Commun Nonlinear Sci Numer Simulat 53 (2017) 111–129

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

journal homepage: www.elsevier.com/locate/cnsns

Review

Exact posterior computation in non-conjugate Gaussian location-scale parameters models

J.A.A. Andrade*, P.N. Rathie

Department of Statistics and Applied Mathematics, Federal University of Ceara, 60455-670, Fortaleza-Ce, Brazil

ARTICLE INFO

Article history:
Received 1 November 2016
Revised 12 April 2017
Accepted 30 April 2017
Available online 4 May 2017

Keywords:
Bayesian computation
Exact posterior distribution
Non-conjugate models
Special functions
H-function

ABSTRACT

In Bayesian analysis the class of conjugate models allows to obtain exact posterior distributions, however this class quite restrictive in the sense that it involves only a few distributions. In fact, most of the practical applications involves non-conjugate models, thus approximate methods, such as the MCMC algorithms, are required. Although these methods can deal with quite complex structures, some practical problems can make their applications quite time demanding, for example, when we use heavy-tailed distributions, convergence may be difficult, also the Metropolis-Hastings algorithm can become very slow, in addition to the extra work inevitably required on choosing efficient candidate generator distributions. In this work, we draw attention to the special functions as a tools for Bayesian computation, we propose an alternative method for obtaining the posterior distribution in Gaussian non-conjugate models in an exact form. We use complex integration methods based on the H-function in order to obtain the posterior distribution and some of its posterior quantities in an explicit computable form. Two examples are provided in order to illustrate the theory.

1. Introduction

A general Bayesian model consists of a random sample distribution, denoted by $\boldsymbol{x} = (x_1, \ldots, x_k)|\boldsymbol{\theta} \overset{d}{\sim} f(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ is a vector of m unknown parameters; the vector of parameters $\boldsymbol{\theta}$ is distributed according to some prior distribution, denoted by $\boldsymbol{\theta} \overset{d}{\sim} p(\boldsymbol{\theta})$. The Bayesian framework bases all the inferences on the posterior distribution, which is a combination of the data and the prior distributions through the Bayes Theorem, that is given by

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_\Theta f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{1}$$

where Θ is the parametric space. The problem arises when the integral in the denominator is not analytically computable, this is usually the case when the parametric space has high dimension. As a consequence, approximated methods are required to compute the posterior quantities, such as means, variances, moments, etc.

In Bayesian analysis there are several methods for computing the posterior quantities, the most widely used are the Markov Chain Monte Carlo (MCMC) algorithms which simulates from posterior distribution in order to obtain a sufficiently large sample that will be used to compute the posterior estimates. The efficiency of these methods certainly represents

* Corresponding author.
E-mail addresses: ailton@ufc.br (J.A.A. Andrade), pushpanrathie@yahoo.com (P.N. Rathie).

http://dx.doi.org/10.1016/j.cnsns.2017.04.036
1007-5704/

a great advancement in Bayesian Statistics and most of the literature has mostly been centred in improving the numerical and the simulation methods. In fact, no alternative approaches have been investigated ever since. However, shown in [4], the MCMC methods has many issues which should, at least, motivate research in other directions, for instance: highly correlated parameters, complex posterior distributions with many parameters, heavy-tailed distributions, etc., all these problems can lead to convergence problems, which will demand some hard computational and mathematical expertise to overcome.

Turning back to the beginning of practical Bayesian analysis, the class of conjugate models were the most natural and elegant way to make posterior inferences. We say that a model is conjugate if the posterior distribution belongs to the same family as the prior distribution. Thus, for a suitably chosen prior distribution, we obtain all the posterior quantities in an exact form. Undoubtedly, obtaining a posterior distribution in an exact form represents a great advantage in the inferential process. Nevertheless, the use of such class of models limits the use of the Bayesian Statistics in practical problems, since the class of conjugate models embraces only a few models. On the other hand, non-conjugate models allow to assign any combination of data and prior distributions, but they will not yield an exact posterior distribution, whence approximated methods are required. In summary, we have the (mathematically nice) class of conjugate models, although of limited use, and the non-conjugate models which allow a greater range of applications, but fails to deliver an elegant analytical Mathematics.

