Stat

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

Continued fraction enhancement of Bayesian computing

Matt P. Wand[a,*] and John T. Ormerod[b]

Received 2 July 2012; Accepted 7 August 2012

The agéd number theoretic concept of continued fractions can enhance certain Bayesian computations. The crux of this claim is due to continued fraction representations of numerically challenging special function ratios that arise in Bayesian computing. Continued fraction approximation via Lentz's Algorithm often leads to efficient and stable computation of such quantities. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: hypergeometric functions; mean field variational Bayes; parabolic cylinder functions; special functions; variable selection; variational approximations

1 Introduction

Bayesian computation is a vibrant area of research that transcends many areas in the statistical and computational sciences. Surveys of various aspects of Bayesian computation are contained in Bishop (2006) and Marin & Robert (2007).

Practicality of Bayesian computation relies on various algorithmic paradigms such as Expectation-Maximization, Markov chain Monte Carlo, junction trees, slice sampling, particle filters, quadrature and coordinate ascent. It is now typical for Bayesian analyses to use a combination of such algorithms. In this article we explain why continued fraction approximation via Lentz's Algorithm deserves to be included on this list because of the enhancement it offers for particular combinations of special functions, especially ratios. Currently, our knowledge of special function approximation via special functions is limited, virtually, to the listings in Cuyt et al. (2008). There may be other approximations, some of which are yet to be developed, that will aid additional Bayesian computation problems.

Section 2 explains continued fractions to readers who have little or no familiarity with this classical branch of number theory. Of particular relevance is Lentz's Algorithm for effective approximation of a real number given its continued fraction expansion. Section 3 catalogues special function ratios that possess simple continued fraction expansions and also arise in Bayesian computing problems. A case study in which continued fractions aid Bayesian variable selection in a regression analysis application is presented in Section 4. Some concluding remarks are made in Section 5.

2 Continued fractions

As a prelude to explaining how continued fractions can enhance Bayesian computation, we briefly describe the concept and associated computational issues here.

[a]School of Mathematical Sciences, University of Technology, Sydney, P.O. Box 123, Broadway, 2007, Australia
[b]School of Mathematics and Statistics, University of Sydney, Sydney, 2006, Australia
***Email: matt.wand@uts.edu.au**

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

The ISI's Journal for the Rapid Dissemination of Statistics Research

2.1. Brief history

Continued fractions are a simple by-product of the greatest common divisor algorithm that the Greek mathematician Euclid devised in approximately 300 B.C. Theory on continued fraction approximation of irrational numbers began with the seminal work of Swiss mathematician Leonhard Euler in the 1700s. Important breakthroughs regarding continued fraction have been taking place ever since. For example, in 1972, American mathematician Bill Gosper first derived practical algorithms for continued fraction arithmetic (Gosper, 1972).

2.2. Simple examples

The most prominent transcendental numbers, e and π, have simple continued fraction representations:

$$e = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{2}{3 + \cfrac{3}{4 + \cdots}}}} \quad \text{and} \quad \pi = \cfrac{4}{1 + \cfrac{1^2}{3 + \cfrac{2^2}{5 + \cfrac{3^2}{7 + \cdots}}}}.$$

2.3. Lentz's algorithm

In Section 5.2 of their famous *Numerical Recipes* book, Press et al. (1992) present a practical algorithm for arbitrarily accurate approximation of a real number given a continued fraction expansion of the number. They use the name *modified Lentz's algorithm* after an algorithm of Lentz (1976), where a particular family of (rather than general) continued fractions is treated. We simply refer to this algorithm as Lentz's Algorithm and, for general continued fractions of the form

$$b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cdots}}}, \tag{1}$$

it is listed as Algorithm 1.

Algorithm 1 Lentz's Algorithm for continued fraction approximation of (1).

Inputs (with defaults): b_0, $a_j, b_j, j \geqslant 1$, ε_1 (10^{-30}), ε_2 (10^{-7}).

$f_{\text{prev}} \leftarrow \varepsilon_1$; $C_{\text{prev}} \leftarrow \varepsilon_2$; $D_{\text{prev}} \leftarrow 0$; $\Delta = 2 + \varepsilon_2$; $j \leftarrow 0$

cycle while $|\Delta - 1| \geqslant \varepsilon_2$:

$j \leftarrow j + 1$; $D_{\text{curr}} \leftarrow b_j + a_j D_{\text{prev}}$; if $D_{\text{curr}} = 0$ then $D_{\text{curr}} \leftarrow \varepsilon_1$

$D_{\text{curr}} \leftarrow 1/D_{\text{curr}}$; $C_{\text{curr}} \leftarrow b_j + a_j/C_{\text{prev}}$; if $C_{\text{curr}} = 0$ then $C_{\text{curr}} \leftarrow \varepsilon_1$

