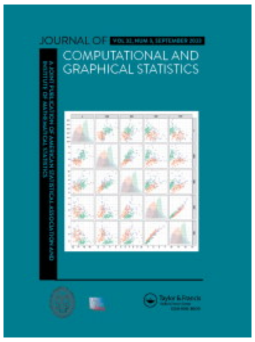

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

On Data Augmentation for Models Involving Reciprocal Gamma Functions

Yasuyuki Hamura, Kaoru Irie & Shonosuke Sugasawa

To cite this article: Yasuyuki Hamura, Kaoru Irie & Shonosuke Sugasawa (2023) On Data Augmentation for Models Involving Reciprocal Gamma Functions, Journal of Computational and Graphical Statistics, 32:3, 908-916, DOI: 10.1080/10618600.2022.2119988

To link to this article: https://doi.org/10.1080/10618600.2022.2119988

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2023, VOL. 32, NO. 3, 908–916
https://doi.org/10.1080/10618600.2022.2119988

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

On Data Augmentation for Models Involving Reciprocal Gamma Functions

Yasuyuki Hamura[a], Kaoru Irie[b], and Shonosuke Sugasawa[c]

[a]Graduate School of Economics, Kyoto University, Yoshida-Honmachi, Sakyo, Kyoto, Japan; [b]Faculty of Economics, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, Japan; [c]Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba, Japan

ABSTRACT

In this article, we introduce a new and efficient data augmentation approach to the posterior inference of the models with shape parameters when the reciprocal gamma function appears in full conditional densities. Our approach is to approximate full conditional densities of shape parameters by using Gauss's multiplication formula and Stirling's formula for the gamma function, where the approximation error can be made arbitrarily small. We use the techniques to construct efficient Gibbs and Metropolis–Hastings algorithms for a variety of models that involve the gamma distribution, Student's t-distribution, the Dirichlet distribution, the negative binomial distribution, and the Wishart distribution. The proposed sampling method is numerically demonstrated through simulation studies. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received March 2022
Accepted August 2022

KEYWORDS
Gauss's multiplication formula; Markov chain Monte Carlo; Reciprocal gamma function; Stirling's formula

1. Introduction

Markov chain Monte Carlo (MCMC) algorithms are now widely adopted in Bayesian posterior computation, where parameters are iteratively sampled from their respective conditional distributions. However, when the models of interest involve the gamma and related distributions, it is computationally costly to sample the shape parameters from their full conditional posteriors. The main difficulty here is that the full conditional densities of the shape parameters involve the reciprocal gamma function, $1/\Gamma(\xi), \xi > 0$, and are not any well-known distributions. Thus, it is not straightforward to construct efficient MCMC algorithms when the shape parameters are also estimated.

Several sampling strategies that have been proposed in the literature are customized for each class of distributions. For gamma distributions, Miller (2019) provided an accurate approximation of the full conditional distribution of the shape parameter. For Student's t-distributions, Fonseca, Ferreira, and Migon (2008) considered the unknown degrees of freedom, at the cost of the complication of the priors. Custom sampling algorithms have also been proposed for the Dirichlet models (Nandram 1998), negative binomial processes (Zhou and Carin 2015), Dirichlet processes (Escobar and West 1995) and Pitman-Yor processes (Bacallado et al. 2022).

Rather than focusing on a particular class of distributions, it is also possible to devise the sampling methods that are applicable to the general class of models with shape parameters, at the cost of efficiency and computational time. For example, the approximation of log-concave densities (Gilks and Wild 1992; Devroye 2012) and the MH acceptance-rejection method (Tierney 1994; Chib and Greenberg 1995) can be used for the posterior inference for models with the reciprocal gamma functions. The latter needs to be further customized to each model, as practiced for Student's t-models in Watanabe (2001). Another approach is the data augmentation scheme, where several latent variables are introduced to simplify the full conditionals of the model. He, Polson, and Xu (2021) proposed a general and efficient data augmentation for models with reciprocal gamma functions, where the simulation from power truncated normal (PTN) distributions become necessary. In this article, we also take the data-augmentation approach, but propose a new augmentation where we only need to simulate from well-known distributions.

Our strategy for deriving an augmented model is 2-fold: (i) using Gauss's multiplication formula for the gamma function to introduce conditionally beta-distributed latent variables and (ii) approximating the augmented densities by Stirling's formula. The full conditionals of the shape parameters and latent variables of the resulting model are all well-known distributions, such as gamma and beta distributions, from which it is easy and fast to simulate. Finally, the accept/reject step is added to justify the sampling algorithm as an independent Metropolis–Hastings (MH) method.

To assess the efficiency of the sampling algorithm based on the proposed augmentation, we evaluate the upper and lower bounds of the approximation error and show that, in many cases, the acceptance probability is close to one. Due to its simplicity, our augmentation scheme can be applied directly to many models with reciprocal gamma functions, including the Student's t-distribution, Dirichlet-multinomial distribution, negative binomial distribution and Wishart distribution.

The remainder of the article is organized as follows. In Section 2, we develop a new data augmentation and approximation of the reciprocal gamma function and illustrate our approach

CONTACT Yasuyuki Hamura yasu.stat@gmail.com Graduate School of Economics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JCGS.

using a simple gamma model. For simplicity, we consider only proper priors for shape parameters as well as other variables, which ensures that full conditional distributions are always proper. In Section 3, we use our approach for a model based on Student's t-distribution. In Section 4, we consider a Dirichlet-multinomial model and apply a generic method. Some concluding remarks are given in Section 5. Proofs and additional results are provided in the supplementary materials.

2. Beta Data Augmentation

2.1. General Ideas

The most important result for our method is the following integral expression, which is based on Gauss's multiplication formula for the gamma function.

Theorem 1. Let $m \in \mathbb{N}$. Then we have

$$\frac{1}{\{\Gamma(\xi)\}^m} = C_m \frac{1}{\xi^{m\xi}} \xi^{m+1/2-1} e^{m\xi} \left\{ \prod_{j=2}^{m} \int_0^1 \rho_j^{\xi+(j-1)/m-1} (1-\rho_j)^{(m-j+1)/m-1} d\rho_j \right\} \frac{(m\xi)^{m\xi-1/2}}{\Gamma(m\xi)e^{m\xi}}, \quad (1)$$

for all $\xi > 0$, where $C_m = 1/\left\{(2\pi)^{(m-1)/2} \prod_{j=2}^{m} \Gamma((m-j+1)/m)\right\}$.

The proof is given in the supplementary materials. By Theorem 1, we can rewrite the mth power of the reciprocal gamma function, $1/\{\Gamma(\xi)\}^m$, by using integrals of $m-1$ beta densities, such that the reciprocal gamma function appears only once in the right-hand side.

Suppose that the target distribution, or the posterior distribution, is the joint density of shape parameter ξ and other variables $\boldsymbol{\vartheta}$ of the form,

$$p(\xi, \boldsymbol{\vartheta}) \propto f(\xi, \boldsymbol{\vartheta}) \frac{1}{\{\Gamma(\xi)\}^m},$$

where typically $f(\xi, \boldsymbol{\vartheta}) \overset{\xi}{\propto} \mathrm{Ga}(\xi | a_1, b_1)$ for some $a_1, b_1 > 0$. This framework covers, for example, the case of n independent observations from a gamma distribution, $x_1, \ldots, x_n \sim \mathrm{Ga}(\alpha, \beta)$; in this case, $m = n$, $(\xi, \boldsymbol{\vartheta}) = (\alpha, \beta)$, and $p(\alpha, \beta)$ is the posterior of (α, β) given $x_1, \ldots, x_n$, or $p(\alpha, \beta) \propto \pi(\alpha, \beta) \times \beta^{n\alpha} \left(\prod_{i=1}^{n} x_i\right)^{\alpha} e^{-\beta \sum_{i=1}^{n} x_i}$, where $\pi(\alpha, \beta)$ is a prior density (see Section 2.2). In general, some of the variables $\boldsymbol{\vartheta}$ may be latent variables introduced based on data augmentation. We are interested in the repeated sampling from the conditional distributions, $p(\xi | \boldsymbol{\vartheta})$ and $p(\boldsymbol{\vartheta} | \xi)$. We assume that it is relatively easy to sample $\boldsymbol{\vartheta}$ from $p(\boldsymbol{\vartheta} | \xi)$, and we focus on the problem of sampling ξ from $p(\xi | \boldsymbol{\vartheta})$ in the following.