With the objective of bringing together the analytical mathematics and the flexibility of the non-conjugate models, Andrade and Rathie [1] proposed a new approach based on the *special functions*, which allowed to obtain exact posterior distributions in a class of single parameter models. In this way the posterior distribution and its quantities are explicitly expressed in a computable form. Gordy [5], although defined a new class of generalised Beta distributions, also points in this direction; the new class is expressed in terms of the confluent hypergeometric function which leads to a natural conjugacy in Bayesian modelling. However his approach is specifically limited to the Gamma model with generalised Beta prior distribution. The theory of special functions is widely applied in many sciences, such as Physics, Engineering and Statistics. In Statistics, applications arise naturally in the *algebra of random variables*, in which we need to assess complex integrals in order to obtain the distribution of the statistics of interest. For some applications, Carter & Springer [3], Springer [11], Provost et al. [8] and Provost & Mabrouk [9].

In this work, we consider a hierarchical Gaussian location-scale parameter model with independent prior distributions belonging to the H-function class. The H-function embraces almost all the probability distributions known in Statistics, hence the theory establishes the exact posterior distribution for a huge variety of non-conjugate models. The method proposes an alternative to the existing thinking that non-conjugate models cannot be obtained in an exact form and hence approximate methods are necessary to make posterior inferences. The theory establishes that, under some regularity conditions, we can write the posterior distribution, the moments and the predictive posterior distribution and its moments, explicitly in a computable form.

In Section 2, we provide some definitions and basic properties of the univariate and bivariate H-functions. In Section 3, we present the results about the exact posterior distribution in non-conjugate Gaussian models under location-scale structures. In fact, only the data distribution is fixed, whereas the prior distributions can assume any form within the H-function family, which is a quite wide class of distributions. In Section 4 we give a couple of examples involving non-conjugate structures, namely the well known conjugate model Normal–Student t–Generalised Gamma and the Normal–Normal–Gamma model. In Section 5, we make some general comments about the theory and future work.

2. The H-function

The H-function is defined in terms of a complex integral, which can be evaluated by the theorem of residues. In this section, we provide the definition and some basic properties, which will be useful to our theory. For a more complete review about H-function, see Srivastava et al. [12], Mathai et al. [7] and Braaksma [2]. Besides, under the context of Bayesian computation, Andrade & Rathie [1] provide the properties and results more relevant to our theory.

Definition 1 (H-function)**.** The H-function is defined by

$$H[z] = H_{p\ q}^{m\ n}\left[z \left|\begin{matrix} {}_1\left(a_j, A_j\right)_p \\ {}_1\left(b_j, B_j\right)_q \end{matrix}\right.\right] = \frac{1}{2\pi i}\int_L \phi(s) z^{-s} ds, \tag{2}$$

where

$$\phi(s) = \frac{\prod_{j=1}^{m}\Gamma(b_j + B_j s)\prod_{j=1}^{n}\Gamma(1 - a_j - A_j s)}{\prod_{j=m+1}^{q}\Gamma(1 - b_j - B_j s)\prod_{j=n+1}^{p}\Gamma(a_j + A_j s)}, \tag{3}$$

$i = \sqrt{-1}$, $z \in \mathbb{C}_0$, $m, n, p, q \in \mathbb{N}_0$ ($0 \le n \le p$ and $0 \le m \le q$, both m and n are not zeros simultaneously), $z, a_j, b_j, \in \mathbb{C}$, A_j, $B_j > 0$ $\forall j$, $z^{-s} = \exp\{-s \ln|z| + i \arg z\}$, $\ln|z|$ represents the natural logarithm of $|z|$ and $\arg z = \text{Arg}(z) + 2\pi\ell$ ($\ell = 0, \pm1, \pm2, \ldots$ and Arg(z) is the principal value of z). L is the Mellin–Barnes contour which separates the poles of $\Gamma(b_j + \beta_j s)$ ($j = 1, \ldots, m$) from the poles of $\Gamma(1 - a_j - A_j s)$ ($j = 1, \ldots, n$).