$\Delta \leftarrow C_{\text{curr}} D_{\text{curr}}$; $f_{\text{curr}} \leftarrow f_{\text{prev}} \Delta$; $f_{\text{prev}} \leftarrow f_{\text{curr}}$; $C_{\text{prev}} \leftarrow C_{\text{curr}}$; $D_{\text{prev}} \leftarrow D_{\text{curr}}$

return $b_0 + f_{\text{curr}}$

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

Lentz's Algorithm, in itself, is not a panacea for efficient computation of every quantity having simple continued fraction representations. In our relatively limited experience with Lentz's Algorithm to date, we have found that convergence can be quite rapid for some continued fractions, but slow for others. An illustration of this spotty behaviour is given in Section 2.4. Finally we note that, for some inputs, Algorithm 1 may not converge and standard safeguarding based on not exceeding a fixed number of iterations is required.

2.4. Special function ratios

A class of numerical problems where continued fractions and Lentz's Algorithm offer attractive solutions are certain ratios involving special functions. An example is the *Mills ratio* function

$$\frac{1-\Phi(x)}{\varphi(x)}$$

where φ and Φ are, respectively, the density and cumulative distribution functions of the $N(0,1)$ distribution. The reciprocal of the Mills ratio function is the standard normal hazard function. For large positive x, both the numerator and denominator become infinitesimally small. This leads to numerical instability if ordinary division is used to compute $\{1-\Phi(x)\}/\varphi(x)$ and is apparent when the following code is entered at the prompt within the R computing environment (R Development Core Team, 2012):

```
xg <- seq(0.5,10,length=1001) ; yg <- (1-pnorm(xg))/dnorm(xg) ; plot(xg,yg,type="l")
```

Fortunately, Laplace (1805) discovered the following simple continued fraction expansion:

$$\frac{1-\Phi(x)}{\varphi(x)} = \cfrac{1}{x+\cfrac{1}{x+\cfrac{2}{x+\cfrac{3}{x+\cdots}}}}, \quad x>0. \tag{2}$$

Figure 1 summarizes key aspects of Lentz's Algorithm evaluation of $\{1-\Phi(x)\}/\varphi(x)$ over the interval $[0.5,10]$ based on (2). The top left panel shows rapid decay of the numerator and denominator towards zero as x increases. In the right-hand panels we see that the *ratio* is well above zero for x as high as 10. Lentz's Algorithm appears to be quite stable. In comparison, direct division, based on the above R code, is seen to be unstable and erroneously returns zero for $x>8.29$. The bottom left panel illustrates the variation in convergence speed of Lentz's Algorithm. For $x \geqslant 2$ convergence is seen to be quite rapid. However, the number of iterations required to attain convergence according to the default tolerance setting in Algorithm 1 sky-rockets as $x \to 0$. Fortunately, there is no numerical instability issue with direct division for low positive x and a strategy for evaluation of $\{1-\Phi(x)\}/\varphi(x)$ such as:

if $0 \leqslant x < 2$ then use direct division;
if $x \geqslant 2$ then use Lentz's Algorithm,

offers stable and efficient evaluation for all $x \geqslant 0$.

Particular special functions and combinations of such functions, such as ratios, admit simple continued fractions. Cuyt et al. (2008) provides a survey of such functions, and is a key reference for the next section. However, ratios of general special functions do not necessarily have simple continued fraction representations.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

