Applied asymptotics

Bayesian and frequentist inference

Nancy Reid

April 20, 2008

Bayesian posterior distribution

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

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

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

... Bayesian posterior distribution

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

- $r = \pm[2\{\ell(\hat\theta) - \ell(\hat\theta_\psi)\}]^{1/2}, \quad \psi \in R$ likelihood root
- $\ell(\theta) = \ell(\psi, \lambda) = \log f(y; \psi, \lambda), \quad y \in R^n$ log-likelihood
- $\hat\theta_\psi = (\psi, \hat\lambda_\psi)$ constrained m.l.e.
- $\ell_{\rm p}(\psi) = \ell(\hat\theta_\psi)$ profile log-likelihood
- $j(\theta) = -\ell''(\theta; y)$ observed information
- $j(\theta) = \left[\begin{array}{cc} j_{\psi\psi}(\theta) & j_{\psi\lambda}(\theta) \\ j_{\lambda\psi})\theta) & j_{\lambda\lambda}(\theta) \end{array}\right]$ partitioned matrix

... Bayesian posterior

$$r_B^* \mathrel{\dot\sim} N(0,1)$$

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

r is the likelihood root

q_B is an adjusted score statistic

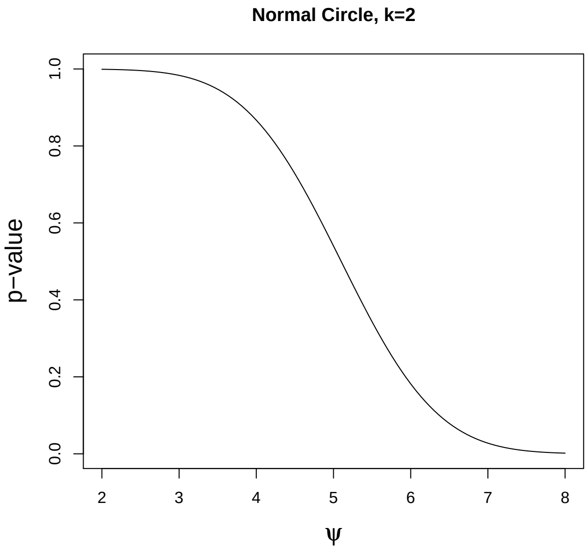
The approximation is very good!

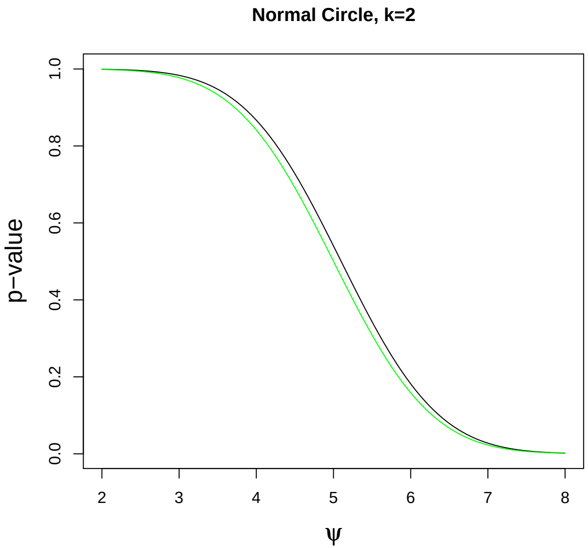
Example: Normal circle

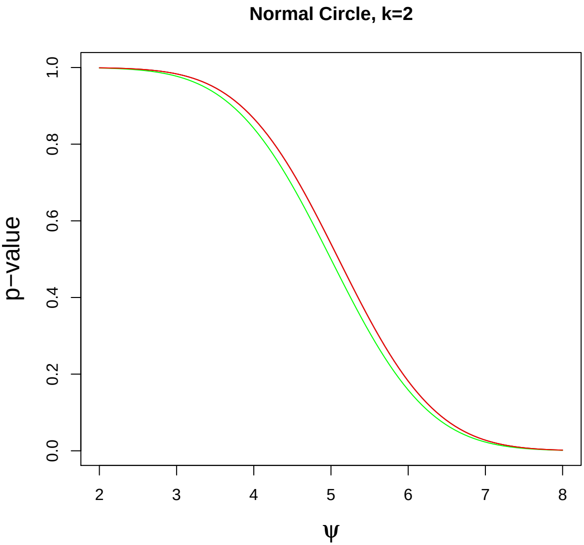
- $y_1 \sim N(\mu_1, 1/n), \ldots, y_k \sim N(\mu_k, 1/n)$
- parameter of interest $\psi = (\mu_1^2 + \cdots + \mu_k^2)^{1/2} = ||\mu||$
- prior $\pi(\mu) = 1$
- Exact marginal posterior $\Pr\{\chi_k^2(n||y||^2) \geq n\psi^2\}$
- Third order

