Applied asymptotics
Bayesian and frequentist inference

Nancy Reid

April 20, 2008
Bayesian posterior distribution

$$Pr_m(\Psi \leq \psi \mid y) = \Phi(r_B^*) = \Phi(r + \frac{1}{r} \log \frac{q_B}{r})$$

$$q_B = -\ell_p'(\psi)j_p(\hat{\psi})^{-1/2} \frac{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta})|^{1/2}} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)}$$

$$r = \pm [2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\}]^{1/2}, \quad \psi \in R$$
... Bayesian posterior distribution

\[
Pr_m(\Psi \leq \psi \mid y) = \Phi(r_B^*) = \Phi\left(r + \frac{1}{r} \log \frac{q_B}{r}\right)
\]

\(\triangleright\) \(r = \pm \left[2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\}\right]^{1/2}, \quad \psi \in \mathbb{R}\) likelihood root

\(\triangleright\) \(\ell(\theta) = \ell(\psi, \lambda) = \log f(y; \psi, \lambda), \quad y \in \mathbb{R}^n\) log-likelihood

\(\triangleright\) \(\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)\) constrained m.l.e.

\(\triangleright\) \(\ell_p(\psi) = \ell(\hat{\theta}_\psi)\) profile log-likelihood

\(\triangleright\) \(j(\theta) = -\ell''(\theta; y)\) observed information

\(\triangleright\) \(j(\theta) = \begin{bmatrix}
j_{\psi\psi}(\theta) & j_{\psi\lambda}(\theta) \\
j_{\lambda\psi}(\theta) & j_{\lambda\lambda}(\theta)
\end{bmatrix}\) partitioned matrix

"... Some asymptotic formulas"
... Bayesian posterior

\[ r_B^* \sim N(0, 1) \]

\[ r_B^* = r + \frac{1}{r} \log \frac{q_B}{r} \]

- \( r \) is the likelihood root
- \( q_B \) is an adjusted score statistic

The approximation is very good!
Example: Normal circle

- $y_1 \sim N(\mu_1, 1/n), \ldots, y_k \sim N(\mu_k, 1/n)$
- Parameter of interest $\psi = (\mu_1^2 + \cdots + \mu_k^2)^{1/2} = ||\mu||$
- Prior $\pi(\mu) = 1$
- Exact marginal posterior $\Pr\{\chi^2_k(n||y||^2) \geq n\psi^2\}$
- Third order

$$r^*_B = \sqrt{n(\hat{\psi} - \psi)} + \frac{1}{\sqrt{n(\hat{\psi} - \psi)}} \log \left\{ \left( \frac{\hat{\psi}}{\psi} \right)^{(k-1)/2} \right\} \sim N(0, 1)$$

- Normal approximation to posterior $\sqrt{n(\hat{\psi} - \psi)} \sim N(0, 1)$
Normal Circle, $k=2$

$p$-value vs $\psi$
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion

Normal Circle, k=2

\(\psi\)

p-value

\(\psi\)

2 3 4 5 6 7 8

0.0 0.2 0.4 0.6 0.8 1.0

Normal Circle, k=2

ψψ

p−value

2 3 4 5 6 7 8

0.0 0.2 0.4 0.6 0.8 1.0

Normal Circle, k=2

ψψ

p−value
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion

Normal Circle, $k=2$
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion
Normal Circle, k=2, 5, 10
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion

Normal Circle, $k=2, 5, 10$

\[ \psi \]

$p$-value

\[ \psi \]

Graph showing the Normal Circle for $k=2, 5, 10$. The $p$-value is plotted against $\psi$. The graph shows curves for different values of $k$, with $k=2$ in green, $k=5$ in black, and $k=10$ in red.
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion
Exact and approximate survivor function for $\psi$, for $n = 1$, $\hat{\psi} = 5$.

