Statistics and Computing
https://doi.org/10.1007/s11222-018-9828-0

Quantifying uncertainty in transdimensional Markov chain Monte Carlo using discrete Markov models

Daniel W. Heck[1] · **Antony M. Overstall**[2] · **Quentin F. Gronau**[3] · **Eric-Jan Wagenmakers**[3]

Received: 9 April 2018 / Accepted: 3 August 2018

Abstract
Bayesian analysis often concerns an evaluation of models with different dimensionality as is necessary in, for example, model selection or mixture models. To facilitate this evaluation, transdimensional Markov chain Monte Carlo (MCMC) relies on sampling a discrete indexing variable to estimate the posterior model probabilities. However, little attention has been paid to the precision of these estimates. If only few switches occur between the models in the transdimensional MCMC output, precision may be low and assessment based on the assumption of independent samples misleading. Here, we propose a new method to estimate the precision based on the observed transition matrix of the model-indexing variable. Assuming a first-order Markov model, the method samples from the posterior of the stationary distribution. This allows assessment of the uncertainty in the estimated posterior model probabilities, model ranks, and Bayes factors. Moreover, the method provides an estimate for the effective sample size of the MCMC output. In two model selection examples, we show that the proposed approach provides a good assessment of the uncertainty associated with the estimated posterior model probabilities.

Keywords Reversible jump MCMC · Product space MCMC · Bayesian model selection · Posterior model probabilities · Bayes factor

1 Introduction

Transdimensional Markov chain Monte Carlo (MCMC) methods provide an indispensable tool for the Bayesian analysis of models with varying dimensionality (Sisson 2005). An important application is Bayesian model selection, where the aim is to estimate posterior model probabilities $p(\mathcal{M}_i \mid \boldsymbol{y})$ for a set of models $\mathcal{M}_i$, $i = 1, \ldots, I$ given the data $\boldsymbol{y}$ (Kass and Raftery 1995). In order to ensure that the Markov chain converges to the correct stationary distribution, transdimensional MCMC methods such as reversible jump MCMC (Green 1995) or the product space approach (Carlin and Chib 1995) match the dimensionality of parameter spaces across different models (e.g., by adding parameters and link functions). Transdimensional MCMC methods have proven to be very useful for the analysis of many statistical models including capture–recapture models (Arnold et al. 2010), generalized linear models (Forster et al. 2012), factor models (Lopes and West 2004), and mixture models (Frühwirth-Schnatter 2001), and are widely used in substantive applications such as selection of phylogenetic trees (Opgen-Rhein et al. 2005), gravitational wave detection in

Daniel W. Heck, Statistical Modeling in Psychology, University of Mannheim, Germany, heck@uni-mannheim.de. R code for all simulations is available at the Open Science Framework at https://osf.io/kjrkz, and the R package `MCMCprecision` is available at https://CRAN.R-project.org/package=MCMCprecision.

✉ Daniel W. Heck
heck@uni-mannheim.de

Antony M. Overstall
A.M.Overstall@soton.ac.uk

Quentin F. Gronau
quentingronau@web.de

Eric-Jan Wagenmakers
ej.wagenmakers@gmail.com

[1] Statistical Modeling in Psychology, University of Mannheim, Mannheim, Germany

[2] School of Mathematical Sciences and Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, UK

[3] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

Published online: 09 August 2018

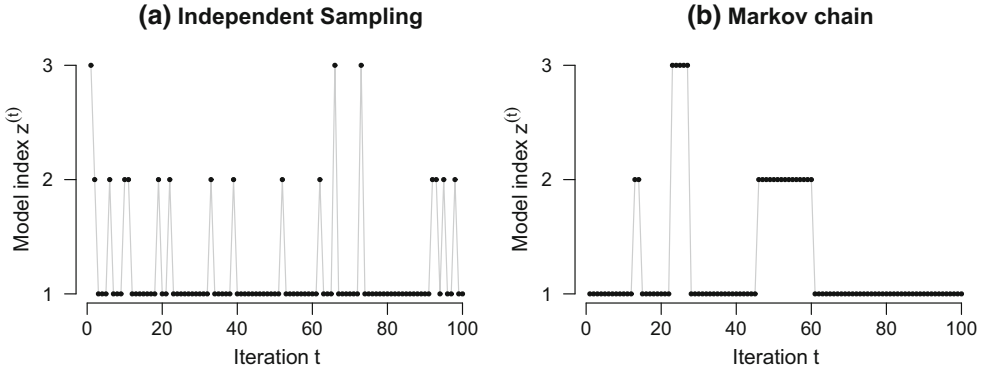

Fig. 1 Illustration of $T = 100$ iterations of a discrete model-indexing variable $z^{(t)}$ that were sampled from **a** independent categorical distributions and **b** a Markov model with positive autocorrelation (cf. Sect. 3). Using the method proposed in Sect. 2.3, the estimated effective sample sizes were $\widehat{T}_{\text{eff}} = 96$ and $\widehat{T}_{\text{eff}} = 8$, respectively

physics (Karnesis 2014), or cognitive models in psychology (Lodewyckx et al. 2011; Heck et al. 2017).

Crucially, transdimensional MCMC methods always include a discrete parameter z with values in $1, \ldots, I$ indexing the competing models. At iteration $t = 1, \ldots, T$, posterior samples are obtained for the indexing variable $z^{(t)}$ and the model parameters, which are usually continuous and differ in dimensionality (for a review, see Sisson 2005). For instance, a Gibbs sampling scheme can be adopted (Barker and Link 2013), in which the indexing variable z and the continuous model parameters are updated in alternating order. Such a sampler switches between models depending on the current values of the continuous parameters, and then updates these parameters in light of the current model $\mathcal{M}_i$ conditionally on the value of $z^{(t)} = i$ (Barker and Link 2013). Given convergence of the MCMC chain, the sequence $z^{(t)}$ follows a discrete stationary distribution with probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_I)^\top$. Due to the design of the sampler, these probabilities are identical to the posterior model probabilities of interest, $\pi_i = p(\mathcal{M}_i \mid \boldsymbol{y})$ and, given uniform model priors $p(\mathcal{M}_i) = 1/I$, also proportional to the marginal likelihoods $p(\boldsymbol{y} \mid \mathcal{M}_i)$. Hence, transdimensional MCMC samplers can be used to directly estimate these posterior probabilities as the relative frequencies of samples $z^{(t)}$ falling into the I categories, $\hat{\pi}_i = 1/T \sum_t \mathbb{I}(z^{(t)} = i)$, where $\mathbb{I}$ is the indicator function. Due to the ergodicity of the Markov chain, this estimator is ensured to be asymptotically unbiased (Green 1995; Carlin and Chib 1995).

Usually, dependencies due to MCMC sampling are taken into account for continuous parameters (Jones et al. 2006; Flegal and Gong 2015; Doss et al. 2014). In contrast, however, the estimate $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \ldots, \hat{\pi}_I)^\top$ based on the sequence of discrete samples $z^{(t)}$ is usually reported without quantifying estimation uncertainty due to MCMC sampling. Often, the samples $z^{(t)}$ are correlated to a substantial, but unknown, degree because of infrequent switching between models. This is illustrated in Fig. 1, which shows a sequence of independent and correlated samples $z^{(t)}$ in Panels A and B, respectively. Inference about the stationary distribution $\boldsymbol{\pi}$ is more reliable in the first case compared to the second case, in which the autocorrelation reduces the amount of information available about $\boldsymbol{\pi}$ (cf. Sect. 3). The standard error $\text{SE}(\hat{\pi}_i) = \sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)/T}$ that assumes independent sampling will obviously underestimate the true variability of the estimate $\hat{\boldsymbol{\pi}}$ if samples are correlated (Green 1995; Sisson 2005). To obtain a measure of precision, Green (1995) proposed running several independent MCMC chains $c = 1, \ldots, C$ and computing the standard deviation of the estimates $\hat{\boldsymbol{\pi}}^{(c)}$ across these independent replications. However, for complex models, this method might require a substantial amount of additional computing time for burn-in and adaption and thus can be infeasible in practice.