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

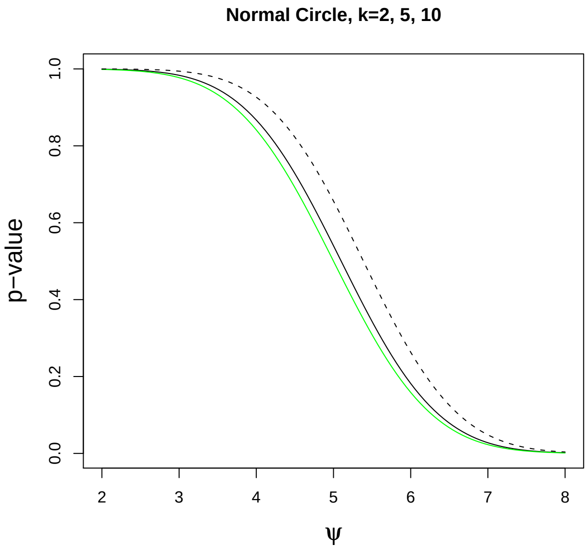
- Normal approximation to posterior $\sqrt{n}(\hat{\psi} - \psi) \dot{\sim} N(0, 1)$

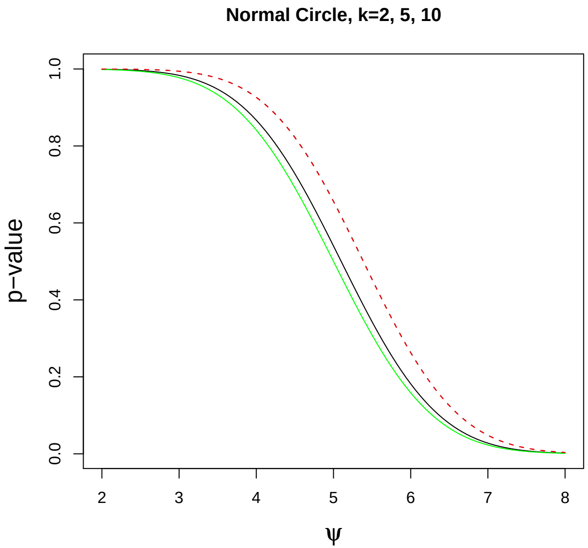

Normal Circle, k=2
p−value
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
8
ψ

