Pseudo-Bayes Factors

Stefano Cabras[1], Walter Racugno[1] and Laura Ventura[2]

[1] Department of Mathematics, University of Cagliari
[2] Department of Statistics, University of Padova

Compstat, 2010

Aims

- Consider a sampling model $p(y; \theta)$, with $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^p$, where

Aims

- Consider a sampling model $p(y;\theta)$, with $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^p$, where
  - ψ is a scalar parameter of interest

Aims

- Consider a sampling model $p(y;\theta)$, with $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^p$, where
  - ψ is a scalar parameter of interest
  - λ is a $p-1$ dimensional nuisance parameter.

Aims

- Consider a sampling model $p(y;\theta)$, with $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^p$, where
  - ψ is a scalar parameter of interest
  - λ is a $p-1$ dimensional nuisance parameter.

Aims

- Consider a sampling model $p(y;\theta)$, with $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^p$, where
  - ψ is a scalar parameter of interest
  - λ is a $p-1$ dimensional nuisance parameter.

We are interested in testing $H_0 : \psi \in \Psi_0$ against $H_1 : \psi \in \Psi_1$

Aims

- Consider a sampling model $p(y;\theta)$, with $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^p$, where
  - ψ is a scalar parameter of interest
  - λ is a $p-1$ dimensional nuisance parameter.

We are interested in testing $H_0 : \psi \in \Psi_0$ against $H_1 : \psi \in \Psi_1$

- Classical approach: use Bayes Factor (BF) based on the ratio of integrated likelihoods. It requires

Aims

- Consider a sampling model $p(y;\theta)$, with $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^p$, where
  - ψ is a scalar parameter of interest
  - λ is a $p-1$ dimensional nuisance parameter.

We are interested in testing $H_0 : \psi \in \Psi_0$ against $H_1 : \psi \in \Psi_1$

- Classical approach: use Bayes Factor (BF) based on the ratio of integrated likelihoods. It requires
  - prior elicitation on λ;

Aims

- Consider a sampling model $p(y;\theta)$, with $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^p$, where
  - ψ is a scalar parameter of interest
  - λ is a $p-1$ dimensional nuisance parameter.

We are interested in testing $H_0 : \psi \in \Psi_0$ against $H_1 : \psi \in \Psi_1$

- Classical approach: use Bayes Factor (BF) based on the ratio of integrated likelihoods. It requires
  - prior elicitation on λ;
  - calculation on a p-dimensional integral

Aims

- Consider a sampling model $p(y;\theta)$, with $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^p$, where
  - ψ is a scalar parameter of interest
  - λ is a $p-1$ dimensional nuisance parameter.

We are interested in testing $H_0 : \psi \in \Psi_0$ against $H_1 : \psi \in \Psi_1$

- Classical approach: use Bayes Factor (BF) based on the ratio of integrated likelihoods. It requires
  - prior elicitation on λ;
  - calculation on a p-dimensional integral
- Proposed approach: use a pseudo-BF based on a pseudo-likelihood $L^*(\psi)$, which is a function of ψ only.

Classical Bayesian Hypothesis Testing (Kass, Raftery, 1995)

- Test

$$H_0 : \psi \in \Psi_0 \textit{ against } H_1 : \psi \in \Psi_1,$$

with the Bayes Factor

$$BF = \frac{\int_{\Psi_0} \int_{\Lambda} L(\psi, \lambda) \pi_0(\lambda|\psi) \pi_0(\psi) \, d\lambda \, d\psi}{\int_{\Psi_1} \int_{\Lambda} L(\psi, \lambda) \pi_1(\lambda|\psi) \pi_1(\psi) \, d\lambda \, d\psi},$$

where

Classical Bayesian Hypothesis Testing (Kass, Raftery, 1995)

- Test

$$H_0 : \psi \in \Psi_0 \textit{ against } H_1 : \psi \in \Psi_1,$$

with the Bayes Factor

$$BF = \frac{\int_{\Psi_0} \int_{\Lambda} L(\psi, \lambda) \pi_0(\lambda|\psi) \pi_0(\psi)\, d\lambda\, d\psi}{\int_{\Psi_1} \int_{\Lambda} L(\psi, \lambda) \pi_1(\lambda|\psi) \pi_1(\psi)\, d\lambda\, d\psi},$$

where

  - $L(\psi, \lambda) = L(\psi, \lambda; \boldsymbol{y})$ is the full likelihood based on data $\boldsymbol{y}$;

Classical Bayesian Hypothesis Testing (Kass, Raftery, 1995)

- Test

$$H_0 : \psi \in \Psi_0 \textit{ against } H_1 : \psi \in \Psi_1,$$

with the Bayes Factor

$$BF = \frac{\int_{\Psi_0} \int_{\Lambda} L(\psi, \lambda)\pi_0(\lambda|\psi)\pi_0(\psi)\, d\lambda\, d\psi}{\int_{\Psi_1} \int_{\Lambda} L(\psi, \lambda)\pi_1(\lambda|\psi)\pi_1(\psi)\, d\lambda\, d\psi},$$

where

  - $L(\psi, \lambda) = L(\psi, \lambda; \boldsymbol{y})$ is the full likelihood based on data $\boldsymbol{y}$;
  - $\pi_k(\psi)$ for $\psi \in \Psi_k$, $k = 0, 1$ are priors under H_0 and H_1 respectively;

Classical Bayesian Hypothesis Testing (Kass, Raftery, 1995)

