

A Mixture-Based Bayesian Model Averaging Method

Georges Nguefack-Tsague¹, Walter Zucchini²

¹Biostatistics Unit, Department of Public Health, Faculty of Medicine and Biomedical Sciences, University of Yaounde 1, Yaounde, Cameroon

²Institute for Statistics and Econometrics, University of Goettingen, Goettingen, Germany
Email: nguefacktsague@yahoo.fr, walter.zucchini@wiwi.uni-goettingen.de

Received 22 January 2016; accepted 18 April 2016; published 21 April 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Bayesian model averaging (BMA) is a popular and powerful statistical method of taking account of uncertainty about model form or assumption. Usually the long run (frequentist) performances of the resulted estimator are hard to derive. This paper proposes a mixture of priors and sampling distributions as a basic of a Bayes estimator. The frequentist properties of the new Bayes estimator are automatically derived from Bayesian decision theory. It is shown that if all competing models have the same parametric form, the new Bayes estimator reduces to BMA estimator. The method is applied to the daily exchange rate Euro to US Dollar.

Keywords

Mixture, Bayesian Model Selection, Bayesian Model Averaging, Bayesian Theory, Frequentist Performance

1. Introduction

Several models are a priori plausible in statistical modeling; it is thus quite common nowadays to apply some model selection procedure to select a single one. For an overview of frequentist model selection criteria, see Leeb and Poetscher [1], Zucchini [2], and Zucchini *et al.* [3]. Alternatively one can give weights to all plausible models and work with the resulting weighted estimator. This can be done either in frequentist approach (frequentist model averaging, FMA) or Bayesian context (Bayesian model averaging, BMA). References on frequentist model averaging include: Nguefack-Tsague [4], Burnham and Anderson [5], Nguefack-Tsague and Zucchini [6], and Nguefack-Tsague [7]-[9]. In the Bayesian context, there are three fundamental factors in decision theory:

- 1) a distribution family of the observation (sampling distribution), $f(x|\mu)$;
 - 2) a prior distribution for the parameter, $\pi(\mu)$;
 - 3) a loss function associated to a decision δ , $L(\mu, \delta)$; with expected loss $\rho(\pi, \delta) = \int_{\Gamma} L(\mu, \delta) \pi(\mu) d\mu$.
- The posterior distribution of μ is given by

$$\pi(\mu|x) = \frac{f(x|\mu)\pi(\mu)}{\int_{\Gamma} f(x|\mu)\pi(\mu)d\mu}. \quad (1)$$

A posterior distribution and a loss function lead to an optimal decision rule (Bayes rule), together with its risk function and its frequentist properties.

1.1. Bayesian Model Selection

Consider a situation in which some quantity of interest, μ , is to be estimated from a sample of observations that can be regarded as realizations from some unknown probability distribution, and that in order to do so, it is necessary to specify a model for the distribution. There are usually many alternative plausible models available and, in general, they each lead to different estimates of μ . Consider a sample of data, x , and a set of K models $M = (M_1, \dots, M_K)$ containing the true model M_i . Each M_k consists of a family of distributions $P(x|\eta_k, M_k)$, where η_k represents a parameter (or vector of parameters). The prior probability that M_k is the true model is denoted by $P(M_k)$ and the prior distribution of the parameters of M_k (given that M_k is true) by $P(\eta_k|M_k)$. Conditioning on the data x and integrating out the parameter η_k , one obtains the following posterior model probabilities:

$$P(M_k|x) = \frac{P(x|M_k)P(M_k)}{\sum_{j=1}^K P(x|M_j)P(M_j)}, \quad (2)$$

where

$$P(x|M_j) = \int P(x|\eta_k, M_k)P(\eta_k|M_k)d\eta_k \quad (3)$$

is the integrated likelihood under M_k . If $P(\eta_k|M_k)$ is a discrete distribution, the integral in (3) is replaced by a sum.