Assessing the precision of the estimate $\hat{\boldsymbol{\pi}}$, which depends on the autocorrelation of the sequence of discrete MCMC samples $z^{(t)}$, is of major importance. In case of model selection, it must be ensured that the estimated posterior probabilities $p(\mathcal{M}_i \mid \boldsymbol{y})$ are sufficiently precise for drawing substantive conclusions. This issue is especially important when estimating a ratio of marginal probabilities, that is, the Bayes factor $B_{ij} = p(\boldsymbol{y} \mid \mathcal{M}_i)/p(\boldsymbol{y} \mid \mathcal{M}_j)$ (Jeffreys 1961). Moreover, it is often of interest to compute the effective sample size defined as the number of independent samples that would provide the same amount of information as the given MCMC output for estimating $\boldsymbol{\pi}$ with $\hat{\boldsymbol{\pi}}$. Besides providing an intuitive measure of precision, a minimum effective sample size can serve as a principled and theoretically justified stopping rule for MCMC sampling (Gong and Flegal 2016). However, standard methods of estimating the effective sample size (e.g., computing the spectral density at zero; Plummer et al. 2006; Heidelberger and Welch 1981) are tailored to continuous parameters. When applied to the model-indexing variable $z^{(t)}$ of a transdimensional

MCMC method, these methods neglect the discreteness of $z^{(t)}$. Depending on the specific numerical labels used for the different models (e.g., $(1, 2, 3, 4)$ vs. $(1, 4, 2, 3)$), spectral decomposition can lead to widely varying and arbitrary estimates for the effective sample size (see Sect. 4).

In summary, transdimensional MCMC is a very important and popular method for Bayesian inference (Sisson 2005). However, little attention has been paid to the analysis of the resulting MCMC output, which requires that one takes into account the autocorrelation as well as the discrete nature of the model-indexing variable. As a solution, we propose to fit a discrete, first-order Markov model to the MCMC output $z^{(t)}$ to assess the precision of the estimated stationary distribution $\hat{\boldsymbol{\pi}}$. Whereas several diagnostics have previously been proposed to assess the *convergence* of transdimensional MCMC samplers (e.g., Brooks and Giudici 2000; Castelloe and Zimmerman 2002; Brooks et al. 2003a; Sisson and Fan 2007), we are unaware of any methods that quantify the *precision* of the point estimate $\hat{\boldsymbol{\pi}}$.

2 Method

2.1 A discrete Markov model for transdimensional MCMC output

The proposed method approximates the output of a transdimensional MCMC method (i.e., the sampled iterations $z^{(t)}$) by a discrete Markov model $\mathcal{M}^{\text{Markov}}$ with transition matrix $\boldsymbol{P}$. This model explicitly accounts for autocorrelation, which in turn allows quantifying estimation uncertainty for the discrete stationary distribution $\boldsymbol{\pi}$. The entries of $\boldsymbol{P}$ are defined as the transition probabilities $p_{ij} = P(z^{(t+1)} = j \mid z^{(t)} = i)$ for all $i, j = 1, \ldots, I$, with rows summing to one, $\sum_{j=1}^{I} p_{ij} = 1$. According to the discrete Markov model, the probability distribution of the indexing variable $z^{(t)}$ at iteration t is given by multiplying the transposed initial distribution $\boldsymbol{\pi}_0^\top$ by the transition matrix t times, $P(z^{(t)} = i) = [\boldsymbol{\pi}_0^\top \boldsymbol{P}^t]_i$. The proposed method estimates the transition matrix $\boldsymbol{P}$ as a free parameter based on the sufficient statistic $\boldsymbol{N}$, the matrix of frequencies n_{ij} counting the observed transitions from $z^{(t)} = i$ to $z^{(t+1)} = j$ (Anderson and Goodman 1957).

Due to the construction of the transdimensional MCMC sampler, the discrete indexing variable $z^{(t)}$ follows a stationary distribution with a constant probability vector $\boldsymbol{\pi}$ (i.e., the posterior model probabilities of interest). Hence, when modeling the sequence $z^{(t)}$ with the discrete Markov model $\mathcal{M}^{\text{Markov}}$, this implies that the transition matrix $\boldsymbol{P}$ must satisfy the condition for stationarity

$$\boldsymbol{\pi}^\top \boldsymbol{P} = 1 \cdot \boldsymbol{\pi}^\top, \tag{1}$$

meaning that the probability vector $\boldsymbol{\pi}$ is the left eigenvector of the matrix $\boldsymbol{P}$ with eigenvalue one (with $\boldsymbol{\pi}$ normalized to sum to one; Anderson and Goodman 1957). Based on the model $\mathcal{M}^{\text{Markov}}$, an estimator for $\boldsymbol{\pi}$ is thus obtained by computing the eigenvector of $\boldsymbol{P}$ with eigenvalue one (Barker and Link 2013).

However, we are less interested in a new estimator $\hat{\boldsymbol{\pi}}$ of the stationary distribution but rather in the precision of this estimate. To quantify estimation uncertainty, we thus fit the model $\mathcal{M}^{\text{Markov}}$ with $\boldsymbol{P}$ as a free parameter in a Bayesian framework by drawing posterior samples $\boldsymbol{P}^{(r)}$ $(r = 1, \ldots, R)$. Similar to a parametric bootstrap, this Bayesian sampling approach has the advantage that we can easily quantify estimation uncertainty (i.e., the dispersion of the posterior distribution of $\boldsymbol{P}$) by computing descriptive statistics of the samples $\boldsymbol{P}^{(r)}$ (e.g., the standard deviation or credibility intervals). Moreover, we can directly quantify the estimation uncertainty of derived quantities such as the posterior model probabilities, model ranks, or Bayes factors (see Sect. 2.2). In the following, it is important to distinguish between the posterior distribution of $\boldsymbol{P}$ given the sufficient statistic $\boldsymbol{N}$, which quantifies the uncertainty of $\boldsymbol{P}$ due to estimation error of the transdimensional MCMC method, and the posterior distribution of the models given the empirical data, which is represented by the constant vector of probabilities $\boldsymbol{\pi}$ for a specific data set.

Next, we define a prior distribution for the parameter $\boldsymbol{P}$ of the model $\mathcal{M}^{\text{Markov}}$. Given that the transition matrix $\boldsymbol{P}$ includes one probability vector $\boldsymbol{p}_i$ for each row i, we assume independent Dirichlet distributions with parameter $\epsilon \geq 0$ for each row,

$$\boldsymbol{p}_i \equiv (p_{i1}, \ldots, p_{iI}) \sim \mathcal{D}(\epsilon, \ldots, \epsilon). \tag{2}$$

Conditional on the MCMC output $\boldsymbol{N}$, the estimation uncertainty of $\boldsymbol{P}$ is approximated by drawing R posterior samples $\boldsymbol{P}^{(r)}$. Since the Dirichlet prior is conjugate to the multinomial distribution, independent samples $\boldsymbol{P}^{(r)}$ can efficiently be drawn from the Dirichlet distribution with parameters

$$\boldsymbol{p}_i^{(r)} \sim \mathcal{D}(n_{i1} + \epsilon, \ldots, n_{iI} + \epsilon). \tag{3}$$

Based on these samples, the estimation uncertainty of the stationary probabilities $\boldsymbol{\pi}$ is assessed by computing the (normalized) eigenvector with eigenvalue one for each sample $\boldsymbol{P}^{(r)}$ (Eq. 1). Algorithm 1 provides an overview of the computational steps of the proposed method as pseudo-code.

With regard to the prior parameter ϵ, small values should be chosen to reduce its influence on the estimation of $\boldsymbol{P}$. In principle, the improper prior $\epsilon = 0$ can be used, which minimizes the impact of the prior on the estimated stationary distribution. This improper prior also ensures that the results do not hinge on the set of models that could possibly be

Algorithm 1 Quantify uncertainty of $\hat{\boldsymbol{\pi}}$ due to transdimensional MCMC sampling.

procedure MARKOV MODEL
  Sampling $z^{(t)}$: T iterations of model-indexing variable z via transdimensional MCMC
  Compute $\boldsymbol{N}$: Observed $I \times I$ transition matrix of $z^{(t)}$ with elements n_{ij}
  Set prior parameter ϵ (default: $\epsilon = 1/I^*$ for the I^* models observed in $z^{(t)}$, $\epsilon = 0$ otherwise)
  for $r = 1, \ldots, R$ **do**
    Initialize posterior sample $\boldsymbol{P}^{(r)}$: $I \times I$ transition matrix with rows $\boldsymbol{p}_i^{(r)}$
    for $i = 1, \ldots, I$ **do**
      Sampling $\boldsymbol{p}_i^{(r)} \sim \text{Dirichlet}(n_{i1} + \epsilon, \ldots, n_{iI} + \epsilon)$
    Initialize posterior sample $\boldsymbol{\pi}^{(r)}$: Posterior model probabilities
    $\boldsymbol{\pi}^{(r)} \leftarrow$ (normalized) eigenvector of $\boldsymbol{P}^{(r)}$ with eigenvalue one
  if (quantify uncertainty) **then**
    Compute summary statistic for all samples $\boldsymbol{\pi}^{(r)}$
    Example: $\text{SD}_{\text{Markov}}(\hat{\pi}_i) \leftarrow \text{SD}(\pi_i^{(r)})$
  if (compute effective sample size) **then**
    Using all $\boldsymbol{\pi}^{(r)}$: Fit Dirichlet parameters $\hat{\alpha}_1, \ldots, \hat{\alpha}_I$ (Minka 2000)
    Compute effect sample size $\widehat{T}_{\text{eff}} \leftarrow \sum_{i=1}^{I} \widehat{\alpha}_i - (I^*)^2\epsilon$

sampled, but were never actually observed in the sequence $z^{(t)}$. For such unsampled models, the corresponding rows and columns of the observed transition matrix $\boldsymbol{N}$ are filled with zeros. With $\epsilon = 0$, the relevant eigenvector of the transition matrix $\boldsymbol{P} \mid \boldsymbol{N}$ is thus identical to that of a reduced matrix $\boldsymbol{P^*} \mid \boldsymbol{N^*}$ that includes only the transitions for the subset of models sampled in $z^{(t)}$. However, in our simulations, this improper Dirichlet prior proved to be numerically unstable and resulted in more variable point estimates than the standard i.i.d. estimate or the proper prior discussed next.

