Likelihood and Asymptotic Theory for Statistical Inference

Nancy Reid

020 7679 1863
reid@utstat.utoronto.ca
n.reid@ucl.ac.uk

http://www.utstat.toronto.edu/reid/ltccF12.html

LTCC London Taught Course Centre
for PhD students in the mathematical sciences

Last week

1. Likelihood – definition, examples, direct inference
2. Derived quantities – score, mle, Fisher information, Bartlett identities
3. Inference from derived quantities – consistency of mle, asymptotic normality
4. Inference via pivotals – standardized score function, standardized mle, likelihood ratio, likelihood root
5. Nuisance parameters and parameters of interest; invariance under parameter transformation
6. Asymptotics for posteriors

This week

1. Bayesian approximation
   1.1 careful statement of asymptotic normality
   1.2 Laplace approximation to posterior density and cumulative distribution function
   1.3 Laplace approximation to marginal posterior density and cdf
   1.4 relation to modified profile likelihood

2. Frequentist inference with nuisance parameters
   2.1 first order summaries; difficulties with profile likelihood
   2.2 marginal and conditional likelihood
   2.3 exponential families
   2.4 transformation families
   2.5 adjustments to profile likelihood

3. Notation

Posterior is asymptotically normal

$$\pi(\theta \mid y) \dot{\sim} N\{\hat{\theta}, j^{-1}(\hat{\theta})\} \quad \theta \in \mathbb{R}, y = (y_1, \ldots, y_n)$$

careful statement

... posterior is asymptotically normal

$$\pi(\theta \mid y) \dot{\sim} N\{\hat{\theta}, j^{-1}(\hat{\theta})\} \quad \theta \in \mathbb{R}, y = (y_1, \ldots, y_n)$$

equivalently $\quad \ell_\pi(\theta) =$

... posterior is asymptotically normal

In fact,

If $\pi(\theta) > 0$ and $\pi'(\theta)$ is continuous in a neighbourhood of θ_0, there exist constants D and n_y s.t.

$$|F_n(\xi) - \Phi(\xi)| < Dn^{-1/2}, \quad \text{for all } n > n_y,$$

on an almost-sure set with respect to $f(y;\theta_0)$, where $y = (y_1, \dots, y_n)$ is a sample from $f(y;\theta_0)$, and θ_0 is an observation from the prior density $\pi(\theta)$.

$$F_n(\xi) = \Pr\{(\theta - \hat{\theta})j^{1/2}(\hat{\theta}) \le \xi \mid y\}$$

Johnson (1970); Datta & Mukerjee (2004)

Laplace approximation

$$\pi(\theta \mid y) \doteq \frac{1}{(2\pi)^{1/2}} |j(\hat{\theta})|^{+1/2} \exp\{\ell(\theta; y) - \ell(\hat{\theta}; y)\} \frac{\pi(\theta)}{\pi(\hat{\theta})}$$

$$\pi(\theta \mid y) = \frac{1}{(2\pi)^{1/2}} |j(\hat{\theta})|^{+1/2} \exp\{\ell(\theta; y) - \ell(\hat{\theta}; y)\} \frac{\pi(\theta)}{\pi(\hat{\theta})} \{1 + O_p(n^{-1})\}$$

$$y = (y_1, \ldots, y_n), \quad \theta \in \mathbb{R}^1$$

$$\pi(\theta \mid y) = \frac{1}{(2\pi)^{1/2}} |j_\pi(\hat{\theta}_\pi)|^{+1/2} \exp\{\ell_\pi(\theta; y) - \ell_\pi(\hat{\theta}_\pi; y)\} \{1 + O_p(n^{-1})\}$$