Bayesian model selection involves selecting the “best” model with some selection criterion; more often the Bayesian information criterion (BIC), also known as the Schwarz criterion [10] is used; it is an asymptotic approximation of the log posterior odds when the prior odds are all equal. More information on Bayesian model selection and applications can be found in Nguefack-Tsague and Ingo [11], Guan and Stephens [12], Nguefack-Tsague [13], Carvalho and Scott [14], Fridley [15], Robert [16], Liang *et al.* [17], and Bernardo and Smith [18]. Other variants of model selection include Nguefack-Tsague and Ingo [11] who used BMA machinery to derive a focused Bayesian information criterion (FoBMA) which selects different models for different purposes, *i.e.* their method depends on the parameter singled out for inferences.

1.2. Bayesian Model Averaging

Let μ be a quantity of interest depending on x , for example a future observation from the same process that generated x . The idea is to use a weighted average of the estimates of μ obtained using each of the alternative models, rather than the estimate obtained using any single model. More precisely, the posterior distribution of μ is given by

$$P(\mu|x) = \sum_{k=1}^K P(\mu|x, M_k)P(M_k|x). \quad (4)$$

Note that $P(\mu|x)$ is a weighted average of the posterior distributions $P(\mu|M_k, x)$, $k = 1, \dots, K$, where the k -th weight, $P(M_k|x)$, is the posterior probability that M_k is the true model.

The posterior distribution of μ , conditioned on M_k being true, is given by

$$P(\mu|x, M_k) = \int P(\mu|\eta_k, M_k) P(\eta_k|x, M_k) d\eta_k. \quad (5)$$

The posterior mean and posterior variance are given by

$$E(\mu|x) = \sum_{k=1}^K E(\mu|x, M_k) P(M_k|x), \quad (6)$$

$$\text{Var}(\mu|x) = \sum_{k=1}^K \left[\text{Var}(\mu|x, M_k) + (E(\mu|x, M_k) - E(\mu|x))^2 \right] P(M_k|x). \quad (7)$$

A classical reference is Hoeting *et al.* [19] with an extensive framework of BMA methodology and applications for different statistical models. Various real data and simulation studies have investigated the predictive performance of BMA (Clyde [20]; Clyde and George [21]). A discussion on the issue of using BMA for dealing with model uncertainty is given in Clyde and George [21]. Nguefack-Tsague [13] uses BMA in the context of estimating a multivariate mean. Others references on BMA include Marty *et al.* [22] for reliable ensemble forecasting, Simmons *et al.* [23] for benchmark dose estimation, Fan and Wang [24] (Autoregressive regression models), Corani and Mignatti [25] (presence-absence data), Tsiotas [26] (quantile regression), Baran [27] (truncated normal components), Lenkoski *et al.* [28] (endogenous variables), Fan *et al.* [29] (regression models), Madadgar [30] (integration of copulas), Koop *et al.* [31] (instrumental variables), and Clyde *et al.* [32] (variable selection).

Clyde and Iversen [33] developed a variant of BMA in which it is not assumed that the true model belongs to competing ones (M-open framework). They developed an optimal weighted scheme and showed that their method provides accurate predictions than any of the proxy models.

An R [34] package for BMA is now available for computational purposes; this package provides ways for carrying out BMA for linear regression, generalized linear models, and survival analysis using Cox proportional hazard models. For computations, Monte Carlo methods, or approximating methods, are used; thus many BMA applications are based on the BIC. As one can realize in deriving BMA, there are no unique statistical model and unique prior distribution associated with BMA, taught these are available for each competing model. This renders frequentist properties of BMA hard to obtain from pure Bayesian decision theory. This was the main motivation of this paper in proposing alternative Bayesian model in which the long run properties of resulted estimators could be automatically obtained from Bayesian decision theory. The present paper is organized as follows. Section 2 introduces the new BMA method; Section 3 provides practical examples while Section 4 provides discussions. The paper ends with concluding remarks.

2. BMA Based on Mixture

2.1. The Model

The purpose of this section is to define a new BMA method. The prior of the quantity of interest can be defined as

$$P(\mu) = \sum_{k=1}^K P(\mu|M_k) P(M_k), \quad (8)$$

where $P(\mu|M_k)$ is the prior distribution of M_k .