- Test

$$H_0 : \psi \in \Psi_0 \textit{ against } H_1 : \psi \in \Psi_1,$$

with the Bayes Factor

$$BF = \frac{\int_{\Psi_0} \int_{\Lambda} L(\psi, \lambda)\pi_0(\lambda|\psi)\pi_0(\psi)\, d\lambda\, d\psi}{\int_{\Psi_1} \int_{\Lambda} L(\psi, \lambda)\pi_1(\lambda|\psi)\pi_1(\psi)\, d\lambda\, d\psi},$$

where

  - $L(\psi, \lambda) = L(\psi, \lambda; \boldsymbol{y})$ is the full likelihood based on data $\boldsymbol{y}$;
  - $\pi_k(\psi)$ for $\psi \in \Psi_k$, $k = 0, 1$ are priors under H_0 and H_1 respectively;
  - $\pi_k(\lambda|\psi)$ for $\psi \in \Psi_k$ are priors on the nuisance parameter given ψ;

Classical Bayesian Hypothesis Testing (Kass, Raftery, 1995)

- Test

$$H_0 : \psi \in \Psi_0 \textit{ against } H_1 : \psi \in \Psi_1,$$

with the Bayes Factor

$$BF = \frac{\int_{\Psi_0} \int_{\Lambda} L(\psi, \lambda)\pi_0(\lambda|\psi)\pi_0(\psi)\, d\lambda\, d\psi}{\int_{\Psi_1} \int_{\Lambda} L(\psi, \lambda)\pi_1(\lambda|\psi)\pi_1(\psi)\, d\lambda\, d\psi},$$

where

  - $L(\psi, \lambda) = L(\psi, \lambda; \boldsymbol{y})$ is the full likelihood based on data $\boldsymbol{y}$;
  - $\pi_k(\psi)$ for $\psi \in \Psi_k$, $k = 0, 1$ are priors under H_0 and H_1 respectively;
  - $\pi_k(\lambda|\psi)$ for $\psi \in \Psi_k$ are priors on the nuisance parameter given ψ;
- This approach

Classical Bayesian Hypothesis Testing (Kass, Raftery, 1995)

- Test

$$H_0 : \psi \in \Psi_0 \textit{ against } H_1 : \psi \in \Psi_1,$$

with the Bayes Factor

$$BF = \frac{\int_{\Psi_0} \int_\Lambda L(\psi, \lambda)\pi_0(\lambda|\psi)\pi_0(\psi)\, d\lambda\, d\psi}{\int_{\Psi_1} \int_\Lambda L(\psi, \lambda)\pi_1(\lambda|\psi)\pi_1(\psi)\, d\lambda\, d\psi},$$

where

  - $L(\psi, \lambda) = L(\psi, \lambda; \boldsymbol{y})$ is the full likelihood based on data $\boldsymbol{y}$;
  - $\pi_k(\psi)$ for $\psi \in \Psi_k$, $k = 0, 1$ are priors under H_0 and H_1 respectively;
  - $\pi_k(\lambda|\psi)$ for $\psi \in \Psi_k$ are priors on the nuisance parameter given ψ;
- This approach
  - requires elicitation of priors $\pi_k(\lambda|\psi)$

Classical Bayesian Hypothesis Testing (Kass, Raftery, 1995)

- Test

$$H_0 : \psi \in \Psi_0 \textit{ against } H_1 : \psi \in \Psi_1,$$

with the Bayes Factor

$$BF = \frac{\int_{\Psi_0} \int_{\Lambda} L(\psi, \lambda)\pi_0(\lambda|\psi)\pi_0(\psi)\, d\lambda\, d\psi}{\int_{\Psi_1} \int_{\Lambda} L(\psi, \lambda)\pi_1(\lambda|\psi)\pi_1(\psi)\, d\lambda\, d\psi},$$

where

  - $L(\psi, \lambda) = L(\psi, \lambda; \boldsymbol{y})$ is the full likelihood based on data $\boldsymbol{y}$;
  - $\pi_k(\psi)$ for $\psi \in \Psi_k$, $k = 0, 1$ are priors under H_0 and H_1 respectively;
  - $\pi_k(\lambda|\psi)$ for $\psi \in \Psi_k$ are priors on the nuisance parameter given ψ;
- This approach
  - requires elicitation of priors $\pi_k(\lambda|\psi)$

Classical Bayesian Hypothesis Testing (Kass, Raftery, 1995)

- Test

$$H_0 : \psi \in \Psi_0 \textit{ against } H_1 : \psi \in \Psi_1,$$

with the Bayes Factor

$$BF = \frac{\int_{\Psi_0} \int_{\Lambda} L(\psi, \lambda)\pi_0(\lambda|\psi)\pi_0(\psi)\, d\lambda\, d\psi}{\int_{\Psi_1} \int_{\Lambda} L(\psi, \lambda)\pi_1(\lambda|\psi)\pi_1(\psi)\, d\lambda\, d\psi},$$

where

  - $L(\psi, \lambda) = L(\psi, \lambda; \boldsymbol{y})$ is the full likelihood based on data $\boldsymbol{y}$;
  - $\pi_k(\psi)$ for $\psi \in \Psi_k$, $k = 0, 1$ are priors under H_0 and H_1 respectively;
  - $\pi_k(\lambda|\psi)$ for $\psi \in \Psi_k$ are priors on the nuisance parameter given ψ;

- This approach
  - requires elicitation of priors $\pi_k(\lambda|\psi)$
    ... critical when p is large and/or λ has not physical meaning;

Classical Bayesian Hypothesis Testing (Kass, Raftery, 1995)