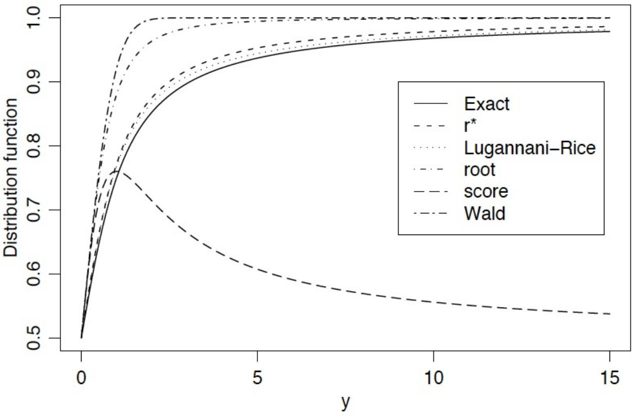
Posterior cdf

$$\int_{-\infty}^{\theta} \pi(\vartheta \mid y) d\vartheta \doteq \int_{-\infty}^{\theta} \frac{1}{(2\pi)^{1/2}} e^{\ell(\vartheta;y)-\ell(\hat{\vartheta};y)} |j(\hat{\vartheta})|^{1/2} \frac{\pi(\vartheta)}{\pi(\hat{\vartheta})} d\vartheta$$

Posterior cdf

$$\int_{-\infty}^{\theta} \pi(\vartheta \mid y) d\vartheta \doteq \int_{-\infty}^{\theta} \frac{1}{(2\pi)^{1/2}} e^{\ell(\vartheta;y)-\ell(\hat{\vartheta};y)} |j(\hat{\vartheta})|^{1/2} \frac{\pi(\vartheta)}{\pi(\hat{\vartheta})} d\vartheta$$

SM, §11.3



BDR, Ch.3, Cauchy with flat prior

Nuisance parameters

$$y = (y_1, \ldots, y_n) \sim f(y; \theta), \quad \theta = (\psi, \lambda)$$

$$\begin{aligned}
\pi_m(\psi \mid y) &= \int \pi(\psi, \lambda \mid y) d\lambda \\
&= \frac{\int \exp\{\ell(\psi, \lambda; y)\pi(\psi, \lambda) d\lambda}{\int \exp\{\ell(\psi, \lambda; y)\pi(\psi, \lambda) d\psi d\lambda}
\end{aligned}$$

... nuisance parameters

$$y = (y_1, \dots, y_n) \sim f(y;\theta), \quad \theta = (\psi, \lambda)$$

$$\pi_m(\psi \mid y) = \int \pi(\psi, \lambda \mid y) d\lambda$$

$$= \frac{\int \exp\{\ell(\psi, \lambda; y)\pi(\psi, \lambda) d\lambda}{\int \exp\{\ell(\psi, \lambda; y)\pi(\psi, \lambda) d\psi d\lambda}$$

$$|j(\hat{\theta})| = |j^{\psi\psi}(\hat{\theta})| \, |j_{\lambda\lambda}(\hat{\theta})|$$

Posterior marginal cdf, $d = 1$

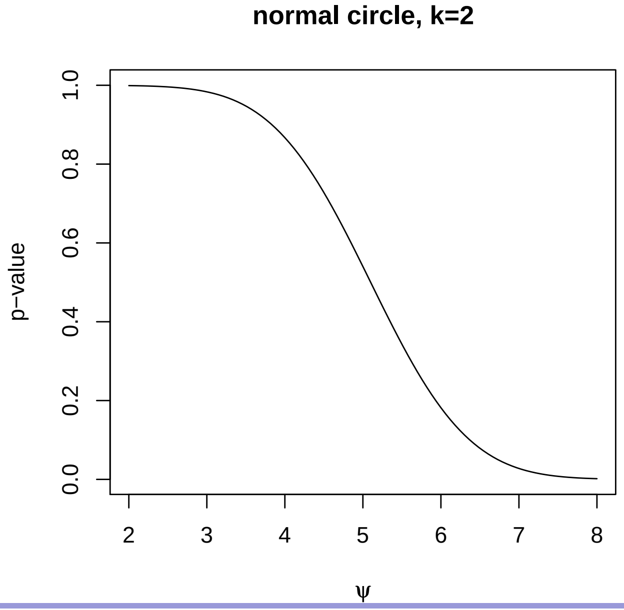
$$
\begin{aligned}
\Pi_m(\psi \mid y) &= \int_{-\infty}^{\psi} \pi_m(\xi \mid y) d\xi \\
&\doteq \int_{-\infty}^{\psi} \frac{1}{(2\pi)^{1/2}} e^{\ell_P(\xi) - \ell_P(\hat{\xi})} j_P^{1/2}(\hat{\xi}) \frac{\pi(\xi, \hat{\lambda}_\xi)}{\pi(\hat{\xi}, \hat{\lambda})} \frac{|j_{\lambda\lambda}(\hat{\xi}, \hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\xi, \hat{\lambda}_\xi)|^{1/2}} d\xi
\end{aligned}
$$