The parametric statistical model $P(x|\mu)$ can also be defined as

$$P(x|\mu) = \sum_{k=1}^K P(x|\mu, M_k) P(M_k); \quad (9)$$

with $P(x|\mu, M_k)$ being the parametric statistical model for model M_k (*i.e.* the sampling distribution of M_k). The use of Bayes rule leads to the posterior of the quantity of interest $P(\mu|x)$ as

$$P(\mu|x) = \frac{P(x|\mu)P(\mu)}{\int_{\Gamma} P(x|\mu)P(\mu)d\mu}. \quad (10)$$

Defining a loss function, Bayesian estimates are then obtained with its long and short run properties known. All

the frequentist properties of Bayes rules now apply, in particular one can find conditions under which there are consistent and admissible. This approach is referred to as *Mixed based Bayesian model averaging* (MBMA).

2.2. Theoretical Properties of MBMA

Proposition 1. Under (8) and (9), assuming that for all k and j , $k \neq j$,

$h_{kj}(x) = \int_{\Gamma} P(x|\mu, M_k)P(\mu|M_j)d\mu < \infty$, the posterior of the quantity of interest in (10) is

$$P(\mu|x) = \frac{\sum_{k=1}^K P(x|M_k)P(\mu|x, M_k)P^2(M_k) + \sum_{k=1}^K \sum_{j=1; j \neq k}^K P(x|\mu, M_k)P(\mu|M_j)P(M_k)P(M_j)}{\sum_{k=1}^K P(x|M_k)P^2(M_k) + \sum_{k=1}^K \sum_{j=1; j \neq k}^K h_{kj}(x)P(M_k)P(M_j)}. \quad (11)$$

Proof.

$$\begin{aligned} P(x|\mu)P(\mu) &= \left\{ \sum_{k=1}^K P(x|\mu, M_k)P(M_k) \right\} \left\{ \sum_{k=1}^K P(\mu|M_k)P(M_k) \right\} \\ &= \sum_{k=1}^K P(x|\mu, M_k)P(M_k)P(\mu|M_k)P(M_k) + \sum_{k=1}^K \sum_{j=1; j \neq k}^K P(x|\mu, M_k)P(M_k)P(\mu|M_j)P(M_j) \\ &= \sum_{k=1}^K P^2(M_k)P(\mu|M_k)P(x|\mu, M_k) + \sum_{k=1}^K \sum_{j=1; j \neq k}^K P(M_k)P(M_j)P(x|\mu, M_k)P(\mu|M_j). \end{aligned}$$

Since, $P(\mu|M_k)P(x|\mu, M_k) = P(x|M_k)P(\mu|x, M_k)$ by Bayes rule,

$$\begin{aligned} P(x|\mu)P(\mu) &= \sum_{k=1}^K P^2(M_k)P(x|M_k)P(\mu|x, M_k) \\ &\quad + \sum_{k=1}^K \sum_{j=1; j \neq k}^K P(M_k)P(M_j)P(x|\mu, M_k)P(\mu|M_j). \end{aligned} \quad (a)$$

$$\begin{aligned} P(x) &= \int_{\Gamma} P(x|\mu)P(\mu)d\mu \\ &= \sum_{k=1}^K P^2(M_k) \int_{\Gamma} P(\mu|M_k)P(x|\mu, M_k)d\mu \\ &\quad + \sum_{k=1}^K \sum_{j=1; j \neq k}^K P(M_k)P(M_j) \int_{\Gamma} P(x|\mu, M_k)P(\mu|M_j)d\mu \\ &= \sum_{k=1}^K P(x|M_k)P^2(M_k) + \sum_{k=1}^K \sum_{j=1; j \neq k}^K h_{kj}(x)P(M_k)P(M_j). \end{aligned} \quad (b)$$

Dividing (a) by (b) yields the result.

Corollary 2. Suppose that all the models have identical sampling distribution, that is $P(x|\mu, M_k) = P(x|\mu, M_j)$ for all k and j , then MBMA reduces to BMA.

Proof.