- Test

$$H_0 : \psi \in \Psi_0 \textit{ against } H_1 : \psi \in \Psi_1,$$

with the Bayes Factor

$$BF = \frac{\int_{\Psi_0} \int_{\Lambda} L(\psi, \lambda)\pi_0(\lambda|\psi)\pi_0(\psi)\, d\lambda\, d\psi}{\int_{\Psi_1} \int_{\Lambda} L(\psi, \lambda)\pi_1(\lambda|\psi)\pi_1(\psi)\, d\lambda\, d\psi},$$

where

  - $L(\psi, \lambda) = L(\psi, \lambda; \boldsymbol{y})$ is the full likelihood based on data $\boldsymbol{y}$;
  - $\pi_k(\psi)$ for $\psi \in \Psi_k$, $k = 0, 1$ are priors under H_0 and H_1 respectively;
  - $\pi_k(\lambda|\psi)$ for $\psi \in \Psi_k$ are priors on the nuisance parameter given ψ;
- This approach
  - requires elicitation of priors $\pi_k(\lambda|\psi)$
    ... critical when p is large and/or λ has not physical meaning;
  - needs p dimensional integrations on (Ψ_k, Λ).

Pseudo likelihoods (Pace, Salvan, 1997; Severini, 2000) in Bayesian inference (Severini, 1999; Ventura *et al.*, 2009; 2010)

- Nuisance parameters λ are eliminated using a pseudo-likelihood $L^*(\psi)$.

Pseudo likelihoods (Pace, Salvan, 1997; Severini, 2000) in Bayesian inference (Severini, 1999; Ventura *et al.*, 2009; 2010)

- Nuisance parameters λ are eliminated using a pseudo-likelihood $L^*(\psi)$.
- Properties of $L^*(\psi)$ are *similar* to those of a genuine likelihood function.

Pseudo likelihoods (Pace, Salvan, 1997; Severini, 2000) in Bayesian inference (Severini, 1999; Ventura *et al.*, 2009; 2010)

- Nuisance parameters λ are eliminated using a pseudo-likelihood $L^*(\psi)$.
- Properties of $L^*(\psi)$ are *similar* to those of a genuine likelihood function.
- Examples of $L^*(\psi)$ are the conditional, marginal, profile and the Modified Profile likelihoods (MPL)

Pseudo likelihoods (Pace, Salvan, 1997; Severini, 2000) in Bayesian inference (Severini, 1999; Ventura *et al.*, 2009; 2010)

- Nuisance parameters λ are eliminated using a pseudo-likelihood $L^*(\psi)$.
- Properties of $L^*(\psi)$ are *similar* to those of a genuine likelihood function.
- Examples of $L^*(\psi)$ are the conditional, marginal, profile and the Modified Profile likelihoods (MPL)
- Using $L^*(\psi)$ as a true likelihood, a posterior distribution of ψ can be considered

$$\pi^*(\psi|\boldsymbol{y}) \propto L^*(\psi)\pi(\psi),$$

where $\pi(\psi)$ is a suitable prior on ψ only.

Pseudo likelihoods (Pace, Salvan, 1997; Severini, 2000) in Bayesian inference (Severini, 1999; Ventura *et al.*, 2009; 2010)

- Nuisance parameters λ are eliminated using a pseudo-likelihood $L^*(\psi)$.
- Properties of $L^*(\psi)$ are *similar* to those of a genuine likelihood function.
- Examples of $L^*(\psi)$ are the conditional, marginal, profile and the Modified Profile likelihoods (MPL)
- Using $L^*(\psi)$ as a true likelihood, a posterior distribution of ψ can be considered

$$\pi^*(\psi|\boldsymbol{y}) \propto L^*(\psi)\pi(\psi),$$

  where $\pi(\psi)$ is a suitable prior on ψ only.
- Advantages in using $\pi^*(\psi|\boldsymbol{y})$:

Pseudo likelihoods (Pace, Salvan, 1997; Severini, 2000) in Bayesian inference (Severini, 1999; Ventura *et al.*, 2009; 2010)

- Nuisance parameters λ are eliminated using a pseudo-likelihood $L^*(\psi)$.
- Properties of $L^*(\psi)$ are *similar* to those of a genuine likelihood function.
- Examples of $L^*(\psi)$ are the conditional, marginal, profile and the Modified Profile likelihoods (MPL)
- Using $L^*(\psi)$ as a true likelihood, a posterior distribution of ψ can be considered

$$\pi^*(\psi|\boldsymbol{y}) \propto L^*(\psi)\pi(\psi),$$

  where $\pi(\psi)$ is a suitable prior on ψ only.
- Advantages in using $\pi^*(\psi|\boldsymbol{y})$:
  - no elicitation on λ;

Pseudo likelihoods (Pace, Salvan, 1997; Severini, 2000) in Bayesian inference (Severini, 1999; Ventura *et al.*, 2009; 2010)

- Nuisance parameters λ are eliminated using a pseudo-likelihood $L^*(\psi)$.
- Properties of $L^*(\psi)$ are *similar* to those of a genuine likelihood function.
- Examples of $L^*(\psi)$ are the conditional, marginal, profile and the Modified Profile likelihoods (MPL)
- Using $L^*(\psi)$ as a true likelihood, a posterior distribution of ψ can be considered

$$\pi^*(\psi|\boldsymbol{y}) \propto L^*(\psi)\pi(\psi),$$

  where $\pi(\psi)$ is a suitable prior on ψ only.
- Advantages in using $\pi^*(\psi|\boldsymbol{y})$:
  - no elicitation on λ;

Pseudo likelihoods (Pace, Salvan, 1997; Severini, 2000) in Bayesian inference (Severini, 1999; Ventura *et al.*, 2009; 2010)

- Nuisance parameters λ are eliminated using a pseudo-likelihood $L^*(\psi)$.
- Properties of $L^*(\psi)$ are *similar* to those of a genuine likelihood function.
- Examples of $L^*(\psi)$ are the conditional, marginal, profile and the Modified Profile likelihoods (MPL)
- Using $L^*(\psi)$ as a true likelihood, a posterior distribution of ψ can be considered

$$\pi^*(\psi|\boldsymbol{y}) \propto L^*(\psi)\pi(\psi),$$

  where $\pi(\psi)$ is a suitable prior on ψ only.
- Advantages in using $\pi^*(\psi|\boldsymbol{y})$:
  - no elicitation on λ;
  - no multi-dimensional integration over Λ.