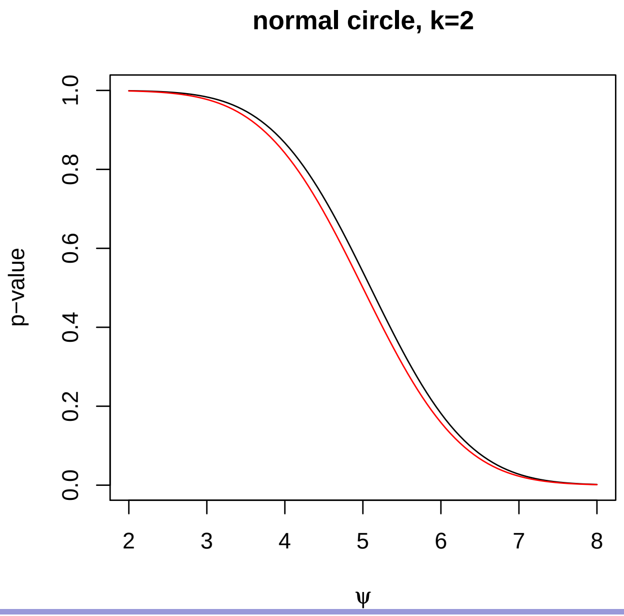
... posterior marginal cdf, $d = 1$

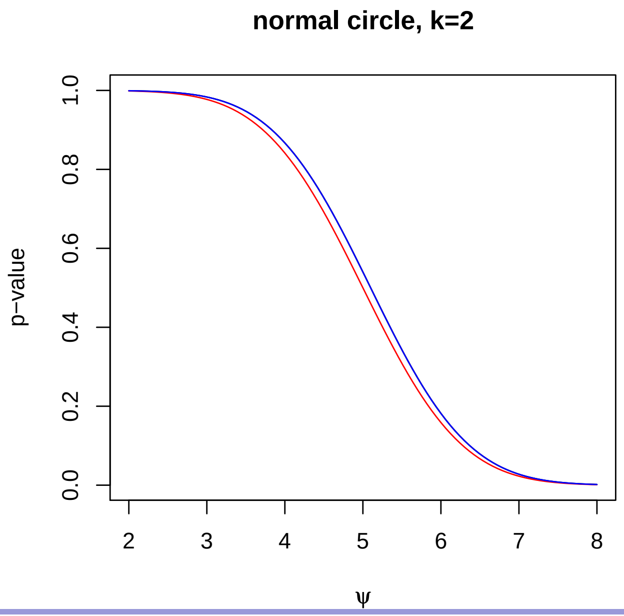
$$\Pi_m(\psi \mid y) \doteq \Phi(r_B^*) = \Phi\{r + \frac{1}{r}\log(\frac{q_B}{r})\}$$