The derivation of the augmented model is a three-step process. First, we rewrite $p(\xi, \boldsymbol{\vartheta})$ as

$$p(\xi, \boldsymbol{\vartheta}) \propto \frac{f(\xi, \boldsymbol{\vartheta})}{\xi^{m\xi}} \xi^{m+1/2-1} e^{m\xi} \left\{ \prod_{j=2}^{m} \int_0^1 \rho_j^{\xi+(j-1)/m-1} (1-\rho_j)^{(m-j+1)/m-1} d\rho_j \right\} \frac{(m\xi)^{m\xi-1/2}}{\Gamma(m\xi)e^{m\xi}}, \quad (2)$$

by using Theorem 1. The mth power, $1/\{\Gamma(\xi)\}^m$, is simplified to a single reciprocal gamma function, $1/\Gamma(m\xi)$, which we further evaluate in the following steps. A set of additional latent variables, $\boldsymbol{\rho} = (\rho_2, \ldots, \rho_m) \in (0,1)^{m-1}$, has the full conditional of the simple form, $\prod_{j=2}^{m} \mathrm{Beta}(\rho_j | \xi + (j-1)/m, (m-j+1)/m)$, from which we can easily sample.

Second, the conditional density of $(\xi, \boldsymbol{\vartheta})$ given $\boldsymbol{\rho}$ is

$$p(\xi, \boldsymbol{\vartheta} | \boldsymbol{\rho}) \propto \frac{f(\xi, \boldsymbol{\vartheta})}{\xi^{m\xi}} \xi^{a_2-1} e^{-b_2\xi} \frac{(m\xi)^{m\xi-1/2}}{\Gamma(m\xi)e^{m\xi}},$$

where $a_2 = m + 1/2$ and $b_2 = -m + \sum_{j=2}^{m} \log(1/\rho_j)$. In the above expression, there are two factors that make it difficult to sample ξ from the full conditional: $1/\xi^{m\xi}$ and $(m\xi)^{m\xi-1/2}/\{\Gamma(m\xi)e^{m\xi}\}$. Here, in order to eliminate $1/\xi^{m\xi}$, we assume that we can make the change of variables $\widetilde{\boldsymbol{\vartheta}} = \varphi(\boldsymbol{\vartheta}; \xi)$ with Jacobian $\xi^{m\xi}$, so that $f(\xi, \boldsymbol{\vartheta}) d(\xi, \boldsymbol{\vartheta}) = \xi^{m\xi} \tilde{f}(\xi, \widetilde{\boldsymbol{\vartheta}}) d(\xi, \widetilde{\boldsymbol{\vartheta}})$, and the density of interest becomes

$$p(\xi, \widetilde{\boldsymbol{\vartheta}} | \boldsymbol{\rho}) \propto \tilde{f}(\xi, \widetilde{\boldsymbol{\vartheta}}) \xi^{a_2-1} e^{-b_2\xi} \frac{(m\xi)^{m\xi-1/2}}{\Gamma(m\xi)e^{m\xi}}.$$

This change-of-variable is available for many models, including the gamma model of Section 2.2. The models for which there is no such change-of-variable, including the Dirichlet-multinomial model of Section 4, are discussed in Section 2.3.

Third, we use the above expression to construct an independent MH algorithm. Let ξ^{old} be a current value of ξ. To generate a new value ξ^{new}, we first sample a proposal ξ^* from the approximate full conditional density proportional to $\tilde{f}(\xi, \widetilde{\boldsymbol{\vartheta}}) \xi^{a_2-1} e^{-b_2\xi}$ and compute

$$p = \min\left\{1, \frac{(m\xi^*)^{m\xi^*-1/2}}{\Gamma(m\xi^*)e^{m\xi^*}} \Big/ \frac{(m\xi^{\text{old}})^{m\xi^{\text{old}}-1/2}}{\Gamma(m\xi^{\text{old}})e^{m\xi^{\text{old}}}}\right\}.$$

Then we set $\xi^{\text{new}} = \xi^*$ with probability p, otherwise $\xi^{\text{new}} = \xi^{\text{old}}$. We note that in all the models considered in this article, proposal distributions corresponding to $\tilde{f}(\xi, \widetilde{\boldsymbol{\vartheta}}) \xi^{a_2-1} e^{-b_2\xi}$ are easy to sample from. The factor dropped in the approximate distribution can be evaluated as

$$\frac{e^{-1/(12\xi)}}{(2\pi)^{1/2}} < \frac{\xi^{\xi-1/2}}{\Gamma(\xi)e^{\xi}} < \frac{1}{(2\pi)^{1/2}}, \quad (3)$$

for any $\xi > 0$ by Stirling's formula. This expression shows that the factor is almost constant when ξ is not extremely small, and that the acceptance probability p is close to one. This can be confirmed by bounding the acceptance probability below as $p \geq e^{-1/(12m\xi^*)} \geq 1 - 1/(12m\xi^*)$, where the lower bound is almost unity unless ξ^* is extremely small.

2.2. An Illustration using a Gamma Model

Here, we consider a simple gamma model for illustration. For this model, several methods for posterior inference are available (e.g., Gilks and Wild 1992). In particular, the method of Miller (2019) is customized for this model and highly efficient.

Suppose that observations $\boldsymbol{x} = (x_1, \ldots, x_n)$ have been independently generated from a gamma distribution $\mathrm{Ga}(\alpha, \beta)$. We assume the independent gamma prior distributions for α and β:

Ga(a, b) and Ga(c, d), respectively. Then the posterior of (α, β) is

$$p(\alpha, \beta|\boldsymbol{x}) \propto \mathrm{Ga}(\alpha|a, b)\beta^{c-1}e^{-d\beta}\frac{\beta^{n\alpha}}{\{\Gamma(\alpha)\}^n}\Big(\prod_{i=1}^n x_i\Big)^\alpha e^{-\beta\sum_{i=1}^n x_i}.$$

Using Theorem 1, we can rewrite the above posterior density as

$$p(\alpha, \beta|\boldsymbol{x}) \propto \mathrm{Ga}(\alpha|a, b)\beta^{c-1}e^{-d\beta}\beta^{n\alpha}\Big(\prod_{i=1}^n x_i\Big)^\alpha e^{-\beta\sum_{i=1}^n x_i}$$
$$\times \frac{1}{\alpha^{n\alpha}}\alpha^{n+1/2-1}e^{n\alpha}\Big\{\prod_{i=2}^n\int_0^1 \rho_i{}^{\alpha+(i-1)/n-1}$$
$$(1-\rho_i)^{(n-i+1)/n-1}d\rho_i\Big\}\frac{(n\alpha)^{n\alpha-1/2}}{\Gamma(n\alpha)e^{n\alpha}}.$$

Now, we consider $\boldsymbol{\rho} = (\rho_2, \dots, \rho_n) \in (0, 1)^{n-1}$ as a set of additional latent variables. Then the conditional distribution of $(\alpha, \beta, \boldsymbol{\rho})$ given $\boldsymbol{x}$ is

$$p(\alpha, \beta, \boldsymbol{\rho}|\boldsymbol{x}) \propto \mathrm{Ga}(\alpha|a, b)\beta^{c-1}e^{-d\beta}\beta^{n\alpha}\Big(\prod_{i=1}^n x_i\Big)^\alpha e^{-\beta\sum_{i=1}^n x_i}$$
$$\times \frac{1}{\alpha^{n\alpha}}\alpha^{n+1/2-1}e^{n\alpha}\Big[\prod_{i=2}^n\{\rho_i{}^{\alpha+(i-1)/n-1}$$
$$(1-\rho_i)^{(n-i+1)/n-1}\}\Big]\frac{(n\alpha)^{n\alpha-1/2}}{\Gamma(n\alpha)e^{n\alpha}}.$$