Pseudo-Bayes Factors BF^*

- We use $\pi^*(\psi|\boldsymbol{y})$ in order to test $H_0 : \psi \in \Psi_0$ *against* $H_1 : \psi \in \Psi_1$.

- We use $\pi^*(\psi|\boldsymbol{y})$ in order to test $H_0 : \psi \in \Psi_0$ *against* $H_1 : \psi \in \Psi_1$.
- In our approach we only need:

Pseudo-Bayes Factors BF^*

- We use $\pi^*(\psi|\boldsymbol{y})$ in order to test $H_0 : \psi \in \Psi_0$ *against* $H_1 : \psi \in \Psi_1$.
- In our approach we only need:
  - a pseudo-likelihood $L^*(\psi)$

- We use $\pi^*(\psi|\boldsymbol{y})$ in order to test $H_0 : \psi \in \Psi_0$ *against* $H_1 : \psi \in \Psi_1$.
- In our approach we only need:
  - a pseudo-likelihood $L^*(\psi)$
  - priors $\pi_k(\psi)$ for $\psi \in \Psi_k$, $k = 0, 1$

- We use $\pi^*(\psi|\boldsymbol{y})$ in order to test $H_0 : \psi \in \Psi_0$ *against* $H_1 : \psi \in \Psi_1$.
- In our approach we only need:
  - a pseudo-likelihood $L^*(\psi)$
  - priors $\pi_k(\psi)$ for $\psi \in \Psi_k$, $k = 0, 1$
- We then define the Pseudo-Bayes Factor:

$$BF^* = \frac{\int_{\Psi_0} L^*(\psi)\pi_0(\psi)\, d\psi}{\int_{\Psi_1} L^*(\psi)\pi_1(\psi)\, d\psi} = \frac{\int_{\Psi_0} \pi_0^*(\psi|\boldsymbol{y})\, d\psi}{\int_{\Psi_1} \pi_1^*(\psi|\boldsymbol{y})\, d\psi},$$

which needs only 1 dimensional integrations over Ψ_k.

Comparison of BF^* versus BF

- The null model H_0 is favored when $BF^* > 1$ o ($BF > 1$).

Comparison of BF^* versus BF

- The null model H_0 is favored when $BF^* > 1$ o ($BF > 1$).
- To evaluate the behavior of BF^* we compare it versus BF in terms of the corresponding the Frequentist Risks, $R^*(\psi, \lambda)$ and $R(\psi, \lambda)$, where

$$
\begin{aligned}
R^*(\psi, \lambda) &= \Pr(BF^*(\boldsymbol{y}) < 1 | H_0) + \Pr(BF^*(\boldsymbol{y}) > 1 | H_1) \\
R(\psi, \lambda) &= \Pr(BF(\boldsymbol{y}) < 1 | H_0) + \Pr(BF(\boldsymbol{y}) > 1 | H_1)
\end{aligned}
$$

Comparison of BF^* versus BF

- The null model H_0 is favored when $BF^* > 1$ o ($BF > 1$).
- To evaluate the behavior of BF^* we compare it versus BF in terms of the corresponding the Frequentist Risks, $R^*(\psi, \lambda)$ and $R(\psi, \lambda)$, where

$$\begin{aligned} R^*(\psi,\lambda) &= \Pr(BF^*(\boldsymbol{y}) < 1|H_0) + \Pr(BF^*(\boldsymbol{y}) > 1|H_1) \\ R(\psi,\lambda) &= \Pr(BF(\boldsymbol{y}) < 1|H_0) + \Pr(BF(\boldsymbol{y}) > 1|H_1) \end{aligned}$$

- If $R^*(\psi, \lambda) < R(\psi, \lambda)$ then, at point (ψ, λ), BF^* outperforms BF.

Comparison of BF^* versus BF

- The null model H_0 is favored when $BF^* > 1$ o ($BF > 1$).
- To evaluate the behavior of BF^* we compare it versus BF in terms of the corresponding the Frequentist Risks, $R^*(\psi, \lambda)$ and $R(\psi, \lambda)$, where

$$\begin{aligned} R^*(\psi,\lambda) &= \Pr(BF^*(\boldsymbol{y}) < 1|H_0) + \Pr(BF^*(\boldsymbol{y}) > 1|H_1) \\ R(\psi,\lambda) &= \Pr(BF(\boldsymbol{y}) < 1|H_0) + \Pr(BF(\boldsymbol{y}) > 1|H_1) \end{aligned}$$

- If $R^*(\psi, \lambda) < R(\psi, \lambda)$ then, at point (ψ, λ), BF^* outperforms BF.
- In the next examples we assume a favorable scenario for BF: $\pi_k(\lambda|\psi) = \pi(\lambda)$ for $k = 0, 1$, with $\pi(\lambda)$ concentrated on the true values.

