

JIRSS (2021)

Vol. 20, No. 01, pp 289-306

DOI:10.52547/jirss.20.1.289

## Matrix-Variate Beta Generator - Developments and Application

Janet van Niekerk <sup>1,2</sup>, Andriëtte Bekker <sup>1</sup>, and Mohammad Arashi <sup>3,1</sup>

<sup>1</sup> University of Pretoria, Department of Statistics, Pretoria, South Africa.

<sup>2</sup> King Abdullah University of Science and Technology, CEMSE Division, Kingdom of Saudi Arabia.

<sup>3</sup> Ferdowsi University of Mashhad, Department of Statistics, Mashhad, Iran.

Received: 24/11/2020, Revision received: 07/02/2021, Published online: 03/04/2021

**Abstract.** Matrix-variate beta distributions are applied in different fields of hypothesis testing, multivariate correlation analysis, zero regression, canonical correlation analysis and etc. A methodology is proposed to generate matrix-variate beta generator distributions by combining the matrix-variate beta kernel with an unknown function of the trace operator. Several statistical characteristics, extensions and developments are presented. Special members are then used in a univariate and multivariate Bayesian analysis setting. These models are fitted to simulated and real datasets, and their fitting and performance are compared to well-established competitors.

**Keywords.** Bayesian Analysis, Binomial, Eigenvalues, Gaussian Sample, Gibbs Sampling, Matrix-Variate Beta

**MSC:** 62H10, 62H15.

---

Janet van Niekerk (janet.vanniekerk@up.ac.za)

Corresponding Author: Andriëtte Bekker (Andriette.Bekker@up.ac.za)

Mohammad Arashi (arashi@um.ac.ir)

## 1 Introduction

The well-known matrix-variate beta distributions, that are used in different fields of hypothesis testing, multivariate correlation analysis, zero regression, have been extended by several authors. The matrix-variate beta type III distribution has been defined, and some of its properties have been studied by Gupta and Nagar (2000), Gupta and Nagar (2009). Nagar et al. (2013) generalized the well-known matrix-variate beta type I distribution, by using an extended matrix-variate beta function. Gupta and Nagar (2006) extended the work of Nadarajah and Kotz (2006) by defining a matrix-variate hypergeometric beta distribution, and a matrix-variate Kummer-beta distribution was defined by Nagar and Gupta (2002). Ehlers (2011) proposed the matrix-variate beta type V distribution, motivated from generalized hypothesis testing in a multivariate setup (see also Bekker et al. (2012), Zine (2012)). Pham-Gia et al. (2020) established expressions for distributions of integral powers of matrix beta variates. To address the need for complex models to solve modeling challenges in real data, Tounsi (2019) defined different versions of extended matrix-variate beta distributions. In terms of bimatrix-variate distributions, Nagar et al. (2017) developed the bimatrix-variate gamma-beta distribution with a matrix-variate beta marginal distribution. These generalizations justify the study of new models within the matrix-variate beta ensemble.

In this paper the matrix-variate beta type I distribution with density function,

$$f(\mathbf{X}) \propto \det(\mathbf{X})^{a-\frac{1}{2}(m+1)} \det(\mathbf{I} - \mathbf{X})^{b-\frac{1}{2}(m+1)}, \quad \mathbf{X} \in I_m;$$

the matrix-variate beta type II distribution with density function,

$$f(\mathbf{X}) \propto \det(\mathbf{X})^{a-\frac{1}{2}(m+1)} \det(\mathbf{I} + \mathbf{X})^{-(a+b)}, \quad \mathbf{X} \in S_m;$$

and matrix-variate beta type III distribution with density function,

$$f(\mathbf{X}) \propto \det(\mathbf{X})^{a-\frac{1}{2}(m+1)} \det(\mathbf{I} - \mathbf{X})^{b-\frac{1}{2}(m+1)} \det(\mathbf{I} + \mathbf{X})^{-(a+b)}, \quad \mathbf{X} \in I_m,$$

form the platform for the development of the matrix-variate beta generator distribution in Section 2. The space  $I_m$  is defined as the space of all square matrices of order  $m$  such that for all  $\mathbf{Y} \in I_m$ , we have  $\mathbf{Y} - \mathbf{X} \in S_m$ , if and only if  $\mathbf{X} \in S_m$ , where  $S_m$  is the space of all positive definite matrices of order  $m$ . Some characteristics of this new generator distribution are derived and visually illustrated in Section 3.

The hypergeometric Wishart generator distribution was studied by Bekker et al. (2017), who emphasized the contribution of this development within the Bayesian

paradigm. Therefore, the footprint of the advances in the matrix-variate beta generator ensemble, in the Bayesian field is illustrated in Section 4.

## 2 Matrix-Variate Beta Generator Distributions

**Definition 2.1.** The random symmetric matrix  $X \in I_m$  is said to have the matrix-variate beta generator distribution with parameters  $a, b$  and  $\Phi$  and shape generator  $h(\cdot)$ , if it has the following density function,

$$f(X) = \zeta_{a,b} \det(X)^{a-\frac{1}{2}(m+1)} \det(I_m - X)^{b-\frac{1}{2}(m+1)} \det(I_m + X)^{-(a+b)} h(\text{tr}(\Phi X)), \quad (2.1)$$

denoted by  $X \sim MBG_m(a, b, h, \Phi)$ , for  $\text{Re}(a) > (m - 1)/2$  and  $\text{Re}(b) > (m - 1)/2$ , where  $\Phi$  is a symmetric complex matrix,  $h(\cdot)$  is a Borel measurable function that admits a Taylor series expansion. Note that

$$(\zeta_{a,b})^{-1} = \sum_{\tau, \kappa, \phi} \frac{(-1)^k (a + b)_\kappa h^{(t)}(0)}{t! k!} \frac{\Gamma_m(b) \Gamma_m(a, \phi)}{\Gamma_m(a + b, \phi)} \left( \frac{\theta^{\kappa, \tau}}{\phi} \right)^2 \frac{C_\phi(I_m)}{C_\tau(I_m)} C_\tau(\Phi), \quad (2.2)$$

where  $C_\tau(\cdot)$  is the zonal polynomial of the ordered partition  $\tau$  of  $t$  (Muirhead, 2005).

