Easily Computed Marginal Likelihoods from Posterior Simulation Using the THAMES Estimator

Statistiques au sommet de Rochebrune

Marie Perrot-Dockès & Sarah Ouadah

MAP5, Université Paris-Cité
MIA Paris-Saclay, AgroParisTech, INRAE, Université Paris-Saclay

27.03.2024

Joint work with

Martin Metodiev
Université Clermont-Auvergne

Nicholas J. Irons
University of Washington

Pierre Latouche
Université Clermont-Auvergne

Adrian E. Raftery
University of Washington

M. Metodiev, M. Perrot-Dockès, S. Ouadah, N. J. Irons, P. Latouche, A. E. Raftery. Easily Computed Marginal Likelihoods from Posterior Simulation Using the THAMES Estimator. *arXiv preprint arXiv :2305.08952*, 2023.

Package R THAMES https://cran.r-project.org/web/packages/thames/index.html

Sommaire

M a parametric model

- **Data** $\mathcal{D}$ with **likelihood** $p_M(\mathcal{D}|\theta) =: L(\theta)$
- **Parameter** θ with **prior** distribution function $\pi(\theta)$

Marginal likelihood :

$$Z := p_M(\mathcal{D}) = \int \pi(\theta) L(\theta)\, d\theta$$

Use of marginal likelihood Z

- **Bayes factors** (Kass and Raftery (1995), Berger et al. (2001)) for Q a quantity of interest :

$$BF(M_1, M_2) = \frac{p_{M_1}(\mathcal{D})}{p_{M_2}(\mathcal{D})} = \frac{Z_1}{Z_2}$$

- **Bayesian Model Averaging** (Jeffreys (1961)) :

$$p(Q|\mathcal{D}) = \sum_{k=1}^{K} p(Q|\mathcal{D}, M_k) p(M_k|\mathcal{D})$$

where $p(M_k|\mathcal{D}) \propto \pi_k Z_k$ and $\sum_k p(M_k|\mathcal{D}) = 1$

Marginal likelihood estimators

- Monte Carlo

$$\hat{p}_1(y) = \frac{1}{T}\sum_{t=1}^{T} p(y|\theta^{(t)}), \quad \theta^{(t)} \sim \pi$$

- Importance Sampling

$$\hat{p}_2(y) = \frac{1}{T}\sum_{t=1}^{T} \frac{p(y|\tilde{\theta}^{(t)})\pi(\tilde{\theta}^{(t)})}{g(\tilde{\theta}^{(t)})}, \quad \tilde{\theta}^{(t)} \sim g$$

- Reciprocal Importance Sampling

$$\hat{p}_3(y) = \left(\frac{1}{T}\sum_{t=1}^{T} \frac{g(\tilde{\theta}^{(t)})}{p(y|\tilde{\theta}^{(t)})\pi(\tilde{\theta}^{(t)})}\right)^{-1}, \quad \tilde{\theta}^{(t)} \sim p(\theta|y)$$

- Bridge Sampling

Exhaustive overview of Monte-Carlo methods in Llorente et al. (2023)

Importance Sampling

- **Idea**

$$p(y) = \int p(y|\theta)\pi(\theta)d\theta$$
$$= \int \frac{p(y|\theta)\pi(\theta)g(\theta)}{g(\theta)}d\theta$$
$$= \mathbb{E}_g\left(\frac{p(y|\theta)\pi(\theta)}{g(\theta)}\right)$$

- **Estimators**

$$\hat{p}(y) = \frac{1}{T}\sum_{t=1}^{T}\frac{p(y|\tilde{\theta}^{(t)})\pi(\tilde{\theta}^{(t)})}{g(\tilde{\theta}^{(t)})}, \quad \tilde{\theta}^{(t)} \sim g$$

- **Difficulties**
  - when g has thinner tails than π
  - choice of the importance sampling function g

Recriprocal Importance Sampling

- **Idea**

$$
\begin{aligned}
\frac{1}{p(y)} &= \int \frac{1}{p(y)} g(\theta) d\theta \\
&= \int \frac{\pi(\theta|y) g(\theta)}{p(y|\theta) p(\theta)} d\theta \\
&= \mathbb{E}_{p(\theta|y)} \left(\frac{g(\theta)}{p(y|\theta)\pi(\theta)} \right)
\end{aligned}
$$