In the numerator of (11),

$$\begin{aligned} &\sum_{k=1}^K \sum_{j=1; j \neq k}^K P(x|\mu, M_k)P(\mu|M_j)P(M_k)P(M_j) \\ &= \sum_{k=1}^K P(\mu|M_k)P(x|\mu, M_k)P(M_k) \sum_{j=1; j \neq k}^K P(M_j), \sum_{j=1; j \neq k}^K P(M_j) = 1 - P(M_k) \\ &= \sum_{k=1}^K P(\mu|M_k)P(x|\mu, M_k)P(M_k)(1 - P(M_k)). \end{aligned}$$

The numerator of (11) is therefore

$$\begin{aligned} &= \sum_{k=1}^K P(x|M_k)P(\mu|x, M_k)P^2(M_k) + \sum_{k=1}^K P(x|M_k)P(\mu|x, M_k)P(M_k)(1 - P(M_k)) \\ &= \sum_{k=1}^K P(x|M_k)P(\mu|x, M_k)P^2(M_k) - \sum_{k=1}^K P(x|M_k)P(\mu|x, M_k)P^2(M_k) \\ &\quad + \sum_{k=1}^K P(x|M_k)P(\mu|x, M_k)P(M_k) \\ &= \sum_{k=1}^K P(x|M_k)P(\mu|x, M_k)P(M_k). \end{aligned}$$

$$h_{kj}(x) = \int_{\Gamma} P(x|\mu, M_k)P(\mu|M_j)d\mu = \int_{\Gamma} P(x|\mu, M_k)P(\mu|M_k)d\mu = P(x|M_k).$$

Therefore, the denominator of (11),

$$\begin{aligned}
P(x) &= \sum_{k=1}^K P(x|M_k)P^2(M_k) + \sum_{k=1}^K \sum_{j=1; j \neq k}^K P(x|M_k)P(M_k)P(M_j) \\
&= \sum_{k=1}^K P(x|M_k)P^2(M_k) + \sum_{k=1}^K P(x|M_k)P(M_k) \sum_{j=1; j \neq k}^K P(M_j) \\
&= \sum_{k=1}^K P(x|M_k)P^2(M_k) + \sum_{k=1}^K P(x|M_k)P(M_k)(1-P(M_k)) \\
&= \sum_{k=1}^K P(x|M_k)P^2(M_k) - \sum_{k=1}^K P(x|M_k)P^2(M_k) + \sum_{k=1}^K P(x|M_k)P(M_k) \\
&= \sum_{k=1}^K P(x|M_k)P(M_k),
\end{aligned}$$

a mixture of marginal distributions.

Therefore

$$\begin{aligned}
P(\mu|x) &= \frac{\sum_{k=1}^K P(x|M_k)P(\mu|x, M_k)P(M_k)}{\sum_{j=1}^K P(x|M_j)P(M_j)} \\
&= \sum_{k=1}^K \left\{ \frac{P(x|M_k)P(M_k)}{\sum_{j=1}^K P(x|M_j)P(M_j)} \right\} P(\mu|x, M_k) \\
&= \sum_{k=1}^K P(\mu|x, M_k)P(M_k|x) = P(\mu|x).
\end{aligned}$$

Thus in this special case, the posteriors mean and variance using the MBMA are those of BMA given in Equations (6) and (7).

2.3. Frequentist (Long Run) Evaluation of MBMA

Evaluating the long run properties of MBMA involves studying frequentist issues, including: asymptotic methods, consistency, efficiency, unbiasedness, and admissibility. Details about derivations for more general Bayes estimates can be found e.g. in Gelman [35] (p. 83). The following are proven in Gelman [35] for any Bayes estimate, in particular for MBMA. Let $J(\mu)$ be the Fisher information, $I(\mu)$ the observed information, Mod the posterior mode; and μ_0 the value of the parameter that makes the model distribution closest (e.g. in the sense of Kullback-Leiber information) to the true distribution.

1) If the sample size is large and the posterior distribution $P(\mu|x)$ is unimodal and roughly symmetric, one can approximate it by a normal distribution centered at Mod with variance $(I(Mod))^{-1}$.

2) If the likelihood ($P(\mu|x)$) is a continuous function of μ and the true parameter value μ_0 is not on the boundary of the parameter space, as the sample size n tends to ∞ , the posterior distribution of μ approaches normality with mean μ_0 and variance $(nJ(\mu_0))^{-1}$ and Mod is consistent for μ_0 .