*Remark 1.* To find the normalizing constant in Definition 2.1, apply Theorem 7.2.10 of Muirhead (2005), Eqs (2.2) and (2.8) of Davis (1979), Eq. (3.28) of Chikuse (1980) and a Taylor series expansion of  $h(\text{tr}[\cdot])$ . Hence,

$$h(\text{tr} \Phi X) = \sum_{t=0}^{\infty} \frac{h^{(t)}(0)}{t!} \text{tr}(\Phi X)^t = \sum_{t=0}^{\infty} \frac{h^{(t)}(0)}{t!} \sum_{\tau} C_\tau(\Phi X). \quad (2.3)$$

*Remark 2.* Similarly beta generator distributions based on the matrix-variate beta type I and type II can be formulated:

- (i) Matrix-variate beta type I generator distribution ( $MBG_m^{(1)}(a, b, h, \Phi)$ ) with density function,

$$f(X) \propto \det(X)^{a-\frac{1}{2}(m+1)} \det(I_m - X)^{b-\frac{1}{2}(m+1)} h(\text{tr}(\Phi X)), \quad (2.4)$$

- (ii) Matrix-variate beta type II generator distribution ( $MBG_m^{(2)}(a, b, h, \Phi)$ ) with density function,

$$f(X) \propto \det(X)^{a-\frac{1}{2}(m+1)} \det(I_m + X)^{-(a+b)} h(\text{tr}(\Phi X)). \quad (2.5)$$

### 3 Characteristics

In this section, we provide some important statistical properties for the matrix-variate beta generator distribution (2.1). The results for type I and type II follows trivially.

**Theorem 3.1.** *Let  $X \sim MBG_m(a, b, h, \Phi)$ . Then it follows that*

$$E(|X|^r) = \frac{\zeta_{a,b}}{\zeta_{a+r,b}}.$$

*Proof.* From Definition 2.1, the result is easily obtained. □

**Theorem 3.2.** *Denote the moment generating function (MGF) of  $X \sim MBG_m(a, b, h, \Phi)$ , by  $\mathcal{M}$ . Then we have*

$$\mathcal{M}(T) = \zeta_{a,b} \sum_{\tau, \kappa, \lambda, \phi} \frac{h^{(t)}(0)}{t!k!} \frac{(-1)^l(a+b)_\lambda}{l!} \theta_\phi^{\kappa, \tau, \lambda} \frac{\Gamma_m(a, \phi)\Gamma_m(b)}{\Gamma_m(a+b, \phi)} C_\phi^{\kappa, \tau, \lambda}(T, \Phi, I),$$

where  $\sum_{\tau, \kappa, \lambda, \phi} \equiv \sum_{t=0}^\infty \sum_\tau \sum_{k=0}^\infty \sum_\kappa \sum_{l=0}^\infty \sum_\lambda \sum_{\phi \in \kappa \cdot \tau \cdot \lambda}$  and  $\theta_\phi^{\kappa, \tau, \lambda} = C_\phi^{\kappa, \tau, \lambda}(\mathbf{I}_m, \mathbf{I}_m, \mathbf{I}_m) / C_\phi(\mathbf{I}_m)$ .

*Proof.* For the MGF of MBG, using (2.3) and Eq. (3.10) of Chikuse (1980), follows that

$$\begin{aligned} \mathcal{M}(T) &= \zeta_{a,b} \int_{I_m} \det(X)^{a-\frac{1}{2}(m+1)} \det(I_m - X)^{b-\frac{1}{2}(m+1)} \det(I_m + X)^{-(a+b)} \\ &\quad h(\text{tr}(\Phi X)) \text{etr}(TX) dX \\ &= \zeta_{a,b} \sum_{t=0}^\infty \sum_\tau \sum_{k=0}^\infty \sum_\kappa \frac{h^{(t)}(0)}{t!k!} \sum_{l=0}^\infty \sum_\lambda \frac{(-1)^l(a+b)_\lambda}{l!} \\ &\quad \int_{I_m} \det(X)^{a-\frac{1}{2}(m+1)} \det(I_m - X)^{b-\frac{1}{2}(m+1)} C_\tau(\Phi X) C_\kappa(TX) C_\lambda(X) dX \\ &= \zeta_{a,b} \sum_{t=0}^\infty \sum_\tau \sum_{k=0}^\infty \sum_\kappa \frac{h^{(t)}(0)}{t!k!} \sum_{l=0}^\infty \sum_\lambda \frac{(-1)^l(a+b)_\lambda}{l!} \sum_{\phi \in \kappa \cdot \tau \cdot \lambda} \theta_\phi^{\kappa, \tau, \lambda} \\ &\quad \int_{I_m} \det(X)^{a-\frac{1}{2}(m+1)} \det(I_m - X)^{b-\frac{1}{2}(m+1)} C_\phi^{\kappa, \tau, \lambda}(TX, \Phi X, X) dX. \end{aligned}$$

Finally applying Eq. (3.28) of Chikuse (1980), yields the result. □

As the final important characteristic, the joint distribution of the eigenvalues for the MBG distributions are given in the next theorem.