Here, we use the weakly informative prior $\epsilon = 1/I$ as a default, which has an impact equivalent to one observation for each row of the observed transition matrix $\boldsymbol{N}$. By putting a small weight on all values of the transition matrix $\boldsymbol{P}$, this prior serves as a regularization of the posterior (Alvares et al. 2018). However, in scenarios where the number of models exceeds the number of iterations of the transdimensional MCMC method (i.e., $I \gg T$), such a regularization assigns substantial probability weight to models that are never observed in $z^{(t)}$. To limit the effect of the prior, we thus set $\epsilon = 1/I^*$ only for those I^* models that were observed in $z^{(t)}$ and $\epsilon = 0$ for the remaining models. Besides reducing the impact of the prior, this choice has the computational advantage that one can draw posterior samples and compute eigenvectors for the reduced matrix $\boldsymbol{P^*} \mid \boldsymbol{N^*}$ that includes only the sampled models. In the two examples in Sects. 4 and 5, this prior has proved to be numerically robust and resulted in point estimates close to the standard i.i.d. estimates.

As a third alternative, the prior can be adapted to the structure of specific transdimensional MCMC implementations, which only implement switches to a small subset of the competing models. For instance, in variable selection, regression parameters are often added or removed one at a time, resulting in a birth-death process (Stephens 2000). For these kinds of samplers, the Dirichlet parameters ϵ_{ij} can be set to zero selectively. However, such adjustments will be dependent on the chosen MCMC sampling scheme. The default choice of $\epsilon = 1/I^*$ for sampled models and $\epsilon = 0$ for nonsampled models provides a good compromise of being very general and numerically robust, while having a small effect on the posterior. However, in general, the choice of ϵ becomes less influential as the number of MCMC samples increases (especially if the row sums of $\boldsymbol{N}$ are large).

2.2 Estimation uncertainty

Based on the posterior samples $\boldsymbol{P}^{(r)}$ of the transition matrix and the derived model probabilities $\boldsymbol{\pi}^{(r)}$, it is straightforward to estimate the stationary distribution by the posterior mean $\hat{\boldsymbol{\pi}}$ (alternatively, the median or mode may be used). More importantly, however, estimation uncertainty due to the transdimensional MCMC method can directly be assessed by plotting the estimated posterior densities for each π_i. To quantify the precision of the estimate $\hat{\boldsymbol{\pi}}$, one can report posterior standard deviations or credibility intervals for the components $\hat{\pi}_i$. These component-wise summary statistics are most useful if the number of models I is relatively small.

An important advantage of drawing posterior samples $\boldsymbol{\pi}^{(r)}$ in a Bayesian framework (instead of using asymptotic approximations for the standard error of $\hat{\boldsymbol{\pi}}$) is that one can directly quantify estimation uncertainty for other quantities of interest. For very large numbers of sampled models, the assessment of estimation uncertainty can be focused on the subset of k models with the highest posterior model probabilities. Within the sampling approach, estimation uncertainty for the k best-performing models can easily be assessed by computing ranks for each of the posterior samples $\boldsymbol{\pi}^{(r)}$. Then, the variability of these model ranks across the R samples can be summarized, for instance, by the percentage of identical rank orders for the k best-performing models, or the percentages of how often each model is included within the subset of the k best-performing models (i.e., has a rank smaller or equal to k).

In case of model selection, dispersion statistics such as the posterior standard deviation are also of interest with respect to the Bayes factor B_{ij} (Kass and Raftery 1995). To

judge the estimation uncertainty for the Bayes factor, one can evaluate the corresponding posterior distribution by computing the derived quantities $B_{ij}^{(r)} = \pi_i^{(r)}/\pi_j^{(r)}$ (given uniform prior model probabilities). Precision can also be assessed for model-averaging contexts when comparing subsets of models against each other (e.g., regression models including a specific effect vs. those not including it). Given such disjoint sets of model indices $M_s \subset \{1, \ldots, I\}$, the posterior probability for each subset of models is directly obtained by summing the posterior samples $\pi_i^{(r)}$ for all $i \in M_s$. Note that it is invalid to aggregate across model subsets or to drop sampled models before applying the proposed Markov approach because functions of discrete Markov chains (e.g., collapsing the I original states into a subset of S states) are not Markovian in general (Burke and Rosenblatt 1958).

2.3 Effective sample size

Besides quantifying estimation uncertainty, the posterior samples $\boldsymbol{\pi}^{(r)}$ can be used to estimate the effective sample size for the transdimensional MCMC output. For this purpose, we consider the benchmark model $\mathcal{M}^{\text{iid}}$ under the ideal scenario of drawing independent samples $\tilde{z}^{(t)}$ from the categorical distribution with probabilities $\tilde{\boldsymbol{\pi}}$. For this model, we assume an improper Dirichlet prior on the stationary distribution, $\tilde{\boldsymbol{\pi}} \sim \mathcal{D}(0, \ldots, 0)$ (whereas the Markov model assumes a Dirichlet prior on the transition probabilities). Since this prior is conjugate to the multinomial distribution, the posterior for the stationary distribution $\tilde{\boldsymbol{\pi}}$ is given by

$$\tilde{\boldsymbol{\pi}} \mid \tilde{\boldsymbol{n}} \sim \mathcal{D}(\tilde{n}_1, \ldots, \tilde{n}_I), \tag{4}$$

conditional on the observed frequencies $\tilde{n}_i = \sum_{t=1}^{T} \mathbb{I}(\tilde{z}^{(t)} = i)$. Note that the transition frequencies are rendered irrelevant in this i.i.d. model, since there are no dependencies in the sampled iterations $\tilde{z}^{(t)}$.

Given the dependent samples $z^{(t)}$ of a transdimensional MCMC chain, we can now compare the empirical posterior distribution of $\boldsymbol{\pi}$ estimated with the model $\mathcal{M}^{\text{Markov}}$ against the theoretically expected posterior distribution of $\tilde{\boldsymbol{\pi}}$ under the hypothetical model $\mathcal{M}^{\text{iid}}$. Essentially, we match the latter distribution to the former to estimate the effective sample size as the total number of independent samples $T_{\text{iid}} = \sum_i \tilde{n}_i$ that would result in a similar dispersion as that estimated by the Markov model. To estimate the $\tilde{n}_i$, the i.i.d. posterior distribution in Eq. 4 is fitted to the posterior distribution of the Markov model by estimating the shape parameters $\alpha_1, \ldots, \alpha_I$ of a Dirichlet distribution given the sampled $\boldsymbol{\pi}^{(r)}$ (which can be achieved by an efficient maximum-likelihood algorithm by Minka 2000, see Appendix). Next, a comparison of the estimated Dirichlet parameters $\widehat{\alpha}_i$ with the conjugate posterior in Eq. 4 yields $\widehat{\tilde{n}} = \widehat{\alpha}_i$, which implies that the dispersion of the posterior model probabilities $\boldsymbol{\pi}^{(r)}$ is equivalent to having observed $\widehat{T}_{\text{iid}} = \sum_i \hat{\alpha}_i$ independent samples. However, the samples $\boldsymbol{\pi}^{(r)}$ are not only informed by the samples $z^{(t)}$ of the transdimensional MCMC sampler, but also by the prior distribution of the Markov model, which is irrelevant for estimating the effective sample size. Hence, to estimate the effective sample size for the transdimensional MCMC sampler, it is necessary to subtract the prior sample size $I^2\epsilon$ of the Markov model (cf. Eq. 2), which reflects the relative weight of the prior, since the Dirichlet shape parameter ϵ occurs I times in each row of the $I \times I$ transition matrix $\boldsymbol{P}$ (Alvares et al. 2018). Overall, it follows that the effective sample size under the assumption of independent sampling from a multinomial distribution is estimated as

$$\widehat{T}_{\text{eff}} = \sum_{i=1}^{I} \widehat{\alpha}_i - I^2\epsilon. \tag{5}$$

Note that it is necessary to replace I by I^* in Eq. 5 if the Markov model uses only those I^* models that were actually sampled in $z^{(t)}$. Importantly, the estimate $\widehat{T}_{\text{eff}}$ takes the discreteness of the indexing variable z into account and does not depend on specific (but arbitrary) numerical values of the model indices.