<table>
<thead>
<tr>
<th></th>
<th>exact</th>
<th>0.99</th>
<th>0.95</th>
<th>0.75</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k = 5$</td>
<td></td>
<td>0.9898</td>
<td>0.9495</td>
<td>0.7491</td>
<td>0.4991</td>
</tr>
<tr>
<td>$k = 10$</td>
<td>$\Phi(r_B^*)$</td>
<td>0.9897</td>
<td>0.9493</td>
<td>0.7486</td>
<td>0.4987</td>
</tr>
<tr>
<td>$k = 20$</td>
<td></td>
<td>0.9899</td>
<td>0.9500</td>
<td>0.7506</td>
<td>0.5012</td>
</tr>
<tr>
<td></td>
<td>exact</td>
<td>0.25</td>
<td>0.05</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>$k = 5$</td>
<td></td>
<td>0.2494</td>
<td>0.0499</td>
<td>0.00997</td>
<td></td>
</tr>
<tr>
<td>$k = 10$</td>
<td>$\Phi(r_B^*)$</td>
<td>0.2494</td>
<td>0.0498</td>
<td>0.00997</td>
<td></td>
</tr>
<tr>
<td>$k = 20$</td>
<td></td>
<td>0.2511</td>
<td>0.0504</td>
<td>0.01012</td>
<td></td>
</tr>
</tbody>
</table>
Frequentist P-value

\[ \text{Pvalue}(\psi) \doteq \Phi(r_F^*) = \Phi(r + \frac{1}{r} \log \frac{q_F}{r}) \]

\[ r = \pm \left[ 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\} \right]^{1/2} \]

\[ q_F = \frac{|\ell; \nabla(\hat{\theta}) - \ell; \nabla(\theta) \ell_\lambda; \nabla(\hat{\theta}_\psi)|}{|\ell_\theta; \nabla(\hat{\theta})|} \frac{|j(\hat{\theta})|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}} \]
Normal circle

- **exact P-value** \( \psi \)

\[
\Pr\{ \chi_k^2(n\psi^2) \geq n||y||^2 \}
\]

- **approx:**

\[
r_F^* = \sqrt{n(\hat{\psi} - \psi)} - \frac{1}{\sqrt{n(\hat{\psi} - \psi)}} \log \left\{ \left( \frac{\hat{\psi}}{\psi} \right)^{(k-1)/2} \right\} \sim N(0, 1)
\]

- **Bayes:**

\[
r_B^* = \sqrt{n(\hat{\psi} - \psi)} + \frac{1}{\sqrt{n(\hat{\psi} - \psi)}} \log \left\{ \left( \frac{\hat{\psi}}{\psi} \right)^{(k-1)/2} \right\} \sim N(0, 1)
\]

- **Bayes vs. Frequentist:**

\[
r_B^* - r_F^* \sim \frac{k-1}{\psi \sqrt{n}}
\]
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion
Sensitivity to priors

\( Y_i \sim N(\mu_i, 1/n), \quad i = 1, \ldots, k: \) flat prior obviously poor

▶ try some more ‘realistic’ priors, cheaply

▶ use \( r^*_B \) with different priors

\[
q_B = -\ell'_p(\psi)j_p(\hat{\psi})^{-1/2} \frac{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}}{|j_{\lambda}(\hat{\theta})|^{1/2}} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)}
\]

\[
r^*_B = \sqrt{n(\hat{\psi} - \psi)} + \frac{1}{\sqrt{n(\hat{\psi} - \psi)}} \log \left\{ \left( \frac{\hat{\psi}}{\psi} \right)^{(k-1)/2} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)} \right\}
\]

▶ Prior 1: \( \mu_i \sim N(0, \tau^2) \)  
Prior 2: \( \mu_i \sim N(a, \tau^2), \quad a \sim N(0, \nu^2) \)