**Theorem 3.3.** Let  $g(\Lambda)$  denote the joint density function of the eigenvalues of  $\mathbf{X}$ ,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m), 1 > \lambda_1 > \dots > \lambda_m > 0$  where  $\mathbf{X} \sim \text{MBG}_m(a, b, \Phi, h)$ . Then it follows that

$$g(\Lambda) = \frac{\pi^{\frac{1}{2}m^2} \zeta_{a,b}}{\Gamma_m\left(\frac{1}{2}m\right)} \sum_{k=0}^{\infty} \sum_{\kappa} \frac{h^{(k)}(0)}{k!} \frac{C_{\kappa}(\Phi)}{C_{\kappa}(\mathbf{I}_m)} \times \prod_{i < j}^m (\lambda_i - \lambda_j) \prod_{i=1}^m \left( \lambda_i^{a-\frac{1}{2}(m+1)} (1 - \lambda_i)^{b-\frac{1}{2}(m+1)} (1 + \lambda_i)^{-(a+b)} \right) C_{\kappa}(\Lambda).$$

*Proof.* From Theorem 3.2.17 of Muirhead (2005), the density function of  $\Lambda$  is given by

$$g(\Lambda) = \frac{\pi^{\frac{1}{2}m^2}}{\Gamma_m\left(\frac{1}{2}m\right)} \prod_{i < j}^m (\lambda_i - \lambda_j) \zeta_{a,b} \det(\Lambda)^{a-\frac{1}{2}(m+1)} \det(\mathbf{I}_m - \Lambda)^{b-\frac{1}{2}(m+1)} \times \det(\mathbf{I} + \mathbf{X})^{-(a+b)} \int_{O(m)} h(\text{tr } \Phi \mathbf{H} \Lambda \mathbf{H}') d\mathbf{H}.$$

Using (2.3) and Eq. (36) of Muirhead (2005), the result follows by noting that  $\det(\Lambda)^{a-\frac{1}{2}(m+1)} = \prod_{i=1}^m \lambda_i^{a-\frac{1}{2}(m+1)}$ ,  $\det(\mathbf{I}_m - \Lambda)^{b-\frac{1}{2}(m+1)} = \prod_{i=1}^m (1 - \lambda_i)^{b-\frac{1}{2}(m+1)}$  and  $\det(\mathbf{I}_m + \Lambda)^{-(a+b)} = \prod_{i=1}^m (1 + \lambda_i)^{-(a+b)}$ . □

*Remark 3.* For  $\Phi = \mathbf{I}_m$ , the result in Theorem 3.3 simplifies to

$$g(\Lambda) = \frac{\pi^{\frac{1}{2}m^2} \zeta_{a,b}}{\Gamma_m\left(\frac{1}{2}m\right)} h(\text{tr}(\Lambda)) \prod_{i < j}^m (\lambda_i - \lambda_j) \prod_{i=1}^m \left( \lambda_i^{a-\frac{1}{2}(m+1)} (1 - \lambda_i)^{b-\frac{1}{2}(m+1)} (1 + \lambda_i)^{-(a+b)} \right).$$

Figure 1 illustrates the joint density function of the eigenvalues of  $\mathbf{X} \sim \text{MBG}_2(a, b, h, \mathbf{I}_2)$ , for different choices of  $h(\cdot)$ . The flexibility provided by these generator distributions are clear from Figure 1.

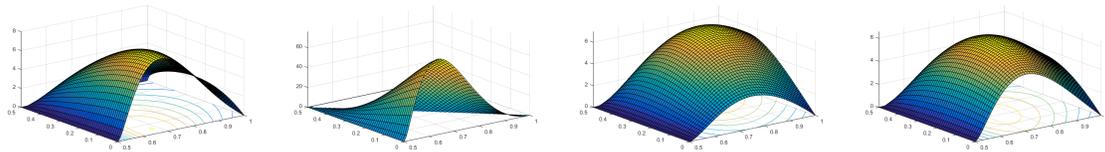


Figure 1: Joint density function of the largest ( $x$ -axis) and smallest ( $y$ -axis) eigenvalues for  $a = 3, b = 2$  and  $h(x) = \exp(-x), \ln(x), x^2, \sqrt{x}$  (left to right)

## 4 Bayesian Application

### 4.1 Univariate Application

In this section, special univariate cases ( $m = 1$ ) of (2.1) and (2.4) are applied as priors for the probability of a success in a binomial experiment. Consider a random sample,  $X_1, \dots, X_n$  of size  $n$  from a binomial experiment with  $s$  independent trials and unknown probability of success  $p$ , i.e.  $X_i \sim \text{Binomial}(s, p)$ ,  $i = 1, \dots, n$ , with likelihood function

$$L(p) \propto p^{\sum_{i=1}^n X_i} (1-p)^{ns - \sum_{i=1}^n X_i}. \quad (4.1)$$

Let  $h(x) = \exp(-x)$  in (2.4) and (2.1), producing the two priors for  $p$  as

$$\pi_1(p) \propto p^{a_1-1} (1-p)^{b_1-1} \exp(-\phi_1 p), \quad (4.2)$$

and

$$\pi_3(p) \propto p^{a-1} (1-p)^{b-1} (1+p)^{-a-b} \exp(-\phi p), \quad (4.3)$$

known as the Kummer-beta type I and III distributions and denoted as  $KBI(a_1, b_1, \phi_1)$  and  $KBIII(a, b, \phi)$ , respectively (Ng and Kotz, 1995). The posterior density functions are obtained from (4.1), (4.2) and (4.3), respectively as

$$q_1(p) \propto p^{a_1 + \sum_{i=1}^n X_i - 1} (1-p)^{b_1 + ns - \sum_{i=1}^n X_i - 1} \exp(-\phi_1 p), \quad (4.4)$$

and

$$q(p) \propto p^{a + \sum_{i=1}^n X_i - 1} (1-p)^{b + ns - \sum_{i=1}^n X_i - 1} (1+p)^{-a-b} \exp(-\phi p). \quad (4.5)$$

From (4.4) and (4.2), and (4.5) and (4.3) it is clear that both priors are conjugate. The Bayes estimators under the squared error loss can then be easily derived.