In order to obtain MCMC samples of $(\alpha, \beta, \boldsymbol{\rho})|\boldsymbol{x}$, we can use the MH within Gibbs sampler. It is easy to sample $\boldsymbol{\rho}$ from its full conditional distribution since $p(\boldsymbol{\rho}|\alpha, \beta, \boldsymbol{x}) = \prod_{i=2}^n \mathrm{Beta}(\rho_i|\alpha + (i-1)/n, (n-i+1)/n)$. Meanwhile, the full conditional of (α, β) is

$$p(\alpha, \beta|\boldsymbol{\rho}, \boldsymbol{x}) \propto \frac{1}{\alpha^{n\alpha}}\alpha^{n-1/2+a-1}\exp\Big\{-\alpha\Big(-\sum_{i=1}^n\log x_i$$
$$+\sum_{i=2}^n\log\frac{1}{\rho_i} - n + b\Big)\Big\}$$
$$\times \beta^{n\alpha+c-1}\exp\Big\{-\beta\Big(\sum_{i=1}^n x_i + d\Big)\Big\}\frac{(n\alpha)^{n\alpha-1/2}}{\Gamma(n\alpha)e^{n\alpha}}.$$

Although the full conditional of β is a gamma distribution, the full conditional density of α does not have a standard form because of the two factors: $g_1(\alpha) = 1/\alpha^{n\alpha}$ and $g_2(\alpha) = (n\alpha)^{n\alpha-1/2}/\{\Gamma(n\alpha)e^{n\alpha}\}$.

First, in order to eliminate $g_1(\alpha)$ from the above expression, we make the change of variables $\gamma = \beta/\alpha$. Then

$$p(\alpha, \gamma|\boldsymbol{\rho}, \boldsymbol{x}) \propto \alpha^{n-1/2+c+a-1}\exp\Big\{-\alpha\Big(-\sum_{i=1}^n\log x_i$$
$$+\sum_{i=2}^n\log\frac{1}{\rho_i} - n + b\Big)\Big\}$$
$$\times \gamma^{n\alpha+c-1}\exp\Big\{-\alpha\gamma\Big(\sum_{i=1}^n x_i + d\Big)\Big\}g_2(\alpha).$$

The full conditional of $\gamma = \beta/\alpha$ is given by $\mathrm{Ga}\big(\gamma\big|n\alpha + c, \alpha\big(\sum_{i=1}^n x_i + d\big)\big)$ and tractable similar to that of the original parameter β.

Next, we use the MH algorithm to update α. The full conditional density of α is given by $p(\alpha|\gamma, \boldsymbol{\rho}, \boldsymbol{x}) \propto \mathrm{Ga}(\alpha|A, B)g_2(\alpha)$, where $A = n - 1/2 + c + a$ and $B = -\sum_{i=1}^n\log x_i + \sum_{i=2}^n\log(1/\rho_i) - n - n\log\gamma + \gamma\big(\sum_{i=1}^n x_i + d\big) + b$. We sample a proposal α^* from $\mathrm{Ga}(\alpha|A, B)$. We accept α^* if an independent standard uniform variable U is less than or equal to $g_2(\alpha^*)/g_2(\alpha^{\mathrm{old}})$, where α^{old} denotes the current value of α. The new value of α, or α^{new}, is set to α^* if α^* is accepted, and to α^{old} otherwise.

The MH within Gibbs sampler is summarized as follows.

Algorithm 1. The variables α, γ, and $\boldsymbol{\rho}$ are updated in the following way.

- Sample $\gamma^* \sim \mathrm{Ga}\big(n\alpha + c, \alpha\big(\sum_{i=1}^n x_i + d\big)\big)$.
- Sample $\boldsymbol{\rho}^* = (\rho_2^*, \dots, \rho_n^*) \sim \prod_{i=2}^n \mathrm{Beta}(\alpha + (i-1)/n, (n-i+1)/n)$.
- Sample $\alpha^* \sim \mathrm{Ga}(A, B)$, where $A = n - 1/2 + c + a$ and

$$B = -\sum_{i=1}^n\log x_i + \sum_{i=2}^n\log\frac{1}{\rho_i^*} - n - n\log\gamma^*$$
$$+\gamma^*\Big(\sum_{i=1}^n x_i + d\Big) + b,$$

and accept α^* with probability

$$\min\Big\{1, \frac{(n\alpha^*)^{n\alpha^*-1/2}}{\Gamma(n\alpha^*)e^{n\alpha^*}}\Big/\frac{(n\alpha)^{n\alpha-1/2}}{\Gamma(n\alpha)e^{n\alpha}}\Big\}.$$

The accuracy of approximation, or the acceptance probability, has already been evaluated in (3). The acceptance probability is, at least, $1 - 1/(12n\alpha^*)$.

2.3. PTN Data Augmentation

The key to the augmentation strategy of Section 2.1 is to find suitable changes of variables $\widetilde{\boldsymbol{\vartheta}} = \varphi(\boldsymbol{\vartheta}; \xi)$ to eliminate the factor $1/\xi^{m\xi}$ in the second step. Because this is not always straightforward, an alternative method is developed in this section. We modify the proposed method of Section 2.1 by introducing additional latent variables. The main tool is the integral expression in the following lemma.

Lemma 1. Let $m \in \mathbb{N}$. Then

$$\frac{1}{\xi^{m\xi}} = (m\xi)^{1/2}e^{m\xi}\frac{(m\xi)^{m\xi-1/2}}{\Gamma(m\xi)e^{m\xi}}\int_0^\infty w^{m\xi-1}e^{-wm\xi^2}dw$$

for all $\xi > 0$.

We assume that $f(\xi, \boldsymbol{\vartheta}) \overset{\xi}{\propto} \mathrm{Ga}(\xi|a_1, b_1)$ for simplicity and consider the conditional density

$$p(\xi|\boldsymbol{\rho}, \boldsymbol{\vartheta}) \propto \frac{1}{\xi^{m\xi}}\xi^{a_3-1}e^{-b_3\xi}\frac{(m\xi)^{m\xi-1/2}}{\Gamma(m\xi)e^{m\xi}},$$

where $a_3 = a_1 + m - 1/2$ and $b_3 = b_1 - m + \sum_{j=2}^{m} \log(1/\rho_j)$. Using Lemma 1, we see that $p(\xi|\boldsymbol{\rho}, \boldsymbol{\vartheta})$ is the marginal density of

$$p(\xi, w|\boldsymbol{\rho}, \boldsymbol{\vartheta}) \propto \xi^{a_3-1/2} e^{-(b_3-m)\xi} \left\{ \frac{(m\xi)^{m\xi-1/2}}{\Gamma(m\xi)e^{m\xi}} \right\}^2 w^{m\xi-1} e^{-wm\xi^2},$$

where $w \in (0, \infty)$ is an additional latent variable. Clearly, $p(w|\xi, \boldsymbol{\rho}, \boldsymbol{\vartheta}) = \mathrm{Ga}(w|m\xi, m\xi^2)$. On the other hand,

$$p(\xi|w, \boldsymbol{\rho}, \boldsymbol{\vartheta}) / \left\{ \frac{(m\xi)^{m\xi-1/2}}{\Gamma(m\xi)e^{m\xi}} \right\}^2 \propto \xi^{c-1} e^{-a\xi^2+b\xi}, \tag{4}$$

where $c = a_3 + 1/2$, $a = mw$, and $b = m \log w + m - b_3$. The right-hand side is proportional to the power truncated normal (PTN) distribution (He, Polson, and Xu 2021) with parameters c, a, and b, which is denoted by $\mathrm{PTN}(c, a, b)$. Since the denominator of the left-hand side in (4) is almost constant as seen in (3), the conditional density $p(\xi|w, \boldsymbol{\rho}, \boldsymbol{\vartheta})$ is approximated by $\mathrm{PTN}(\xi|c, a, b)$. Then, we generate a proposal, $\xi^* \sim \mathrm{PTN}(c, a, b)$, and accept it with probability

$$\min \left\{ 1, \left\{ \frac{(m\xi^*)^{m\xi^*-1/2}}{\Gamma(m\xi^*)e^{m\xi^*}} \Big/ \frac{(m\xi^{\mathrm{old}})^{m\xi^{\mathrm{old}}-1/2}}{\Gamma(m\xi^{\mathrm{old}})e^{m\xi^{\mathrm{old}}}} \right\}^2 \right\},$$

where ξ^{old} is the current state of ξ.