▶ Prior 3: \( \mu_i \sim N(0, \sigma^2), \quad 1/\sigma^2 \sim \Gamma(\alpha, \beta) \)
Prior 4: \( \mu_i \sim N(a, \sigma^2), \quad a \sim N(0, \nu^2), \quad 1/\sigma^2 \sim \Gamma(\alpha, \beta) \)
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion

go to preview
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion

\[ \psi \]

\[ P\text{-value} \]

\[ \tau = 1 \]

\[ \tau = 5 \]

\[ \tau = 10 \]

\[ \tau = 100 \]
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion

hyperprior $N(0, \nu^2)$

- $\tau = 1$
- $\tau = 5$
- $\tau = 10$
- $\tau = 100$
Flat priors for vector parameters

- poor sampling behaviour for normal circle
- prior needs to be targeted on parameter of interest, e.g.
  \[ \pi(\mu) \propto ||\mu||^{k-1} \]
- e.g. (Cox & Hinkley, 1974): \[ Y_i \sim N(\mu_i, \sigma^2) \]
- \[ \mu_i = \beta_0 + \beta_1 \rho^{x_0 + ja} \]
- prior \[ \pi(\beta_0, \beta_1 \rho, \sigma) \propto d\beta_0 d\beta_1 d\log \sigma d\rho, \quad 0 \leq \rho \leq 1 \]
- marginal posterior for \( \rho \) zero, except \( \to \infty \) at \( \rho = 0, 1 \) (improper)
Logistic regression

- $\Pr(Y_i = 1 \mid x_i) = p_i = \alpha + \beta x_i$
- $Y_i \sim \text{Ber}(p_i)$
- $\psi = -\alpha/\beta$
<table>
<thead>
<tr>
<th></th>
<th>left</th>
<th>center</th>
<th>right</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi(r)$</td>
<td>0.0711</td>
<td>0.9189</td>
<td>0.0100</td>
</tr>
<tr>
<td>$\Phi(r_F^*)$</td>
<td>0.0454</td>
<td>0.9485</td>
<td>0.0061</td>
</tr>
<tr>
<td>$\Phi(r_B^*)$</td>
<td>0.0987</td>
<td>0.8864</td>
<td>0.0149</td>
</tr>
</tbody>
</table>
Multiple logistic regression

- $Pr(Y_i = 1 \mid x_i) = p_i = \beta_0 + \beta_1 x_{1i} + \ldots \beta_6 x_{6i}$
- $Y_i \sim Ber(p_i)$
- $\psi = Pr(Y = 1 \mid x^*) = \exp(\beta' x^*) / \{1 + \exp(\beta' x^*)\}$
Some asymptotic formulas

Checking the priors

Matching priors

Conclusion
Checking the priors

Matching priors

Conclusion
<table>
<thead>
<tr>
<th></th>
<th>$Pr(y = 1 \mid x^*) = 0.86$</th>
<th></th>
<th>$Pr(y = 1 \mid x^*) = 0.21$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>left</td>
<td>center</td>
<td>right</td>
</tr>
<tr>
<td>$\Phi(r)$</td>
<td>0.0228</td>
<td>0.9185</td>
<td>0.0587</td>
</tr>
<tr>
<td>$\Phi(r^*_F)$</td>
<td>0.0297</td>
<td>0.9460</td>
<td>0.0243</td>
</tr>
<tr>
<td>$\Phi(r^*_{B1})$ (flat)</td>
<td>0.0178</td>
<td>0.8612</td>
<td>0.1210</td>
</tr>
<tr>
<td>$\Phi(r^*_{B2})$ (matching)</td>
<td>0.0275</td>
<td>0.9450</td>
<td>0.0275</td>
</tr>
</tbody>
</table>
Strong matching

- $s(\psi) = \Phi(r + \frac{1}{r} \log \frac{q_B}{r})$: Bayesian survivor value

- $p(\psi) = \Phi(r + \frac{1}{r} \log \frac{q_F}{r})$: Frequentist $p$-value

\[
q_B = -\ell_p'(\psi) j_p(\hat{\psi})^{-1/2} \left| \frac{j_{\lambda\lambda}(\hat{\theta}_\psi)}{|j_{\lambda\lambda}(\hat{\theta})|^{1/2}} \right| \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)}
\]