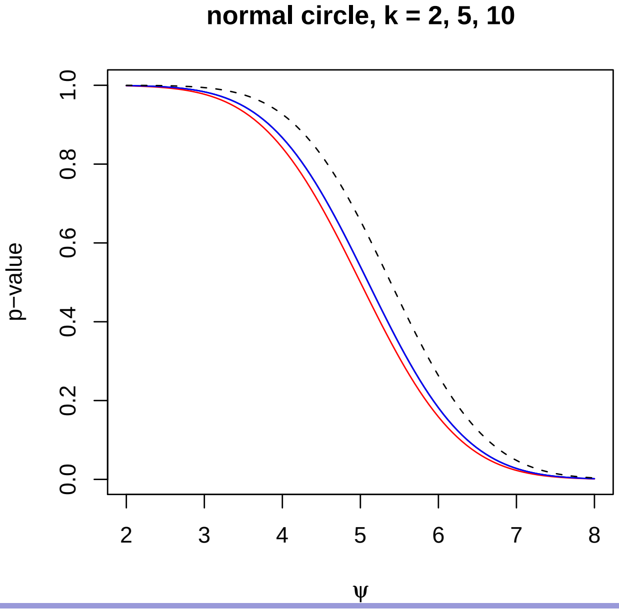
$$r = r(\psi) =$$

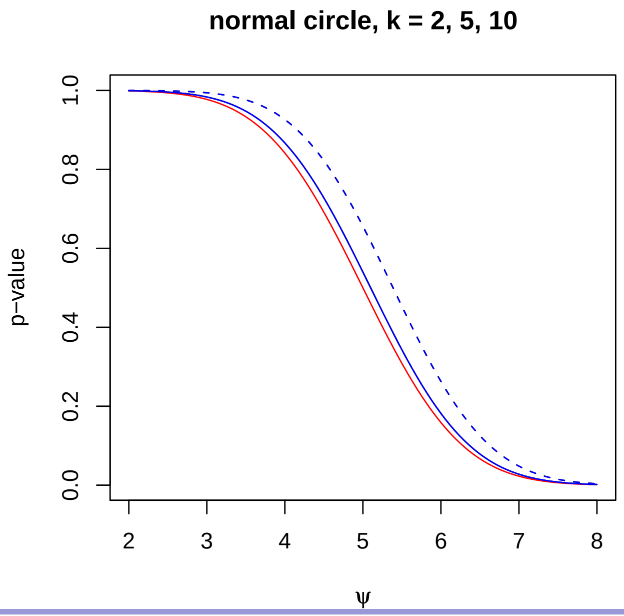
$$q_B = q_B(\psi) =$$

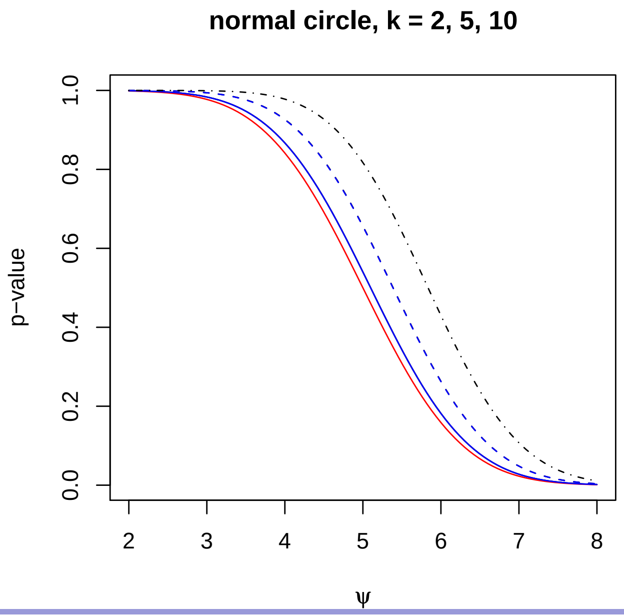

normal circle, k=2
p−value
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
8
ψ