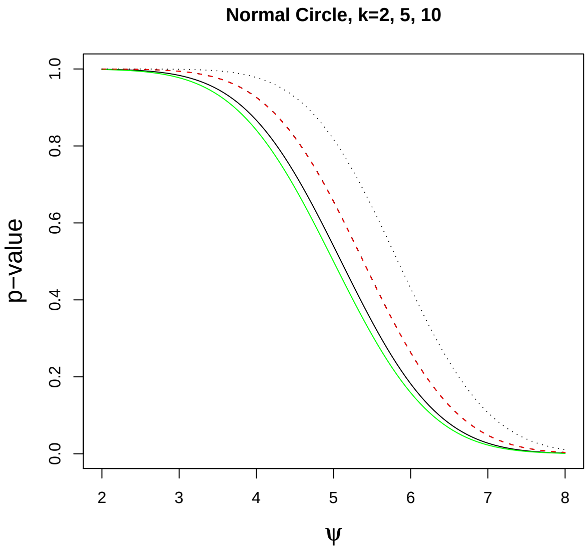

Normal Circle, k=2
p−value
0.0
0.2
0.4
0.6
0.8
1.0
2
3
4
5
6
7
8
ψ

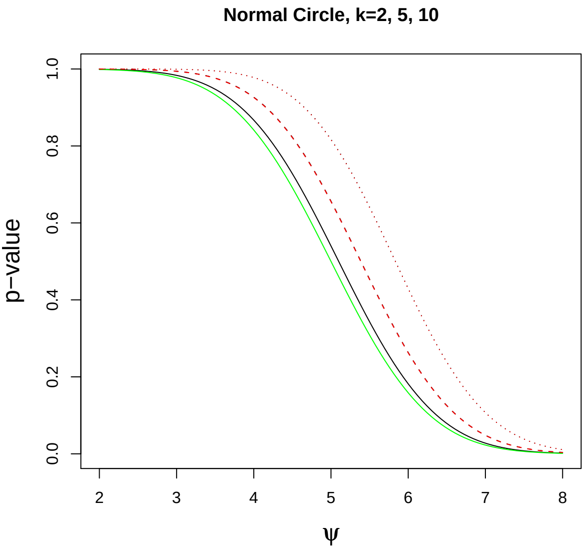


Normal Circle, k=2
p−value
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
8
ψ


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


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


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


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

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

	exact	0.99	0.95	0.75	0.5
$k = 5$		0.9898	0.9495	0.7491	0.4991
$k = 10$	$\Phi(r_B^*)$	0.9897	0.9493	0.7486	0.4987
$k = 20$		0.9899	0.9500	0.7506	0.5012

	exact	0.25	0.05	0.01
$k = 5$		0.2494	0.0499	0.00997
$k = 10$	$\Phi(r_B^*)$	0.2494	0.0498	0.00997
$k = 20$		0.2511	0.0504	0.01012

Frequentist P-value

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

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

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

Normal circle

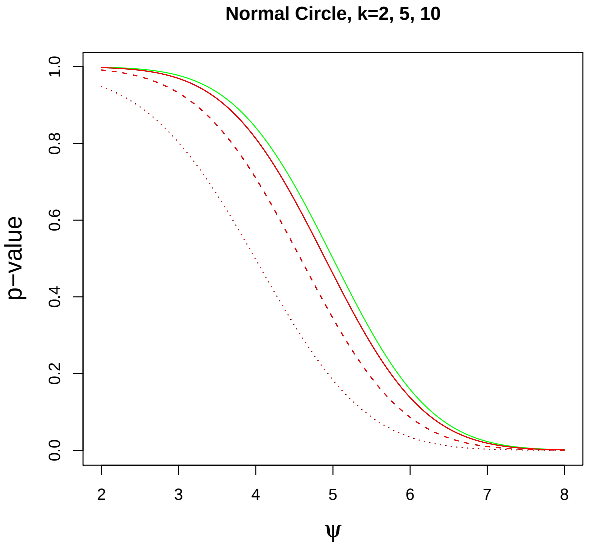
- exact P-value(ψ) $\Pr\{\chi^2_k(n\psi^2) \geq n||y||^2\}$
- approx:

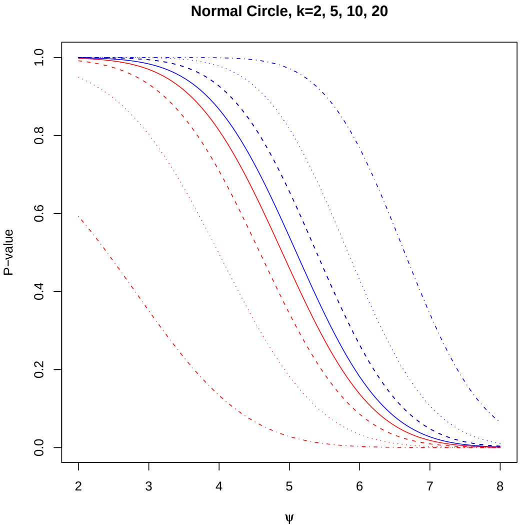
$$r_F^* = \sqrt{n}(\hat{\psi} - \psi) - \frac{1}{\sqrt{n}(\hat{\psi} - \psi)} \log\left\{\left(\frac{\hat{\psi}}{\psi}\right)^{(k-1)/2}\right\} \dot{\sim} N(0,1)$$

- Bayes:

