(r) \left( \frac{1}{r} - \frac{1}{q} \right)$$

$$= \Phi(r + \frac{1}{r} \log \frac{q}{r})$$

$$r = \pm \sqrt{2\{\ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi, \hat{\lambda}_\psi)\}}$$
based on likelihood ratio test
easily related to Bayesian solution
limits scope of solutions (Skovgaard, 99)
2. Steps in solution

1. $p^*$ or something similar to get a density supported on $R^k$ instead of $R^n$ (i.e. condition on ancillary)

2. transform this to joint density of a one-dimensional pivotal statistic for $\psi$, and something else

3. integrate out something else

4. integrate density up to (or beyond) observed data point

Note that 1. and 3. are also needed for problems with no nuisance parameters, and that 2. can be obtained via 1.
3. Details

Step 1. \( n \downarrow k \)

A: \( y \leftrightarrow (\hat{\theta}, a) \); need \( p(a; \theta) \equiv p(a) \)

Embed \( f(y; \theta) \) in a family with a \( k + d \) dimensional sufficient statistic and \( k + d \) dimensional parameter, \( f(y; \theta, \eta) \), say.

Successively test \( \eta_j = 0, j = 1, \ldots, d \) using \( r^* \) type statistics (standard normal to high order)


\[
p^*(\hat{\theta}|a; \theta) = c|j(\hat{\theta})|^{1/2} \exp\{\ell(\theta) - \ell(\hat{\theta})\}
\]
B: For \( p \)-values, the only information needed about \( a \) is \( \partial \ell(\theta; y)/\partial V(y) \), where \( V \) is a set of vectors tangent to the ancillary surface at the observed data point.

These can be computed using a location model approximation to \( f(y; \theta) \) locally near \((y^0, \tilde{\theta}^0)\)

The approximate \( k \)-dimensional model thus obtained has the structure of an exponential family model; called a tangent exponential model

\[
p_{TEM}(s|a; \theta) = \tilde{c}|j(\varphi)|^{-1/2} \exp[\ell(\theta; y^0) - \ell(\tilde{\theta}^0; y^0) + \{\varphi(\theta) - \varphi(\tilde{\theta}^0)\}^T s] \\
\varphi(\theta) = \varphi(\theta; y^0) = \ell; V(\theta; y^0) \\
j(\varphi) = -\frac{\partial^2 \ell(\theta; y^0)}{\partial \varphi^2} \{\theta(\varphi); y^0\}
\]
Example

\[ f(y; \theta) = \frac{e^{y-\theta}}{\{1 + e^{y-\theta}\}^2} e^{\theta(y-\theta)-c(\theta)}, \quad -1 \leq \theta \leq 1 \]

\[ c(\theta) = \log\{\pi\theta/sin(\pi\theta)\} \]

\( n = 2 \)
A: Need to find coordinates \((\tilde{\theta}, a) \leftrightarrow (y_1, y_2)\) where \( \ell'(\tilde{\theta}) = 0; p(a; \theta) \) free of \( \theta \) (approx)

B: Trace out the curve in \((y_1, y_2)\) from \((y_1^0, y_2^0)\) by finding, at \(i\)th iteration

\[ v_1^i = -\frac{F_\theta(y_1, \theta)}{f(y_1, \theta)} \bigg|_{y_1^{i-1}, \tilde{\theta}^{i-1}} \]

\[ v_2^i = -\frac{F_\theta(y_2, \theta)}{f(y_2, \theta)} \bigg|_{y_2^{i-1}, \tilde{\theta}^{i-1}} \]

Curve is \(y_1^0, y_1^1 = y_0^0 + \delta v_0^0, \ldots\)
$f(y_1, y_2; \theta)$
Step 2. Fix \( a \); find a pivotal statistic for \( \psi \):

If \( \psi \) is fixed, a pivotal for \( \psi \) is an ancillary statistic for the nuisance parameter \( \lambda \); apply Step 1. to find this pivotal.

Leads to dimension reduction from \( k \) to 1.