### 4.2 Multivariate Application

In this section we develop a Bayesian analysis framework for a  $m$ -variate Gaussian sample  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  of size  $n$ , i.e.  $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}_{m \times 1}, \boldsymbol{\Sigma}_{m \times m})$ .

The sample covariance matrix is the frequentist maximum likelihood estimator (MLE) of  $\boldsymbol{\Sigma}$ , hence  $\widehat{\boldsymbol{\Sigma}}_{\text{MLE}} = \mathbf{S}$ ,  $\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$ . Now we propose a Bayesian framework to estimate  $\boldsymbol{\Sigma}$ , using a special case of the proposed matrix variate beta type II generator (2.5).

Consider an improper objective prior for  $\boldsymbol{\mu}$ ,  $\pi(\boldsymbol{\mu}) = 1$ , and the  $MBG_m^{(2)}(a, b, \exp(\cdot), \boldsymbol{\Phi})$  distribution as the prior for  $\boldsymbol{\Sigma}$ . Then the joint posterior density function is

$$q(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{X}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}+a-\frac{m+1}{2}} |\mathbf{I}_m + \boldsymbol{\Sigma}|^{-a-b} \text{etr} \left[ -\frac{1}{2} \boldsymbol{\Sigma}^{-1} \left[ \mathbf{V} + n(\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})' \right] \right] \text{etr}(\boldsymbol{\Phi} \boldsymbol{\Sigma}), \quad (4.6)$$

with  $\mathbf{V} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$  and  $\bar{\mathbf{X}}$  is the sample mean.

### 4.2.1 Gibbs Sampling Algorithm

We use a Metropolis-Hastings algorithm inside our Gibbs sampling algorithm to sample from the posteriors of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ .

#### Algorithm

1. Initialize  $\boldsymbol{\mu}_0$  and  $\boldsymbol{\Sigma}_0$ .
2. Repeat the following steps for  $t = 1, \dots, 10000$  times:
  - (a) Generate  $\boldsymbol{\mu}_t \sim N_p(\bar{\mathbf{X}}, \frac{1}{n} \boldsymbol{\Sigma}_{t-1})$ .
  - (b) Calculate  $\mathbf{A}_t = \mathbf{V} + n(\bar{\mathbf{X}} - \boldsymbol{\mu}_t)(\bar{\mathbf{X}} - \boldsymbol{\mu}_t)'$ .
  - (c) Metropolis-Hastings algorithm:
    - i. **Proposals:** Generate the random matrices  $\boldsymbol{\Sigma}_1^* \sim W_m(\boldsymbol{\Phi}_1, \phi_1)$  (Wishart distribution) and  $\boldsymbol{\Sigma}_2^* \sim W_m^{-1}(\mathbf{I}_m + \boldsymbol{\Phi}_2, \phi_2)$  (inverted Wishart distribution) such that  $E[\boldsymbol{\Sigma}_1^*] = cE[\boldsymbol{\Sigma}_2^*]$  ( $E$  stands for expected value).
    - ii. Calculate  $\boldsymbol{\Sigma}^* = w\boldsymbol{\Sigma}_1^* + (1-w)\boldsymbol{\Sigma}_2^*$  for some  $0 < w < 1$ .
    - iii. If  $\min\left(\frac{f^*[\boldsymbol{\Sigma}^*|\boldsymbol{\mu}_t]}{f^*[\boldsymbol{\Sigma}_{t-1}|\boldsymbol{\mu}_t]}, 1\right) > u$  where  $u$  is a random uniform(0, 1) variate, then  $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}^*$  else  $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1}$ , with  $f^*[\boldsymbol{\Sigma}|\boldsymbol{\mu}_t] \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}+a-\frac{m+1}{2}} |\mathbf{I}_m + \boldsymbol{\Sigma}|^{-a-b} \text{etr} \left( -\frac{1}{2} \boldsymbol{\Sigma}^{-1} \left[ \mathbf{V} + n(\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})' \right] \right) \text{etr}(\boldsymbol{\Phi} \boldsymbol{\Sigma})$ .
3. **Burn-in:** Discard the first couple of observations, i.e. the posterior observations are  $\boldsymbol{\mu}_{1000}, \dots, \boldsymbol{\mu}_{100000}$  and  $\boldsymbol{\Sigma}_{1000}, \dots, \boldsymbol{\Sigma}_{100000}$ .

The matrix variate samples of  $\boldsymbol{\Sigma}$  should then be checked for convergence and we use the methods proposed in Bekker et al. (2017), namely the determinant, trace and largest eigenvalue of each simulation.

## 5 Numerical Studies

### 5.1 Univariate Application

In this section, priors (4.2) and (4.3) are applied and their performance is evaluated against the maximum likelihood estimator (MLE) and the Bayes estimator when assuming the well-known beta prior.

Therefore, the three priors considered in the subsequent Bayesian analysis are:

- Beta prior, i.e.  $p \sim \text{Beta}(a_0, b_0)$ , with  $f(p) \propto p^{a_0-1}(1-p)^{b_0-1}$ .
- Kummer beta type I prior, i.e.  $p \sim \text{KBI}(a_1, b_1, \phi_1)$  as in (4.2).
- Kummer beta type III prior, i.e.  $p \sim \text{KBIII}(a, b, \phi)$  as in (4.3).

The hyperparameters are chosen according to a prior belief  $p = p_0$ . Due to the conjugacy of the prior, the posterior density functions and Bayes estimates can be calculated analytically using (4.4) and (4.5), without the need for approximation algorithms like Gibbs sampling or INLA.