Comparison of BF^* versus BF

- The null model H_0 is favored when $BF^* > 1$ o ($BF > 1$).
- To evaluate the behavior of BF^* we compare it versus BF in terms of the corresponding the Frequentist Risks, $R^*(\psi, \lambda)$ and $R(\psi, \lambda)$, where

$$\begin{aligned} R^*(\psi,\lambda) &= \Pr(BF^*(\boldsymbol{y}) < 1|H_0) + \Pr(BF^*(\boldsymbol{y}) > 1|H_1) \\ R(\psi,\lambda) &= \Pr(BF(\boldsymbol{y}) < 1|H_0) + \Pr(BF(\boldsymbol{y}) > 1|H_1) \end{aligned}$$

- If $R^*(\psi,\lambda) < R(\psi,\lambda)$ then, at point (ψ,λ), BF^* outperforms BF.
- In the next examples we assume a favorable scenario for BF: $\pi_k(\lambda|\psi) = \pi(\lambda)$ for $k = 0, 1$, with $\pi(\lambda)$ concentrated on the true values.
- Both $R^*(\psi,\lambda)$ and $R(\psi,\lambda)$ are approximated by Monte Carlo simulations.

Examples

- We discuss two examples:

Examples

- We discuss two examples:
  1. stress-strength model;

Examples

- We discuss two examples:
  1. stress-strength model;
  2. logistic regression.

Examples

- We discuss two examples:
  1. stress-strength model;
  2. logistic regression.
- We mainly focus on the Modified Profile Likelihood (Severini, 2000)

$$L_{mp}(\psi) = L_p(\psi) \frac{|j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)|^{1/2}}{|I(\psi, \hat{\boldsymbol{\lambda}}_\psi; \hat{\boldsymbol{\theta}})|} ,$$

where $L_p(\psi) = L(\psi, \hat{\boldsymbol{\lambda}}_\psi)$, $\hat{\boldsymbol{\theta}}$ is the MLE for $\boldsymbol{\theta}$, $\hat{\boldsymbol{\lambda}}_\psi$ is the conditional MLE of $\boldsymbol{\lambda}$ and

$$I(\psi, \boldsymbol{\lambda}; \boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0}(\ell_{\boldsymbol{\lambda}}(\psi, \boldsymbol{\lambda})\ell_{\boldsymbol{\lambda}}(\psi_0, \boldsymbol{\lambda}_0)^T) ,$$

with $\boldsymbol{\theta}_0 = (\psi_0, \boldsymbol{\lambda}_0)$ and $\ell_{\boldsymbol{\lambda}}(\psi, \boldsymbol{\lambda}) = \partial\ell(\psi, \boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$.

Example 1. Stress-strength model (Kotz *et al.*, 2003)

- Let $X \sim Exp(\alpha)$ and $Y \sim Exp(\beta)$ be random variables. Interest is on
$$\psi = \Pr(X < Y) = \alpha/(\alpha + \beta)$$
with $\lambda = \alpha$.

Example 1. Stress-strength model (Kotz *et al.*, 2003)

- Let $X \sim Exp(\alpha)$ and $Y \sim Exp(\beta)$ be random variables. Interest is on

$$\psi = \Pr(X < Y) = \alpha/(\alpha + \beta)$$

with $\lambda = \alpha$.

- Given random samples of size n and m from from X and Y respectively, we test $H_0 : \psi < 1/2$ against $H_1 : \psi > 1/2$, $\forall \lambda > 0$ assuming the following priors

Example 1. Stress-strength model (Kotz *et al.*, 2003)

- Let $X \sim Exp(\alpha)$ and $Y \sim Exp(\beta)$ be random variables. Interest is on
$$\psi = \Pr(X < Y) = \alpha/(\alpha + \beta)$$
with $\lambda = \alpha$.
- Given random samples of size *n* and *m* from from *X* and *Y* respectively, we test $H_0 : \psi < 1/2$ against $H_1 : \psi > 1/2, \forall \lambda > 0$ assuming the following priors
  - $\pi_0(\psi) = U(0, 1/2)$

Example 1. Stress-strength model (Kotz *et al.*, 2003)

- Let $X \sim Exp(\alpha)$ and $Y \sim Exp(\beta)$ be random variables. Interest is on
$$\psi = \Pr(X < Y) = \alpha/(\alpha + \beta)$$
with $\lambda = \alpha$.
- Given random samples of size n and m from from X and Y respectively, we test $H_0 : \psi < 1/2$ against $H_1 : \psi > 1/2, \forall\lambda > 0$ assuming the following priors
  - $\pi_0(\psi) = U(0, 1/2)$
  - $\pi_1(\psi) = U(1/2, 1)$

Example 1. Stress-strength model (Kotz *et al.*, 2003)

- Let $X \sim Exp(\alpha)$ and $Y \sim Exp(\beta)$ be random variables. Interest is on
$$\psi = \Pr(X < Y) = \alpha/(\alpha + \beta)$$
with $\lambda = \alpha$.
- Given random samples of size n and m from from X and Y respectively, we test $H_0 : \psi < 1/2$ against $H_1 : \psi > 1/2, \forall \lambda > 0$ assuming the following priors
  - $\pi_0(\psi) = U(0, 1/2)$
  - $\pi_1(\psi) = U(1/2, 1)$
  - $\pi(\lambda) = Gamma(1, 1)$

- Let $X \sim Exp(\alpha)$ and $Y \sim Exp(\beta)$ be random variables. Interest is on
$$\psi = \Pr(X < Y) = \alpha/(\alpha+\beta)$$
with $\lambda = \alpha$.
- Given random samples of size n and m from from X and Y respectively, we test $H_0 : \psi < 1/2$ against $H_1 : \psi > 1/2$, $\forall \lambda > 0$ assuming the following priors
    - $\pi_0(\psi) = U(0, 1/2)$
    - $\pi_1(\psi) = U(1/2, 1)$
    - $\pi(\lambda) = Gamma(1, 1)$
- The MPL is $L_{mp}(\psi) = -(n+m-2)\log\left(s_x + s_y\frac{1-\psi}{\psi}\right) + m\log\frac{1-\psi}{\psi}$, where $s_x = \sum x_i$ and $s_y = \sum y_i$.