A: Leads to \( r_{\psi}^* \) and the calculation

\[
p^*(\theta) \rightarrow p^*(r_{\psi}^*, \bar{\lambda}_\psi)
\]

thence to

\[
p_{m}(r_{\psi}^*) \overset{d}{=} N(0, 1)
\]

\[
r_{\psi}^* = r_{\psi} + \frac{1}{r} \log \left( \frac{u}{r} \right)
\]

B: Ancillary exists by construction in Step 1., so

\[
p_{TEM}(s; \theta|a) = p_1(s; \theta|a_{\psi}) p_2(a_{\psi})
\]

where both \( p_{TEM} \) and \( p_1 \) have exponential family form, take ratio to get \( p_2(a_{\psi}) \)
... Step 2

In both cases the approximations are of “Laplace type” and cdf easily found to be of the form

\[ \Phi(r^*_\psi) \equiv \Phi\{r + (1/r) \log(u/r)\} \]

\[ \equiv \Phi(r) + \phi(r)(1/r - 1/q) \]

\[ r = \pm [2\{\ell(\tilde{\psi}, \tilde{\lambda}) - \ell(\psi, \tilde{\lambda}_\psi)\}] \]

\[ u = \frac{\left| \ell; \hat{\theta}(\hat{\theta}) - \ell; \hat{\theta}(\hat{\theta}_\psi) \right| \ell_\lambda; \hat{\theta}(\hat{\theta}_\psi)}{\left| \ell_\theta; \hat{\theta}(\hat{\theta}) \right|} \frac{|j_{\theta\theta}(\hat{\theta})|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}} \]

\[ q = \frac{\left| \ell; V(\hat{\theta}) - \ell; V(\hat{\theta}_\psi) \right| \ell_\lambda; V(\hat{\theta}_\psi)}{\left| \ell_\theta; V(\hat{\theta}) \right|} \frac{|j_{\theta\theta}(\hat{\theta})|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}} \]

\[ u \text{ is a ‘dual score’ statistic} \]

\[ q = (\tilde{v} - \tilde{v}_\psi) / \hat{\sigma}_v \text{ is a Wald statistic for a derived parameter} \]
\[ \nu(\theta) = e^T \varphi(\theta) \]
\[ \hat{\sigma}_\nu = \frac{|j_{(\lambda \lambda)}(\tilde{\theta}_\psi)|^{1/2}}{|j_{(\theta \theta)}(\tilde{\theta})|^{1/2}} \]
\[ |j_{(\theta \theta)}(\tilde{\theta})| = |j_{\theta \theta}(\tilde{\theta})||\varphi_{\theta}(\tilde{\theta})|^{-2} \]
\[ |j_{(\lambda \lambda)}(\tilde{\theta}_\psi)| = |j_{\lambda \lambda}(\tilde{\theta}_\psi)||\varphi_{\lambda}(\tilde{\theta}_\psi)|^{-2} \]
\[ \varphi(\theta) = \ell;V(\theta; y^0) \]
\[ e_{\psi} = \psi_{\varphi}(\tilde{\theta}_\psi)/|\psi_{\varphi}(\tilde{\theta}_\psi)| \]
4. Bayesian versions and matching priors

1. Bayesian version

Derivations similar, but easier because inferential basis prescribed, also variable of integration is $\psi$, not $\hat{\psi}$

Result is

$$\Pr_m(\psi \geq \psi | y) = \Phi(r^*)$$

where

$$q = q_B = \ell_\psi(\hat{\theta}_\psi) \frac{|j \lambda \lambda(\hat{\theta}_\psi)|^{1/2}}{|j(\hat{\theta})|^{1/2}} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)}$$
strong matching sets $q_B = q_f$, which implies that

$$\frac{\pi(\hat{\theta}_\psi)}{\pi(\hat{\theta})} = \ldots$$

This prior is flat in an approximate location parametrization, with an adjustment for nuisance parameters.