#### 5.1.1 Simulation Study

For the purpose of this study, a sample of size 10 was simulated from a binomial distribution with 10 trials and 0.3 probability of success, i.e.  $X \sim \text{Binomial}(s = 10, p = 0.3)$ . The aim is to estimate  $p$  as accurately as possible. To achieve this, we employ the Beta, KBI and KBIII priors and compare the resulting estimates with the MLE.

Three combinations of hyperparameter values are used as summarized in Table 1.

The coverage probability for each prior is calculated based on 10000 repetitions and given in Table 2, together with the median credible interval widths.

In Figure 2, the MSE and bias for the different estimates are given based on 50 repetitions with combination 3.

It is clear from Table 2 and Figure 2 that the Kummer beta type 1 and type 3 priors perform very well when compared to the beta prior and the MLE.

Table 1: Hyperparameter values for the beta (...), KBI (- - -) and KBIII (-.-) priors

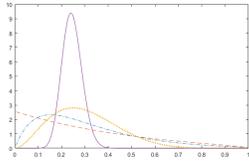
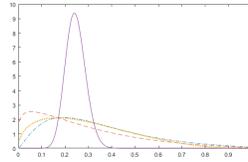
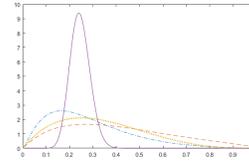
Combination	1	2	3
Beta prior	$a_0 = 3, b_0 = 7$	$a_0 = 1.5, b_0 = 3.5$	$a_0 = 2, b_0 = 4$
KBI prior	$a_1 = 1, b_1 = 1.5, \phi_1 = 1.5$	$a_1 = 1.2, b_1 = 1.5, \phi_1 = 3$	$a_1 = 2, b_1 = 1.5, \phi_1 = 3$
KBIII prior	$a = 2, b = 1.2, \phi = 3.5$	$a = 2.3, b = 1.2, \phi = 3.5$	$a = 2.3, b = 2, \phi = 3$
The likelihood function is given by the — curve			

Table 2: Coverage probabilities (median credible interval widths) for the three priors and combinations

Prior	Beta	KBI	KBIII
Combination 1	94.1(0.14)	94.6(0.145)	95.2(0.14)
Combination 2	90.3(0.14)	92.4(0.14)	92.2(0.145)
Combination 3	90.7(0.145)	90.8(0.145)	91.7(0.145)

### 5.1.2 Space Shuttle Dataset

After the explosion of the USA Space Shuttle Challenger on 28 January 1986, an investigation ensued into the reliability of the shuttle’s propulsion system. The explosion was eventually traced to the failure of one of the three field joints on one of the two solid booster rockets. Each of these six field joints includes two O-rings, designated as primary and secondary, which fail when phenomena called erosion and blowby both occur.

The night before the launch a decision had to be made regarding launch safety. The discussion among engineers and managers leading to this decision included concern that the probability of failure of the O-rings were too high. Based on the 17 previous launches under the same conditions, the failure probability should be estimated.

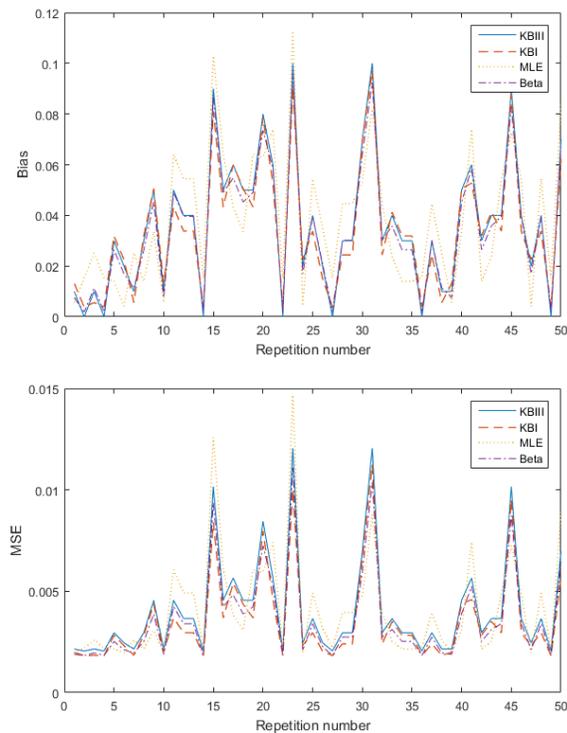


Figure 2: Bias (left) and MSE (right) of the Bayes estimates and the MLE for 50 repetitions

A Bayesian approach is now used to estimate the failure probability using the following three priors:

- Beta prior, i.e.  $p \sim \text{Beta}(0.3, 9)$ .
- Kummer beta type I prior, i.e.  $p \sim \text{KBI}(0.3, 5, 5)$  as in (4.2).
- Kummer beta type III prior, i.e.  $p \sim \text{KBIII}(0.2, 1.8, 5)$  as in (4.3).

The prior belief was set as  $p = 0.03$  from previous studies. The three priors under consideration and the likelihood are visually illustrated in Figure 3, together with the resulting posterior density functions.

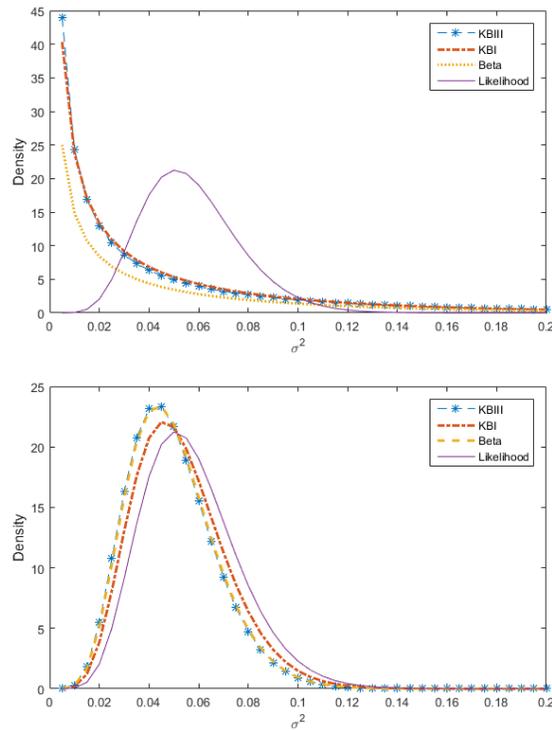


Figure 3: Likelihood, prior and posterior density functions

Figure 3 illustrates that the resulting posterior distributions are more concise than the priors, confirming the applicability of these priors to the data. The Bayes estimates and the MLE of the failure probability are given in Table 3.