- Let $X \sim Exp(\alpha)$ and $Y \sim Exp(\beta)$ be random variables. Interest is on
$$\psi = \Pr(X < Y) = \alpha/(\alpha+\beta)$$
with $\lambda = \alpha$.
- Given random samples of size n and m from from X and Y respectively, we test $H_0 : \psi < 1/2$ against $H_1 : \psi > 1/2$, $\forall \lambda > 0$ assuming the following priors
  - $\pi_0(\psi) = U(0, 1/2)$
  - $\pi_1(\psi) = U(1/2, 1)$
  - $\pi(\lambda) = Gamma(1, 1)$
- The MPL is $L_{mp}(\psi) = -(n+m-2)\log\left(s_x + s_y\frac{1-\psi}{\psi}\right) + m\log\frac{1-\psi}{\psi}$, where $s_x = \sum x_i$ and $s_y = \sum y_i$.
- Both BF^* and BF are evaluated numerically.

Example 1. Stress-strength model (Kotz *et al.*, 2003)

	R		R^*	
	$\lambda = 0.5$	$\lambda = 2.5$	$\lambda = 0.5$	$\lambda = 2.5$
$\psi = 0.4$	10%	33 %	21%	23 %
$\psi = 0.6$	29%	22 %	21%	26 %

Table: Values of risks (with $n = m = 5$).

- For some (ψ, λ), $R < R^*$ just because prior $\pi(\lambda) = Gamma(1, 1)$ gives high probability to the "true" values of λ.

	R		R^*	
	$\lambda = 0.5$	$\lambda = 2.5$	$\lambda = 0.5$	$\lambda = 2.5$
$\psi = 0.4$	10%	33 %	21%	23 %
$\psi = 0.6$	29%	22 %	21%	26 %

Table: Values of risks (with $n = m = 5$).

- For some (ψ, λ), $R < R^*$ just because prior $\pi(\lambda) = Gamma(1, 1)$ gives high probability to the "true" values of λ.
- However, R is much more sensitive to λ than R^*.

Example 1. Stress-strength model (Kotz *et al.*, 2003)

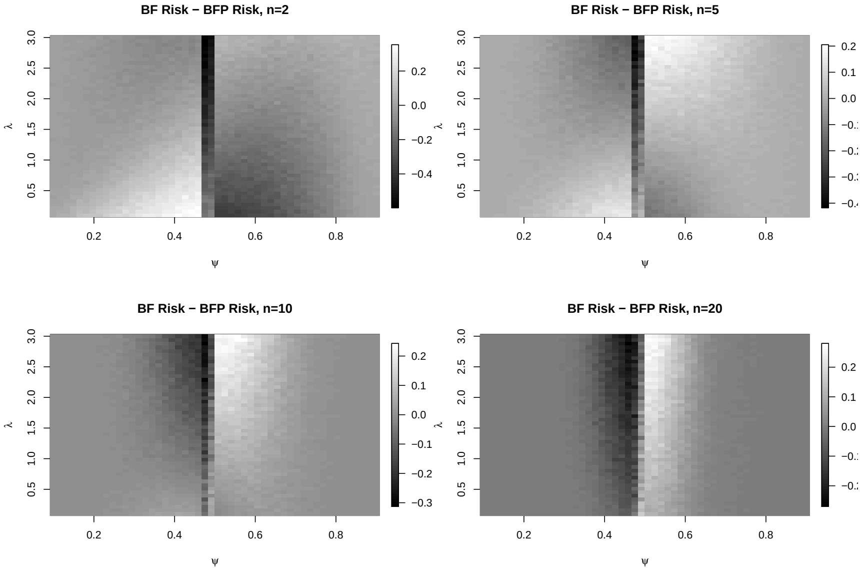

Figure: Values of $R(\psi, \lambda) - R^{*}(\psi, \lambda)$ for $n = m = 2, 5, 10, 20$.

Example 2. Logistic Regression

- The logistic regression model has likelihood function

$$L(\beta) = \exp\left\{ \sum_{i=1}^{n} y_i \sum_{j=1}^{p} \beta_j x_{ij} - \sum_{i=1}^{n} \log\left(1 + e^{\sum_{j=1}^{p} \beta_j x_{ij}}\right) \right\}$$

with $\beta = (\beta_1, \ldots, \beta_p)$ unknown regression coefficient and x_{ij} fixed constants, $i = 1, \ldots, n$ and $j = 1, \ldots, p$.

Example 2. Logistic Regression

- The logistic regression model has likelihood function

$$L(\beta) = \exp\left\{ \sum_{i=1}^{n} y_i \sum_{j=1}^{p} \beta_j x_{ij} - \sum_{i=1}^{n} \log\left(1 + e^{\sum_{j=1}^{p} \beta_j x_{ij}}\right) \right\}$$

  with $\beta = (\beta_1, \ldots, \beta_p)$ unknown regression coefficient and x_{ij} fixed constants, $i = 1, \ldots, n$ and $j = 1, \ldots, p$.
- Assume $\psi = \beta_p$ and let $\lambda = (\beta_1, \ldots, \beta_{p-1})$ be the nuisance parameter.

