Note on Posterior Inference for the Bingham Distribution

Mike G. Tsionas*

6th January 2017

Abstract

The properties of high-dimensional Bingham distributions have been studied by Kume and Walker (2014). Fallaize and Kypraios (2016) propose Bayesian inference for the Bingham distribution and they use developments in Bayesian computation for distributions with doubly intractable normalising constants (Møller et al. 2006; Murray et al. 2006). However, they rely heavily on two Metropolis updates that they need to tune. In this paper we propose instead model selection with the marginal likelihood.

Key words: Bingham distribution; Bayesian; Markov Chain Monte Carlo; marginal likelihood.

*Lancaster University Management School, LA1 4YK, U.K. & Athens University of Economics and Business, Greece, m.tsionas@lancaster.ac.uk

1 Introduction

The properties of high-dimensional Bingham distributions have been studied by Kume and Walker (2014). Bee, Benedetti and Espa (2017) have considered approximate maximum likelihood estimation of the Bingham distribution. Fallaize and Kypraios (2016) propose Bayesian inference for the Bingham distribution and they use developments in Bayesian computation for distributions with doubly intractable normalising constants (Møller et al. 2006; Murray et al. 2006). However, they rely heavily on two Metropolis updates that they need to tune. In this paper we rely on Walker (2014) but we avoid his reversible-jump MCMC via model selection with the marginal likelihood.

2 Methods

2.1 Fixed A

The Bingham distribution is

$$p(x; A) = \frac{1}{c(A)} \exp(-x'Ax),\ x \in \Re^q,\ x'x = 1, \tag{1}$$

where $c(A)$ is an intractable integration constant and A is a $q \times q$ symmetric matrix. The standard form of the distribution is:

$$p(x; \Lambda) = \frac{1}{c(\Lambda)} \exp\left\{-\sum_{i=1}^{q-1} \lambda_i x_i^2\right\}, \tag{2}$$

with respect to a uniform measure on the sphere and

$$c(\Lambda) = \int_{S^{q-1}} \exp\left\{-\sum_{i=1}^{q-1} \lambda_i x_i^2\right\} dS^{q-1}(x), \tag{3}$$

where S^{q-1} is the unit sphere in $\Re^q$, and$\Lambda = diag(\lambda_1, ..., \lambda_q)$ under the identifiability constraint

$$\lambda_1 \geq ... \geq \lambda_{q-1} \geq \lambda_q = 0. \tag{4}$$

We need these constraints as the density does not change if we add a positive constant to the λ_is, see Bingham (1974).

If $A = V\Lambda V'$ where V is orthonormal, if X follows a Bingham distribution with density $p(x; A)$ then $Y = VX$ follows a Bingham distribution with density $p(x; \Lambda)$, see Kume and Walker (2006). The maximum likelihood estimator of V is the matrix of eigenvectors of $\sum_{t=1}^{T} x_t x_t'$.

The joint density from a random sample $X = \{x_t; t = 1, ..., T\}$ gives rise to the likelihood:

$$L(\Lambda; X) = c(\Lambda)^{-T} \exp\left\{-\sum_{i=1}^{q-1} \lambda_i Q_i\right\}, \tag{5}$$

where $Q_i = \sum_{t=1}^{T} x_{ti}^2$. Given a prior, say $p(\Lambda)$ we have the posterior:

$$p(\Lambda|X) \propto c(\Lambda)^{-T} \exp\left\{-\sum_{i=1}^{q-1} \lambda_i Q_i\right\} p(\Lambda). \tag{6}$$

It is possible to introduce an auxiliary variable v so that

$$p(\Lambda, v|X) \propto v^{T-1} \exp\{-c(\Lambda)v\} \exp\left\{-\sum_{i=1}^{q-1} \lambda_i Q_i\right\} p(\Lambda). \tag{7}$$