- **Estimators**

$$
\hat{p}(y)^{-1} = \frac{1}{T} \sum_{t=1}^{T} \frac{g(\theta^{\star(t)})}{p(y|\theta^{\star(t)})\pi(\theta^{\star(t)})}, \quad \theta^{\star(t)} \sim p(\theta|y)
$$

- **Difficulties**
  - when g has fatter tails than π

- **Idea**

$$1 = \frac{\int p(y|\theta)\pi(\theta)g(\theta)h(\theta)d\theta}{\int p(y|\theta)\pi(\theta)g(\theta)h(\theta)d\theta}$$

$$p(y) = \frac{\int p(y|\theta)\pi(\theta)g(\theta)h(\theta)d\theta}{\int \frac{p(y|\theta)\pi(\theta)}{p(y)} g(\theta)h(\theta)d\theta}$$

$$= \frac{\mathbb{E}_g\left(p(y|\theta)\pi(\theta)h(\theta)\right)}{\mathbb{E}_{p(\theta|y)}\left(g(\theta)h(\theta)\right)}$$

- **Estimators**

$$\hat{p}(y) = \frac{\frac{1}{T_1}\sum_{t=1}^{T_1} p(y|\tilde{\theta}^{(t)})\pi(\tilde{\theta}^{(t)})h(\tilde{\theta}^{(t)})}{\frac{1}{T_2}\sum_{t=1}^{T_2} g(\theta^{*(t)})h(\theta^{*(t)})}, \quad \tilde{\theta}^{(t)} \sim g \, , \theta^{*(t)} \sim p(\theta|y)$$

- **Difficulties**
  - simulations from 2 models
  - iterative algorithm : optimal h depending on $p(y)$

Sommaire

A glimpse of geography

- Source : Thames Head (Gloucestershire)
- Longueur : 346 km
- Fin de course : Mer du Nord

A glimpse of history

- **1950s** : the Thames has been declared biologically dead
- **2004-** : ZSL improves the ecology and biodiversity

A glimpse of history

- **1950s** : the Thames has been declared biologically dead
- **2004-** : ZSL improves the ecology and biodiversity
  Now : Seals, Sharks, Hippocampus !

Seal Population

Year	Survey lead	Harbour seal population estimate (95% CI)	Grey seal population estimate (95% CI)
2003	SMRU	250 (205 – 333)	402 (336 – 500)
2008	SMRU	443 (363 – 591)	669 (559 – 833)
2010	SMRU	526 (431 – 702)	1644 (1374 – 2047)
2013	ZSL, BA	654 (535 – 872)	812 (678 – 1010)
2014	ZSL, BA	679 (556 – 906)	1879 (1570 – 2339)
2015	ZSL, BA	626 (513 – 835)	1900 (1587 – 2365)
2016	ZSL, BA	964 (789 – 1285)	2013 (1682 – 2505)
2017	ZSL, BA	1104 (903 – 1472)	2406 (2010 – 2995)
2018	SMRU	1026 (840 – 1369)	2490 (2080 – 3099)
2019	ZSL, BA	932 (763 – 1243)	3243 (2710 – 4036)

Back to framework

M a parametric model

- **Data** $\mathcal{D}$ with **likelihood** $p_M(\mathcal{D}|\theta) =: L(\theta)$
- **Parameter** θ with **prior** distribution function $\pi(\theta)$

Marginal likelihood $p_M(\mathcal{D}) = \int L(\theta)\pi(\theta)\, d\theta := Z$

Bayes' rule

$$Z = \frac{L(\theta)\pi(\theta)}{p(\theta|\mathcal{D})}$$

Key identity

$$Z^{-1} = \mathbf{E}\left[\left.\frac{h(\theta)}{L(\theta)\pi(\theta)}\right| \mathcal{D}\right]$$

where $h(\theta)$ is a pdf over the posterior support.