The ISI's Journal for the Rapid Dissemination of Statistics Research

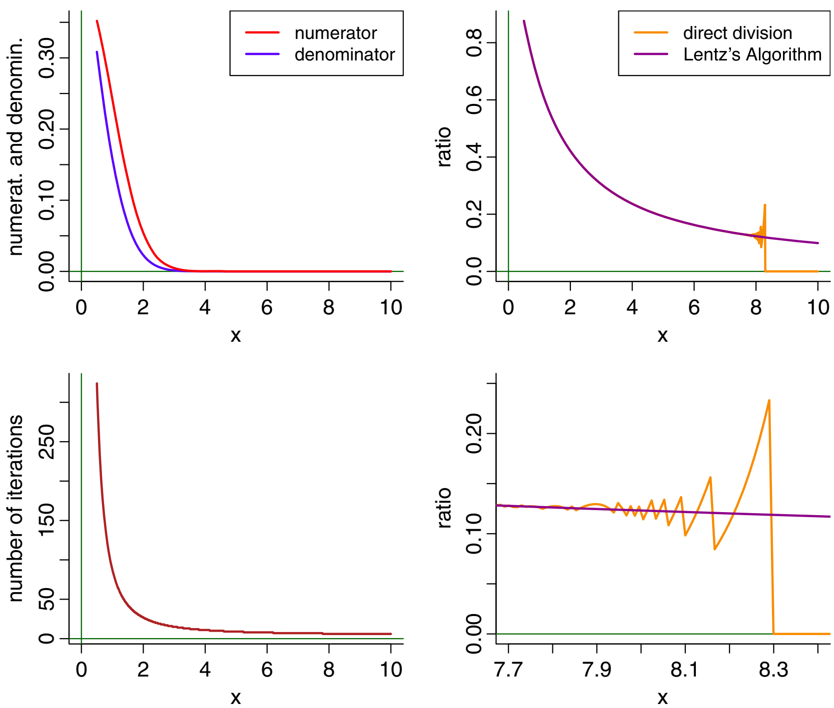

Figure 1. Continued fraction evaluation of the Mills ratio function. Top left panel: the Mills ratio numerator and denominator function. Bottom left panel: number of iterations required for convergence of Lentz's Algorithm, using the default tolerance value $\varepsilon_2 = 10^{-7}$ in Algorithm 1, based on Laplace's continued fraction expansion (2). Top right panel: comparison of Mills ratio values via both direct division and Lentz's Algorithm. Bottom right panel: zoomed version of top right panel.

3 Special function ratios arising in Bayesian computing

Our primary goal in this section is to start a catalogue, of sorts, of continued fractions for special function ratios that arise in Bayesian computing. Table I is a list of special function ratios and their continued fraction approximations. As discussed later in this section, each arises in existing Bayesian computation algorithms. The special functions in Table I are:

$K_\nu(x)$	modified Bessel function of the second kind of order ν,
$E_1(x)$	exponential integral function of order 1,
$D_\nu(x)$	parabolic cylinder function of order ν,
${}_2F_1(\alpha,\beta;\gamma;x)$	Gauss's hypergeometric function of order (α,β,γ),

and, for each of these functions, we follow the definitions used in Gradshteyn & Ryzhik (1994). The notation

$$\mathcal{A}_j(\nu_1,\nu_2) \equiv \frac{\{2\nu_1+2j-1+(-1)^j(2\nu_1-3)\}\{2\nu_1-2j-4\nu_2+5-(-1)^j(2\nu_1-3)\}}{16(\nu_2+j-1)(\nu_2+j-2)} \quad (3)$$

is also used in Table I.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

Table I. Definition and values of b_0, a_j and b_j, $j \geqslant 1$, for some special function ratios that arise in Bayesian computing.

special function ratio	b_0	a_1	$a_j,\ j>1$	$b_j,\ j>0$
$\frac{1-\Phi(x)}{\varphi(x)}$	0	1	$j-1$	x
$\frac{K_{\nu+1}(x)}{K_\nu(x)},\ \nu\in\mathbb{R}$	$\frac{2\nu+2x+1}{2x}$	$\frac{\nu^2-\frac{1}{4}}{x}$	$\nu^2-\frac{1}{4}(2j-1)^2$	$2(x+j)$
$\frac{E_1(x)}{\exp(-x)}$	0	1	$-(j+1)^2$	$x+2j-1$
$\frac{D_{-\nu-2}(x)}{D_{-\nu-1}(x)},\ \nu>0$	0	1	$\nu+j$	x
$\frac{{}_2F_1(\nu_1,2;\nu_2+1;x)}{{}_2F_1(\nu_1,1;\nu_2;x)},\ \nu_1\in\mathbb{R},\nu_2>0$	0	1	$\mathcal{A}_j(\nu_1,\nu_2)\,x$	1

At the time of this writing, Table I comprises almost all of the special function ratios that we have noticed in the Bayesian literature, but it is likely there are others still to be added. We say "almost" since, courtesy of various recurrence formulae, there are some closely related ratios that can be written in terms of the Table I ratios. For example

$$\frac{D_{-\nu-3}(x)}{D_{-\nu-1}(x)}=(\nu+4)^{-1}\left\{1-\frac{x\,D_{-\nu-2}(x)}{D_{-\nu-1}(x)}\right\},\qquad \nu>0,$$

arises in Algorithm 5 of Neville et al. (2012).

3.1. Instances of Table I functions in Bayesian computing

We now link Table I to Bayesian computing, by listing instances where such special function ratios arise.

3.1.1. The special function ratio $\{1-\Phi(x)\}/\varphi(x)$.

The Mills ratio function $\{1-\Phi(x)\}/\varphi(x)$ arises in mean field variational Bayesian approximate inference for probit regression (Girolami & Rogers, 2006; Consonni & Marin, 2007). Algorithm 4 of Ormerod & Wand (2010) is succinct summary of this procedure. The second update can be written in terms of the Mills ratio function after elementary algebraic manipulations.

3.1.2. The special function ratio $K_{\nu+1}(x)/K_\nu(x)$.

Armagan et al. (2011) present a mean field variational Bayes algorithm for approximate inference in the sparse signal regression with shrinkage via three-parameter Beta prior distributions. The updates involve repeated evaluation of the ratio between two modified Bessel functions of the second kind, with orders differing by one.