2.4 Remarks

The proposed method quantifies estimation uncertainty by fitting a discrete Markov model to transdimensional MCMC output. For this purpose, a simplifying assumption is made that is not guaranteed to hold. Whereas samples of the full model space $(z^{(t)}, \boldsymbol{\theta}^{(t)})$ necessarily follow a Markov process by construction, this does not imply that the samples $z^{(t)}$ follow a Markov chain marginally (Brooks et al. 2003b; Lodewyckx et al. 2011). In practice, the iterations of the model-indexing variable $z^{(t)}$ might have higher-order dependencies since transition probabilities depend on the exact state of the MCMC sampler in each of the models' parameter spaces. However, in Sects. 4 and 5 we show in two empirical examples that the proposed simplification (i.e., fitting a Markov chain of order one) is sufficient to account for autocorrelations in the samples $z^{(t)}$ in practice. Moreover, the approximation by a first-order Markov chain provides a trade-off between ignoring dependencies completely (i.e., assuming i.i.d. samples) and accounting for any higher-order dependencies (which will likely increase the computational burden especially for large numbers of models). Note that it is a common practice to rely on simplifying assumptions for the analysis of simulation output; for instance, a standard approach of estimating the effective sample size for continuous parameters assumes that the output sequence can be modeled as a covariance stationary process with a smooth log spectrum (Heidelberger and Welch 1981).

The proposed method of fitting a discrete Markov model is very general and can be applied irrespective of specific transdimensional MCMC implementations. Moreover, it requires only the sampled sequence $z^{(t)}$ of the discrete parameter or the matrix $\boldsymbol{N}$ with the observed frequency of transitions. If output from multiple independent chains $c = 1, \ldots, C$ is available, the transition frequency matrices $\boldsymbol{N}^{(1)}, \ldots, \boldsymbol{N}^{(C)}$ can simply be summed before applying the method. This follows directly from Bayesian updating of the stationary distribution $\boldsymbol{\pi}$. Essentially, each chain provides independent evidence for the posterior of the transition matrix $\boldsymbol{P}$, which is reflected by using the sums $\sum_c n_{ij}^{(c)}$ for the conjugate Dirichlet prior in Eq. 3. Note that this feature can be used to compare the efficiency of many short versus few long MCMC chains.

In the R package `MCMCprecision` (Heck et al. 2018), we provide an implementation that relies on the efficient computation of eigenvectors in the C++ library `Armadillo` (Sanderson and Curtin 2016), accessible in R via the package `RcppArmadillo` (Eddelbuettel and Sanderson 2014). On a notebook with an Intel® i7-7700HQ processing unit, drawing $R = 5000$ samples from the posterior distribution for 10 (100) sampled models requires approximately 150 ms (28 s). Similar to any MCMC or bootstrap approach, the choice of the number of samples R depends on the summary statistic used to quantify uncertainty. Whereas more samples are required to approximate the density distribution (e.g., $R \geq 5000$), less samples (e.g., $R \approx 1000$) are sufficient to approximate the SD of the estimated posterior model probabilities. Since the samples $\boldsymbol{\pi}^{(r)}$ are independently drawn and SDs are usually sufficient to quantify uncertainty, the choice $R = 1000$ is often sufficient in practice (however, for the simulations below, we use $R = 5000$).

3 Illustration: effect of autocorrelation

Before applying the proposed method to actual output of transdimensional MCMC samplers, we first illustrate its use in an idealized setting, where the interest is in approximating the posterior model probabilities $\boldsymbol{\pi} = (0.85, 0.13, 0.02)^\top$ by drawing random samples $z^{(t)}$. To investigate the effect of independent versus dependent sampling, we generated sequences $z^{(t)}$ from the Markov model $\mathcal{M}^{\text{Markov}}$ with the stationary distribution $\boldsymbol{\pi}$. To induce autocorrelation, we defined a mixture process for each iteration t. With probability β, the discrete indexing variable was identical to the current model, $z_{t+1} = z_t$. In contrast, with probability $1 - \beta$, the value z_{t+1} was sampled from the given stationary distribution $\boldsymbol{\pi}$. Thereby, increasing values of β resulted in a larger autocorrelation of the sequence $z^{(t)}$ as illustrated for $\beta = 0$ and $\beta = 0.8$ in Fig. 1a, b, respectively.

For varying levels of $\beta = 0, 0.1, \ldots, 0.8$, we sampled 500 replications with $T = 1000$ iterations each, applied the proposed method (with $R = 5000$) and computed the precision of the estimate $\hat{\boldsymbol{\pi}}$. The main interest is in the posterior SD and in the coverage probability, defined as the probability that the data-generating values $\boldsymbol{\pi}$ are in the 90% credibility interval defined by the 5% and 95% quantiles. As a benchmark, we also computed these summary statistics under the (false) assumption that the samples $z^{(t)}$ were independently drawn by fitting the model $\mathcal{M}^{\text{iid}}$ with the Dirichlet posterior distribution in Eq. 4. Note that the latter uncertainty estimate is equivalent to the standard Monte Carlo error that assumes independent sampling.

Figure 2 shows the results of this simulation. In Fig. 2a, the three panels correspond to the estimation uncertainty (i.e., the posterior SD) of the three posterior model probabilities $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)^\top$. The estimated posterior SD of the Markov model indicated increasing uncertainty for larger values of β, thus taking the increasing autocorrelation into account. In contrast, the model $\mathcal{M}^{\text{iid}}$ assumes independence a priori, and thus, the posterior uncertainty was independent of β. As a result of this, the corresponding 90% credibility interval was less likely to include the data-generating value $\boldsymbol{\pi}$ for increasing values of β (see Fig. 2b), whereas the Markov model provided an accurate description of the estimation uncertainty for any degree of dependence.

4 Variable selection in logistic regression

In the following, we apply the proposed method to the problem of selecting variables in a logistic regression, an example introduced by Dellaportas et al. (2000) to highlight the implementation of transdimensional MCMC in `BUGS` (see also Dellaportas et al. 2002; Ntzoufras 2002). Table 1 shows the frequencies of deaths and survivals conditional on severity and whether patients received treatment (i.e., antitoxin medication; Healy 1988). To emphasize the importance of considering estimation uncertainty for the posterior model probabilities, we compare the efficiency of two transdimensional MCMC approaches, which can both be implemented in `JAGS` (Plummer 2003).

The full logistic regression model assumes a binomial distribution $\mathcal{B}$ of the survival frequencies y_{jl} and a linear model on the logit-transformed survival probabilities p_{jl},

$$y_{jl} \sim \mathcal{B}(p_{jl}, n_{jl}) \tag{6}$$

$$\log\left(\frac{p_{jl}}{1 - p_{jl}}\right) = \beta_0 + \beta_1 a_j + \beta_2 b_l + \beta_3 (ab)_{jl}, \qquad j, l = 1, 2 \tag{7}$$

where n_{jl} are the total number of patients in condition jl and $\boldsymbol{\beta}$ the regression coefficient for the effect-coded variables a_j, b_l, and $(ab)_{jl}$. Variable selection is required to choose

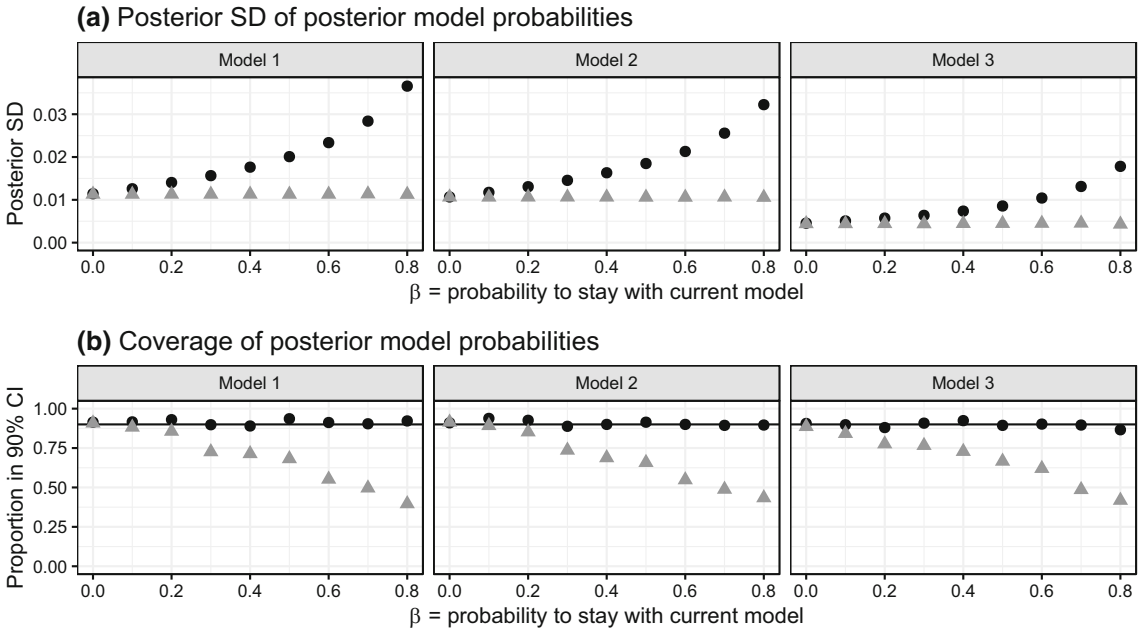

Fig. 2 Estimation uncertainty for the stationary distribution $\boldsymbol{\pi}$. **a** The Markov method (black dots) correctly indicated that estimation error of the posterior model probabilities increased as autocorrelation increased. When assuming i.i.d. sampling (gray triangles), the estimated precision did not depend on the autocorrelation. **b** Proportion of 500 replications for which the 90% CI intervals included the data-generating stationary distribution $\boldsymbol{\pi}$