RIS Gelfand and Dey (1994)) :

1. Simulate $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(T)} \sim p(\theta|\mathcal{D})$ using MCMC
2. Estimator :

$$\hat{Z}^{-1} = \frac{1}{T}\sum_{t=1}^{T}\frac{h(\theta^{(t)})}{L(\theta^{(t)})\pi(\theta^{(t)})}$$

Need to

- overlaps with the area of the parameter space (for efficiency)
- have thin enough tails (for finite variance) :

$$\int \frac{h(\theta)^2}{L(\theta)\pi(\theta)} d\theta < \infty$$

Harmonic mean estimator

Newton and Raftery (1994) : $h(\theta) = \pi(\theta)$

$$\hat{Z}^{-1} = \frac{1}{T}\sum_{t=1}^{T}\frac{h(\theta^{(t)})}{L(\theta^{(t)})\pi(\theta^{(t)})} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{L(\theta^{(t)})}$$

- often infinite variance

Uniform h on a convex hull estimator

Robert and Wraith (2009) : $h(\theta)$ the uniform distribution on the convex hull of simulated MCMC parameters values in the α-HPD region

$$\hat{Z}^{-1} = \frac{1}{T}\sum_{t=1}^{T}\frac{h(\theta^{(t)})}{L(\theta^{(t)})\pi(\theta^{(t)})} = \frac{1}{V(B_\alpha)T}\sum_{\theta^{(t)}\in B}\frac{1}{L(\theta^{(t)})\pi(\theta^{(t)})}$$

- choice of α
- compute the volume of a convex hull (hard when $d > 2$)

Truncated normal h estimator

DiCiccio et al. (1997) : h the truncated normal $TN_A(\hat{\theta}, \hat{\Sigma})$ to the set A

$$\hat{Z}^{-1} = \frac{1}{T} \sum_{t=1}^{T} \frac{h(\theta^{(t)})}{L(\theta^{(t)})\pi(\theta^{(t)})} = \frac{1}{T} \sum_{t=1}^{T} \frac{TN_A(\theta^{(t)}; \hat{\theta}, \hat{\Sigma})}{L(\theta^{(t)})\pi(\theta^{(t)})}$$

where $A = \{\theta : (\theta - \hat{\theta})^T \hat{\Sigma}^{-1} (\theta - \hat{\theta}) < c^2\}$ and $\hat{\theta}, \hat{\Sigma}$ estimates of posterior mode or mean and covariance matrix

- sensitive to specification of c

Truncated HArmonic Mean EStimator

THAMES : RIS estimator with h the uniform distribution on A

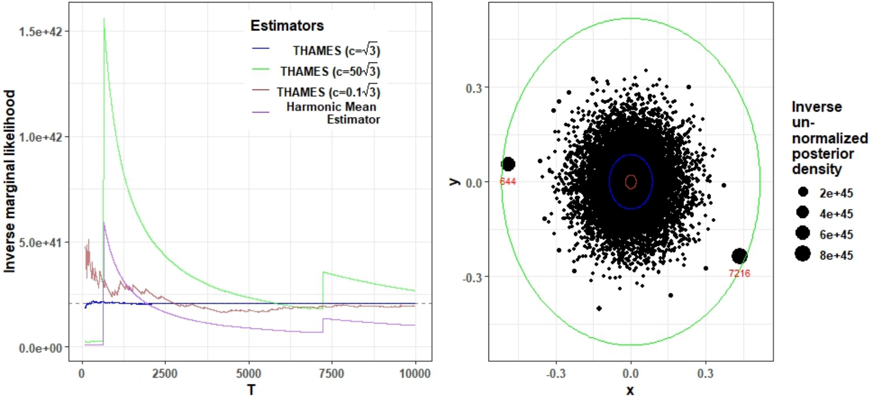
$$\hat{Z}^{-1} = \frac{1}{V(A)T} \sum_{t=1,\ \theta^{(t)} \in A}^{T} \frac{1}{L(\theta^{(t)})\pi(\theta^{(t)})}$$

where

- $A = \{\theta : (\theta - \hat{\theta})^T \hat{\Sigma}^{-1}(\theta - \hat{\theta}) < c^2\}$
- $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(T)} \sim p(\theta|\mathcal{D})$
- $\hat{\Sigma}$ (resp $\hat{\theta}$) estimated variance (resp mean) of $(\theta^{(t)})$.