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

- $r_B^* - r_F^* \dot{\sim} \frac{k-1}{\psi\sqrt{n}}$

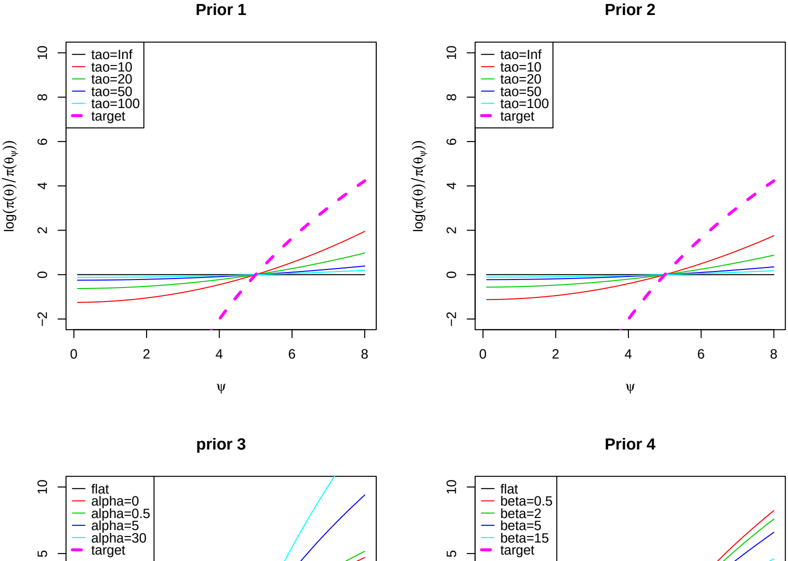

Normal Circle, k=2, 5, 10
p−value
ψ
0.0
0.2
0.4
0.6
0.8
1.0
2
3
4
5
6
7
8

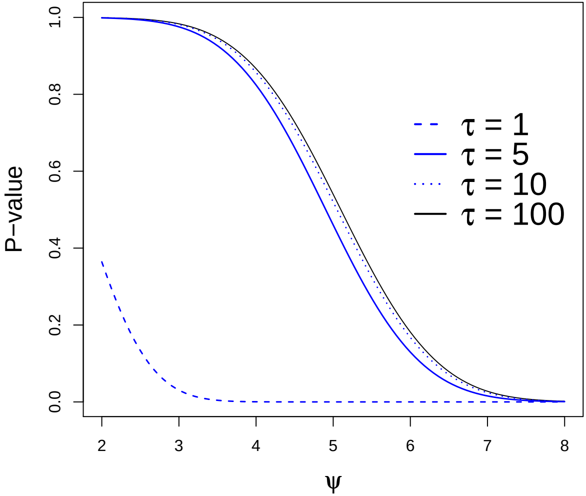

Normal Circle, k=2, 5, 10, 20
P−value
ψ
0.0
0.2
0.4
0.6
0.8
1.0
2
3
4
5
6
7
8

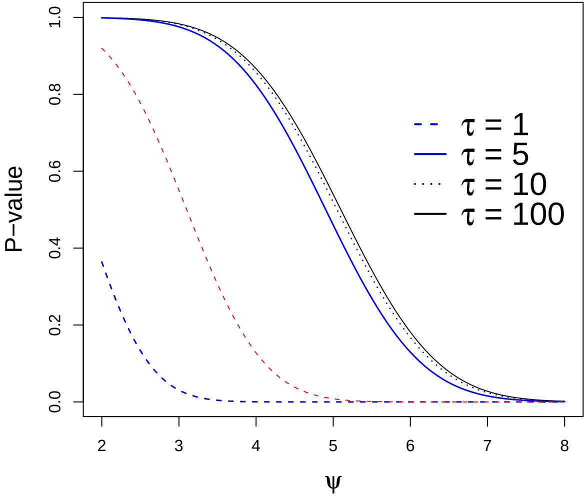
Sensitivity to priors

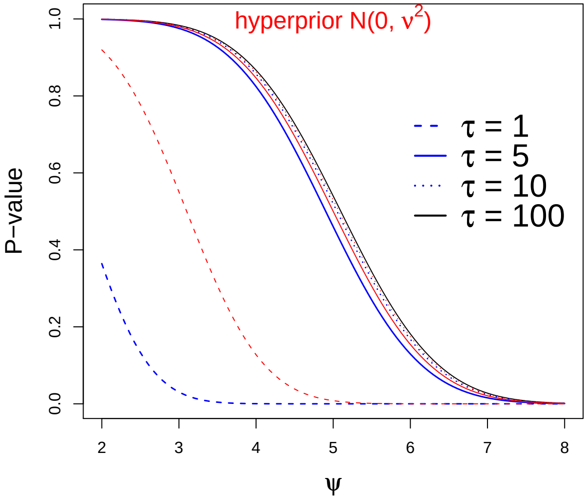
- $Y_i \sim N(\mu_i, 1/n), \quad i = 1, \ldots, k$: flat prior obviously poor
- try some more 'realistic' priors, cheaply
- use r_B^* with different priors
-
$$q_B = -\ell_p'(\psi) j_p(\hat{\psi})^{-1/2} \frac{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta})|^{1/2}} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)}$$
-
$$r_B^* = \sqrt{n}(\hat{\psi} - \psi) + \frac{1}{\sqrt{n}(\hat{\psi} - \psi)} \log \left\{ \left(\frac{\hat{\psi}}{\psi} \right)^{(k-1)/2} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)} \right\}$$
- Prior 1: $\mu_i \sim N(0, \tau^2)$ Prior 2: $\mu_i \sim N(a, \tau^2), a \sim N(0, \nu^2)$
- Prior 3: $\mu_i \sim N(0, \sigma^2), \quad 1/\sigma^2 \sim \Gamma(\alpha, \beta)$
  Prior 4: $\mu_i \sim N(a, \sigma^2), \quad a \sim N(0, \nu^2), \quad 1/\sigma^2 \sim \Gamma(\alpha, \beta)$