3.1.3. The special function ratio $E_1(x)/\exp(-x)$.

This ratio involving the exponential integral function of order 1:

$$E_1(x)\equiv\int_x^\infty \frac{e^{-t}}{t}\,dt,\quad x\in\mathbb{R},\ x\neq 0$$

arises in Neville et al. (2012), where mean field variational Bayesian approximate inference for sparse signal regression with a Horseshoe prior distribution (Carvalho et al., 2010) is treated.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

The ISI's Journal for the Rapid Dissemination of Statistics Research

3.1.4. The special function ratio $D_{-\nu-2}(x)/D_{-\nu-1}(x)$. Another continuous sparse signal shrinkage distribution treated in Neville et al. (2012) is the Normal-Exponential-Gamma distribution (Griffin & Brown, 2011). The mean field variational Bayes updates for the single auxiliary variable representation of the Normal-Exponential-Gamma distribution involves repeated computation of the parabolic cylinder function ratio $D_{-\nu-2}(x)/D_{-\nu-1}(x)$ with $\nu > 0$.

3.1.5. The special function ratio ${}_2F_1(\nu_1, 2; \nu_2 + 1; x)/{}_2F_1(\nu_1, 1; \nu_2; x)$. Liang et al. (2008) provide a prescription for Bayesian variable selection in the linear regression model with mixtures of g priors. Section 4 contains a case study on this approach to variable selection and the use of continued fractions, so we will postpone the details. As explained in Section 4, the special function ratio ${}_2F_1(\nu_1, 2; \nu_2 + 1; x)/{}_2F_1(\nu_1, 1; \nu_2; x)$ arises in the Liang et al. (2008) variable selection procedure.

3.2. Cursory convergence assessment

As demonstrated in Figure 1 for the Mills ratio function, convergence of Lentz's Algorithm can be rapid for some values of the special function ratio argument, but slow for other arguments. In this section we perform convergence checks for members of the families of special function ratios listed in Table I. Apart from $E_1(x)/\exp(-x)$, each of the families are parametrized by continuums of order parameters, so a comprehensive convergence assessment is too big a task to attempt here.

Figure 2 shows the behaviour of Lentz's Algorithm for approximation of

$$\frac{K_{2.4}(x)}{K_{1.4}(x)}, \quad \frac{E_1(x)}{\exp(-x)}, \quad \frac{D_{-2.1}(x)}{D_{-1.1}(x)} \quad \text{for } x > 0 \quad \text{and} \quad \frac{{}_2F_1(\nu_1, 2; \nu_2 + 1; x)}{{}_2F_1(\nu_1, 1; \nu_2; x)} \quad \text{for } 0 < x < 0.99.$$

The middle row of plots in Figure 2 show that the special function ratios can assume moderate values, even though the numerators and denominators are both very small or very large. For the ${}_2F_1$ ratio, the numerator and denominator are extremely large for x close to 1. Overflow is unavoidable for larger values of ν_1 and ν_2.

The number of iterations required for convergence of Lentz's Algorithm, using the default tolerance value is low to moderate for those values of x where direct computation is prone to instability. However, since this study involves just one value of the order parameters for each of the special functions it is, admittedly, quite limited and a fuller study is needed to make general recommendations about the use of continued fractions and Lentz's algorithms for computing functions of this type.

4 Case study

As a case study on the use of continued fractions in Bayesian computing, we present a Bayesian regression analysis using mixtures of g priors for variable selection (Liang et al., 2008).

4.1. Setting

Liang et al. (2008) consider the setting

$$\boldsymbol{Y}|\boldsymbol{\mu}, \sigma^2 \sim N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}),$$

where $\boldsymbol{Y}$ is an $n \times 1$ response vector and a set of candidate predictor variables $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p$ is available. We assume that each of these predictors are centred in that $\boldsymbol{X}_j^T \mathbf{1}_n = \mathbf{0}$, where $\mathbf{1}_n$ is the $n \times 1$ vector with each entry equal to 1. We will now outline the mixtures of g priors approach to selection of the $\boldsymbol{X}_j$. Fuller details are given in Liang et al. (2008).