Table 1 Logistic regression data set by Healy (1988)

Condition (A)	Antitoxin (B)	Death	Survival
More severe	Yes	15	6
	No	22	4
Less severe	Yes	5	15
	No	7	5

between $I = 5$ models: the intercept-only model I, the three main effect models A, B, and A+B, and the model AB that includes the interaction. For comparability, we use the same priors as Dellaportas et al. (2000) and assume a centered Gaussian prior with variance $\sigma^2 = 8$ for each regression parameter, $\beta_k \sim \mathcal{N}(0, 8)$. Moreover, the model probabilities were set to be uniform, $p(\mathcal{M}_i) = 1/5$. Note that nonuniform prior probabilities might be used to protect against multiple testing issues (i.e., Bayes multiplicity; Scott and Berger 2010).

One of the two implemented transdimensional MCMC approaches uses unconditional priors (Kuo and Mallick 1998, KM98) and includes indicator variables $\gamma_{ik} \in \{0, 1\}$ for each regression coefficient β_k in model $\mathcal{M}_i$. The parameter $\boldsymbol{\gamma}_i$ determines which regression coefficients are included by removing some of the additive terms of the linear model in Eq. 7. Details about the full and conditional posterior distributions are provided by Dellaportas et al. (2000, p. 7).

As a second transdimensional MCMC approach, we implemented the method of Carlin and Chib (1995; CC95), which stacks up all model parameters into a new parameter $\boldsymbol{\theta} = (z, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_I)$, where $\boldsymbol{\beta}_i$ is the vector of regression parameters of model $\mathcal{M}_i$. Thereby, this approach samples a total of 12 regression parameters along with the indexing variable z. Note that the method of Carlin and Chib (1995) uses pseudo-priors $p(\boldsymbol{\beta}_i \mid \mathcal{M}_j)$, $i \neq j$, that do not influence the statistical inference about $p(\boldsymbol{y} \mid \mathcal{M}_i)$ and $p(\boldsymbol{\beta}_i \mid \boldsymbol{y}, \mathcal{M}_i)$. However, these pseudo-priors determine the conditional proposal probabilities $p(z \mid \boldsymbol{y}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_I)$ of switching between the models and thereby affect the efficiency of the MCMC chain. In substantive applications, these pseudo-priors should be chosen to match the posterior $p(\boldsymbol{\beta}_i \mid \mathcal{M}_i)$ in order to ensure high probabilities of switching between the models (cf. Carlin and Chib 1995; Barker and Link 2013). Here, however, we did not optimize the sampling scheme and used $\boldsymbol{\beta}_{ik} \mid \mathcal{M}_j \sim \mathcal{N}(0, 8)$ for the pseudo-priors to illustrate that our method can correctly detect the lower precision resulting from this suboptimal choice.

Figure 3 shows the estimated posterior distribution ($R = 5000$) of the posterior model probabilities using one Markov chain with 11,000 iterations (including 1000 burn-in samples). The vertical black lines show the reference values for $\boldsymbol{\pi}$, approximated with very high accuracy by the KM98 approach using eight independent chains and one million samples each. As expected, the (incorrect) assumption that

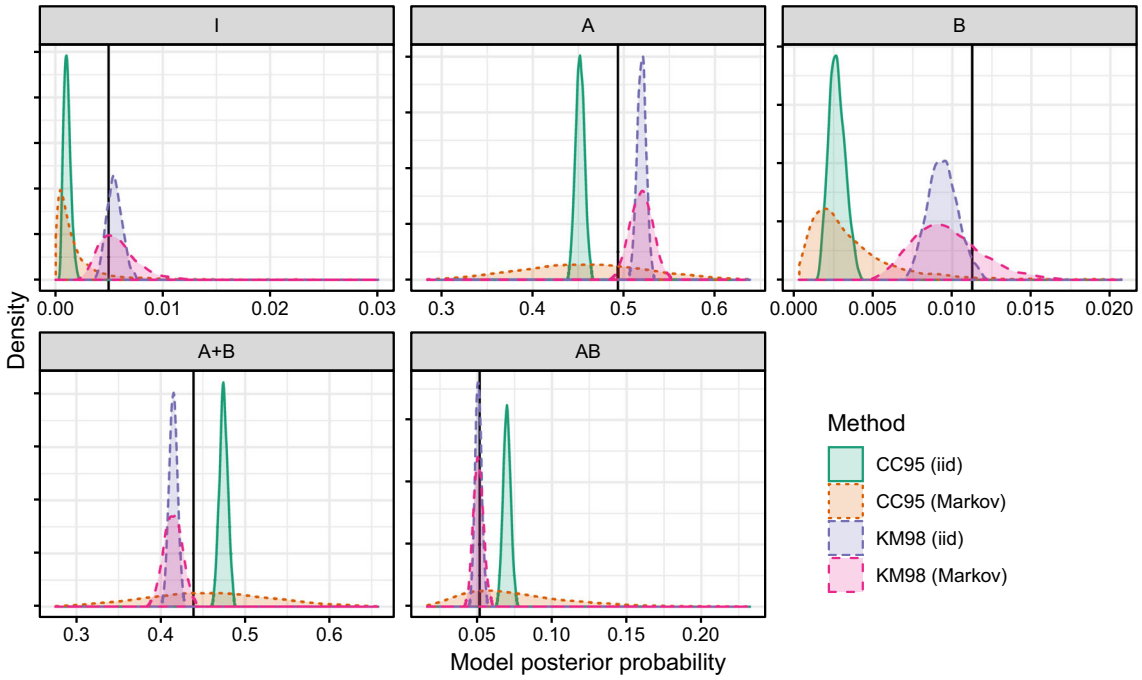

Fig. 3 Five panels show the estimation uncertainty of the posterior model probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_5)^\top$ for the five logistic regression models I (intercept only), A, B, and A+B (only main effects), and AB (two main effects and interaction). For both transdimensional MCMC samplers (CC95 = Carlin and Chib 1995; KM98 = Kuo and Mallick 1998), the posterior distribution of the Markov model included the correct reference values (vertical black lines) with high probability. In contrast, the i.i.d. model underestimated estimation uncertainty and posterior distributions did not include the target values with high probability

$z^{(t)}$ are sampled independently resulted in overconfidence in the point estimates of the CC95 approach. For all models, the corresponding posterior distributions missed the correct value and did not identify this estimation uncertainty. This shows the importance of assessing the dependency in the samples $z^{(t)}$ in order to judge the estimation uncertainty for the estimated posterior model probabilities. As a remedy, the proposed Markov approach resulted in a posterior distribution that covered the target values with high probability. Moreover, the novel estimation method revealed that the KM98 implementation had a higher precision compared to the CC95 approach, which was likely due to the (intentionally not optimized) choice of the pseudo-priors in the latter method. Hence, the Markov model allows comparison of the estimation uncertainty of different transdimensional MCMC methods for the model probabilities $\boldsymbol{\pi}$.