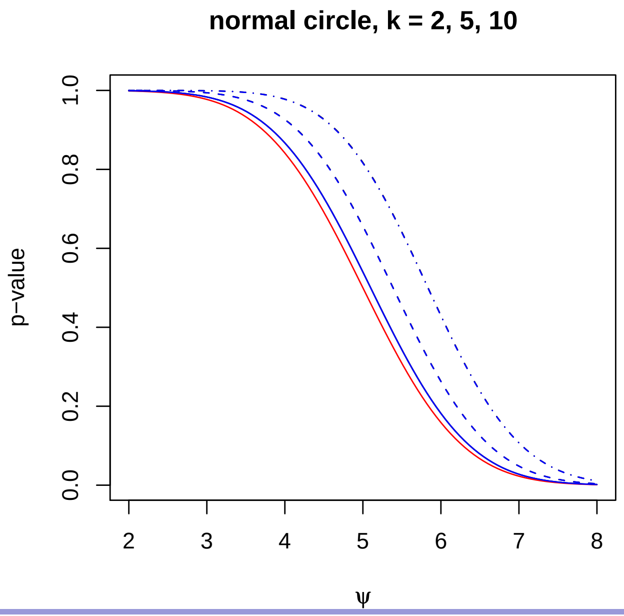

normal circle, k=2
p−value
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
8
ψ


normal circle, k=2
p−value
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
8
ψ


normal circle, k = 2, 5, 10
p−value
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
8
ψ


normal circle, k = 2, 5, 10
p−value
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
8
ψ


normal circle, k = 2, 5, 10
1.0
0.8
0.6
0.4
0.2
0.0
p−value
2
3
4
5
6
7
8
ψ


normal circle, k = 2, 5, 10
p−value
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
8
ψ

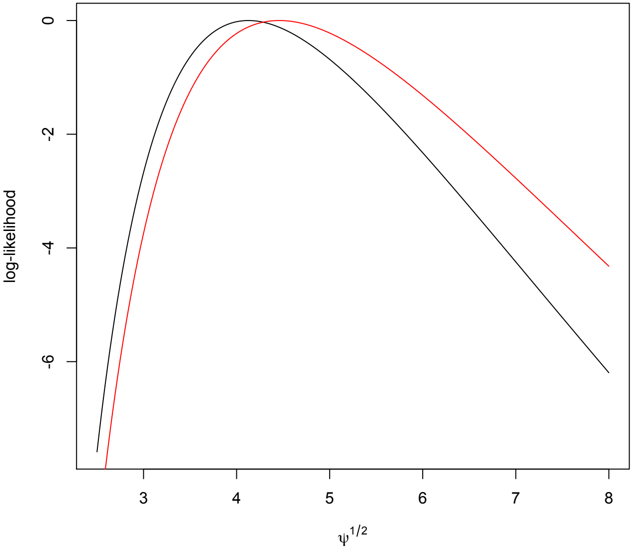
Posterior marginal and adjusted log-likelihoods

$$\pi_m(\psi \mid y) \doteq \frac{1}{(2\pi)^{d/2}} e^{\ell_P(\xi)-\ell_P(\hat{\xi})} j_P^{1/2}(\hat{\xi}) \frac{\pi(\xi, \hat{\lambda}_\xi)}{\pi(\hat{\xi}, \hat{\lambda})} \frac{|j_{\lambda\lambda}(\hat{\xi}, \hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\xi, \hat{\lambda}_\xi)|^{1/2}}$$

$$\Pi_m(\psi \mid y) =$$

Frequentist inference, nuisance parameters

- first-order pivotal quantities
- $r_u(\psi) = \ell_P'(\psi) j_P(\hat{\psi})^{1/2} \dot{\sim} N(0,1)$,
- $r_e(\psi) = (\hat{\psi} - \psi) j_P(\hat{\psi})^{1/2} \dot{\sim} N(0,1)$,
- $r(\psi) = \text{sign}(\hat{\psi} - \psi) 2\{\ell_P(\hat{\psi}) - \ell_P(\psi)\} \dot{\sim} N(0,1)$
- all based on treating profile log-likelihood as a one-parameter log-likelihood
- example $y = X\beta + \epsilon, \quad \epsilon \sim N(0, \psi)$
- $\hat{\psi} = (y - X\hat{\beta})^T (y - X\hat{\beta})/n$



log-likelihood
0
-2
-4
-6
3
4
5
6
7
8
$\psi^{1/2}$

Eliminating nuisance parameters

- by using marginal density
- $f(y; \psi, \lambda) \propto f_m(t_1; \psi) f_c(t_2 \mid t_1; \psi, \lambda)$
- Example
  $N(X\beta, \sigma^2 I)$: $\quad f(y; \beta, \sigma^2) \propto f_m(RSS; \sigma^2) f_c(\hat{\beta} \mid RSS; \beta, \sigma^2)$
- by using conditional density
- $f(y; \psi, \lambda) \propto f_c(t_1 \mid t_2; \psi) f_m(t_2; \psi, \lambda)$
- Example
  $N(X\beta, \sigma^2 I)$: $\quad f(y; \beta, \sigma^2) \propto f_c(RSS \mid \hat{\beta}; \sigma^2) f_m(\hat{\beta}; \beta, \sigma^2)$