- optimal choice of c
- $V(A)$ easily computable

Impact of the ellipsoid choice



- optimal choice of c for minimizing the variance

Criterion for choice of c

- THAMES estimator

$$\hat{Z}^{-1} = \frac{1}{V(A)T} \sum_{t=1,\ \theta^{(t)} \in A}^{T} \frac{1}{L(\theta^{(t)})\pi(\theta^{(t)})}$$

$$A = \{\theta : (\theta - \hat{\theta})^T \hat{\Sigma}^{-1} (\theta - \hat{\theta}) < c^2\}$$

- THAMES variance under the assumption of independent draws of posterior simulation

$$Var(\hat{Z}^{-1}|\mathcal{D}) = \frac{1}{T} \cdot \frac{1}{Z^2} \cdot \left(\frac{1}{V(A)^2} \int_A \frac{1}{p(\theta|\mathcal{D})}\, d\theta - 1 \right)$$

- optimal radius c minimizes the variance

Variance minimization according to c under the (oracle) Gaussian case

- Gaussian posterior : $\mathrm{p}(\theta|\mathcal{D}) = \mathcal{N}_d(m_n, \Sigma)$
- Ellipsoid : $A_{or} = \{\theta : (\theta - m_n)^T \Sigma^{-1} (\theta - m_n) < c^2\}$
- THAMES variance $\propto$

$$\frac{1}{V(A)^2} \int_A \frac{1}{p(\theta|\mathcal{D})} \, d\theta = \frac{|\Sigma|}{V(A_{or})^2} \int_{T(A_{or}, m, \Sigma)} (2\pi)^{d/2} \exp\left(\frac{x^t x}{2}\right) dx$$
$$= \dots$$
$$= \frac{\kappa_d}{c^{2d}} \int_0^c \exp\left(\frac{r^2}{2}\right) r^{d-1} dr$$

Optimal c "exact" solution

- Variance derivative :

$$-2d\frac{1}{c_d^2}f(d,c_d)+\exp\left(\frac{1}{2}c_d^2\right)=0$$

where

$$f(d,c)=\begin{cases}\frac{1}{c}\sqrt{\frac{\pi}{2}}\text{erfi}\left(\frac{c}{\sqrt{2}}\right) & d=1,(\text{erfi}(x)=\int_0^x e^{t^2}dt)\\ \exp(\frac{c^2}{2})-1 & d=2.\end{cases}$$

with recursion

$$f(d,c)=\exp\left(\frac{1}{2}c^2\right)-(d-2)\frac{1}{c^2}f(d-2,c)$$

- solved numerically

Optimal choice of c

Theorem

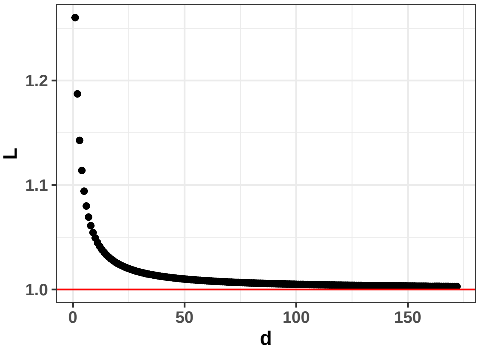
Assumptions

- *Independance of MCMC simulations*
- *Posterior distributions* $p(\theta|\mathcal{D}) = \mathcal{N}(m, \Sigma)$
- $A_{or} := \{\theta : (\theta - m)^T \Sigma^{-1} (\theta - m) < c^2\}$ *with known* Σ

There exists a unique radius $c_d \in (0, \infty)$ *such that the ellipse* A_{or} *with radius* c_d *minimizes the variance of the THAMES :*

$$c_d = \sqrt{d + L_d}, \text{ where } L_d \geq 0, \frac{L_d}{d} \overset{d \to \infty}{\to} 0.$$