2.4. Additional Data Augmentation for the PTN Distribution

In order to sample from the PTN distribution (4), one can use the accept/reject algorithm described by He, Polson, and Xu (2021). In this article, we consider other approaches so that we do not necessarily need to use accept/reject algorithms. Our approaches also have potential flexibility that they are easily extended to the case where $f(\xi, \boldsymbol{\vartheta})$ is proportional to a generalized-inverse-Gaussian density as a function of ξ.

Let $M > 0$ be a constant possibly dependent on w, $\boldsymbol{\rho}$, and $\boldsymbol{\vartheta}$ such that $M > b$. (A convenient choice is $M = 1 + \max\{0, b\}$.) Then, the PTN density is written as

$$\mathrm{PTN}(\xi|c, a, b) \propto \xi^{c-1} e^{-a\xi^2 - b'\xi} e^{M\xi},$$

where c, a, $b' = M - b$, and M are all positive.

The exponential term $e^{M\xi}$ can be augmented in two ways. The first approach is based on the following expression:

$$\begin{aligned} \mathrm{PTN}(\xi|c, a, b) &\propto \xi^{c-1} e^{-a\xi^2 - b'\xi} \sum_{\zeta=0}^{\infty} \frac{M^\zeta \xi^\zeta}{\zeta!} \\ &= \sum_{\zeta=0}^{\infty} \frac{M^\zeta \xi^\zeta}{\zeta!} \xi^{c-1} e^{-a\xi^2} \\ &\quad \int_0^\infty \frac{1}{\sqrt{2\pi}} \eta^{1/2-1} e^{-\eta/2} e^{-(b')^2\xi^2/(2\eta)} d\eta, \end{aligned}$$

where we consider $\zeta \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ and $\eta \in (0, \infty)$ as additional latent variables. Then the full conditional distributions of ζ and η are $\mathrm{Po}(\zeta|M\xi)$ and $\mathrm{GIG}(\eta|1/2, 1, (b')^2\xi^2)$, respectively. The full conditional density of ξ divided by $\{(m\xi)^{m\xi-1/2}/\Gamma(m\xi)e^{m\xi}\}^2$ is proportional to

$$\xi^{\zeta+c-1} e^{-\{a+(b')^2/(2\eta)\}\xi^2}.$$

We can easily sample from the above distribution since it is simply the square root of a gamma variable.

The second approach uses the integral expression based on the normal density.

Lemma 2. For all $\xi > 0$, we have

$$e^\xi = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\xi}} e^{-\theta^2/(4\xi)+\theta} d\theta.$$

By this lemma, we have

$$\begin{aligned} \mathrm{PTN}(\xi|c, a, b) &\propto \xi^{c-1} e^{-a\xi^2 - b'\xi} \int_{-\infty}^{\infty} \xi^{1/2-1} e^{-\theta^2/(4M\xi)+\theta} d\theta \\ &\propto \int_{-\infty}^{\infty} e^\theta \xi^{c-1/2-1} e^{-a\xi^2} \\ &\quad \left[\int_0^\infty \eta^{1/2-1} e^{-\eta/2} e^{-\{b'\xi+\theta^2/(4M\xi)\}^2/(2\eta)} d\eta \right] d\theta, \end{aligned}$$

where we consider $\theta \in (-\infty, \infty)$ and $\eta \in (0, \infty)$ as additional latent variables. Sampling from the full conditional of (η, θ) can be done in a compositional way; we sample θ (with η marginalized out) from $\mathrm{N}(\theta|2M\xi, 2M\xi)$, then $(\eta|\theta)$ from $\mathrm{GIG}(\eta|1/2, 1, \{b'\xi + \theta^2/(4M\xi)\}^2)$. The full conditional density of ξ divided by $\{(m\xi)^{m\xi-1/2}/\Gamma(m\xi)e^{m\xi}\}^2$ is proportional to

$$\xi^{c-1/2-1} e^{-\{a+(b')^2/(2\eta)\}\xi^2} e^{-\{\theta^4/(32M^2\eta)\}/\xi^2},$$

which is the square root of a generalized-inverse-Gaussian distribution.

3. Student's *t*-Distribution

3.1. Sampling Algorithm

Student's t-distribution is widely adopted in Bayesian inference to handle outliers in samples or heavy-tailed properties of data generating processes (e.g., Geweke 1993; Fonseca, Ferreira, and Migon 2008; Villa and Rubio 2018; da Silva, Prates, and Goncalves 2020). A typical problem in using Student's t-distribution is that the posterior inference of the degrees of freedom is not straightforward since its full conditional distribution has a complicated form. However, we can use our data-augmentation approach. We here consider the simplest case where the means of all observations are the same, and we use the normal-scale-mixture representation of Student's t-distribution, under which the degrees-of-freedom parameter is regarded as the shape parameter in the gamma distribution.

Suppose that for $i = 1, \dots, n$,

$$\begin{aligned} x_i &\sim \mathrm{t}(x_i|(\theta, \tau), 2\alpha) = \frac{\Gamma(\alpha + 1/2)}{\sqrt{2\pi}\tau^{1/2}\alpha^{1/2}\Gamma(\alpha)} \Big/ \left\{ 1 + \frac{(x_i - \theta)^2}{2\tau\alpha} \right\}^{\alpha+1/2}, \\ (\theta, \tau) &\sim p(\theta, \tau), \quad \alpha \sim p(\alpha), \end{aligned}$$

where $x_i \in \mathbb{R}$, $\theta \in \mathbb{R}$, and $\tau \in (0, \infty)$. Then the posterior distribution $p(\theta, \tau, \alpha|\boldsymbol{x})$ is obtained as the marginal distribution of

$$p(\theta, \tau, \alpha, \boldsymbol{w}, \boldsymbol{\rho}|\boldsymbol{x}) \propto \frac{p(\theta, \tau)}{\tau^{n/2}} p(\alpha) \alpha^{n+1/2-1} e^{n\alpha}$$

$$\left(\prod_{i=1}^{n}[w_i{}^{\alpha+1/2-1}e^{-w_i\{\alpha+(x_i-\theta)^2/(2\tau)\}}]\right) \times \left[\prod_{i=2}^{n}\{\rho_i{}^{\alpha+(i-1)/n-1}(1-\rho_i)^{(n-i+1)/n-1}\}\right] \times \frac{(n\alpha)^{n\alpha-1/2}}{\Gamma(n\alpha)e^{n\alpha}}, \tag{5}$$

where $\boldsymbol{w} = (w_1, \ldots, w_n) \in (0, \infty)^n$ and $\boldsymbol{\rho} = (\rho_2, \ldots, \rho_n) \in (0, 1)^{n-1}$ are additional latent variables. The above expression is derived in Section S6 of the supplementary materials by using Theorem 1.

If we use the priors $p(\theta, \tau) = \mathrm{N}(\theta|b, \tau/a)\mathrm{IG}(\tau|c, d)$ and $p(\alpha) = \mathrm{Ga}(\alpha|a_0, b_0)$ for $a, c, d \in (0, \infty)$ and $b \in \mathbb{R}$ and $a_0, b_0 \in (0, \infty)$, we can use the following algorithm to generate posterior samples.

Algorithm 2. The variables θ, τ, α, $\boldsymbol{w}$, and $\boldsymbol{\rho}$ are updated in the following way.