3) Suppose the normal approximation for the posterior distribution ($P(\mu|x)$), $Mod \rightarrow \mu_0$ and the true data distribution is included in the class of models, then $[I(Mod)]^{1/2}(\mu - Mod) \sim N(0, I)$, where I is the identity matrix.

4) When the truth is included in the family of models ($P(\mu|x)$) being fitted, the posteriors mode, mean and median are consistent and asymptotically unbiased and efficient under mild regularity conditions.

5) If a prior distribution ($P(\mu)$) is strictly positive with finite Bayes risk and the risk function is continuous, MBMA is admissible.

2.4. Predictive Performance of MBMA

One measure of predictive performance is the Good's logarithm score rule [36]. From the nonnegativity of Kullback-Leiber information divergence, it follows that if f and g two probabilities distribution functions,

$$E_f(\log f(X)) \geq E_f(\log g(X)).$$

Applying this to MBMA leads to

$$E[\log P(\mu|x)] \geq E[\log P(\mu|x, M_k)], \quad k=1, \dots, K. \quad (12)$$

MBMA provides thus better predictive performance than any single model.

3. Applications

Laplace distribution, $\text{Laplace}(\theta, \sigma^2)$ and normal distribution, $N(\theta, \sigma^2)$ are often used in finance.

$$X \sim N(\theta, \sigma^2), \quad f_1(x|\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\theta)^2/2\sigma^2}, \quad x \in R, \quad E(X) = \theta, \quad \text{Var}(X) = \sigma^2$$

$$X \sim \text{Laplace}(\theta, \sigma^2), \quad f_2(x|\theta, \sigma^2) = \frac{1}{2\sigma} e^{-|x-\theta|/\sigma}, \quad x \in R,$$

$$E(X) = \theta, \quad \text{Var}(X) = 2\sigma^2.$$

Laplace distribution (the double exponential) is symmetric with fat tails (much fatter than the normal). It is not bell-shaped (it has a peak at $x = \theta$).

Suppose that the mean is known and the quantity of interest is $\mu = \sigma^2$. The data are the daily foreign exchange rates Euros versus US Dollars from January 3 2000 till June 15 2006 (the aim being their return value). The prior for σ^2 is $\text{Gamma}(\sigma^2, \alpha, \beta)$ for both models (the idea remains the same for different priors, e.g., uniform priors). Model probabilities are assigned for each model.

Table 1 shows the properties of the competing models, BMA, and MBMA. Starting from equal prior for M_1 and M_2 , *i.e.* 0.5 each; after observing data, M_1 is more likely to be true (0.83) than M_2 (0.17). While M_1 , M_2 and MBMA have priors (over the parameter of interest) and statistical models; BMA does not have. This implies that the frequentist properties of MBMA can be automatically derived from Bayesian decision theory (see Sub-section); this is not possible for BMA. The bayesian estimates (conditional on the observations) of these models are very similar, with MBMA having the smaller conditional variance (0.03).

4. Discussion

In general, as Bayes estimate, the form of the posteriors mean and variance for MBMA are not known in advance; in a special case, the properties of MBMA are those of BMA and are given in Equations (6) and (7). Posterior distributions of MBMA are very complex, thus a major challenge is in computing. MBMA estimate is thus computationally demanding (but feasible) since the posterior $P(\mu|x)$ involves many sums, especially if the number K of models is large. This is not new as BMA faces the same drawback, though nowadays program exist for complex computations (e.g. R [34]). Another problem is the selection of priors both for models and parameters (common to any Bayesian model). In most cases, uniform priors are used for each model, *i.e.* $P(M_k) = 1/K$, $k = 1, 2, \dots, K$. When the number of models is large, model search strategies are sometimes used to reduce the set of models (e.g. Occam's window method, Hoeting *et al.* [19]), by eliminating those that seem comparatively less compatible with the data. Most currently Bayesian mixtures are based either in the priors or on the statistical model, not both as the new MBMA described in this paper. For example Abd and Al-Zaydi [37] [38] used statistical mixtures model for order statistics; Al-Hussaini and Hussein [39] for exponential components; Ley and Steel [40] used a prior of mixtures with economic applications. Other Bayesian mixtures include Schäfer *et al.* [41] (spatial clustering), Yao [42] (Bayesian labeling), Sabourin and Naveau [43] (extremes), and Rodriguez and Walker [44] kernel estimation). Programming codes are under development for performing model averaging using MBMA with real data and simulations, and will be available as an add-on package on R [34].