To test the validity of the proposed method more rigorously, we replicated the previous analysis 500 times. Thereby, the estimated precision can be compared against the actual sampling variability of the estimated model probabilities. For both transdimensional MCMC methods, Table 2 shows the mean estimated model probabilities in percent. Across replications, the point estimates (posterior means) from the Markov and the i.i.d. approach were very similar with a median absolute difference of 0.03% and 0.31% for the KM98 and CC95 implementations, respectively. To judge whether the estimated precision (i.e., the mean posterior standard deviations $\overline{\text{SD}}_{\text{iid}}$ and $\overline{\text{SD}}_{\text{Markov}}$) is valid, Table 2 shows the empirical SD of the estimates $\hat{\boldsymbol{\pi}}$ across replications. The results show that the assumption of independent samples $z^{(t)}$ leads to an overconfident assessment of the precision for the estimated model probabilities, $\overline{\text{SD}}_{\text{iid}} \ll \text{SD}(\hat{\boldsymbol{\pi}})$, which is especially severe for the less efficient CC95 implementation. In contrast, the Markov approach provided good estimates of the actual estimation uncertainty, $\overline{\text{SD}}_{\text{Markov}} \approx \text{SD}(\hat{\boldsymbol{\pi}})$. Moreover, for the MCMC method by Carlin and Chib (1995), the larger SDs indicate a smaller efficiency compared to the unconditional prior approach by Kuo and Mallick (1998). This theoretically expected result is due to the suboptimal choice of pseudo-priors. However, note that this difference in efficiency may be overlooked when merely computing relative proportions based on the sampled indexing variable $z^{(t)}$ (i.e., when implicitly assuming independent samples).

The higher efficiency of the KM98 approach becomes even clearer when assessing the median of the estimated effective sample size, which was 2043 for the KM98 approach compared to only 65 for the CC95 method. As discussed above, commonly used estimators of effective sample size for continuous parameters (e.g., Plummer et al. 2006) should not be applied to the discrete model-indexing vari-

Table 2 Estimated posterior model probabilities in percent

Model	Kuo and Mallick (1998)				Carlin and Chib (1995)			
	Mean($\hat{\boldsymbol{\pi}}$)	SD($\hat{\boldsymbol{\pi}}$)	$\overline{\text{SD}}_{\text{iid}}$	$\overline{\text{SD}}_{\text{Markov}}$	Mean($\hat{\boldsymbol{\pi}}$)	SD($\hat{\boldsymbol{\pi}}$)	$\overline{\text{SD}}_{\text{iid}}$	$\overline{\text{SD}}_{\text{Markov}}$
1	0.51	0.24	0.07	0.16	0.57	0.35	0.06	0.39
A	49.28	1.38	0.50	1.22	48.55	7.14	0.49	6.92
B	1.14	0.44	0.10	0.26	1.26	0.63	0.10	0.73
A + B	43.85	1.25	0.50	1.10	43.61	7.41	0.49	7.19
AB	5.22	0.37	0.22	0.34	6.00	3.38	0.21	3.82

Posterior model probability estimates $\hat{\boldsymbol{\pi}}$ are shown in percent. Mean($\hat{\boldsymbol{\pi}}$) and SD($\hat{\boldsymbol{\pi}}$) were computed across 500 replications. As a measure for the estimated precision, means of the posterior SD are shown ($\overline{\text{SD}}_{\text{iid}}$ assumes independent sampling; $\overline{\text{SD}}_{\text{Markov}}$ assumes a Markov chain model)

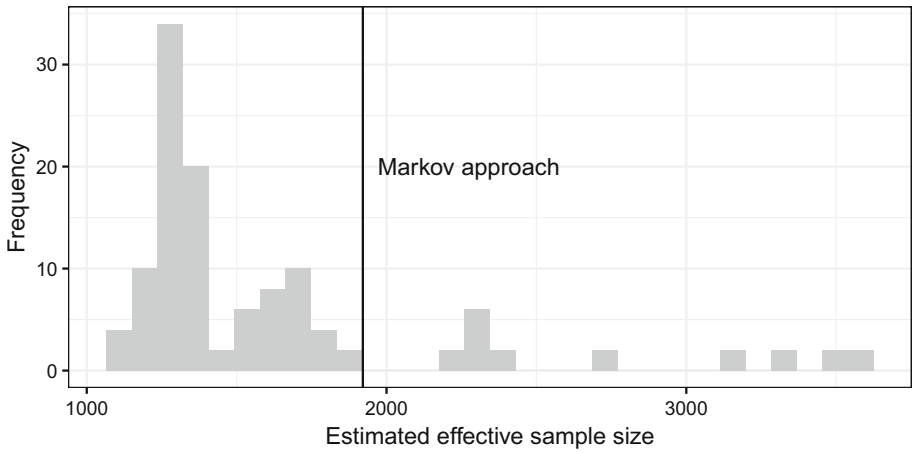

Fig. 4 Effective sample size as estimated by the spectral density at zero (Plummer et al. 2006) for all permutations of the model indices for a given MCMC output $z^{(t)}$ (based on 10,000 samples of the method by Kuo and Mallick 1998)

able z because they depend on the arbitrary numerical labels used for the models. If such methods are applied nevertheless, the resulting estimate for the effective sample size cannot be interpreted because it is not invariant under permutations of the arbitrary model indices used for the discrete parameter z. To illustrate this, Fig. 4 shows the distribution of the estimated effective sample size when applying the spectral decomposition available in the R package `coda` (Plummer et al. 2006) to all 120 permutations of the model indices $(1, \ldots, 5)$ for a fixed sequence $z^{(t)}$. Since this method incorrectly assumes that the discrete variable z is continuous, the estimated effective sample size was not invariant, but varied considerably depending on the specific labeling of the models (gray histogram). In contrast, the proposed Markov approach resulted in a well-defined, invariant estimate $\widehat{T}_{\text{eff}} = 1921$ (vertical black line) by explicitly accounting for the discreteness of z.

Finally, we show that the posterior samples $\boldsymbol{\pi}^{(t)}$ of the model $\mathcal{M}^{\text{Markov}}$ can directly be used to assess the uncertainty of Bayes factor estimates. For instance, substantive applications could be interested in testing whether to include the interaction term of condition (A) and treatment (B) in a logistic regression model. Given the output of a single MCMC run with 10,000 samples, Fig. 5 shows the resulting posterior distribution of the Bayes factor $B_{\text{A+B,AB}}$ in favor for the absence of an interaction. Similar to the posterior model probabilities, the i.i.d. approach resulted in overconfidence regarding the estimate and most of the probability mass excluded the correct value 8.51 (approximated with a precision of SD = 0.020). In contrast, the Markov approach corrected for dependencies in the samples $z^{(t)}$ and included the correct value. The same pattern emerged across the 500 replications, that is, the mean estimated SD of the Bayes factor approximated the corresponding empirical SD of the Bayes factor estimates (KM98: 0.56 vs. 0.60; CC95: 74.7 vs. 114.3). When using transdimensional MCMC, Bayes factors cannot be expected to be reliably estimated if models are never or very infrequently sampled (e.g., Model 1 in Table 2). For instance, the Bayes factor $B_{\text{A,B}} \approx 43.8$ was estimated very imprecisely even in the KM98 approach (mean SD = 13.0; empirical SD = 24.3). To obtain more precise Bayes factor estimates in the presence of infrequently sampled models, it is recommended to rerun the transdimensional MCMC chain including only the two relevant models of interest (Lodewyckx et al. 2011).

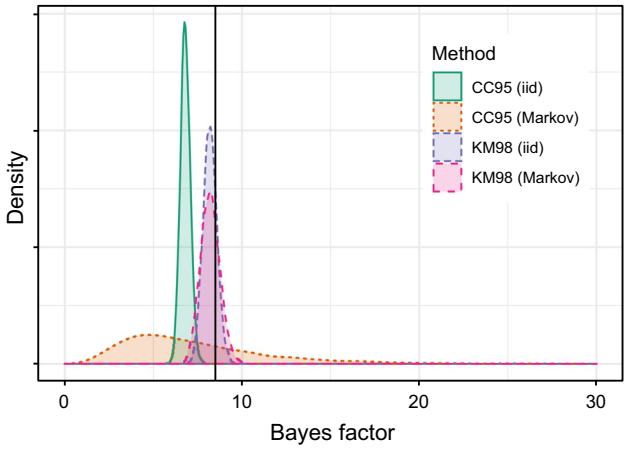

Fig. 5 Posterior distribution for the Bayes factor in favor of Model A + B (only main effects) versus AB (two main effects and interaction). The vertical black line shows the target value estimated using two different transdimensional MCMC samplers (CC95 = Carlin and Chib 1995; KM98 = Kuo and Mallick 1998). In contrast to the Markov model, the i.i.d. model incorrectly assumes independence and thus overestimated estimation uncertainty

5 Log-linear models for a 2^6 contingency table

The application of the proposed method is also feasible in realistic scenarios with hundreds of sampled models. To illustrate this, we reanalyzed the 2^6 complete contingency table by Edwards and Havránek (1985), which includes six risk factors for coronary heart disease (i.e., smoking, strenuous mental work, strenuous physical work, systolic blood pressure, ratio of α and β lipoproteins, and family anamnesis of coronary heart disease). We are interested in finding the most parsimonious log-linear model, which accounts for the cell frequencies y_j of cell j ($j = 1, \ldots, 2^6$) by assuming a Poisson distribution with mean μ_j and

$$\log \mu_j = \phi + \boldsymbol{x}_j^\top \boldsymbol{\beta}, \tag{8}$$

where ϕ is the intercept, $\boldsymbol{\beta}$ the vector of regression parameters, and $\boldsymbol{x}_j^\top$ the (transposed) design vector, which selects the elements of $\boldsymbol{\beta}$ included for modeling cell j. Here, we consider the class of hierarchical log-linear models that only allow the inclusion of an interaction if the corresponding marginal effects and lower interaction terms are included in the model as well (e.g., Overstall and King 2014b).