– Sample $\tau^* \sim \mathrm{IG}(c', d')$, where $c' = n/2 + c$ and

$$d' = \frac{1}{2}\left\{ab^2 + \sum_{i=1}^{n} w_i x_i{}^2 - \frac{\left(ab + \sum_{i=1}^{n} w_i x_i\right)^2}{a + \sum_{i=1}^{n} w_i}\right\} + d.$$

– Sample $\theta^* \sim \mathrm{N}(b', \tau^*/a')$, where $a' = a + \sum_{i=1}^{n} w_i$ and

$$b' = \frac{ab + \sum_{i=1}^{n} w_i x_i}{a + \sum_{i=1}^{n} w_i}.$$

– Sample $\boldsymbol{w}^* = (w_1^*, \ldots, w_n^*) \sim \prod_{i=1}^{n} \mathrm{Ga}(\alpha + 1/2, \alpha + (x_i - \theta^*)^2/(2\tau^*))$.
– Sample $\boldsymbol{\rho}^* = (\rho_2^*, \ldots, \rho_n^*) \sim \prod_{i=2}^{n} \mathrm{Beta}(\alpha + (i-1)/n, (n - i + 1)/n)$.
– Sample $\alpha^* \sim \mathrm{Ga}(a_0{}', b_0{}')$, where $a_0{}' = a_0 + n - 1/2$ and

$$b_0{}' = b_0 - n + \sum_{i=1}^{n}(w_i^* - \log w_i^*) + \sum_{i=2}^{n} \log \frac{1}{\rho_i^*},$$

and accept α^* with probability

$$\min\left\{1, \frac{(n\alpha^*)^{n\alpha^*-1/2}}{\Gamma(n\alpha^*)e^{n\alpha^*}} \Big/ \frac{(n\alpha)^{n\alpha-1/2}}{\Gamma(n\alpha)e^{n\alpha}}\right\}.$$

Since we introduce the additional latent variables $\rho_2, \ldots, \rho_n$, our method is less efficient than an alternative method in terms of the effective sample size for an MCMC sequence of a fixed number of parameter values. However, since we do not need to use numerical approximation, our method takes less time. These are confirmed in Section 3.2.

We remark that our method is flexible and we can use many other types of priors. For example, we can use a scale mixture of gamma distributions as a prior for α. We can use a truncated gamma prior for α and this case is considered in the second half of Section 3.2. Also, for $a_0, b_0, c_0 \in (0, \infty)$, we can use the beta-type prior $p(\alpha) \propto \alpha^{a_0-1}(1 - \alpha/c_0)^{b_0-1}\chi_{(0,c_0)}(\alpha)$.

3.2. Simulation Study

Here, we compare the performance of our method based on data augmentation (DA) with the performance of an alternative method based on the approximation proposed by Miller (2019) (A-MH). See Section S7 of the supplementary materials for details of the A-MH method.

First, we set either $n = 10$, $n = 30$, or $n = 100$ and use the conjugate prior $p(\theta, \tau) = \mathrm{N}(\theta|0, \tau/(1/10)) \times \mathrm{IG}(\tau|1/10, 1/10)$ and the gamma prior $p(\alpha) = \mathrm{Ga}(\alpha|1/10, 1/10)$. We generate x_i from $\mathrm{t}(x_i|(3, 1), 2\alpha_0)$. We consider the cases $2\alpha_0 = 1/10$, $2\alpha_0 = 1$, and $2\alpha_0 = 10$. Then, for each of the two methods, we generate 4,000 posterior samples after discarding the first 1000 samples. We use $(\varepsilon, M) = (10^{-8}, 10)$ for the convergence tolerance and the maximum number of iterations for the A-MH method as recommended in Miller (2019). We repeat this simulation 100 times.

Boxplots of the ratios of the effective sample sizes for α, τ, and θ to the computation times for the two methods are shown in Figure 1 for $n = 10$. (For the boxplots for $n = 30$ and $n = 100$, see Figures S1 and S2 of the supplementary materials.) Table 1 reports the averages over the simulations of the ratios (sESS) and the original effective sample sizes (ESS), as well as the mean squared error (MSE) ratios of the estimators of α, τ, and θ, where the MSE ratio is defined as the MSE of the alternative method divided by that of our proposed method. In terms of MSE, there is little difference between the two methods in many cases including those in the supplementary materials. In terms of sESS, our method is better especially for θ and τ when $n = 10$. When $n = 30$, the alternative method becomes better in terms of α and competitive in terms of τ and θ. When $n = 100$, the alternative method is clearly better than ours. This increase of sESS of the DA method for large n is most likely due to the increased number of latent parameters $\rho_{2:n}$, affecting both efficiency and computational time. For example, the ESSs of center parameter θ are almost unchanged (or even improve) when n increases from 30 to 100, hence, the decrease of the sESSs for θ is mainly due to the increased computational time.

Thus, when n is large, our method benefits rather from its simplicity and applicability to more complicated models. To see this point, we consider additional scenarios where a truncated gamma prior is used for the shape parameter; $p(\alpha) \propto \mathrm{Ga}(\alpha|1/10, 1/10)\chi_{(\underline{\alpha},\infty)}(\alpha)$, where $\underline{\alpha} > 0$. With this truncated priors, the method of Miller (2019) must evaluate the expected values of truncated gamma distributions, taking longer time for posterior computation. In contrast, no complication is needed for our method to use the truncated prior, except that we now need to sample from truncated distributions. We set $2\alpha_0 = 10$ and conduct the same simulation study for the truncated gamma prior with $2\underline{\alpha} = 1, 3$.

Boxplots of sESSs are shown in Figure 2 for $n = 10$ (and in Figures S3 and S4 for $n = 30$ and $n = 100$, respectively), and Table 2 lists the averages of ESSs, sESSs and the ratios of MSEs computed in this experiment. In these scenarios, our method becomes more competitive even for large n. In particular, our method outperforms the A-MH method in terms of sESS for θ and τ when $n = 10$ and $n = 30$, and for θ when $n = 100$.

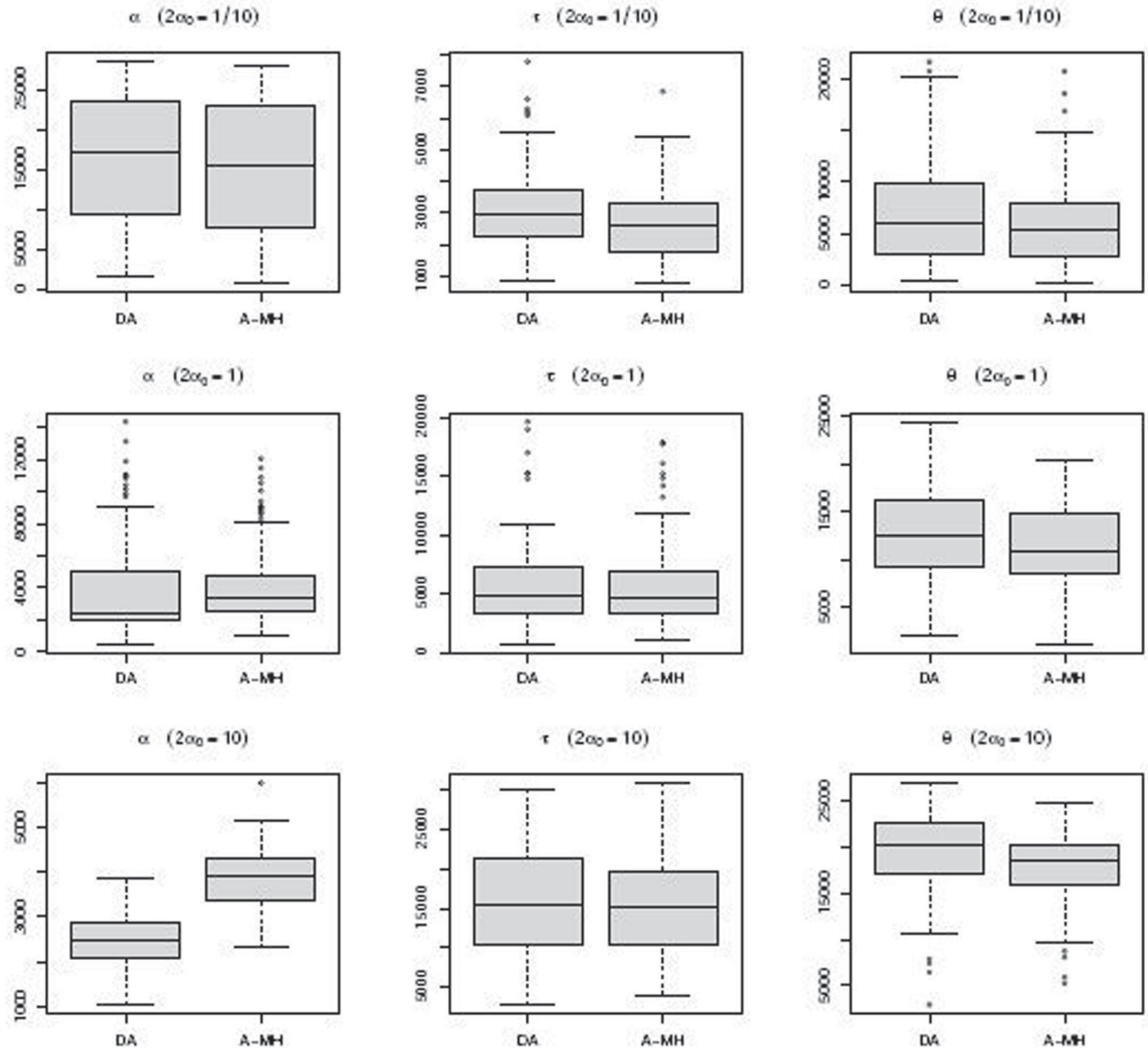

Figure 1. Boxplots of the effective sample sizes standardized by the computation times for the proposed method (DA) and the alternative method (A-MH) for $n = 10$.