Table 3: Bayes estimates and MLE of the failure probability

MLE	Beta	KBI	KBIII
0.0527	0.0507	0.0491	0.0496

To investigate the plausibility of these estimates, a sample of size 17 is simulated using the estimates of  $p$  in Table 3. The empirical cumulative distribution function

(ecdf) of each sample as well as the original dataset, is obtained and visually displayed in Figure 4. The ecdf which is closest to the ecdf of the original data indicates the most plausible estimate of  $p$ .

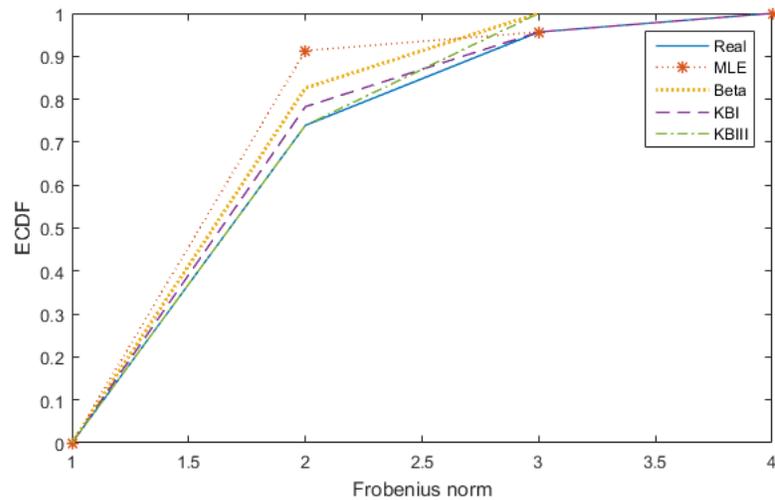


Figure 4: The ecdf's of the simulated samples based on the respective estimates and the original dataset

It is clear from Figure 4 that the KBI and KBIII priors produces more plausible estimates than the beta prior and the MLE.

## 5.2 Multivariate Application

The simulation algorithm 4.2.1 is quite complex and simulates  $p$ -variate vectors and square matrices at each index. This can be quite computationally intensive. For this study we used a Macbook Pro 2019 with a 2.3 GHz i5 processor and 16 GB LPDDR3 with dedicated graphics as well as a Linux server with 64 cores with hyperthreading. The computer code for the simulation study and data application is based on the algorithm explained in Section 4.2.1.

### 5.2.1 Simulation Study

For brevity, we only present the results of one scenario for  $\Sigma$ . Other hyperparameter values were considered as well.

We simulate a sample of size  $n = 30$  such that  $\mathbf{X} \sim N_3(\mathbf{0}_{3 \times 1}, \Sigma = \mathbf{I}_3)$ . We propose the prior (2.5) for  $\Sigma$  as follows,  $\Sigma \sim MBG_3^{(2)}(\frac{n}{2} + \frac{3}{2} + 2, 1, \exp(\cdot), \mathbf{I}_3)$ . Indeed, for rigorous hyperparameters' selection, an empirical or hierarchical Bayesian analysis can be used.

In Figure 5, these measures are illustrated for  $\{\Sigma_1, \dots, \Sigma_T\}, T = 100000$  simulated using Algorithm 4.2.1. It is clear from Figure 5, that the chain converged and the proposed algorithm provides stable samples.

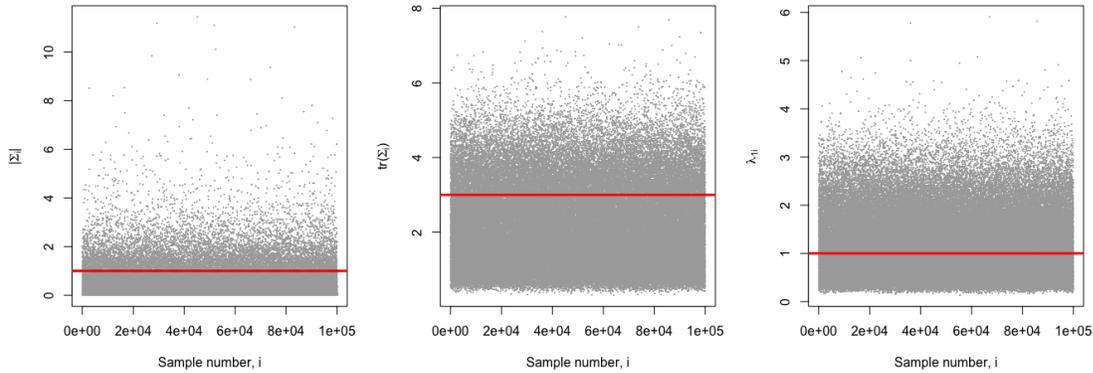


Figure 5: The proposed multivariate convergence measures for a sample of size 100000, determinant, trace and largest eigenvalue (from left to right) with the true value indicated by the solid line