To select between all 7.8 million possible hierarchical log-linear models (Dellaportas and Forster 1999), we used the reversible jump algorithm proposed by Forster et al. (2012), which is implemented in the R package `conting` (Overstall and King 2014a). Assuming a unit information prior (Ntzoufras et al. 2003), we sampled 100,000 iterations, discarded 10,000 as burn-in, and applied the proposed Markov chain method by drawing $R = 5000$ samples for the posterior model probabilities of the I^* sampled models. To assess whether the estimated uncertainty accurately quantifies sampling variability, we ran 200 replications initialized with randomly chosen models.

Across replications, 5805 models were sampled (on average, 562.7 per replication). Table 3 shows the results for the 10 models with the highest posterior probabilities. All of these 10 models included the six main effects (A: smoking, B: strenuous mental work, C: strenuous physical work, D: systolic blood pressure, E: ratio of α and β lipoproteins, F: family anamnesis of coronary heart disease) and the first-order interactions AC, AD, AE, BC, and DE, but differed with respect to including the remaining interactions. Despite the large number of iterations, the estimation uncertainty (i.e., the posterior SD) of the posterior model probabilities was relatively large, indicating that the samples $z^{(t)}$ were autocorrelated to a substantial degree. This is also reflected by the effective sample size, which was estimated to be $\widehat{T}_{\text{eff}} = 4259$ on average (SD = 181), approximately 5% of the number of iterations after burn-in.

Table 3 also shows means and standard deviations of the sampled model rank τ for the models with the highest posterior probability, indicating that estimation uncertainty (i.e., the posterior SD) increased for models with smaller posterior probabilities. Moreover, the proportion of replications is shown for which the sampled rank τ was identical to the model index ($\tau = \#$) and smaller than or equal to 10 ($\tau \leq 10$). According to these proportions, exact ranks were estimated precisely only for the two best models, whereas the set of the 10 models with highest posterior probabilities was relatively stable across posterior samples (with the exception of model 10). Importantly, the Markov approach provided mean estimated probabilities $\overline{P(\tau = \#)}$ and $\overline{P(\tau \leq 10)}$ that matched the corresponding empirical proportions across replications.

Note that these results regarding estimation uncertainty are in line with our expectations—if models have small posterior probabilities, they are also sampled infrequently, which in turn results in estimation uncertainty. To quantify this variability, the proposed Markov chain approach provides an estimate for the achieved precision to assess the quality of the results and to find an appropriate stopping rule for MCMC sampling.

6 Conclusion

We proposed a novel approach for estimating the precision of transdimensional MCMC output. Essentially, a first-order Markov model is fitted to the observed model-indexing variable $z^{(t)}$ to quantify estimation uncertainty of the corresponding stationary distribution. We showed that the method accounts for autocorrelation in a given sequence $z^{(t)}$

Table 3 Models with the highest posterior probability for the 2^6 contingency table

#	Model	Posterior model probabilities $\boldsymbol{\pi}$				Rank τ						
		Mean($\hat{\boldsymbol{\pi}}$)	SD($\hat{\boldsymbol{\pi}}$)	$\overline{\text{SD}}_{\text{iid}}$	$\overline{\text{SD}}_{\text{Markov}}$	Mean(τ)	SD(τ)	$\overline{\text{SD}(\tau)}$	$\tau = \#$	$\overline{P(\tau = \#)}$	$\tau \leq 10$	$\overline{P(\tau \leq 10)}$
1	CE	18.78	1.34	0.13	1.02	1.00	0.00	0.03	1.00	1.00	1.00	1.00
2	BE	11.92	0.94	0.11	0.84	2.00	0.00	0.04	1.00	1.00	1.00	1.00
3	BE + CE	7.12	1.11	0.09	0.43	3.34	0.61	0.37	0.72	0.78	1.00	1.00
4	BF + CE	6.57	1.20	0.08	0.52	3.94	0.84	0.42	0.71	0.75	1.00	1.00
5	BE + BF	4.20	0.85	0.07	0.41	5.42	1.59	0.21	0.92	0.93	0.96	0.99
6	CE + EF	2.77	0.50	0.06	0.33	6.80	1.71	0.58	0.62	0.65	0.94	1.00
7	BE + BF + CE	2.53	0.60	0.05	0.24	8.24	5.64	0.54	0.58	0.66	0.92	1.00
8	CE + ADE	1.88	0.30	0.05	0.25	8.72	1.35	0.80	0.47	0.56	0.95	0.95
9	BE + EF	1.76	0.38	0.04	0.26	9.43	3.21	0.88	0.45	0.54	0.92	0.93
10	BE + ADE	1.19	0.22	0.04	0.19	12.05	3.11	1.40	0.32	0.39	0.39	0.56

All of the 10 models include the six main effects, A: smoking, B: strenuous mental work, C: strenuous physical work, D: systolic blood pressure, E: ratio of α and β lipoproteins, F: family anamnesis of coronary heart disease, and the first-order interactions AC, AD, AE, BC, and DE. Posterior model probabilities $\boldsymbol{\pi}$ are shown in percent. Mean($\hat{\boldsymbol{\pi}}$), SD($\hat{\boldsymbol{\pi}}$), Mean(τ), and SD(τ) were computed across 200 replications. The columns $\tau = \#$ and $\tau \leq 10$ refer to the proportion of replications for which the model rank τ was (a) equal to the model index # or (b) smaller than or equal to 10

and provides a good assessment of estimation uncertainty. Importantly, the method does not require output of multiple independent MCMC chains and thus reduces the computational costs for adaption and burn-in. Besides being useful for transdimensional MCMC output, the method provides an estimate of the precision and effective sample size of discrete parameters in MCMC samplers in general. Thereby, researchers can easily decide whether the obtained precision is sufficiently high for substantive applications of interest.

Acknowledgements Daniel W. Heck was supported by the research training group *Statistical Modeling in Psychology* (GRK 2277), funded by the German Research Foundation (DFG).

Appendix: Estimating the shape parameters of a Dirichlet distribution

In the following, we outline the fixed-point algorithm proposed by Minka (2000) to estimate the vector of shape parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I)^\top$ of a Dirichlet distribution. Given a set of R probability vectors $\boldsymbol{\pi}^{(r)}$ (in the proposed method, these are the derived samples of the posterior model probabilities), the likelihood function of the shape parameters $\boldsymbol{\alpha}$ is

$$L(\boldsymbol{\alpha}) = \prod_{r=1}^{R} \left[\frac{\Gamma\left(\sum_i \alpha_i\right)}{\prod_i \Gamma(\alpha_i)} \prod_i \left(\pi_i^{(r)}\right)^{\alpha_i - 1} \right]. \tag{9}$$

To maximize this likelihood function, Minka (2000) developed an efficient fixed-point algorithm and proved its convergence to the unique maximum likelihood estimate $\hat{\boldsymbol{\alpha}}$. The computational steps are outlined in Algorithm 2. At its core, the current estimates α_i are updated in line 8 by using the digamma function Ψ and its inverse Ψ^{-1}. As remarked by Minka (2000), the algorithm converges very fast even for a large number of shape parameters I (e.g., 80 ms on an Intel® i7-7700HQ for $I = 1000$).

Algorithm 2 Estimating the shape parameters $\boldsymbol{\alpha}$ of a Dirichlet distribution.