Example 2. Logistic Regression

- The logistic regression model has likelihood function

$$L(\beta) = \exp\left\{ \sum_{i=1}^{n} y_i \sum_{j=1}^{p} \beta_j x_{ij} - \sum_{i=1}^{n} \log\left(1 + e^{\sum_{j=1}^{p} \beta_j x_{ij}}\right) \right\}$$

  with $\beta = (\beta_1, \ldots, \beta_p)$ unknown regression coefficient and x_{ij} fixed constants, $i = 1, \ldots, n$ and $j = 1, \ldots, p$.
- Assume $\psi = \beta_p$ and let $\lambda = (\beta_1, \ldots, \beta_{p-1})$ be the nuisance parameter.
- Interest on testing $H_0 : \psi > 0$ against $H_1 : \psi < 0$, $\forall \lambda$ assuming the following priors:

Example 2. Logistic Regression

- The logistic regression model has likelihood function

$$L(\beta) = \exp\left\{ \sum_{i=1}^{n} y_i \sum_{j=1}^{p} \beta_j x_{ij} - \sum_{i=1}^{n} \log\left(1 + e^{\sum_{j=1}^{p} \beta_j x_{ij}}\right) \right\}$$

with $\beta = (\beta_1, \ldots, \beta_p)$ unknown regression coefficient and x_{ij} fixed constants, $i = 1, \ldots, n$ and $j = 1, \ldots, p$.

- Assume $\psi = \beta_p$ and let $\lambda = (\beta_1, \ldots, \beta_{p-1})$ be the nuisance parameter.
- Interest on testing $H_0 : \psi > 0$ against $H_1 : \psi < 0$, $\forall\lambda$ assuming the following priors:
  - $\pi_0(\psi) = TruncNormal(0, 1, \underline{\psi} = 0, \overline{\psi} = \infty)$;

Example 2. Logistic Regression

- The logistic regression model has likelihood function

$$L(\beta) = \exp\left\{\sum_{i=1}^{n} y_i \sum_{j=1}^{p} \beta_j x_{ij} - \sum_{i=1}^{n} \log\left(1 + e^{\sum_{j=1}^{p} \beta_j x_{ij}}\right)\right\}$$

  with $\beta = (\beta_1, \ldots, \beta_p)$ unknown regression coefficient and x_{ij} fixed constants, $i = 1, \ldots, n$ and $j = 1, \ldots, p$.
- Assume $\psi = \beta_p$ and let $\lambda = (\beta_1, \ldots, \beta_{p-1})$ be the nuisance parameter.
- Interest on testing $H_0 : \psi > 0$ against $H_1 : \psi < 0$, $\forall \lambda$ assuming the following priors:
  - $\pi_0(\psi) = \mathit{TruncNormal}(0, 1, \underline{\psi} = 0, \overline{\psi} = \infty)$;
  - $\pi_1(\psi) = \mathit{TruncNormal}(0, 1, \underline{\psi} = -\infty, \overline{\psi} = 0)$;

- The logistic regression model has likelihood function

$$L(\beta) = \exp\left\{\sum_{i=1}^{n} y_i \sum_{j=1}^{p} \beta_j x_{ij} - \sum_{i=1}^{n} \log\left(1 + e^{\sum_{j=1}^{p} \beta_j x_{ij}}\right)\right\}$$

  with $\beta = (\beta_1, \ldots, \beta_p)$ unknown regression coefficient and x_{ij} fixed constants, $i = 1, \ldots, n$ and $j = 1, \ldots, p$.
- Assume $\psi = \beta_p$ and let $\lambda = (\beta_1, \ldots, \beta_{p-1})$ be the nuisance parameter.
- Interest on testing $H_0 : \psi > 0$ against $H_1 : \psi < 0$, $\forall \lambda$ assuming the following priors:
  - $\pi_0(\psi) = \mathit{TruncNormal}(0, 1, \underline{\psi} = 0, \overline{\psi} = \infty)$;
  - $\pi_1(\psi) = \mathit{TruncNormal}(0, 1, \underline{\psi} = -\infty, \overline{\psi} = 0)$;
  - $\pi(\boldsymbol{\lambda}) = \pi(\beta_1) \cdot \ldots \cdot \pi(\beta_{p-1})$, $\pi(\beta_j) = N(0, 10^{12})$, $j = 1, 2, \ldots, p-1$.

Example 2. Logistic Regression - Urine data (Davison, Hinkley, 1997).

- For illustration, we analyze the presence/absence of calcium oxalate crystals in urine samples Y together with the values of $p = 6$ quantitative covariates on $n = 77$ individuals.

Example 2. Logistic Regression - Urine data (Davison, Hinkley, 1997).

- For illustration, we analyze the presence/absence of calcium oxalate crystals in urine samples Y together with the values of $p = 6$ quantitative covariates on $n = 77$ individuals.
- Assume that ψ is the coefficient of the effect of the variable urea concentration.

- For illustration, we analyze the presence/absence of calcium oxalate crystals in urine samples Y together with the values of $p = 6$ quantitative covariates on $n = 77$ individuals.
- Assume that ψ is the coefficient of the effect of the variable urea concentration.
- The weights of evidence (Good, 1985), $W = \log BF$ and $W^* = \log BF^*$, are

	W and W^*
BF based on $L(\beta)$	4.2
BF^* based on $L_{mp}(\psi)$	4.2

According to the Jeffreys' scale, the evidence is substantial in favor of positive effect of urea concentration.