Linear exponential families

- conditional density free of nuisance parameter
- $f(y_i; \psi, \lambda) = \exp\{\psi^T s(y_i) + \lambda^T t(y_i) - k(\psi, \lambda)\} h(y_i)$
- $f(y; \psi, \lambda) =$

  $s =$ $t =$

- $f(s, t; \psi, \lambda) =$
- $f(s \mid t; \psi) =$

Saddlepoint approximation in linear exponential families

- no nuisance parameters $f(y_i;\theta) = \exp\{\theta^T s(y_i) - k(\theta)\} h(y_i)$
- $f(s;\theta) = \exp\{\theta^T s - nk(\theta)\}\tilde{h}(s)$
- $\ell(\theta; s) = \theta^T s - nk(\theta)$
- $f(s;\theta) \doteq$
- $f(\hat{\theta};\theta) \doteq$

Saddlepoint approximation to conditional density

- $f(y_i; \psi, \lambda) = \exp\{\psi^T s(y_i) + \lambda^T t(y_i) - k(\psi, \lambda)\} h(y_i)$

- $f(s \mid t; \psi) =$

- $f(\hat{\psi} \mid t; \psi) \doteq c |j_P(\hat{\psi})|^{1/2} e^{\ell_P(\psi) - \ell_P(\hat{\psi})}$

SM §12.3

Approximating distribution function

- $f(\hat{\theta};\theta) \doteq c|j(\hat{\theta})|^{1/2}\exp\{\ell(\theta;\hat{\theta}) - \ell(\hat{\theta};\hat{\theta})\}$

- $\int_{-\infty}^{\hat{\theta}} f(\hat{\vartheta};\theta)d\hat{\vartheta} \doteq$

Summary

- No nuisance parameters
  - Bayesian p-value $\Phi(r_B^*)$
  - $r_B^* = r + \frac{1}{r}\log\frac{q_B}{r}$
  - Exponential family p-value $\Phi(r^*)$
  - $r^* = r + \frac{1}{r}\log\frac{q}{r}$
- Nuisance parameters
  - Bayesian p-value $\Phi(r_B^*)$
  - $r_B^* = r + \frac{1}{r}\log\frac{q_B}{r}$
  - Exponential family p-value $\Phi(r^*)$
  - $r^* = r + \frac{1}{r}\log\frac{q}{r}$

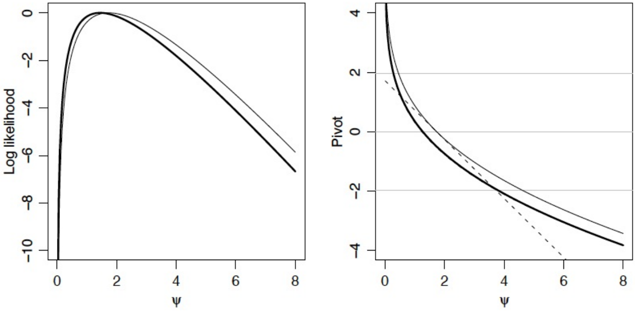


Figure 2.3: Inference for shape parameter ψ of gamma sample of size $n = 5$. Left: profile log likelihood ℓ_{p} (solid) and the log likelihood from the conditional density of u given v (heavy). Right: likelihood root $r(\psi)$ (solid), Wald pivot $t(\psi)$ (dashes), modified likelihood root $r^*(\psi)$ (heavy), and exact pivot overlying $r^*(\psi)$. The horizontal lines are at $0, \pm 1.96$.