Table 1. Comparison of inference for the scale, $\mu = \sigma^2$.

Model	$P(M)$	$P(x \mu, M)$	$P(\mu M)$	$P(M x)$	$E(\mu x, M)$	$\text{Var}(\mu x, M)$
M_1	0.5	$N(\mu, \sigma^2)$	$\text{Ga}(\alpha_1, \beta_1)$	0.83	0.373	0.033
M_2	0.5	$\text{Laplace}(\mu, \sigma^2)$	$\text{Ga}(\alpha_2, \beta_2)$	0.17	0.345	0.052
MBMA	NA	Mixture	Mixture	NA	0.323	0.030
BMA	NA	NA	NA	NA	0.368	0.036

5. Concluding Remarks

This paper proposes a new method (with application) for model averaging in Bayesian context (MBMA) when the main focus of a data analyst is on the long run (frequentist) performances of the Bayesian estimator. The method is based on using a mixture of priors and sampling distributions for model averaging. When conditioning on data at hand, the well popular Bayesian model averaging (BMA) should be preferable, given the complexity in computing of MBMA. MBMA is especially useful when exploiting the well known frequentist properties within the framework of Bayesian decision theory.

Acknowledgements

We thank the editor and the referee for their comments on earlier versions of this paper.

References

- [1] Leeb, H. and Pötscher, B.M. (2009) Model Selection. In: *Handbook of Financial Time Series*, Springer, Berlin Heidelberg, 889-925. http://dx.doi.org/10.1007/978-3-540-71297-8_39
- [2] Zucchini, W. (2000) An Introduction to Model Selection. *Journal of Mathematical Psychology*, **44**, 41-61. <http://dx.doi.org/10.1006/jmps.1999.1276>
- [3] Zucchini, W., Claeskens, G. and Nguefack-Tsague, G. (2011) Model Selection. In: *International Encyclopedia of Statistical Science*, Springer, Berlin Heidelberg, 830-833. http://dx.doi.org/10.1007/978-3-642-04898-2_373
- [4] Nguefack-Tsague, G. (2013) An Alternative Derivation of Some Commons Distributions Functions: A Post-Model Selection Approach. *International Journal of Applied Mathematics and Statistics*, **42**, 138-147.
- [5] Burnham, K.P. and Anderson, D.R. (2013) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, Cambridge.
- [6] Nguefack-Tsague, G. and Walter, Z. (2011) Post-Model Selection Inference and Model Averaging. *Pakistan Journal of Statistics and Operation Research*, **7**, 347-361. <http://dx.doi.org/10.18187/pjsor.v7i2-Sp.292>
- [7] Nguefack-Tsague, G. (2013) On Bootstrap and Post-Model Selection Inference. *International Journal of Mathematics and Computation*, **21**, 51-64.
- [8] Nguefack-Tsague, G. (2014) Estimation of a Multivariate Mean under Model Selection Uncertainty. *Pakistan Journal of Statistics and Operation Research*, **10**, 131-145. <http://dx.doi.org/10.18187/pjsor.v10i1.449>
- [9] Nguefack-Tsague, G. (2014) On Optimal Weighting Scheme in Model Averaging. *American Journal of Applied Mathematics and Statistics*, **2**, 150-156. <http://dx.doi.org/10.12691/ajams-2-3-9>
- [10] Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, **6**, 461-465. <http://dx.doi.org/10.1214/aos/1176344136>
- [11] Nguefack-Tsague, G. and Ingo, B. (2014) A Focused Bayesian Information Criterion. *Advances in Statistics*, **2014**, Article ID: 504325. <http://dx.doi.org/10.1155/2014/504325>
- [12] Guan, Y. and Stephens, M. (2011) Bayesian Variable Selection Regression for Genome-Wide Association Studies, and Other Large-Scale Problems. *Annals of Applied Statistics, Series B*, **3**, 1780-1815. <http://dx.doi.org/10.1214/11-AOAS455>
- [13] Nguefack-Tsague, G. (2011) Using Bayesian Networks to Model Hierarchical Relationships in Epidemiological Studies. *Epidemiology and Health*, **33**, e201100633. <http://dx.doi.org/10.4178/epih/e2011006>
- [14] Carvalho, C.M. and Scott, J.G. (2009) Objective Bayesian Model Selection in Gaussian Graphical Models. *Biometrika*, **96**, 497-515. <http://dx.doi.org/10.1093/biomet/asp017>
- [15] Fridley, B.L. (2009) Bayesian Variable and Model Selection Methods for Genetic Association Studies. *Genetic Epidemiology, Series B*, **33**, 27-37. <http://dx.doi.org/10.1002/gepi.20353>
- [16] Robert, C. (2007) *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer, New York.
- [17] Liang, F., Paulo, R., Molina, G., Clyde, M.A. and Berger, J.O. (2008) Mixtures of g Priors for Bayesian Variable Selection. *Journal of the American Statistical Association*, **103**, 174-200. <http://dx.doi.org/10.1198/016214507000001337>
- [18] Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory*. Wiley, New York. <http://dx.doi.org/10.1002/9780470316870>
- [19] Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999) Bayesian Model Averaging: A Tutorial (with Discussions). *Statistical Science*, **14**, 382-417.