go to preview




P−value
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
8
ψ
τ = 1
τ = 5
τ = 10
τ = 100


1.0
0.8
0.6
0.4
0.2
0.0
P−value
2
3
4
5
6
7
8
ψ
τ = 1
τ = 5
τ = 10
τ = 100

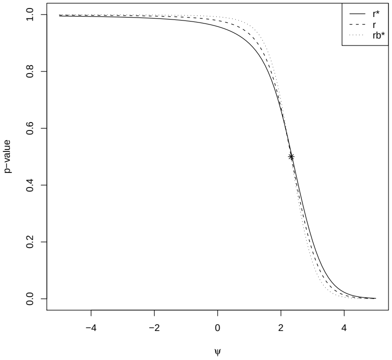

hyperprior N(0, ν²)
τ = 1
τ = 5
τ = 10
τ = 100
P−value
ψ
0.0
0.2
0.4
0.6
0.8
1.0
2
3
4
5
6
7
8

Flat priors for vector parameters

- poor sampling behaviour for normal circle
- prior needs to be targetted 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 ρ zero, except $\rightarrow \infty$ at $\rho = 0, 1$ (improper)

Logistic regression

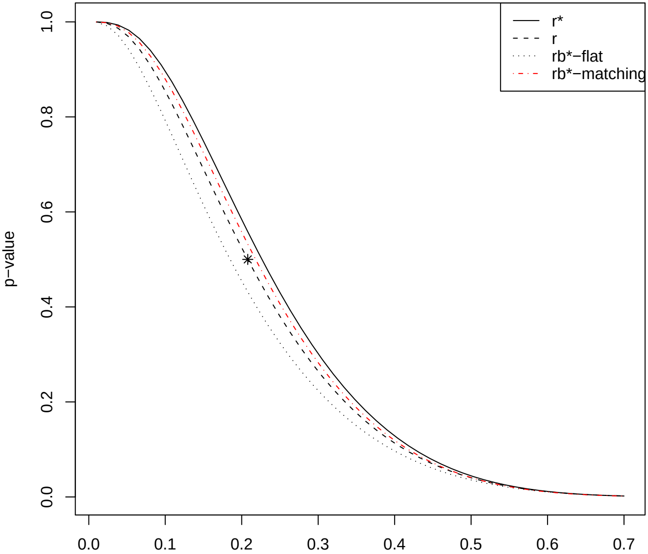
- $Pr(Y_i = 1 \mid x_i) = p_i = \alpha + \beta x_i$
- $Y_i \sim Ber(p_i)$
- $\psi = -\alpha/\beta$