As a method to assess the similarity to the true value  $\Sigma = \mathbf{I}_3$  we consider the Frobenius norm between the estimator and the true value  $FN_{MLE} = \|\widehat{\Sigma}_{MLE} - \mathbf{I}_3\|_F$  and  $FN_B = \|\widehat{\Sigma}_B - \mathbf{I}_3\|_F$ , with  $\widehat{\Sigma}_{MLE}$  and  $\widehat{\Sigma}_B$  are the ML and Bayes estimators of  $\Sigma$ , respectively. A smaller value of the Frobenius norm would then imply an estimate that is more similar to the true value. The estimated results for this setup are

$$\widehat{\Sigma}_{MLE} = \begin{bmatrix} 1.046 & -0.121 & 0.096 \\ -0.121 & 0.604 & 0.325 \\ 0.096 & 0.325 & 2.517 \end{bmatrix}, \quad FN_{MLE} = 1.938,$$

and

$$\widehat{\Sigma}_B = \begin{bmatrix} 1.264 & -0.016 & 0.007 \\ -0.016 & 1.229 & 0.024 \\ 0.007 & 0.024 & 1.367 \end{bmatrix}, \quad FN_B = 0.399.$$

It is clear from the results that the Bayesian estimator of  $\Sigma$  is more accurate than the MLE, as expected from a small sample size. We thus propose the consideration of the  $MBG_3^{(2)}(a, b, \exp(\cdot), \Phi)$  as the prior distribution for  $\Sigma$ , in the application of forthcoming section.

### 5.2.2 Application

The Fisher iris dataset has been used extensively in multivariate analysis and we use it here as an illustration of a proposed member of the beta generator (see (2.5)) as the prior for the covariance matrix of a multivariate Gaussian distribution.

The data set contains the measurements (in centimeters) of the  $X_1 =$  sepal length,  $X_2 =$  sepal width,  $X_3 =$  petal length and  $X_4 =$  petal width, respectively, for 150 flowers from three different species of iris (50 flowers from each species). The species are Iris setosa, Iris versicolor, and Iris virginica.

We assume the following model with an improper prior for  $\mu$ :

$$\begin{aligned} \mathbf{X} &= \{X_1 \ X_2 \ X_3 \ X_4\} \sim N_4(\mu, \Sigma), \\ \Sigma &\sim MBG_4^{(2)}\left(\frac{n}{2} + 6, 1, \exp(\cdot), I_4\right). \end{aligned}$$

A Markov chain of length 100000 were produced for  $\mu$  and  $\Sigma$  and thinning was applied to these chains as to include every 9<sup>th</sup> sample after a burn-in of 10% for a final sample of size 10000 from Algorithm 4.2.1. The sample estimates and Bayesian estimates for  $\Sigma$  are as follows:

$$\widehat{\Sigma}_{MLE} = \begin{bmatrix} 0.68 & -0.04 & 1.27 & 0.52 \\ -0.04 & 0.19 & -0.33 & -0.12 \\ 1.27 & -0.33 & 3.12 & 1.29 \\ 0.52 & -0.12 & 1.29 & 0.58 \end{bmatrix} \quad \text{and} \quad \widehat{\Sigma}_B = \begin{bmatrix} 0.58 & -0.03 & 0.78 & 0.34 \\ -0.03 & 0.32 & -0.26 & -0.09 \\ 0.78 & -0.26 & 2.13 & 0.79 \\ 0.34 & -0.09 & 0.79 & 0.63 \end{bmatrix}.$$

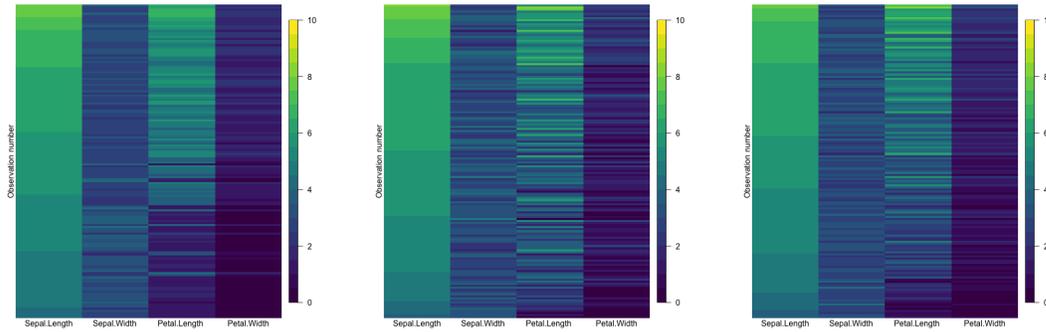


Figure 6: Original sample (right), and samples simulated form the model estimated by MLE (middle) and Bayes (left)

Since we cannot compare the results using the Frobenius norm between the estimate and the target as before (since we do not know the true values), we simulate a sample from the estimated models (MLE and Bayes). We can then compare the original sample with the simulated samples and inspect visually, through an image, the similarity in the dependency structure between the original and simulated samples. From Figure 6 we can see that the Bayesian framework produces an estimated model that closely matches the observed data.

## 6 Further developments

The distributions of quadratic forms can be of interest for inferential purposes, hence the following theorem.

**Theorem 6.1.** Suppose that  $Y = (\Omega - \Psi)^{\frac{1}{2}} X (\Omega - \Psi)^{\frac{1}{2}} + \Psi$  with  $X \sim MBG_m^{(nc)}(a, b, h, \Phi), \Psi, \Omega \in S_m$  are some known matrix parameters and under the meaning of partial löwner ordering,  $\Omega > \Psi$ . Then we have

$$f(Y) = \zeta_{a,b} \det(Y - \Psi)^{a - \frac{1}{2}(m+1)} \det(\Omega - Y)^{b - \frac{1}{2}(m+1)} \det(\Omega + Y - 2\Psi)^{-(a+b)} \det(\Omega - \Psi)^{\frac{1}{2}(m+1)} h(\text{tr } \Theta(Y - \Psi)),$$

where  $\Psi < Y < \Omega$  and  $\Theta = (\Omega - \Psi)^{-\frac{1}{2}} \Phi (\Omega - \Psi)^{-\frac{1}{2}}$ .