Table 1. The averages of the effective sample sizes (ESS) for the proposed method (DA) and the alternative method (A-MH) by Miller (2019), the averages of those standardized by computation time (sESS), and the ratios of the mean squared errors (MSE) of the A-MH method to those of the DA method.

			ESS			sESS			MSE ratio		
n	$2\alpha_0$	Method	θ	τ	α	θ	τ	α	θ	τ	α
10	0.1	DA	874	382	1972	7148	3121	16058	–	–	–
10	0.1	A-MH	805	361	2097	5887	2643	15408	1.14	8.87	1.06
10	1	DA	1506	686	466	12749	5815	3926	–	–	–
10	1	A-MH	1514	755	581	11252	5648	4285	1.00	1.05	0.93
10	10	DA	2276	1862	288	19406	15875	2458	–	–	–
10	10	A-MH	2316	1989	505	17634	15193	3854	1.00	1.01	0.84
30	0.1	DA	796	188	2068	4836	1135	12525	–	–	–
30	0.1	A-MH	834	178	2103	5115	1095	12914	1.10	0.79	1.02
30	1	DA	1408	444	510	9019	2837	3252	–	–	–
30	1	A-MH	1440	473	645	9114	2991	4072	1.00	1.00	1.00
30	10	DA	2371	715	122	15235	4570	782	–	–	–
30	10	A-MH	2448	921	220	16451	6235	1479	1.02	1.02	1.03
100	0.1	DA	904	109	1770	3200	384	6270	–	–	–
100	0.1	A-MH	895	105	1802	4040	472	8120	1.00	0.77	1.00
100	1	DA	1359	386	527	4889	1388	1888	–	–	–
100	1	A-MH	1364	391	640	6251	1794	2926	1.02	1.01	1.03
100	10	DA	2711	284	56	10380	1093	214	–	–	–
100	10	A-MH	2721	407	100	13793	2075	507	1.01	1.04	0.98

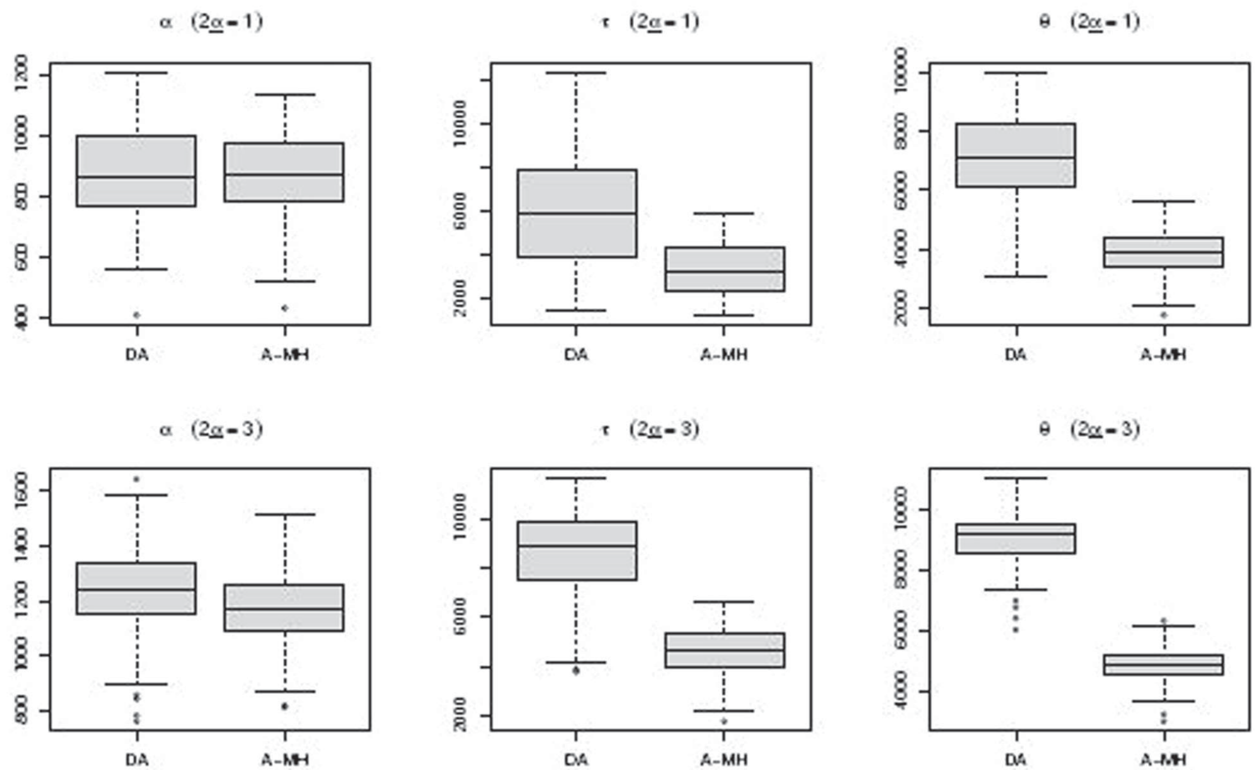

Figure 2. Boxplots of the effective sample sizes standardized by the computation times for the proposed method (DA) and the alternative method (A-MH) for $2\underline{\alpha} = 1, 3$ for $n = 10$.

Table 2. The averages of the effective sample sizes (ESS) for the proposed method (DA) and the alternative method (A-MH), the averages of those standardized by computation time (sESS), and the ratios of the mean squared errors (MSE) of the A-MH method to those of the DA method for $2\underline{\alpha} = 1, 3$.

			ESS			sESS			MSE ratio		
n	$2\alpha_0$	Method	θ	τ	α	θ	τ	α	θ	τ	α
10	1	DA	2376	2033	293	7029	6013	867	–	–	–
10	1	A-MH	2371	2092	533	3866	3419	868	1.01	0.98	0.96
10	3	DA	2885	2735	396	8963	8494	1230	–	–	–
10	3	A-MH	2893	2761	695	4877	4660	1171	1.00	1.00	0.99
30	1	DA	2341	601	126	6354	1627	342	–	–	–
30	1	A-MH	2371	767	217	4074	1325	373	1.01	0.99	0.89
30	3	DA	2710	1057	141	7376	2871	383	–	–	–
30	3	A-MH	2702	1206	251	4567	2051	424	1.00	1.03	1.03
100	1	DA	2660	247	59	5595	521	123	–	–	–
100	1	A-MH	2678	346	106	4278	552	168	0.99	1.00	0.94
100	3	DA	2679	283	63	5670	600	134	–	–	–
100	3	A-MH	2719	381	108	4300	604	170	1.00	1.03	1.04

4. The Dirichlet-Multinomial Distribution

4.1. Sampling Algorithm

Dirichlet-multinomial distribution is useful for modeling multi-label variables, as used in topic modeling (e.g., Blei, Ng, and Jordan 2003). Since the full conditional distribution of the shape parameters of the Dirichlet distribution includes the reciprocal gamma function, their posterior sampling is typically not straightforward (e.g., Nandram 1998). Although in this section our focus is the estimation of the shape parameters of the Dirichlet-multinomial distribution, our result is also relevant in the context of finite mixture modeling (e.g., Frühwirth-Schnatter 2006).