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

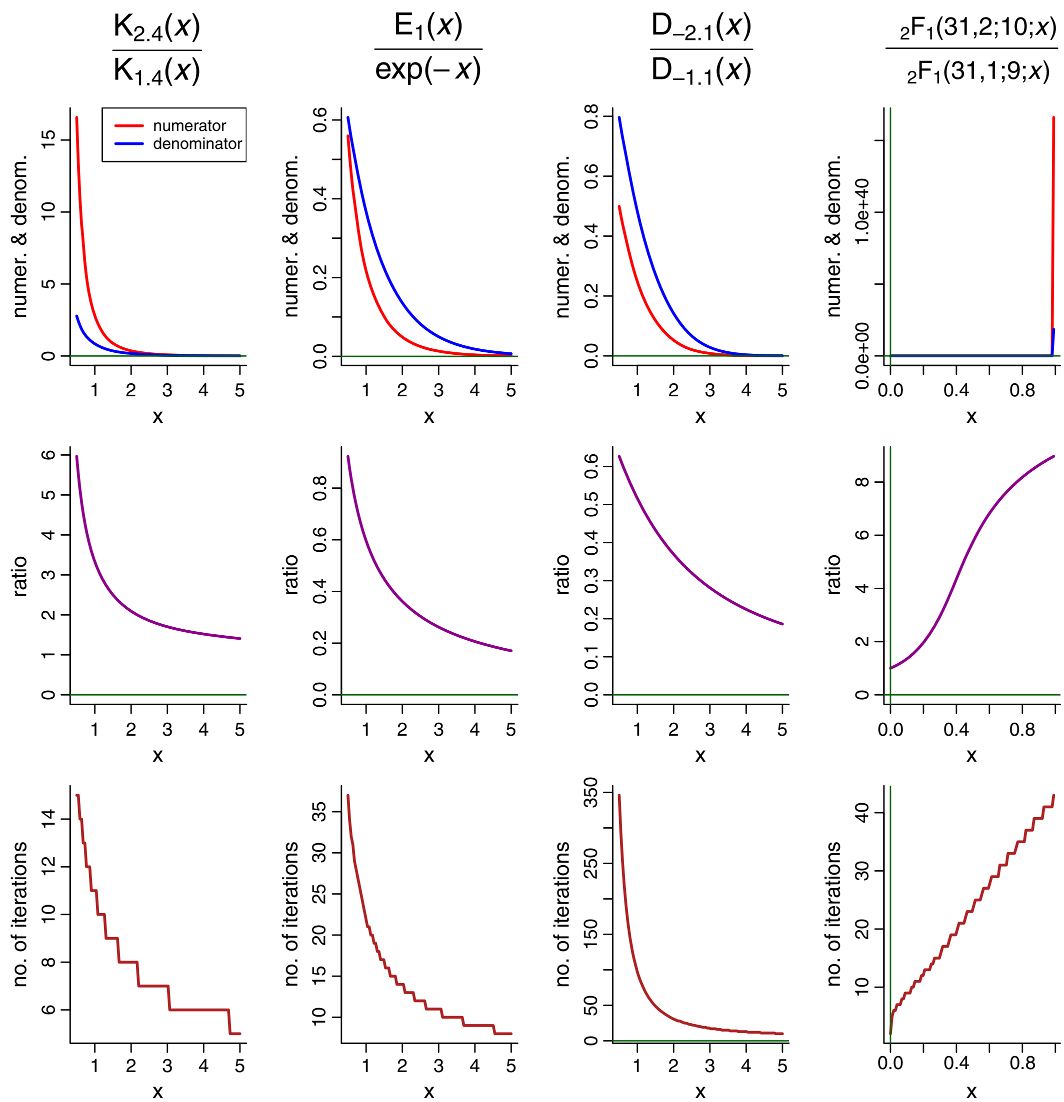

Figure 2. Convergence assessment for Lentz's Algorithm approximation of special function ratios from Table I. First row: formula of special function ratio being assessed. Second row: plots of numerator and denominator functions. Third row: plots of ratio functions. Fourth row: number of iterations required for convergence of Lentz's Algorithm, using the default tolerance value $\varepsilon_2 = 10^{-7}$ in Algorithm 1, based on the continued fraction expansion listed in Table I.

Index a model space by

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix}$$

where $\gamma_j = 1$ if $\boldsymbol{X}_j$ is included in the model and $\gamma_j = 0$ if $\boldsymbol{X}_j$ is excluded. Then define

$$\mathcal{M}_{\boldsymbol{\gamma}}: \quad \boldsymbol{\mu} = \mathbf{1}_n \alpha + \boldsymbol{X}_{\boldsymbol{\gamma}} \boldsymbol{\beta}_{\boldsymbol{\gamma}}$$

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

The ISI's Journal for the Rapid Dissemination of Statistics Research

where $\boldsymbol{X}_{\boldsymbol{\gamma}}$ is a design matrix containing those $\boldsymbol{X}_j$ for which $\gamma_j = 1$ and $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is defined analogously.