However, since $c(\Lambda)$ is intractable this is not much help. Following Walker (2011) we can introduce variables $(k, s_1, ..., s_k)$ to have:

$$p(\Lambda, v, k, s^{(k)}|X) \propto \frac{v^{T+k-1}\exp(-v)}{k!} \exp\left\{-\sum_{i=1}^{q-1}\lambda_i Q_i\right\} \prod_{j=1}^{k}\left[1-\tilde{p}(s_j;\Lambda)\right] p(\Lambda), \tag{8}$$

where $\tilde{p}(s;\Lambda) = \exp\left\{-\sum_{j=1}^{q-1} s_j^2 \lambda_j\right\}$, $s^{(k)} \equiv (s_1, ..., s_k) \in (0,1)$, for all k. Further latent variables $u_1, ..., u_k$ can be introduced and they contribute a term:

$$\prod_{j=1}^{k} \mathbf{1}\left(u_j < 1 - \tilde{p}(s_j;\Lambda)\right).$$

The conditional posterior of v is $gamma(T+k, 1)$. The conditional posterior for u_j is uniform in $(0,\ 1-\tilde{p}(s_j;\Lambda))$. The conditional posterior for s_j is uniform in $\{s \in (0,1):\ \tilde{p}(s;\Lambda) < 1 - u_j\}$. If we integrate out v we get:

$$p(\Lambda, k, s^{(k)}|X) \propto \left(\begin{array}{c} T+k-1 \\ k \end{array}\right) \prod_{j=1}^{k}\left[1-\tilde{p}(s_j;\Lambda)\right] \exp\left\{-\sum_{i=1}^{q-1}\lambda_i Q_i\right\} p(\Lambda). \tag{9}$$

The conditional posterior for Λ is:

$$p(\Lambda|k, s^{(k)}, X) \propto \exp\left\{-\sum_{i=1}^{q-1}\lambda_i Q_i\right\} p(\Lambda)\mathbf{1}(\theta \in A), \tag{10}$$

where $A = \{\theta:\ p(s_j;\Lambda) < 1 - u_j,\ \forall j = 1, ..., k\}$.

Walker (2011) proposed an infinite mixture for $\left(k, s^{(k)}, u\right)$ along the lines of Godsill (2001). Here, we propose to consider values of $k = 1, ..., \overline{k}$ and choose the one that maximizes the marginal likelihood which, for fixed k, can be obtained from (9) using the method of Perrakis, Ntzoufras and Tsionas (2014). Walker (2014) used the ideas above but resorted to reversible jump MCMC (Green, 1995), a fact that was shown that can be bypassed in Fallaize and Kypraios (2016). The method we propose here, avoids both the use of reversible jump MCMC as in Walker (2013) and also the Metropolis updates in Fallaize and Kypraios (2016).

2.2 Posterior inference for A

The conditional posterior of A, if we assign normal priors $N(0, \underline{h}^2)$ to its elements is given by:

$$p(A|.) \propto \frac{\exp\left\{-\sum_{t=1}^{T} y_t' A y_t - \frac{1}{2\underline{h}^2} a' a\right\}}{c(A)^n} = \frac{\exp\left\{-\mathrm{det} AQ - \frac{1}{2\underline{h}^2} a' a\right\}}{c(A)^n}, \tag{11}$$

where a is the vector of distinct elements of A and $Q = \sum_{t=1}^{T} y_t y_t'$. We can use the same "trick" as in (7) and introduce an auxiliary variable v so that:

$$p(A|\cdot) \propto v^{T-1} \exp\left\{-AQ - \frac{1}{2\underline{h}^2} a' a - c(A) v\right\}, \tag{12}$$

but since $c(A)$ is unknown this is not very helpful. However, we can proceed by introducing additional latent variables K and $S^{(k)}$, as we did in (8). We avoid, again, the infinite mixture construction and reversible-jump MCMC by using model selection for K via the marginal likelihood. It can be shown that each different element of A in the final conditional posterior distribution follows a normal distribution subject to constraints similar to those that we provided previously for Λ.

3 Applications

We apply our techniques to two artificial data sets as described in Fallaize and Kypraios (2016). For both sets we use 15,000 iterations the first 5,000 of which are discarded to mitigate start up effects.

DATA SET 1. We consider a sample of $n = 100$ unit vectors which result approximately in the pair of sufficient statistics $(\tau_1, \tau_2) = (0.30,\ 0.32)$ where $\tau_i = \sum_{t=1}^{n} y_{ti}^2$. We assign independent exponential prior distributions with mean 100 to the parameters λ_1 and λ_2. Mardia and Zemroch (1977) report maximum likelihood estimates of 0.588 and 0.421.

DATA SET 2. We consider an artificial dataset of 100 vectors which result approximately in the pair of sufficient statistics $(\tau_1, \tau_2) = (0.02,\ 0.40)$for which the maximum likelihood estimates are 25.31 and 0.762 as reported in Mardia and Zemroch (1977). We use the same priors as in data set 1.