- [20] Clyde, M.A. (1999) Bayesian Model Averaging and Model Search Strategies. In: Bernardo, J.M., Dawid, A.P., Berger, J.O. and Smith, A.F.M., Eds., *Bayesian Statistics 6*, Oxford University Press, Oxford, 157-188.
- [21] Clyde, M.A. and George, E.I. (2004) Model Uncertainty. *Statistical Science*, **19**, 81-94. <http://dx.doi.org/10.1214/088342304000000035>
- [22] Marty, R., Fortin, V., Kuswanto, H., Favre, A.-C. and Parent, E. (2015) Combining the Bayesian Processor of Output with Bayesian Model Averaging for Reliable Ensemble Forecasting. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **64**, 75-92. <http://dx.doi.org/10.1111/rssc.12062>
- [23] Simmons, S.J., Chen, C., Li, X., Wang, Y., Piegorsch, W.W., Fang, Q., Hu, B. and Dunn, G.E. (2015) Bayesian Model Averaging for Benchmark Dose Estimation. *Environmental and Ecological Statistics*, **22**, 5-16. <http://dx.doi.org/10.1007/s10651-014-0285-4>
- [24] Fan, T.-H. and Wang, G.-T. (2015) Bayesian Model Averaging in Longitudinal Regression Models with AR (1) Errors with Application to a Myopia Data Set. *Journal of Statistical Computation and Simulation*, **85**, 1667-1678. <http://dx.doi.org/10.1080/00949655.2014.891205>
- [25] Corani, G. and Mignatti, A. (2015) Robust Bayesian Model Averaging for the Analysis of Presence-Absence Data. *Environmental and Ecological Statistics*, **22**, 513-534. <http://dx.doi.org/10.1007/s10651-014-0308-1>
- [26] Tsiotas G. (2015) A Quasi-Bayesian Model Averaging Approach for Conditional Quantile Models. *Journal of Statistical Computation and Simulation*, **85**, 1963-1986. <http://dx.doi.org/10.1080/00949655.2014.913044>
- [27] Baran S. (2014) Probabilistic Wind Speed Forecasting Using Bayesian Model Averaging with Truncated Normal Components. *Computational Statistics and Data Analysis*, **75**, 227-238. <http://dx.doi.org/10.1016/j.csda.2014.02.013>
- [28] Lenkoski, A., Eicher, T.S. and Raftery, A.E. (2014) Two-Stage Bayesian Model Averaging in Endogenous Variable Models. *Econometric Reviews*, **33**, 122-151. <http://dx.doi.org/10.1080/07474938.2013.807150>
- [29] Fan, T.-H., Wang, G.-T. and Yu, J.-H. (2014) A New Algorithm in Bayesian Model Averaging in Regression Models. *Communications in Statistics-Simulation and Computation*, **43**, 315-328. <http://dx.doi.org/10.1080/03610918.2012.700750>
- [30] Madadgar, S. and Moradkhani, H. (2014) Improved Bayesian Multimodeling: Integration of Copulas and Bayesian Model Averaging. *Water Resources Research*, **50**, 9586-9603. <http://dx.doi.org/10.1002/2014WR015965>
- [31] Koop, G., Leon-Gonzalez, R. and Strachan, R. (2012) Bayesian Model Averaging in the Instrumental Variable Regression Model. *Journal of Econometrics*, **171**, 237-250. <http://dx.doi.org/10.1016/j.jeconom.2012.06.005>