Optimal choice of c



- $c_d = \sqrt{d + L_d}$
- $c_d \approx \sqrt{d+1}$ (depends only on d)
- for $A_{or} := \{\theta : (\theta - m)^T \Sigma^{-1} (\theta - m) < c_d^2\}$ in Gaussian case $\mathbb{P}(\theta^{(t)} \in A_{or}) = F_{\chi^2(d)}(c_d^2) \overset{d \to \infty}{\to} 0.5$

THAMES properties

Proposition

- $\hat{Z}^{-1}$ *is unbiased and asymptotically normal*
- $\hat{Z}$ *is* ***asymptotically unbiased*** *and*

$$\sqrt{T}\left(log(\hat{Z}) - log(Z)\right) \xrightarrow{L} \mathcal{N}(0, V_d) \text{ with } V_d = O(\sqrt{d})$$

- **Sample-splitting** :
  1. $(\hat{\theta}, \hat{\Sigma})$ estimation using the first $T/2$ posterior samples
  2. $\hat{Z}^{-1}$ calculation based on $A = \{\theta : (\theta - \hat{\theta})^T \hat{\Sigma}^{-1}(\theta - \hat{\theta}) < d + 1\}$ using the last $T/2$ posterior samples

- **Constrained parameter space** : adjust the volume using a Monte Carlo approximation
  1. $\nu^{(1)}, \dots, \nu^{(N)} \sim \mathcal{U}(A)$
  2. Approximate $V(A \cap \text{supp}(\theta|\mathcal{D}))/V(A)$:

$$\hat{R} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{\theta|\pi(\theta)L(\theta)>0\}}(\nu^{(i)}).$$

Input : Data $\mathcal{D}$ and posterior samples $(\theta^{(i)})_{i\in[\![1,T]\!]}$.

Sample splitting : Calculate the empirical mean $\hat{\theta}$ and sample covariance matrix $\hat{\Sigma}$ based on the first $T/2$ posterior samples $(\theta^{(i)})_{i\in[\![1,T/2]\!]}$.

Standardization : $\tilde{\theta}^{(i)} = (\theta^{(i)} - \hat{\theta})\hat{\Sigma}^{-1/2}$ for $i \in [\![T/2+1, T]\!]$.

Truncation subset : $\mathcal{S} = \{i : \|\tilde{\theta}^{(i)}\|_2^2 < d+1\}$.

Calculate THAMES estimator :

$$\hat{Z}^{-1} = \frac{1}{V(A)T/2} \sum_{i\in S} \frac{1}{L(\theta^{(i)})\pi(\theta^{(i)})},$$

$$A = \{\theta : (\theta - \hat{\theta})^T \hat{\Sigma}^{-1}(\theta - \hat{\theta}) < d+1\}.$$

if constrained parameters $\hat{Z}^{-1} \leftarrow \hat{R}^{-1}\hat{Z}^{-1}$

Output : THAMES estimator $\hat{Z}^{-1}$.

Sommaire

- $Y \in \mathbb{R}^n$ target vector
- $X \in \mathcal{M}_{n\times(d-1)}$ design matrix

$$Y|X, \beta, \sigma^2 \sim \mathrm{MVN}_n(X\beta, \sigma^2 I_n)$$

with Zellner's prior

$$p(\beta|\sigma^2) = \mathrm{MVN}_{d-1}(\beta; 0_{d-1}, g\sigma^2(X^TX)^{-1}),$$
$$p(\sigma^2) = \mathsf{InvGamma}\left(\sigma^2; \frac{1}{2}\nu_0, \frac{1}{2}\sigma_0^2\nu_0\right),$$

with $g, \nu_0, \sigma_0^2 > 0$.

Posterior distribution

$$p(\beta|\sigma^2, \mathcal{D}) = \mathrm{MVN}_{d-1}\left(\beta; \frac{g}{g+1}m_n, \frac{g}{g+1}\sigma^2(X^TX)^{-1}\right)$$

$$p(\sigma^2|\mathcal{D}) = \mathsf{InvGamma}\left(\sigma^2; \frac{1}{2}(\nu_0 + n), \frac{1}{2}(\nu_0\sigma_0^2 + s_n)\right)$$

with $m_n = (X^TX)^{-1}X^T\mathbf{y}$, $s_n = \mathbf{y}^T\mathbf{y} - \frac{g}{g+1}\mathbf{y}^TXm_n$ and $\mathbf{y} \in \mathbb{R}^n$ the *observed* vector of target variables associated with Y.