Depends on $y$, although we could take expectations.
\[
\pi_m(y) = \frac{\int e^{\ell(\psi, \lambda)} \pi(\psi, \lambda) d\lambda}{\int \int e^{\ell(\psi, \lambda)} \pi(\psi, \lambda) d\lambda d\psi} \\
= \frac{e^{\ell(\psi, \hat{\lambda}_\psi)} |j_{\lambda}(\psi, \hat{\lambda}_\psi)|^{-1/2} \pi(\psi, \hat{\lambda}_\psi) \sqrt{(2\pi)^k-1}}{e^{\ell(\tilde{\psi}, \hat{\lambda})} |j(\tilde{\psi}, \hat{\lambda})|^{-1/2} \pi(\tilde{\psi}, \hat{\lambda}) \sqrt{(2\pi)^k}} \\
= \frac{1}{\sqrt{(2\pi)}} e^{\ell_a(\hat{\psi})-\ell_a(\psi)} \frac{\pi(\psi, \hat{\lambda}_\psi)}{\pi(\tilde{\psi}, \hat{\lambda})}
\]

\[
\int_{\psi}^{\infty} \pi(\psi|y) d\psi = \int \\
= \int \frac{1}{\sqrt{(2\pi)}} e^{-r^2/2} \frac{r}{\ell'_a(\psi)} \pi(\psi, \hat{\lambda}_\psi) d\psi \\
= \int \frac{1}{\sqrt{(2\pi)}} e^{-r^2/2} \left( \frac{r}{q_B} \right) d\psi \\
= \Phi(r) + \int r\phi(r) \left( \frac{1}{q_B} - \frac{1}{r} \right) d\psi \\
= \Phi(r) + \phi(r) \left( \frac{1}{q_B} - \frac{1}{r} \right) + \ldots
\]
2. using expansions

- Edgeworth expansion for posterior density, leading to Cornish-Fisher expansion for posterior \((1 - \alpha)\) quantile, a series in \(\hat{\theta}\) and observed log-likelihood derivatives

- Taylor series expansion for frequentist coverage, using asymptotic distribution of log-likelihood derivatives

- equate frequentist coverage to \(1 - \alpha\)

- leads to differential equation involving \(\pi\) and expected log-likelihood derivatives
1. to $O(n^{-1})$ (Welch and Peers, 1963):

scalar parameter $\theta$:

$$\frac{\pi'(\theta)}{\pi(\theta)} - \frac{1}{2} \frac{i'(\theta)}{i(\theta)} = 0 \implies \pi(\theta) \propto i^{1/2}(\theta)$$

with nuisance parameters:

$$\sum_j \frac{\partial}{\partial \theta_j} \left[ \{i^{11}(\theta)\}^{-1/2} i_{1j}(\theta) \pi(\theta) \right] = 0 \quad (1)$$

Peers, 1965, Bka; Ghosh and Mukerjee, 1997

– in general has infinitely many solutions; e.g. suppose $i_{1j}(\theta) = 0, j = 2, \ldots, p$:

$$\frac{\partial}{\partial \theta_1} \{i^{1/2}(\theta) \pi(\theta)\} = 0$$

$$\pi(\theta) \propto i^{1/2}_{11}(\theta) g(\theta_2, \ldots, \theta_p)$$

Tibshirani, 1989, Bka
- require (1) for each component in turn; leads to no solutions (in general)

2. to $O(n^{-3/2})$: - Depends on the model structure

$$T_2(\pi, \theta) = 0 \iff \frac{d}{d\theta} \left[ \frac{E \left( \frac{\partial \ell}{\partial \theta} \right)^3}{\{i(\theta)\}^{3/2}} \right] = 0$$

in scalar parameter case, and a similar condition in the case of nuisance parameters

Muikerjee & Ghosh, 1997, Bka
- In the orthogonal parameter case, this condition is

\[
\frac{1}{6} g(\theta(2)) D_1(i_{11}^{-3/2} i_{1,1,1}) \\
+ \sum_{v=2}^{p} \sum_{s=2}^{p} D_v \{ i_{11}^{-1/2} i_{11s} i_{sv} g(\theta(2)) \} = 0
\]

where \( i_{1,1,1} = E(\ell_1)^3 \) and \( i_{11s} = E(\ell_{11s}) \)
...4

Conclusions

- matching to $O(n^{-1})$ doesn’t lead to a unique solution

- to next order depends both on matching criteria and on model

- strong matching approach may provide insight into this

- Bayes/frequentist inference ‘can’t’ agree to $O(n^{-3/2})$ because of sample space differentiation