Marginal posterior densities of $\lambda_1$and λ_2 and their autocorrelation functions from MCMC are reported in Figure 1 for data set 1 and in Figure 2 for data set 2. The marginal posteriors are visually very close to the ones reported in Fallaize and Kypraios (2016) but MCMC autocorrelations are substantially smaller.

In Figure 3 we present normalized marginal likelihoods for selection of k in the two data sets (we normalize the marginal likelijhood to 1 when $k = 1$). In both cases the marginal likelihood favours k=20.

DATA SET 3. To illustrate inference for a general matrix A we use the data of Bingham (1974) as in Fallaize and Kypraios (2016). The data consist of n= 150 measurements on the c-axis of calcite grains from theTaconic Mountains of New York state. We use $\underline{h}$=10. For the diagonal components of Λ, Fallaize and Kypraios (2016) obtain posterior median values $\lambda_1 = 3.631$ and $\lambda_2 = 1.963$. We obtain 3.60 and 1.942 respectively. For V, the orthonormal component matrix of A, our posterior means are:

$$\overline{V} = \begin{bmatrix} 0.1751 & -0.4423 & 0.8812 \\ 0.1375 & 0.8932 & 0.4251 \\ -0.9681 & 0.0463 & 0.2195 \end{bmatrix},$$

which in in broad agreement with the results in Fallaize and Kypraios (2016).

References

Bee, M., R. Benedetti and G. Espa (2017). Approximate maximum likelihood estimation of the Bingham distribution. Computational Statistics and Data Analysis 108, 84-96.

Bingham, C. (1974). An antipodally symmetric distribution on the sphere. Annals of Statistics 2, 1201–1225.

Fallaize, C.J., Kypraios, T. (2016). Exact Bayesian inference for the Bingham distribution. Statistics and Computing 26, 349–360.

Godsill, S. J. (2001). On the relationship between Markov chain Monte Carlo methods for model uncertainty. Journal of Computational and Graphical Statistic 10:230–248.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Kume, A., Walker, S.G. (2006). Sampling from compositional and directional distributions. Statistics 16, 261–265.

Kume, A., Walker, S.G. (2014). On the Bingham distribution with large dimension. Journal of Multivariate Analysis 124, 345–352.

Mardia,K.V., Zemroch, P.J.: Table of maximum likelihood estimates for the bingham distribution. Statist. Comput. Simul. 6, 29–34 (1977)

Møller, J., Pettitt, A.N., Reeves, R., Berthelsen, K.K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. Biometrika 93(2), 451–458.

Murray, I., Ghahramani, Z., MacKay, D.J.C. (2006). MCMC for doubly intractable distributions. In Proceedings of the 22nd annual conference on uncertainty in artificial intelligence (UAI-06), pages 359– 366. AUAI Press.