Analytical expression for Z the marginal likelihood

Hoff (2009, Chapter 9) :

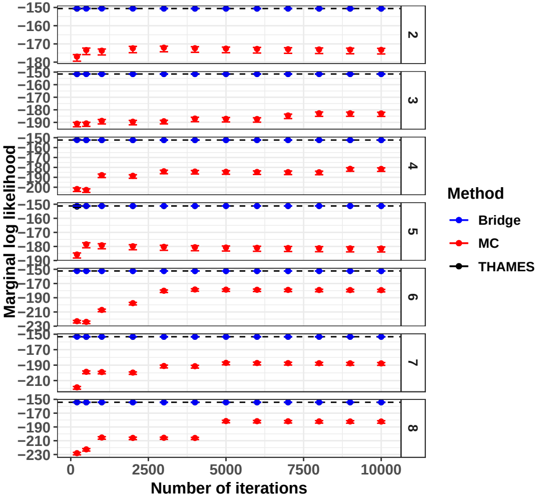
$$p(\mathbf{y}|X) = \frac{(g+1)^{-(d-1)/2}}{\pi^{n/2}} \cdot \frac{\Gamma(\frac{1}{2}(\nu_0 + n))}{\Gamma(\frac{1}{2}\nu_0)} \cdot \left(\frac{\nu_0\sigma_0^2}{\nu_0\sigma_0^2 + s_n}\right)^{\nu_0/2}$$

Prostate data

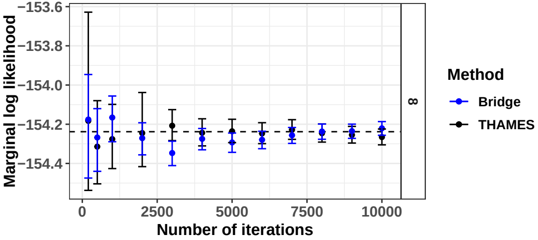
- Y the level of a prostate-specific antigen
- X eight clinical measure
  log cancer volume (lcavol), log prostate weight (lweight), age, log of the amount of benign prostatic hyperplasia (lbph), seminal vesicle invasion (svi), log of capsular penetration (lcp), Gleason score (gleason), and percent of Gleason scores 4 or 5 (pgg45)

Choice of the hyperparameters : $(g, \nu_0, \sigma_0^2) = (\sqrt{n}, 4, 1)$ (Porwal and Raftery (2022)).

Estimators comparison



Zoom on Bridge sampling vs THAMES



Netherlands schools dataset

Data : Netherlands schools dataset of Snijders and Bosker (1999) :
Language test scores of

- 2,287 eighth-grade pupils
- from 133 classes in the Netherlands.

M_0 : All classes perform the same

M_1 : There is variation at the class level

All classes perform the same

M_0 :

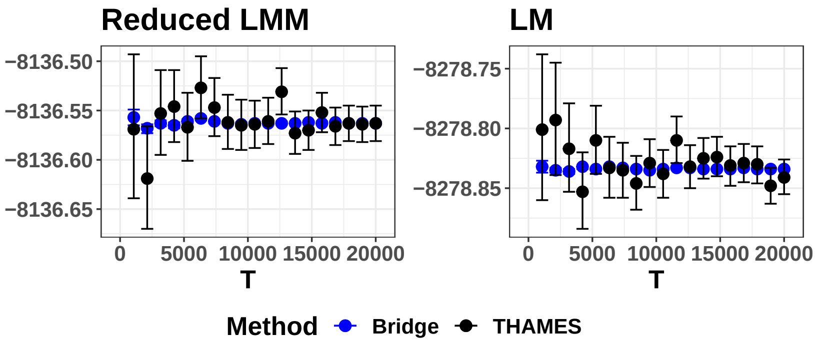
$$
\begin{aligned}
y_{ij} &= \mu + \varepsilon_{ij}, \quad j \in \{1, \dots, J\}, i \in \{1, \dots, n_j\}, \sum_j n_j = n \\
\varepsilon_{ij} &\overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2), \\
\mu &\sim N(\hat{\mu}, \hat{\sigma}_\mu^2), \\
\sigma_\varepsilon^2 &\sim \text{InverseGamma}(\hat{\nu}_\varepsilon, \hat{\beta}_\varepsilon).
\end{aligned}
$$