The g prior for $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is

$$\boldsymbol{\beta}_{\boldsymbol{\gamma}} | \sigma^2, \mathcal{M}_{\boldsymbol{\gamma}} \sim N\left(\mathbf{0}, g\,\sigma^2 \left(\boldsymbol{X}_{\boldsymbol{\gamma}}^T \boldsymbol{X}_{\boldsymbol{\gamma}}\right)^{-1}\right).$$

There are numerous options for specification of g although the one favoured by Liang et al. (2008) is

$$\frac{g}{1+g} \sim \text{Beta}\left(1, \frac{a}{2} - 1\right)$$

where $a > 0$ is a user-specified hyperparameter. Liang et al. (2008) recommend choosing a from the interval $(2, 4]$. The posterior mean of μ under selection of a model $\mathcal{M}_{\boldsymbol{\gamma}} \neq \mathcal{M}_N$ is

$$E(\boldsymbol{\mu}|\boldsymbol{Y}) = \mathbf{1}_n \overline{Y} + \sum_{\boldsymbol{\gamma}:\mathcal{M}_{\boldsymbol{\gamma}} \neq \mathcal{M}_N} p(\mathcal{M}_{\boldsymbol{\gamma}}|\boldsymbol{Y})\, E\left(\frac{g}{1+g}\middle| \mathcal{M}_{\boldsymbol{\gamma}}, \boldsymbol{Y}\right) \boldsymbol{X}_{\boldsymbol{\gamma}} \left(\boldsymbol{X}_{\boldsymbol{\gamma}}^T \boldsymbol{X}_{\boldsymbol{\gamma}}\right)^{-1} \boldsymbol{X}_{\boldsymbol{\gamma}}^T \boldsymbol{Y} \tag{4}$$

where $\mathcal{M}_N$ is the null model, i.e.

$$\mathcal{M}_N : \quad \boldsymbol{\mu} = \mathbf{1}_n \alpha.$$

Liang et al. (2008) point out that (4) can be expressed in terms of ratios of the ${}_2F_1$ function, since

$$E\left(\frac{g}{1+g}\middle| \mathcal{M}_{\boldsymbol{\gamma}}, \boldsymbol{Y}\right) = \frac{2}{p_{\boldsymbol{\gamma}} + a} \frac{{}_2F_1\left(\frac{1}{2}(n-1), 2; \frac{1}{2}(p_{\boldsymbol{\gamma}} + a) + 1; R_{\boldsymbol{\gamma}}^2\right)}{{}_2F_1\left(\frac{1}{2}(n-1), 1; \frac{1}{2}(p_{\boldsymbol{\gamma}} + a); R_{\boldsymbol{\gamma}}^2\right)} \tag{5}$$

and

$$p(\mathcal{M}_{\boldsymbol{\gamma}}|\boldsymbol{Y}) = \frac{p(\mathcal{M}_{\boldsymbol{\gamma}})\, {}_2F_1\left(\frac{1}{2}(n-1), 1; \frac{1}{2}(p_{\boldsymbol{\gamma}} + a); R_{\boldsymbol{\gamma}}^2\right) / \left(p_{\boldsymbol{\gamma}} + a - 2\right)}{\sum_{\boldsymbol{\gamma}'} p(\mathcal{M}_{\boldsymbol{\gamma}'})\, {}_2F_1\left(\frac{1}{2}(n-1), 1; \frac{1}{2}(p_{\boldsymbol{\gamma}'} + a); R_{\boldsymbol{\gamma}'}^2\right) / \left(p_{\boldsymbol{\gamma}'} + a - 2\right)}. \tag{6}$$

Here $p_{\boldsymbol{\gamma}}$ is the number of non-intercept coefficients and $R_{\boldsymbol{\gamma}}^2$ is the squared correlation of the observed and fitted values corresponding to $\mathcal{M}_{\boldsymbol{\gamma}}$.

Ostensibly, (5) and (6) imply that (4) can be evaluated exactly provided a routine for ${}_2F_1$ (e.g. Hankin, 2007) is available. However, Liang et al. (2008) point out numerical instability problems if the ratios are computed directly. In particular, ${}_2F_1$ can be extremely large if n is large and $R_{\boldsymbol{\gamma}}^2$ is close to 1. Since (5) is a special case of the ${}_2F_1$ ratio appearing in Table I, we can use Algorithm 1 to achieve efficient and stable evaluation of $E\left(\frac{g}{1+g}\middle| \mathcal{M}_{\boldsymbol{\gamma}}, \boldsymbol{Y}\right)$.

Unfortunately, there is no simple continued fraction expansion for the ${}_2F_1$ ratios that arise in $p(\mathcal{M}_{\boldsymbol{\gamma}}|\boldsymbol{Y})$. Additionally, numerical problems could arise due to one term in the numerator summation being extremely dominant. Because of this, we recommend working the following re-expression of (6):

$$p(\mathcal{M}_{\boldsymbol{\gamma}}|\boldsymbol{Y}) = \frac{p(\mathcal{M}_{\boldsymbol{\gamma}}) \exp\left[\log\left\{{}_2F_1\left(\frac{1}{2}(n-1), 1; \frac{1}{2}(p_{\boldsymbol{\gamma}} + a); R_{\boldsymbol{\gamma}}^2\right)\right\} - \text{MAX}\right] / (p_{\boldsymbol{\gamma}} + a - 2)}{\sum_{\boldsymbol{\gamma}'} p(\mathcal{M}_{\boldsymbol{\gamma}'}) \exp\left[\log\left\{{}_2F_1\left(\frac{1}{2}(n-1), 1; \frac{1}{2}(p_{\boldsymbol{\gamma}'} + a); R_{\boldsymbol{\gamma}'}^2\right)\right\} - \text{MAX}\right] / (p_{\boldsymbol{\gamma}'} + a - 2)} \tag{7}$$

where

$$\text{MAX} \equiv \max_{\boldsymbol{\gamma}'} \left[\log\left\{{}_2F_1\left(\frac{1}{2}(n-1), 1; \frac{1}{2}(p_{\boldsymbol{\gamma}'} + a); R_{\boldsymbol{\gamma}'}^2\right)\right\}\right].$$