The derived density function is of a noncentral form, and can thus be viewed as a noncentral generator distribution. It is of interest to find the distribution of the

determinant where the matrix-variate has this distribution leading to *generalized Wilks'* statistics.

**Theorem 6.2.** *Let  $X \sim MBG_m^{(nc)}(a, b, \Phi, h)$ . Then  $y = \det(X)$  has the following density function*

$$f(y) = \zeta_{a,b} \Gamma_m(b) \sum_{\tau, \kappa, \phi} \frac{(-1)^k h^{(t)}(0) \left(\theta_{\phi}^{\kappa, \tau}\right)^2}{t! k! C_{\tau}(\mathbf{I}_m)} C_{\tau}(\Phi) G_{2m, 2m}^{2m, 0} \left( y \mid \begin{matrix} c_1, \dots, c_{2m} \\ d_1, \dots, d_{2m} \end{matrix} \right),$$

where

$$\begin{aligned} c_j &= \begin{cases} a + b - 1 + \phi_{\frac{i+1}{2}} - \frac{1}{4}(i - 1), & i = 1, 3, 5, \dots, 2m - 1 \\ a + b - 1 - \frac{1}{4}(i - 2), & i = 2, 6, 10, \dots, 2m, \end{cases} \\ d_j &= \begin{cases} a + b - 1 + k_{\frac{i+1}{2}} - \frac{1}{4}(i - 1), & i = 1, 3, 5, \dots, 2m - 1 \\ a - 1 + \phi_{\frac{i}{2}} - \frac{1}{4}(i - 2), & i = 2, 6, 10, \dots, 2m, \end{cases} \end{aligned}$$

where  $G(\cdot)$  denotes the Meijer's G function.

## 7 Conclusion

In this paper a matrix-variate beta generator distributions was constructed from the well-known matrix-variate beta type III distributions by employing a Borel measurable function. Similar generators based on the multivariate beta type I and type II can be proposed. Some characteristics of this generator distribution were derived. The joint density function of the eigenvalues of such a chase was visually illustrated for various  $h(\cdot)$ . The applicability of these generator distributions were shown through a Bayesian analysis of the binomial model, and in the multivariate sense the Bayesian analysis of a multivariate Gaussian sample.

## Acknowledgments

The authors would like to hereby acknowledge the support of the StatDisT group. This work is based upon research supported by the National Research foundation of South Africa, Reference: SRUG 190308422768 grant number 120839, IFR170227223754 grant number 109214 and SARCHI Research Chair UID: 71199. The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

## References

- Bekker, A., Roux, J. J. J., Ehlers, R., and Arashi, M. (2012), Distribution of the product of determinants of noncentral bimatrix beta variates. *Journal of Multivariate Analysis*, **109**, 73-87.
- Bekker, A., Van Niekerk, J., and Arashi, M. (2017), Wishart distributions: Advances in theory with Bayesian application. *Journal of Multivariate Analysis*, **155**, 272-283.
- Chikuse, Y. (1980), Invariant polynomials with matrix arguments and their applications. *Multivariate Statistical Analysis*, **1**, 54-68.
- Davis, A. W. (1979), Invariant polynomials with two matrix arguments extending the zonal polynomials: Applications to multivariate distribution theory. *Annals of the Institute of Statistical Mathematics*, **31**(A), 465-485.
- Ehlers, R. (2011), *Bimatrix Variate Distributions of Wishart Ratios With Application*. Unpublished PhD Dissertation, University of Pretoria, South Africa.
- Gupta, A.K. and Nagar, D.K. (2000), Matrix-variate beta distribution. *International Journal of Mathematical Sciences*, **24**(7), 449-459.
- Gupta, A. K., and Nagar, D. K. (2006), A generalized matrix-variate beta distribution. *International Journal of Applied Mathematical Sciences*, **31**(1), 21-36.
- Gupta, A. K., and Nagar, D. K. (2009), Properties of matrix-variate beta type 3 distribution. *International Journal of Mathematical Sciences*, <http://dx.doi.org/10.1155/2009/308518>.
- Muirhead, R. J. (2005), *Aspects of Multivariate Statistical Theory*. New York: Wiley.
- Nadarajah, S., and Kotz, S. (2006), Some beta distributions. *Bulletin of the Brazilian Mathematical Society*, **31**(1), 103-125.
- Nagar, D. K., and Gupta, A. K. (2002), Matrix-variate Kummer-beta distribution. *Journal of the Australian Mathematical Society*, **73**(1), 11-26.
- Nagar, D. K., Roldá-Correa, A., and Gupta, A. K. (2013), Extended matrix-variate gamma and beta functions. *Journal of Multivariate Analysis*, **122**, 53-69.
- Nagar, D. K., Arashi, M., and Nadarajah, S. (2017), Bimatrix variate gamma-beta distributions. *Communications in Statistics - Theory and Methods*, **46**(9), 4464-4483.

- Ng, K. W., and Kotz, S. (1995), *Kummer-gamma and Kummer-beta univariate and multivariate distributions*. Research report, The University of Hong Kong, Hong Kong.
- Pham-Gia, T., Phong, D. T., and Thanh, D. N. (2020), Distributions of powers of the central beta matrix variates and applications. *Statistical Methods and Applications*, **29**(3), 651-668.
- Tounsi, M. (2019), The Extended Matrix-Variate Beta Probability Distribution on Symmetric Matrices. *Methodology and Computing in Applied Probability*, 1-30.
- Zine, R. (2012), On the matrix-variate beta distribution. *Communications in Statistics-Theory and Methodology*, **41**(9), 1569-1582.