Figure 1: Data Set 1

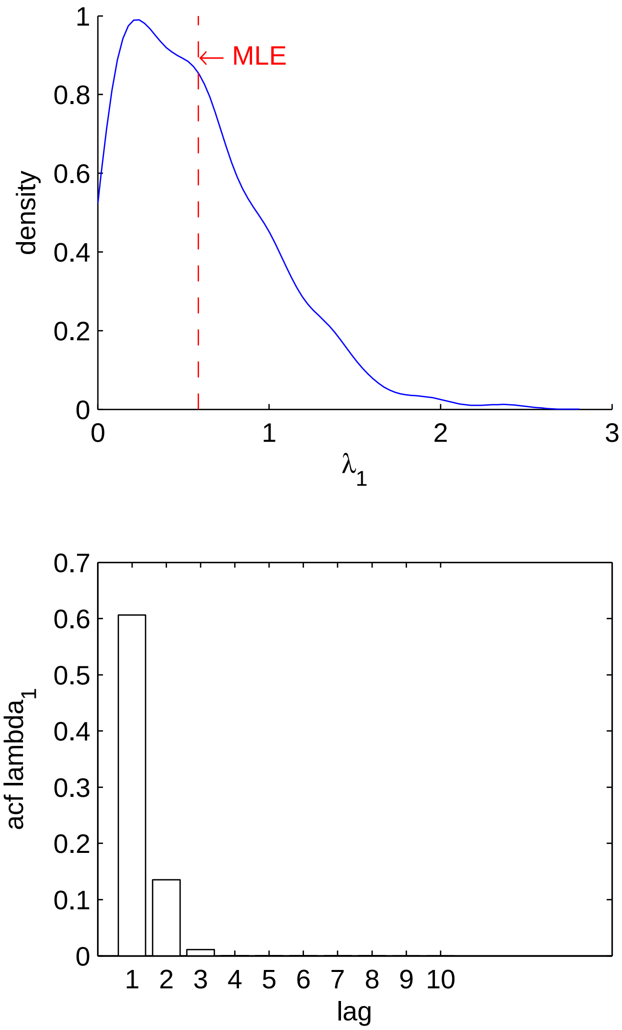


Figure 2: Data Set 2

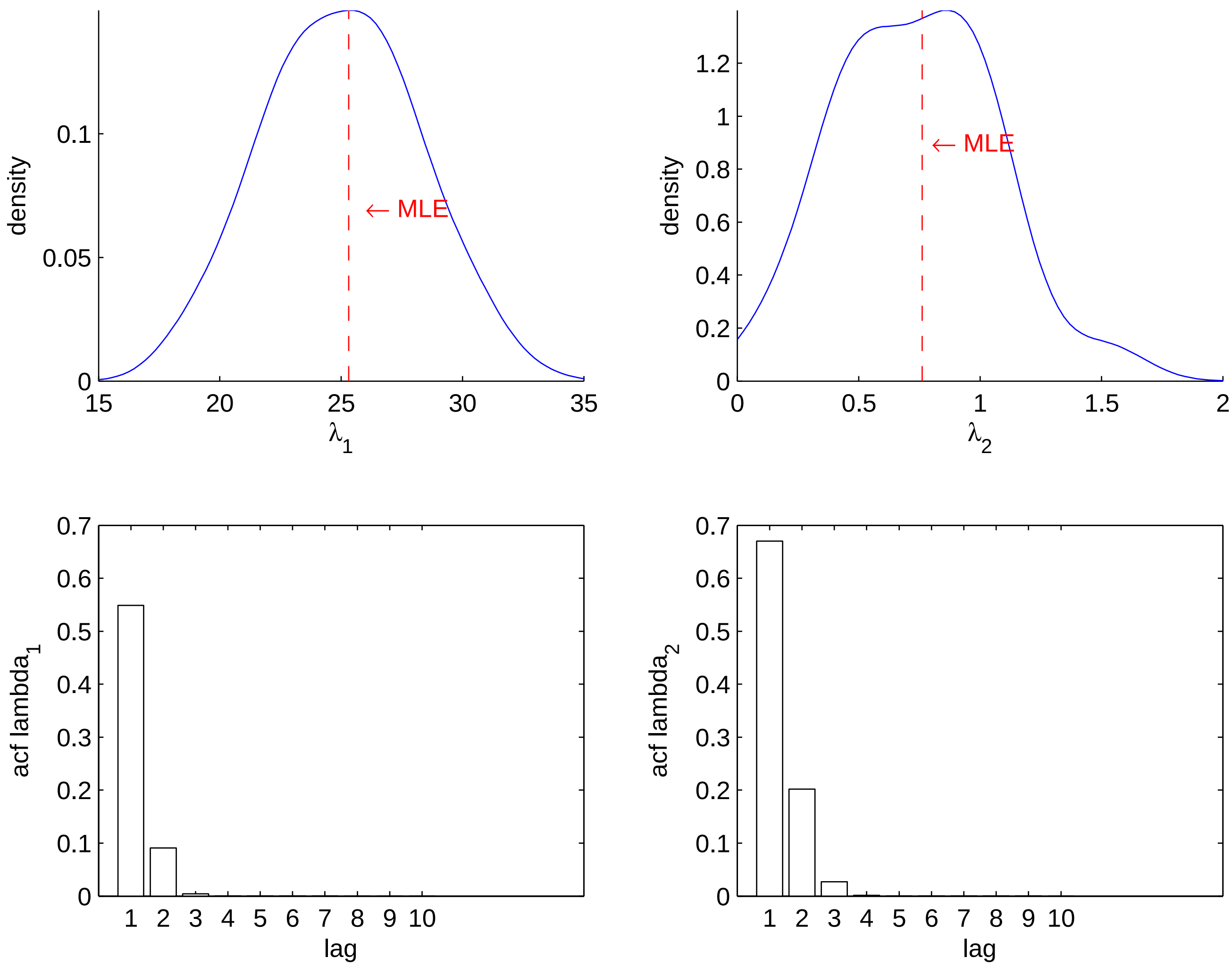
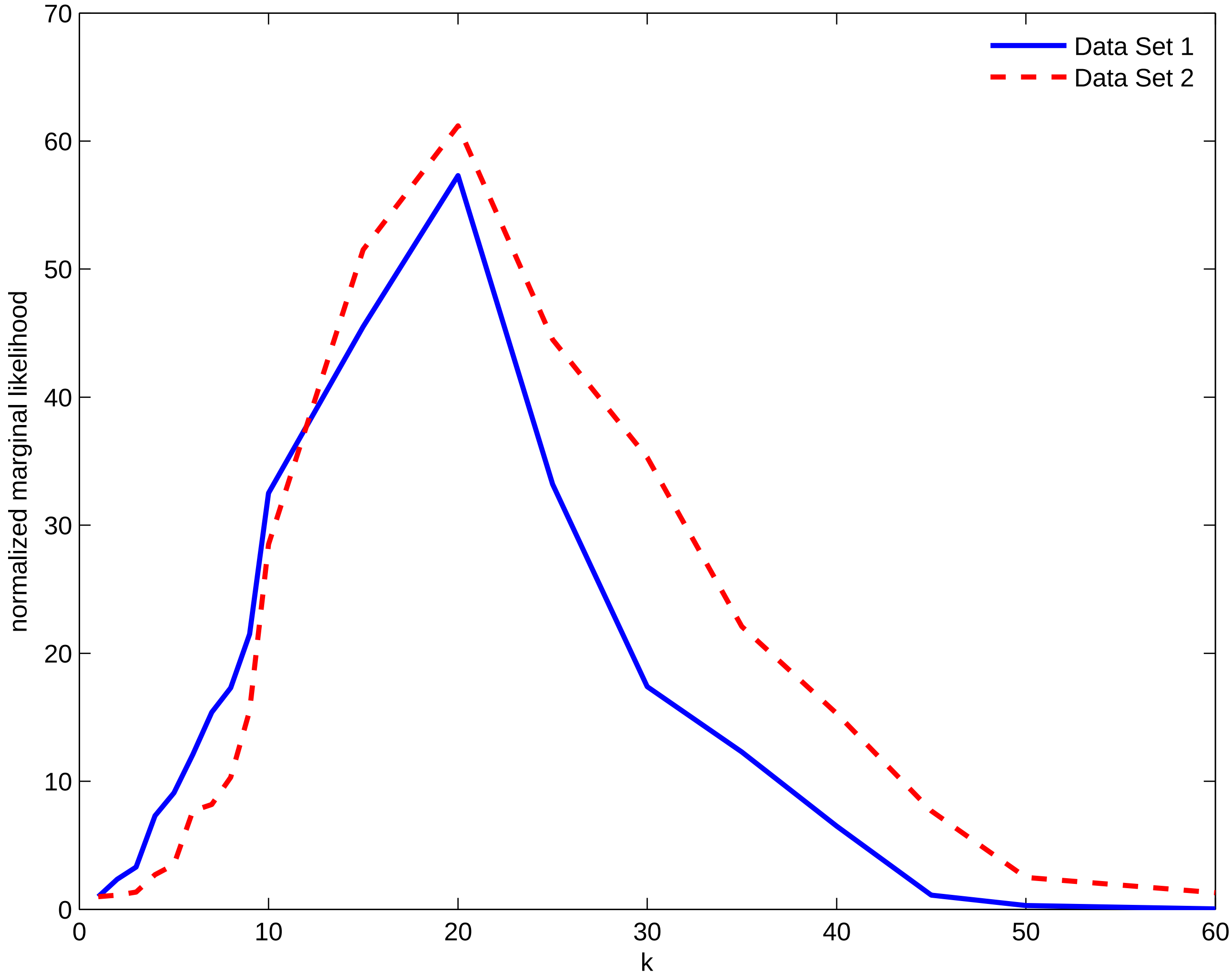


Figure 3: Normalized marginal likelihood to select k



Perrakis, K., Ntzoufras, I. and Tsionas, E. (2014). On the use of marginal posteriors in marginal likelihood estimation via importance-sampling. Computational Statistics and Data Analysis, 77, 54-69.

Walker, S.G. (2012). Posterior sampling when the normalizing constant is unknown. Communications in Statistics - Simulation and Computation, 40:5, 784-792.

Walker, S.G. (2014). Bayesian estimation of the Bingham distribution. Brazilian Journal of Probability and Statistics, 28(1), 61–72.