In Section 4.2. we describe continued fraction evaluation of $\log\{{}_2F_1(a, b; c; x)\}$ which allows stable evaluation of (7).

Stat

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

4.2. Continued fraction evaluation of $\log\{{}_2F_1(a,b;c;x)\}$

A continued fraction expansion of ${}_2F_1(a,b;c;x)$ (Trott, 2012) is

$$ {}_2F_1(a,b;c;x) = 1 + \cfrac{\frac{abx}{c}}{1 + \cfrac{\frac{-(a+1)(b+1)}{2(c+1)}x}{1 + \frac{(a+1)(b+1)}{2(c+1)}x + \cfrac{\frac{-(a+2)(b+2)}{3(c+1)}x}{1 + \frac{(a+2)(b+2)}{3(c+2)}x + \ldots}}}. $$

In the notation of (1), this corresponds to

$$ b_0 = 1, \quad a_1 = \frac{abx}{c}, \quad b_1 = 1, \quad \text{and} $$
$$ a_j = \frac{-(a+j)(b+j)x}{(j+1)(c+j)}, \quad b_j = 1 + \frac{(a+j)(b+j)x}{(j+1)(c+j)}, \quad j = 2, 3, \ldots. $$

Algorithm 2 is a modification of Lentz algorithm for the calculation the logarithm of a positive-valued continued fraction. It is an adaptation of Algorithm 1 that, essentially, replaces products by sums of logarithms. It may also require safeguarding against non-convergence.

Algorithm 2 Lentz's Algorithm for approximating the logarithm of (1) when $b_0 > 0$.

Inputs (with defaults): $b_0 > 0$, $a_j, b_j, j \geqslant 1$, ε_1 (10^{-30}), ε_2 (10^{-7}).

$\text{logf}_{\text{prev}} \leftarrow \log(b_0)$; $C_{\text{prev}} \leftarrow b_0$; $D_{\text{prev}} \leftarrow 0$; $\Delta = 2 + \varepsilon_2$; $\Delta_{\text{prod}} \leftarrow 1$; $j \leftarrow 0$

cycle while $\Delta_{\text{prod}} > 0$ and $|\Delta - 1| \geqslant \varepsilon_2$

$j \leftarrow j + 1$; $D_{\text{curr}} \leftarrow b_j + a_j D_{\text{prev}}$; if $D_{\text{curr}} = 0$ then $D_{\text{curr}} \leftarrow \varepsilon_1$

$D_{\text{curr}} \leftarrow 1/D_{\text{curr}}$; $C_{\text{curr}} \leftarrow b_j + a_j/C_{\text{prev}}$; if $C_{\text{curr}} = 0$ then $C_{\text{curr}} \leftarrow \varepsilon_1$

$\Delta \leftarrow C_{\text{curr}} D_{\text{curr}}$; $\Delta_{\text{prod}} \leftarrow \Delta \Delta_{\text{prod}}$

if $\Delta_{\text{prod}} > 0$ then

$\text{logf}_{\text{curr}} \leftarrow \text{logf}_{\text{prev}} + \log(\Delta_{\text{prod}})$; $\Delta_{\text{prod}} = 1$

$\text{logf}_{\text{prev}} \leftarrow \text{logf}_{\text{curr}}$; $C_{\text{prev}} \leftarrow C_{\text{curr}}$; $D_{\text{prev}} \leftarrow D_{\text{curr}}$

return $\text{logf}_{\text{curr}}$

4.3. Application

We now apply the Liang et al. (2008) technology, with Lentz's Algorithm evaluation of ${}_2F_1$ ratios (with ε_1 and ε_2 set to their default values), to the `VietNamI` data set from the R package `Ecdat` (Croissant, 2011). The data set details the medical expenses of a cross-section of 27,765 individuals from 1997 in Vietnam as described in Cameron & Trivedi (2005). The aim is to predict the log of total medical expenditure based on the remaining 10 variables (the categorical variable `commune` was excluded from the analysis). The results of the analysis are summarized in Table II.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

The ISI's Journal for the Rapid Dissemination of Statistics Research

Table II. Posterior means of the indicator variables γ_j and the coefficients β_j for the Liang et al. (2008) Bayesian variable selection procedure as described in Section 4.1. applied to the `VietNamI` data set from the R package `Ecdat` (Croissant, 2011).