Suppose that for $i = 1, \dots, n$,

$$\boldsymbol{x}_i \sim \mathrm{Multin}_L(\boldsymbol{x}_i | N_i, \boldsymbol{p}_i) = \frac{N_i!}{\prod_{l=0}^{L}(x_{i,l}!)} \prod_{l=0}^{L} {p_{i,l}}^{x_{i,l}},$$

$$\boldsymbol{p}_i \sim \mathrm{Dir}_L(\boldsymbol{p}_i | \boldsymbol{\alpha}) = \frac{\Gamma\big(\sum_{l=0}^{L} \alpha_l\big)}{\prod_{l=0}^{L} \Gamma(\alpha_l)} \prod_{l=0}^{L} {p_{i,l}}^{\alpha_l - 1},$$

$$\boldsymbol{\alpha} \sim p(\boldsymbol{\alpha}),$$

where $\boldsymbol{x}_i = (x_{i,1}, \dots, x_{i,L})$, $x_{i,0} = N_i - \sum_{l=1}^{L} x_{i,l}$, $\boldsymbol{p}_i = (p_{i,1}, \dots, p_{i,L}) \in D_L = \big\{(\tilde{p}_1, \dots, \tilde{p}_L) \in (0,1)^L \big| \tilde{p}_1, \dots, \tilde{p}_L > 0, \sum_{l=1}^{L} \tilde{p}_l < 1\big\}$, $p_{i,0} = 1 - \sum_{l=1}^{L} p_{i,l}$, and $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_L)$. Let $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ and $\boldsymbol{p} = (\boldsymbol{p}_1, \dots, \boldsymbol{p}_n)$. Since we have been unable to find good changes of variables to perform the second step of Section 2.1, we use the flexible method of Section 2.3. In Section S6 of the supplementary materials, we prove that the posterior distribution $p(\boldsymbol{p}, \boldsymbol{\alpha}|\boldsymbol{x})$ is obtained as the marginal distribution of

$$
\begin{aligned}
p(\boldsymbol{p}, \boldsymbol{\alpha}, \boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\rho}|\boldsymbol{x}) \propto p(\boldsymbol{\alpha}) &\Big\{ \prod_{l=0}^{L} ({\alpha_l}^n e^{2n\alpha_l}) \Big\} \Big(\prod_{i=1}^{n} \prod_{l=0}^{L} {p_{i,l}}^{x_{i,l}+\alpha_l-1} \Big) \\
&\Big\{ \prod_{i=1}^{n} ({z_i}^{\sum_{l=0}^{L} \alpha_l - 1} e^{-z_i}) \Big\} \\
&\times \Big\{ \prod_{l=0}^{L} ({w_l}^{n\alpha_l-1} e^{-w_l n {\alpha_l}^2}) \Big\} \\
&\Big[\prod_{i=2}^{n} \prod_{l=0}^{L} \{ {\rho_{i,l}}^{\alpha_l+(i-1)/n-1} (1-\rho_{i,l})^{(n-i+1)/n-1} \} \Big] \\
&\times \prod_{l=0}^{L} \Big\{ \frac{(n\alpha_l)^{n\alpha_l-1/2}}{\Gamma(n\alpha_l) e^{n\alpha_l}} \Big\}^2, \qquad (6)
\end{aligned}
$$

where $\boldsymbol{z} = (z_1, \dots, z_n) \in (0, \infty)^n$, $\boldsymbol{w} = (w_0, \dots, w_L) \in (0, \infty)^{L+1}$, and $\boldsymbol{\rho} = (\boldsymbol{\rho}_2, \dots, \boldsymbol{\rho}_n) = ((\rho_{2,0}, \dots, \rho_{2,L}), \dots, (\rho_{n,0}, \dots, \rho_{n,L})) \in (0,1)^{(n-1)(L+1)}$ are additional latent variables.

If we use the prior $p(\boldsymbol{\alpha}) = \prod_{l=0}^{L} \mathrm{Ga}(\alpha_l|a, b)$ for example, we can use the following algorithm to generate posterior samples.

Algorithm 3. The variables $\boldsymbol{p}$, $\boldsymbol{\alpha}$, $\boldsymbol{z}$, $\boldsymbol{w}$, and $\boldsymbol{\rho}$ are updated in the following way.

- Sample $\boldsymbol{p}^* = ((p_{1,0}^*, \dots, p_{1,L}^*), \dots, (p_{n,0}^*, \dots, p_{n,L}^*)) \sim \prod_{i=1}^{n} \mathrm{Dir}_L(\boldsymbol{x}_i + \boldsymbol{\alpha})$.
- Sample $\boldsymbol{z}^* = (z_1^*, \dots, z_n^*) \sim \big\{ \mathrm{Ga}\big(\sum_{l=0}^{L} \alpha_l, 1 \big) \big\}^n$.
- Sample $\boldsymbol{w}^* = (w_0^*, \dots, w_L^*) \sim \prod_{l=0}^{L} \mathrm{Ga}(n\alpha_l, n{\alpha_l}^2)$.
- Sample $\boldsymbol{\rho}^* = ((\rho_{2,0}^*, \dots, \rho_{2,L}^*), \dots, (\rho_{n,0}^*, \dots, \rho_{n,L}^*)) \sim \prod_{i=2}^{n} \prod_{l=0}^{L} \mathrm{Beta}(\alpha_l + (i-1)/n, (n-i+1)/n)$.
- For $l = 0, \dots, L$, let $c_l = n + a$, $a_l = nw_l^*$, and

$$
\begin{aligned}
b_l = &-\sum_{i=1}^{n} \log \frac{1}{p_{i,l}^*} + \sum_{i=1}^{n} \log z_i^* + 2n + n \log w_l^* \\
&- \sum_{i=2}^{n} \log \frac{1}{\rho_{i,l}^*} - b
\end{aligned}
$$

and sample α_l^* in one of the following three ways and accept it with probability

$$
\min \Big\{ 1, \Big\{ \frac{(n\alpha_l^*)^{n\alpha_l^*-1/2}}{\Gamma(n\alpha_l^*) e^{n\alpha_l^*}} \Big\}^2 \Big/ \Big\{ \frac{(n\alpha_l)^{n\alpha_l-1/2}}{\Gamma(n\alpha_l) e^{n\alpha_l}} \Big\}^2 \Big\}.
$$

(i) Sample $\alpha_l^* \sim \mathrm{PTN}(c_l, a_l, b_l)$ by using the PTN sampler developed by He, Polson, and Xu (2021).

(ii) Let $M_l = 1 + \max\{0, b_l\}$ and ${b_l}' = M_l - b_l$.

  - Sample $\zeta_l^* \sim \mathrm{Po}(M_l \alpha_l)$.
  - Sample $\eta_l^* \sim \mathrm{GIG}(1/2, 1, ({b_l}')^2 {\alpha_l}^2)$.
  - Sample $\tilde{\alpha}_l^* \sim \mathrm{Ga}((\zeta_l^* + c_l)/2, a_l + ({b_l}')^2/(2\eta_l^*))$ and set $\alpha_l^* = (\tilde{\alpha}_l^*)^{1/2}$.

(iii) Let $M_l = 1 + \max\{0, b_l\}$ and ${b_l}' = M_l - b_l$.

  - Sample $\theta_l^* \sim \mathrm{N}(2M_l\alpha_l, 2M_l\alpha_l)$.
  - Sample $\eta_l^* \sim \mathrm{GIG}(1/2, 1, \{{b_l}'\alpha_l + {\theta_l}^2/(4M_l\alpha_l)\}^2)$.
  - Sample $\tilde{\alpha}_l^* \sim \mathrm{GIG}(c_l/2 - 1/4, 2a_l + ({b_l}')^2/\eta_l^*, (\theta_l^*)^4/(16{M_l}^2\eta_l^*))$ and set $\alpha_l^* = (\tilde{\alpha}_l^*)^{1/2}$.

4.2. Simulation Study

In this section, we conduct a simulation study—the posterior inference of Dirichlet shape parameters—to compare our method and the method of He, Polson, and Xu (2021). Both methods are based on data augmentation but in different ways. Many other standard methods, including one by Miller (2019), are not directly applicable.