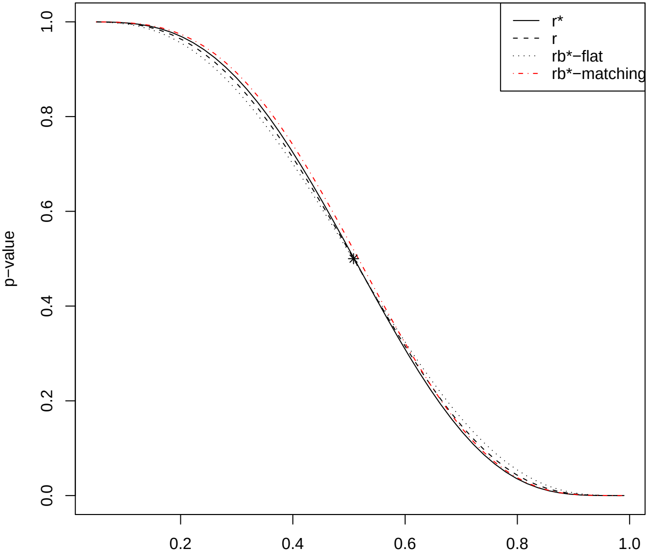


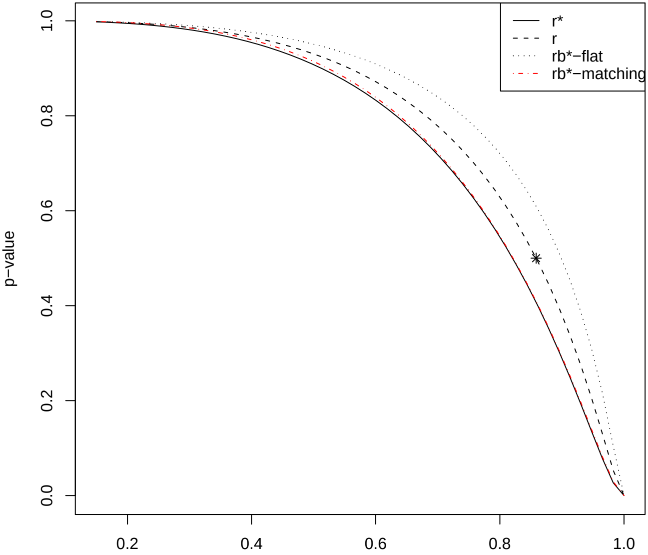
	left	center	right
$\Phi(r)$	0.0711	0.9189	0.0100
$\Phi(r_F^*)$	0.0454	0.9485	0.0061
$\Phi(r_B^*)$	0.0987	0.8864	0.0149

Multiple logistic regression

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


r*
r
rb*−flat
rb*−matching
p−value
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7


r*
r
rb*−flat
rb*−matching
p−value
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.4
0.6
0.8
1.0


r*
r
rb*−flat
rb*−matching
p−value
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.4
0.6
0.8
1.0

	$Pr(y=1 \mid x^*) = 0.86$			$Pr(y=1 \mid x^*) = 0$		
	left	center	right	left	center	
$\Phi(r)$	0.0228	0.9185	0.0587	0.0356	0.9315	0
$\Phi(r_F^*)$	0.0297	0.9460	0.0243	0.0265	0.9515	0
$\Phi(r_{B1}^*)$ (flat)	0.0178	0.8612	0.1210	0.0477	0.9045	0
$\Phi(r_{B2}^*)$ (matching)	0.0275	0.9450	0.0275	0.0252	0.9507	0

	$Pr(y=1 \mid x^*) = 0.21$		
	left	center	right
$\Phi(r)$	0.0535	0.9220	0.0245
$\Phi(r_F^*)$	0.0240	0.9487	0.0273
$\Phi(r_{B1}^*)$ (flat)	0.1027	0.8755	0.0218
$\Phi(r_{B2}^*)$ (matching)	0.0315	0.9419	0.0266

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
- $q_B = -\ell_p'(\psi) j_p(\hat{\psi})^{-1/2} \frac{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta})|^{1/2}} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)}$
- $q_F = \frac{|\ell_{;V}(\hat{\theta}) - \ell_{;V}(\theta) \quad \ell_{\lambda;V}(\hat{\theta}_\psi)|}{|\ell_{\theta;V}(\hat{\theta})|} \frac{|j(\hat{\theta})|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}}$
- $q_B = q_F \Leftrightarrow \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)} = \ldots$
- default prior along the curve $\theta = \hat{\theta}_\psi$ F&R, 2002

Example: normal circle $y_i \sim N(\mu_i, 1)$

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

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

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

Logistic regression

$$\frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)} = \frac{i_{\psi\psi}^{1/2}(\hat{\theta})}{i_{\psi\psi}^{1/2}(\hat{\theta}_\psi)} \qquad \theta = (\psi, \lambda);\ \lambda \perp \psi;\ \psi = \beta_p \text{ or } \psi = p_1(x^*)$$