1: **procedure** DIRICHLET ESTIMATION (MINKA 2000)
2: Compute $\boldsymbol{\mu}$: $\mu_i \leftarrow \frac{1}{R} \sum_{r=1}^{R} \log \pi_i^{(r)}$
3: Set starting values $\boldsymbol{\alpha}$ with $\alpha_i > 0$ for all $i = 1, \ldots, I$
4: Set absolute tolerance $\epsilon > 0$ and $\delta \leftarrow \infty$
5: **while** $\delta > \epsilon$ **do**
6: $\boldsymbol{\alpha}' \leftarrow \boldsymbol{\alpha}$
7: **for** $i = 1, \ldots, I$ **do**
8: $\alpha_i \leftarrow \Psi^{-1}\left(\Psi(\sum_j \alpha_j') + \mu_i\right)$
9: $\delta \leftarrow ||\boldsymbol{\alpha}' - \boldsymbol{\alpha}||$
10: **return** $\boldsymbol{\alpha}$

References

Alvares, D., Armero, C., Forte, A.: What does objective mean in a Dirichlet-multinomial process? Int. Stat. Rev. **86**, 106–118 (2018). https://doi.org/10.1111/insr.12231

Anderson, T.W., Goodman, L.A.: Statistical inference about Markov chains. Ann. Math. Stat. **28**, 89–110 (1957). https://doi.org/10.1214/aoms/1177707039

Arnold, R., Hayakawa, Y., Yip, P.: Capture–recapture estimation using finite mixtures of arbitrary dimension. Biometrics **66**, 644–655 (2010). https://doi.org/10.1111/j.1541-0420.2009.01289.x

Barker, R.J., Link, W.A.: Bayesian multimodel inference by RJMCMC: a Gibbs sampling approach. Am. Stat. **67**, 150–156 (2013). https://doi.org/10.1080/00031305.2013.791644

Brooks, S.P., Giudici, P.: Markov chain Monte Carlo convergence assessment via two-way analysis of variance. J. Comput. Graph. Stat. **9**, 266–285 (2000). https://doi.org/10.1080/10618600.2000.10474880

Brooks, S., Giudici, P., Philippe, A.: Nonparametric convergence assessment for MCMC model selection. J. Comput. Graph. Stat. **12**, 1–22 (2003a). https://doi.org/10.1198/1061860031347

Brooks, S.P., Giudici, P., Roberts, G.O.: Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **65**, 3–39 (2003b). https://doi.org/10.1111/1467-9868.03711

Burke, C.J., Rosenblatt, M.: A Markovian function of a Markov chain. Ann. Math. Stat. **29**, 1112–1122 (1958). https://doi.org/10.1214/aoms/1177706444

Carlin, B.P., Chib, S.: Bayesian model choice via Markov chain Monte Carlo methods. J. R. Stat. Soc. Ser. B (Methodol.) **57**, 473–484 (1995)

Castelloe, J.M., Zimmerman, D.L.: Convergence assessment for reversible jump MCMC samplers. Technical Report 313, Department of Statistics and Actuarial Science, University of Iowa (2002)

Dellaportas, P., Forster, J.J.: Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. Biometrika **86**, 615–633 (1999). http://www.jstor.org/stable/2673658

Dellaportas, P., Forster, J.J., Ntzoufras, I.: Bayesian variable selection using the Gibbs sampler. In: Dey, D.K., Ghosh, S.K., Mallick, B.K. (eds.) Generalized Linear Models: A Bayesian Perspective, pp. 273–286. Marcel Dekker Inc, New York (2000)

Dellaportas, P., Forster, J.J., Ntzoufras, I.: On Bayesian model and variable selection using MCMC. Stat. Comput. **12**, 27–36 (2002). https://doi.org/10.1023/A:1013164120801

Doss, C.R., Flegal, J.M., Jones, G.L., Neath, R.C.: Markov chain Monte Carlo estimation of quantiles. Electron. J. Stat. **8**, 2448–2478 (2014). https://doi.org/10.1214/14-EJS957

Eddelbuettel, D., Sanderson, C.: RcppArmadillo: accelerating R with high-performance C++ linear algebra. Comput. Stat. Data Anal. **71**, 1054–1063 (2014). https://doi.org/10.1016/j.csda.2013.02.005

Edwards, D., Havránek, T.: A fast procedure for model search in multidimensional contingency tables. Biometrika **72**, 339–351 (1985). https://doi.org/10.2307/2336086

Flegal, J.M., Gong, L.: Relative fixed-width stopping rules for markov chain Monte Carlo simulations. Stat. Sin. **25**, 655–675 (2015). http://www.jstor.org/stable/24311039

Forster, J.J., Gill, R.C., Overstall, A.M.: Reversible jump methods for generalised linear models and generalised linear mixed models. Stat. Comput. **22**, 107–120 (2012). https://doi.org/10.1007/s11222-010-9210-3

Frühwirth-Schnatter, S.: Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. J. Am. Stat. Assoc. **96**, 194–209 (2001). https://doi.org/10.1198/016214501750333063

Gong, L., Flegal, J.M.: A practical sequential stopping rule for high-dimensional Markov Chain Monte Carlo. J. Comput. Graph. Stat. **25**, 684–700 (2016). https://doi.org/10.1080/10618600.2015.1044092

Green, P.J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82**, 711–732 (1995). https://doi.org/10.1093/biomet/82.4.711

Healy, M.J.R.: GLIM: An Introduction. Claredon Press, Oxford (1988)

Heck, D.W., Hilbig, B.E., Moshagen, M.: From information processing to decisions: formalizing and comparing probabilistic choice models. Cogn. Psychol. **96**, 26–40 (2017). https://doi.org/10.1016/j.cogpsych.2017.05.003

Heck, D.W., Gronau, Q.F., Overstall, A.M., Wagenmakers, E.J.: MCMCprecision: precision of discrete variables in trans-dimensional MCMC (2018). https://CRAN.R-project.org/package=MCMCprecision

Heidelberger, P., Welch, P.D.: A spectral method for confidence interval generation and run length control in simulations. Commun. ACM **24**, 233–245 (1981). https://doi.org/10.1145/358598.358630

Jeffreys, H.: Theory of Probability. Oxford University Press, New York (1961)

Jones, G.L., Haran, M., Caffo, B.S., Neath, R.: Fixed-width output analysis for Markov chain Monte Carlo. J. Am. Stat. Assoc. **101**, 1537–1547 (2006). https://doi.org/10.1198/016214506000000492

Karnesis, N.: Bayesian model selection for LISA pathfinder. Phys. Rev. D **89**, 062001 (2014). https://doi.org/10.1103/PhysRevD.89.062001

Kass, R.E., Raftery, A.E.: Bayes factors. J. Am. Stat. Assoc. **90**, 773–795 (1995). https://doi.org/10.1080/01621459.1995.10476572

Kuo, L., Mallick, B.: Variable selection for regression models. Sankhyā Indian J. Stat. Ser. B **60**, 65–81 (1998). http://www.jstor.org/stable/25053023

Lodewyckx, T., Kim, W., Lee, M.D., Tuerlinckx, F., Kuppens, P., Wagenmakers, E.J.: A tutorial on Bayes factor estimation with the product space method. J. Math. Psychol. **55**, 331–347 (2011). https://doi.org/10.1016/j.jmp.2011.06.001

Lopes, H.F., West, M.: Bayesian model assessment in factor analysis. Stat. Sin. **14**, 41–67 (2004). http://www.jstor.org/stable/24307179

Minka, T.P.: Estimating a Dirichlet distribution. Technical Report, MIT, Cambridge, MA (2000). https://tminka.github.io/papers/dirichlet/

Ntzoufras, I.: Gibbs variable selection using BUGS. J. Stat. Softw. **7**, 1–19 (2002). https://doi.org/10.18637/jss.v007.i07

Ntzoufras, I., Dellaportas, P., Forster, J.J.: Bayesian variable and link determination for generalised linear models. J. Stat. Plan. Inference **111**, 165–180 (2003). https://doi.org/10.1016/S0378-3758(02)00298-7

Opgen-Rhein, R., Fahrmeir, L., Strimmer, K.: Inference of demographic history from genealogical trees using reversible jump Markov chain Monte Carlo. BMC Evolut. Biol. **5**, 6 (2005). https://doi.org/10.1186/1471-2148-5-6

Overstall, A., King, R.: Conting: an R package for Bayesian analysis of complete and incomplete contingency tables. J. Stat. Softw. **58**, 1–27 (2014a). https://doi.org/10.18637/jss.v058.i07

Overstall, A.M., King, R.: A default prior distribution for contingency tables with dependent factor levels. Stat. Methodol. **16**, 90–99 (2014b). https://doi.org/10.1016/j.stamet.2013.08.007

Plummer, M.: JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing, vol. 124, p. 125. Vienna (2003)

Plummer, M., Best, N., Cowles, K., Vines, K.: CODA: convergence diagnosis and output analysis for MCMC. R News **6**, 7–11 (2006)

Sanderson, C., Curtin, R.: Armadillo: a template-based C++ library for linear algebra. J. Open Source Softw. **1**, 26 (2016). https://doi.org/10.21105/joss.00026

Scott, J.G., Berger, J.O.: Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. Ann. Stat. **38**, 2587–2619 (2010). https://doi.org/10.1214/10-AOS792

Sisson, S.A.: Transdimensional Markov Chains. J. Am. Stat. Assoc. **100**, 1077–1089 (2005). https://doi.org/10.1198/016214505000000664

Sisson, S.A., Fan, Y.: A distance-based diagnostic for trans-dimensional Markov chains. Stat. Comput. **17**, 357–367 (2007). https://doi.org/10.1007/s11222-007-9025-z

Stephens, M.: Bayesian analysis of mixture models with an unknown number of components- an alternative to reversible jump methods. Ann. Stat. **28**, 40–74 (2000). http://www.jstor.org/stable/2673981