There is variation at the class level

M_1 :

$$
\begin{aligned}
y_{ij} &= \mu + \varepsilon_{ij}, \\
\varepsilon_{ij} &\sim N(0, \sigma_\varepsilon^2 + \sigma_\alpha^2), \\
\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'j}) &= \sigma_\alpha^2, \quad i, i' \in \{1, \ldots, n_j\}, i \neq i', \\
\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) &= 0, \quad j \neq j', \\
\mu &\sim N(\hat{\mu}, \hat{\sigma}_\mu^2), \\
\sigma_\varepsilon^2 &\sim \mathrm{InverseGamma}(\hat{\nu}_\varepsilon, \hat{\beta}_\varepsilon), \\
\sigma_\alpha^2 &\sim \mathrm{InverseGamma}(\hat{\nu}_\alpha, \hat{\beta}_\alpha).
\end{aligned}
$$

Nlschools marginal likelihood



Computation time

Table – Average CPU times (in seconds per 1,000 posterior draws) for producing the previous estimates

	M_1	M_1
Bridge	1.6482	0.0723
MC	0.0002	0.0002
THAMES	0.0030	0.0027

Bayes factor

$$\begin{aligned}\log(BF(M_0, M_1)) &= \log(Z_0) - \log(Z_1)\\ &= -8278.842 + 8136.561\\ &= -142.281\end{aligned}$$

- decisive evidence in favor of the full linear model (Kass and Raftery (1995))

Sommaire

Message

A method for computing marginal likelihoods from MCMC

- a version of Reciprocal Importance Sampling (RIS)
- combines ideas of Robert and Wraith (2009) and DiCiccio et al. (1997)
- simple to compute with available confidence intervals
- finite variance, consistent, asymptotically unbiased and normal

Easy implementation

- uses sample splitting
- corrects for the presence of bounded parameters

Accurate estimates in Gaussian and non-Gaussian settings

Merci

Dans la session je THAMES moi non plus

- Modèle de mélange
- SBM
- Topic bloc model (etude de gros réseaux sociaux COP28)
- Présentation de l'utilisation de THAMES
  - Dans le cadre des Gabriels
  - Dans le cadre d'une Marie
  - autres propositions ?

References I

Berger, J. O., Pericchi, L. R., Ghosh, J. K., Samanta, T., Santis, F. D., Berger, J. O., and Pericchi, L. R. (2001). Objective bayesian methods for model selection : Introduction and comparison. *Lecture Notes-Monograph Series*, 38 :135–207.

DiCiccio, T. J., Kass, R. E., Raftery, A. E., and Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, 92 :903–915.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice : asymptotics and exact calculations. *Journal of the Royal Statistical Society : Series B (Methodological)*, 56 :501–514.

Hoff, P. (2009). *A First Course in Bayesian Statistical Methods*. Springer Texts in Statistics. Springer New York.

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford, U.K., 3rd edition.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430) :773–795.

Llorente, F., Martino, L., Delgado, D., and Lopez-Santiago, J. (2023). Marginal likelihood computation for model selection and hypothesis testing : An extensive review. *SIAM Review*, 65 :3–58.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society : Series B (Methodological)*, 56 :3–26.

Porwal, A. and Raftery, A. E. (2022). Comparing methods for statistical inference with model uncertainty. *Proceedings of the National Academy of Sciences*, 119(16) :e2120737119.

Robert, C. P. and Wraith, D. (2009). Computational methods for Bayesian model choice. In *AIP conference proceedings*, volume 1193, pages 251–262. American Institute of Physics.

Snijders, T. A. B. and Bosker, R. J. (1999). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modelling*. Sage.