	$p_1(x^*) = 0.86$			$p_1(x^*) = 0.51$		
	left	center	right	left	center	right
$\Phi(r_F^*)$	0.0297	0.9460	0.0243	0.0265	0.9515	0.0220
$\Phi(r_{B1}^*)$ (flat)	0.0178	0.8612	0.1210	0.0477	0.9045	0.0478
$\Phi(r_{B2}^*)$ (matching)	0.0275	0.9450	0.0275	0.0252	0.9507	0.0241

	$p_1(x^*) = 0.21$		
	left	center	right
$\Phi(r_F^*)$	0.0240	0.9487	0.0273
$\Phi(r_{B1}^*)$ (flat)	0.1027	0.8755	0.0218
$\Phi(r_{B2}^*)$ (matching)	0.0315	0.9419	0.0266

Flat priors for high dimensional parameters are a bad idea

- Another example: Heinrich, J. (2005) Proceedings of Phystat05
- n events from Poisson with rate $\epsilon s + b$, with s of interest and additional Poisson measurements of b and ϵ
- 'flat' priors for s, ϵ, b

Heinrich, 2005

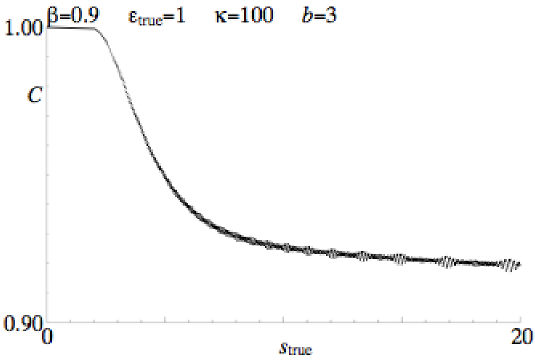


Fig. 2. Typical single channel case. Coverage for 90% credibility level upper limits, acceptance uncertainty = 10%, background uncertainty = zero.

Heinrich, 2005

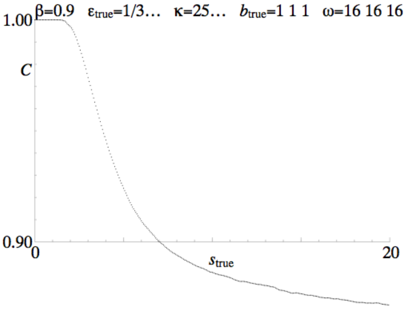


Fig. 6. 3 independent channels. Coverage for 90% credibility level upper limits, acceptance uncertainty = 34%/channel, background uncertainty = 25%/channel.

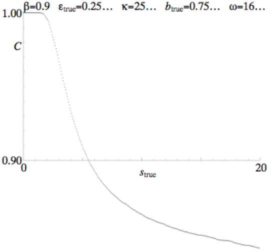


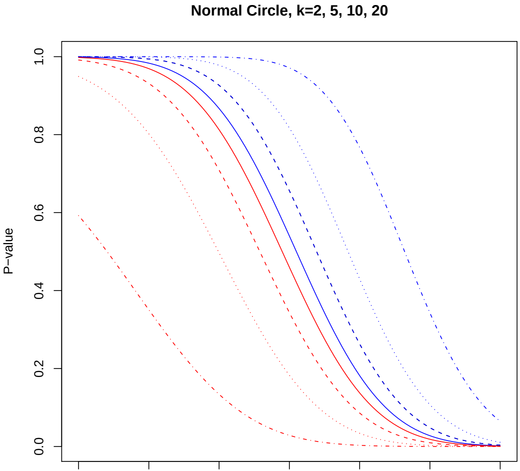
Fig. 7. 4 independent channels. Coverage for 90% credibility level upper limits, acceptance uncertainty = 40%/channel, Background uncertainty = 29%/channel.

... 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 $\mu, \underline{\alpha}, \sigma_\beta, \sigma_\gamma$” (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

- Fraser, D.A.S. and Reid, N. (2002). Strong matching of frequentist and Bayesian inference. *J. Statist. Plan. Infer.* **103**, 263–285.
- Staicu, A.M. and Reid, N. (2007). Uniqueness of matching priors. submitted
- Staicu, A.M. (2007). On some aspects of likelihood methods with applications in biostatistics. PhD Thesis, U Toronto.
- Datta, G.S. and Ghosh, M. (1995). Some remarks on noninformative priors. *J. Amer. Statist. Assoc.* **90**, 1357–1363.
- Dawid, A.P., Stone, M., and Zidek, J.V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc.* B **35**, 189–233.
- Bernardo, J.M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc.* B **41**, 113–147 (with discussion).