- [32] Clyde, M.A., Ghosh, J. and Littman, M.L. (2011) Bayesian Adaptive Sampling for Variable Selection and Model Averaging. *Journal of Computational and Graphical Statistics*, **20**, 80-101. <http://dx.doi.org/10.1198/jcgs.2010.09049>
- [33] Clyde, M.A. and Iversen, E.S. (2015) Bayesian Model Averaging in the M-Open Framework. In: Damien, P., Dellaportas, P., Polson, N.G. and Stephens, D.A., Eds., *Bayesian Theory and Applications*, Oxford University Press, Oxford, 483-498.
- [34] R Development Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
- [35] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2014) Bayesian Data Analysis. Chapman and Hall/CRC, London.
- [36] Good, I.J. (1952) Rational Decisions. *Journal of the Royal Statistical Society, Series B*, **14**, 107-114.
- [37] Abd, E.B.A. and Al-Zaydi, A.M. (2015) Bayesian Prediction of Future Generalized Order Statistics from a Class of Finite Mixture Distributions. *Open Journal of Statistics*, **5**, 585-599. <http://dx.doi.org/10.4236/ojs.2015.56060>
- [38] Abd, E.B.A. and Al-Zaydi, A.M. (2013) Inferences under a Class of Finite Mixture Distributions Based on Generalized Order Statistics. *Open Journal of Statistics*, **3**, 231-244. <http://dx.doi.org/10.4236/ojs.2013.34027>
- [39] Al-Hussaini, E.K. and Hussein, M. (2012) Estimation under a Finite Mixture of Exponentiated Exponential Components Model and Balanced Square Error Loss. *Open Journal of Statistics*, **2**, 28-38. <http://dx.doi.org/10.4236/ojs.2012.21004>
- [40] Ley, E. and Steel, M.F.J. (2012) Mixtures of g-Priors for Bayesian Model Averaging with Economic Applications. *Journal of Econometrics*, **171**, 251-266. <http://dx.doi.org/10.1016/j.jeconom.2012.06.009>
- [41] Schäfer, M., Radon, Y., Klein, T., Herrmann, S., Schwender, H., Verveer, P.J. and Ickstadt, K. (2015) A Bayesian Mixture Model to Quantify Parameters of Spatial Clustering. *Computational Statistics and Data Analysis*, **92**, 163-176. <http://dx.doi.org/10.1016/j.csda.2015.07.004>
- [42] Yao, W. (2012) Model Based Labeling for Mixture Models. *Statistics and Computing*, **22**, 337-347. <http://dx.doi.org/10.1007/s11222-010-9226-8>

- [43] Sabourin, A. and Naveau, P. (2014) Bayesian Dirichlet Mixture Model for Multivariate Extremes: A Re-Parametrization. *Computational Statistics and Data Analysis*, **71**, 542-567. <http://dx.doi.org/10.1016/j.csda.2013.04.021>
- [44] Rodriguez, C.E. and Walker, S.G. (2014) Univariate Bayesian Nonparametric Mixture Modeling with Unimodal Kernels. *Statistics and Computing*, **24**, 35-49. <http://dx.doi.org/10.1007/s11222-012-9351-7>