predictor	`pharvis`	`age`	`sex`	`married`	`educ`
$E(\gamma_j\|\mathbf{Y})$	0.998	1.000	0.058	1.000	1.000
$E(\beta_j\|\mathbf{Y})$	0.013	0.064	0.000	−0.087	0.075
predictor	`illness`	`injury`	`illdays`	`actdays`	`insurance`
$E(\gamma_j\|\mathbf{Y})$	1.000	0.054	0.952	0.131	1.000
$E(\beta_j\|\mathbf{Y})$	−0.062	0.000	−0.003	−0.001	0.147

Due to the large number of observations ($n = 27,765$) a näive implementation via the function `hyperg_2F1()` from the R package `gsl` (Hankin, 2007) results in divide-by-zero or overflow problems when calculating (4).

5 Conclusion

Ratios of special functions for which the numerators and denominators can be very large or very small, but for which the ratio may be moderate, are common in Bayesian inference. As is well-known, the normalizing factor of the Beta distribution involves ratios of Gamma functions and working with the log-Gamma function is generally recommended. However, the Gamma function appears to be a rare example of a special function for which software for its logarithm is readily available. As we have demonstrated, continued fractions expansions of special function ratios, combined with Lentz's Algorithm, can often resolve the problem. As shown in Section 4, the logarithm version of Lentz's Algorithm (Algorithm 2) can also be called upon for stable ratio computation.

Acknowledgement

This research was partially supported by Australian Research Council Discovery Project DP110100061.

References

Armagan, A, Dunson, DB & Clyde, M. (2011). 'Generalized beta mixtures of Gaussians', In *Advances in Neural Information Processing Systems 24*, Shawe-Taylor, J, Zemel, RS, Bartlett, P, Pereira, F & Weinberger, KQ (eds.), pp. 523–531.

Bishop, CM (2006), *Pattern Recognition and Machine Learning*, *Springer*, New York.

Cameron, AC & Trivedi, PK (2005), *Microeconometrics: Methods and Applications*, *Cambridge University Press*, New York.

Carvalho, CM, Polson, NG & Scott, JG (2010), 'The horseshoe estimator for sparse signals', *Biometrika*, **97**, 465–480.

Consonni, G & Marin, J-M (2007), 'Mean-field variational approximate Bayesian inference for latent variable models', *Computational Statistics and Data Analysis*, **52**, 790–798.

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.4

Croissant, Y (2011), Ecdat 0.1-6.1. Data sets for econometrics. R package, http://cran.r-project.org.

Cuyt, A, Petersen, VB, Verdonk, B, Waadeland, H & Jones, WB (2008), *Handbook of Continued Fractions for Special Functions*, *Springer*, New York.

Girolami, M & Rogers, S (2006), 'Variational Bayesian multinomial probit regression', *Neural Computation*, **18**, 1790–1817.

Gosper, RW (1972), Item 101a in Beeler, M., Gosper, R.W. and Schroeppel, R. HAKMEM, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Memo AIM–239. pp. 37–39.

Gradshteyn, IS & Ryzhik, IM (1994), *Tables of Integrals, Series, and Products*, 5th Edition, *Academic Press*, San Diego, California.

Griffin, JE & Brown, PJ (2011), 'Bayesian hyper-lassos with non-convex penalization', *Australian and New Zealand Journal of Statistics*, **53**, 423–442.

Hankin, RKS (2007), gsl 1.9. Wrapper for the Gnu Scientific Library. R package, http://cran.r-project.org.

Laplace, PS (1805), *Traîté de Mécanique Céleste (Volume 4)*, *Duprat*, Paris.

Lentz, WJ (1976), 'Generating Bessel functions in Mie scattering calculations using continued fractions', *Applied Optics*, **3**, 668–671.

Liang, F, Paulo, R, Molina, G, Clyde, MA & Berger, JO (2008), 'Mixtures of g priors for Bayesian variable selection', *Journal of the American Statistical Association*, **103**, 410–423.

Marin, J–M & Robert, CP (2007), *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*, *Springer*, New York.

Neville, SE, Ormerod, JT & Wand, MP (2012), Mean field variational Bayes for continuous sparse signal shrinkage: pitfalls and remedies. Submitted for publication.

Ormerod, JT & Wand, MP (2010), 'Explaining variational approximations', *The American Statistician*, **64**, 140–153.

Press, W, Teukolosky, S, Vetterling, W & Flannery, B (1992), *Numerical Recipes: The Art of Scientific Computing*, 2nd Edition, *Cambridge University Press*, New York.

R Development Core Team (2012), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Trott, M (2012), The Wolfram Functions Site - A Wolfram Web Resource. http://functions.wolfram.com/HypergeometricFunctions/Hypergeometric2F1/.