Example 2. Logistic Regression - p large.

In order to assess the behaviour of BF^*, with respect to large p, we evaluate the corresponding W^* in 1000 data sets with $p = 20$ coefficients

Example 2. Logistic Regression - p large.

In order to assess the behaviour of BF^*, with respect to large p, we evaluate the corresponding W^* in 1000 data sets with $p = 20$ coefficients, with 5 positive, 5 negative and 10 zero coefficients.

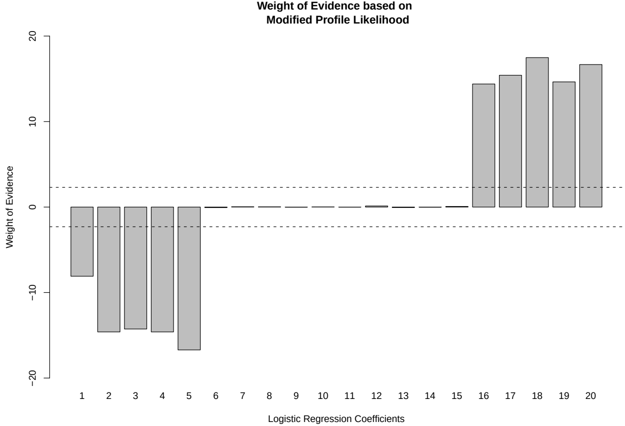

Figure: Empirical mean W^* for the sign of 20 coefficients.
(Horizontal dashed lines are the levels of strong evidence).

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.
- For the analyzed examples we obtained that:

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.
- For the analyzed examples we obtained that:
  - in general, $R^*(\psi, \lambda) < R(\psi, \lambda)$

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.
- For the analyzed examples we obtained that:
  - in general, $R^*(\psi, \lambda) < R(\psi, \lambda)$
  - R^* is almost constant with respect to λ

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.
- For the analyzed examples we obtained that:
  - in general, $R^*(\psi, \lambda) < R(\psi, \lambda)$
  - R^* is almost constant with respect to λ
- Main advantages of BF^*:

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.
- For the analyzed examples we obtained that:
  - in general, $R^*(\psi, \lambda) < R(\psi, \lambda)$
  - R^* is almost constant with respect to λ
- Main advantages of BF^*:
  - avoid elicitation of priors on nuisance parameters

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.
- For the analyzed examples we obtained that:
  - in general, $R^*(\psi, \lambda) < R(\psi, \lambda)$
  - R^* is almost constant with respect to λ
- Main advantages of BF^*:
  - avoid elicitation of priors on nuisance parameters
  - no multidimensional integrations.

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.
- For the analyzed examples we obtained that:
  - in general, $R^*(\psi, \lambda) < R(\psi, \lambda)$
  - R^* is almost constant with respect to λ
- Main advantages of BF^*:
  - avoid elicitation of priors on nuisance parameters
  - no multidimensional integrations.
- BF^* may be also obtained for semi-parametric models or complex models, when the $L(\psi, \lambda)$ is difficult or even impossible to compute,

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.
- For the analyzed examples we obtained that:
  - in general, $R^*(\psi, \lambda) < R(\psi, \lambda)$
  - R^* is almost constant with respect to λ
- Main advantages of BF^*:
  - avoid elicitation of priors on nuisance parameters
  - no multidimensional integrations.
- BF^* may be also obtained for semi-parametric models or complex models, when the $L(\psi, \lambda)$ is difficult or even impossible to compute,

Conclusions

- The use of pseudo-Bayes factor BF^* may be of potential interest in situations with many nuisance parameters having a weak physical meaning.
- For the analyzed examples we obtained that:
  - in general, $R^*(\psi, \lambda) < R(\psi, \lambda)$
  - R^* is almost constant with respect to λ
- Main advantages of BF^*:
  - avoid elicitation of priors on nuisance parameters
  - no multidimensional integrations.
- BF^* may be also obtained for semi-parametric models or complex models, when the $L(\psi, \lambda)$ is difficult or even impossible to compute, in fact, it is possible to resort to quasi-profile likelihoods (Ventura *et al.*, 2010), empirical likelihoods and composite likelihoods (Pauli *et al.*, 2010).

Some references

- BRAZZALE, A.R., DAVISON, A.C. and REID, N. (2007): *Applied Asymptotics*. Cambridge University Press, Cambridge.
- DAVISON, A.C. and HINKLEY, D.V. (1997): *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.
- GOOD, I.J. (1985): Weight of evidence: A brief survey. *Bayesian Statistics 2, 249–269*.
- KASS, R.E. and RAFTERY, A.E. (1995): Bayes factors. *JASA 90 (430), 773–795*.
- KOTZ, S., LUMELSKII, Y. and PENSKY, M. (2003): *The Stress-Strength Model and its Generalizations*. World Scientific, Singapore.
- PACE, L. and SALVAN, A. (1997): *Principles of Statistical Inference from a Neo-Fisherian Perspective*. World Scientific, Singapore.
- PAULI, F., VENTURA, L., and RACUGNO, W. (2010): Bayesian composite marginal likelihoods, *Stat. Sinica, to appear*.
- SEVERINI, T.A. (2000): *Likelihood Methods in Statistics*. Oxford University Press.
- VENTURA, L., CABRAS, S. and RACUGNO, W. (2009): Prior distributions from pseudo-likelihoods in the presence of nuisance parameters, *JASA 104 (486), 768–777*.
- VENTURA, L., CABRAS, S. and RACUGNO, W. (2010): Default prior distributions from quasi- and quasi-profile likelihoods. *JSPI* to appear.