Following He, Polson, and Xu (2021), we set $L + 1 = 10$ and $N_1 = \cdots = N_n = 500$ and use the prior $p(\boldsymbol{\alpha}) = \prod_{j=0}^{9} \mathrm{Ga}(\alpha_l|b/10, b)$ with $b = 1$. We generate $\boldsymbol{p}_i$ from $\mathrm{Dir}_9(\boldsymbol{\alpha}_0)$ and then $\boldsymbol{x}_i$ from $\mathrm{Multin}_9(500, \boldsymbol{p}_i)$. We consider the cases $n = 100$ and $n = 1000$. For each of these cases, we consider two scenarios: (I) $\boldsymbol{\alpha}_0 = (1/10, \dots, 1/10)$ (equal case) and (II) $\boldsymbol{\alpha}_0 = (1/10, 2/10, \dots, 10/10)$. Other scenarios are also considered and reported in the supplementary materials. We generate 4000 posterior samples after discarding the first 1000 samples. We repeat this simulation 100 times. The method of He, Polson, and Xu (2021) requires sampling from the exponential reciprocal gamma (ERG) distribution, for which they gave three methods. We use the first method because it is the easiest to implement. Setting N equal to a large value in (16) of He, Polson, and Xu (2021) makes their approximation accurate. We set $N = 3$, so that their approximation is sufficiently accurate.

We consider the proposed method based on (iii), (ii), and (i) of Algorithm 3 (DA-N, DA-P and DA-PT, respectively), as well as the method of He, Polson, and Xu (2021) (ERG). Using these methods, we calculate the averages over the simulations of

Table 3. The average effective sample size (ESS), the average computation time (CT), the standardized effective sample size by the computation time (sESS), and the mean squared error (MSE) for the proposed data-augmentation method with normal latent variables (DA-N), Poisson latent variables (DA-P), and the PTN sampler of He, Polson, and Xu (2021) (DA-PT) and the original method proposed by He, Polson, and Xu (2021) (ERG).

n	Scenario	Method	ESS	CT	sESS	MSE
100	(I)	DA-N	863	2.1	414	0.82
		DA-P	856	1.8	480	0.83
		DA-PT	1199	1.8	667	0.82
		ERG	1808	81.4	22	0.82
100	(II)	DA-N	580	2.0	287	5.70
		DA-P	668	1.7	389	5.68
		DA-PT	846	1.8	478	5.73
		ERG	1315	79.5	17	5.66
1000	(I)	DA-N	846	7.7	110	0.80
		DA-P	846	7.4	115	0.80
		DA-PT	1192	7.4	162	0.80
		ERG	1800	819.1	2	0.80
1000	(II)	DA-N	582	7.6	76	5.06
		DA-P	671	7.2	93	5.18
		DA-PT	834	7.3	115	5.15
		ERG	1317	835.4	2	5.17

NOTE: These values are averaged over $\alpha_0, \dots, \alpha_9$. MSE values under $n = 100$ and $n = 1000$ are multiplied by 10^3 and 10^4, respectively.

the means of the effective sample sizes for $\alpha_0, \dots, \alpha_9$ (ESS), the averages over the simulations of the computation times (CT), and the averages over the simulations of the ratios of the means of the effective sample sizes to the computation times (sESS). We also calculate the mean squared errors (MSE) of the estimators of $\alpha_0, \dots, \alpha_9$.

The results are reported in Table 3. In all scenarios, the ERG method has the largest ESS but the longest CT. In contrast, the DA-N, DA-P, and DA-PT methods are less competitive than the ERG method in ESSs, but significantly outperform it in computational time. Consequently, all of our methods have much larger sESSs than the state-of-the-art ERG method. Among the proposed methods, the DA-PT method has the best sESS. The other two methods cost computational efficiency for the simplicity of their algorithms, as noted in Section 2.3. In terms of MSE, no significant difference can be seen in the four methods.

Thus, the difference of the method of He, Polson, and Xu (2021) and ours in computational efficiency critically depends on the computational time. For the fairness of comparison, it should be noted that the computation by the ERG method can speed-up by using parallelization, and could be competitive as our methods in some computational environments that enable such parallelization. Other than the efficiency, the advantage of our method to be emphasized is its simplicity; no explicit parallelization is needed in implementing our method. In addition, our method is tuning parameter free, while the ERG method requires tuning N.

5. Concluding Remarks

The data augmentation approach proposed in this article can be applicable to any posterior inference if the conditional posterior involves the reciprocal gamma functions. Examples of such models include the one-parameter Dirichlet, negative binomial and Wishart models, in addition to the gamma, Student's t and Dirichlet-multinomial models considered in the previous sections. The sampling algorithms for those models can be derived straightforwardly and are provided in Section S1 of the supplementary materials.

A remaining issue related to the proposed approach is that our method is likely to be less efficient for extremely small n. In that case, the data augmentation in Theorem S1 should be customized for the model of interest. For example, if $n = 1$ and $0 < \alpha_l \ll 1$ in the Dirichlet-multinomial model, we could improve the proposed augmentation; see Section S5 of the supplementary materials.

Supplementary Materials

Online Supplementary Materials: All proofs and additional simulation experiments are included. The applications to the one-parameter Dirichlet, negative-binomial and Wishart models are also discussed in detail. (pdf)

R-code for t and Dirichlet models: The R-codes to implement the posterior computation for the t model in Section 3.2 and the Dirichlet models in Section 4.2 are publicly available on GitHub repository (*https://github.com/sshonosuke/Gamma-DA*).

Acknowledgments

We would like to thank the editor, the associate editor, and the reviewer for many valuable comments and helpful suggestions that led to an improved version of this article.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Research of the authors was supported in part by JSPS KAKENHI Grant Number 20J10427, 19K11852, 17K17659, and 21H00699 from Japan Society for the Promotion of Science.

References

Bacallado, S., Favaro, S., Power, S., and Trippa, L. (2022), "Perfect Sampling of the Posterior in the Hierarchical Pitman–Yor Process," *Bayesian Analysis*, 17, 685–709. [908]

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022. [914]

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis–Hastings Algorithm," *The American Statistician*, 49, 327–335. [908]

da Silva, N. B., Prates, M. O., and Goncalves, F. B. (2020), "Bayesian Linear Regression Models with Flexible Error Distributions," *Journal of Statistical Computation and Simulation*, 90, 2571–2591. [911]

Devroye, L. (2012), "A Note on Generating Random Variables with Log-concave Densities," *Statistics and Probability Letters*, 82, 1035–1039. [908]

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference using Mixtures," *Journal of the American Statistical Association*, 90, 577–588. [908]

Fonseca, T. C., Ferreira, M. A., and Migon, H. S. (2008), "Objective Bayesian Analysis for the Student-t Regression Model," *Biometrika*, 95, 325–333. [908,911]

Frühwirth-Schnatter, S. (2006), *Finite mixture and Markov switching models*, New York: Springer-Verlag. [914]

Geweke, J. (1993), "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics*, 8, 19–40. [911]

Gilks, W. R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Journal of the Royal Statistical Society*, Series C, 41, 337–348. [908,909]

He, J., Polson, N., and Xu, J. (2021), "Bayesian Inference for Gamma Models," arXiv preprint arXiv:2106.01906. [908,911,915,916]

Miller, J. W. (2019), "Fast and Accurate Approximation of the Full Conditional for Gamma Shape Parameters," *Journal of Computational and Graphical Statistics*, 28, 476–480. [908,909,912,913,915]

Nandram, B. (1998), "A Bayesian Analysis of the Three-Stage Hierarchical Multinomial Model," *Journal of Statistical Computation and Simulation*, 61, 97–126. [908,914]

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," (with Discussion), *The Annals of Statistics*, 22, 1701–1728. [908]

Villa, C., and Rubio, F. J. (2018), "Objective Priors for the Number of Degrees of Freedom of a Multivariate t Distribution and the t-copula," *Computational Statistics and Data Analysis*, 124, 197–219. [911]

Watanabe, T. (2001), "On Sampling the Degree-of-Freedom of Student's-t Disturbances," *Statistics & Probability Letters*, 52, 177–181. [908]

Zhou, M., and Carin, L. (2015), "Negative Binomial Process Count and Mixture Modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37, 307–320. [908]