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

- $q_B = q_F \iff \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)} = ...$

- Default prior along the curve $\theta = \hat{\theta}_\psi$  
  F&R, 2002
Example: normal circle $y_i \sim N(\mu_i, 1)$

$$r_F^* = \sqrt{n}(\hat{\psi} - \psi) - \frac{1}{\sqrt{n}(\hat{\psi} - \psi)} \log \left\{ \left( \frac{\hat{\psi}}{\psi} \right)^{(k-1)/2} \right\} \sim N(0, 1)$$

$$r_B^* = \sqrt{n}(\hat{\psi} - \psi) + \frac{1}{\sqrt{n}(\hat{\psi} - \psi)} \log \left\{ \left( \frac{\hat{\psi}}{\psi} \right)^{(k-1)/2} \right\} \sim N(0, 1)$$

$$r_B^* = r_F^* \iff \pi(\mu) d\mu \propto \|\mu\|^{-(k-1)} d\mu \quad (\psi = \|\mu\|)$$
Logistic regression

\[
\frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)} = \frac{i_{\psi\psi}^{1/2}(\hat{\theta})}{i_{\psi\psi}^{1/2}(\hat{\theta}_\psi)}
\]

\[
\theta = (\psi, \lambda); \lambda \perp \psi; \psi = \beta_p \text{ or } \psi = p_1(x^*)
\]

<table>
<thead>
<tr>
<th>(p_1(x^*) = 0.86)</th>
<th>(p_1(x^*) = 0.51)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>left</strong></td>
<td><strong>center</strong></td>
</tr>
<tr>
<td>(\Phi(r_F^*))</td>
<td>0.0297</td>
</tr>
<tr>
<td>(\Phi(r_{B1}^*)) (flat)</td>
<td>0.0178</td>
</tr>
<tr>
<td>(\Phi(r_{B2}^*)) (matching)</td>
<td>0.0275</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(p_1(x^*) = 0.21)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>left</strong></td>
</tr>
<tr>
<td>(\Phi(r_F^*))</td>
</tr>
<tr>
<td>(\Phi(r_{B1}^*)) (flat)</td>
</tr>
<tr>
<td>(\Phi(r_{B2}^*)) (matching)</td>
</tr>
</tbody>
</table>
Flat priors for high dimensional parameters are a bad idea

- $n$ events from Poisson with rate $\epsilon s + b$, with $s$ of interest and additional Poisson measurements of $b$ and $\epsilon$
- ‘flat’ priors for $s, \epsilon, b$
Fig. 2. Typical single channel case. Coverage for 90% credibility level upper limits, acceptance uncertainty = 10%, background uncertainty = zero.
Fig. 6. 3 independent channels. Coverage for 90% credibility level upper limits, acceptance uncertainty = 34%/channel, background uncertainty = 25%/channel.

Fig. 7. 4 independent channels. Coverage for 90% credibility level upper limits, acceptance uncertainty = 40%/channel, Background uncertainty = 29%/channel.
... not flat priors

- Priors that give ‘good’ frequentist properties need to be **targetted** on the parameter of interest.
- Otherwise, the coverage properties of posterior credible intervals cannot be guaranteed.
- Reference priors are targetted on the parameter of interest: a different reference prior is needed for each parameter of interest.
- “Matching” priors are also targetted on the parameter of interest: explicit goal to give good frequentist coverage.
- (This can only be done approximately, to $O(n^{-1})$ in regular models.)
- This is a lot of work!!
- “Non-informative uniform priors on $\mu, \alpha, \sigma_\beta, \sigma_\gamma$” (Gelman et al, 2007, JASA; Jones et al, 2008 Stat Sci)
... not flat priors

marginalization to curved parameters using flat priors may lead to poorly calibrated inferences
Some references


