The Thirty-Third International
FLAIRS Conference (FLAIRS-33)

MCMC-Based Learning of Finite Bivariate Beta Mixture Models

Maryam Rasti, Narges Manouchehri, Nizar Bouguila
Concordia Institute for Information Systems Engineering, Concordia University
Montreal, Quebec, Canada - H3G 1M8
narges.manouchehri@mail.concordia.ca
nizar.bouguila@concordia.ca

Abstract

In this paper, we present a Bayesian approach for finite mixture models based on three-parameter bivariate Beta distributions. The estimation of the parameters is based on the Monte Carlo simulation technique of Gibbs sampling mixed with a Metropolis-Hastings step. The performance of our Bayesian algorithm is verified by several synthetic datasets and at the end, the feasibility of the proposed method is demonstrated by experimenting on some real datasets in which, the results are compared with those obtained by implementing the same approach using Gaussian mixture model.

Introduction

Technological advances led to generating huge amounts of various types of complex data and analyzing such valuable resources of data and extraction of the latent pattern is a topic of interest in different areas of science and technology. Thus, data mining and machine learning techniques witnessed tremendous improvement. One of the main attention-grabbing approaches is clustering method, specifically mixture model. To describe data by this approach, we need to define a distribution and Gaussian distribution has been widely used in mixture models. However, various types of data may have non-Gaussian and asymmetric nature and other distributions could better describe data. To use mixture models, parameter estimation is a major task and several approaches such as expectation-maximization have been applied. However, they have some disadvantages (McLachlan and Krishnan 1997) such as optimization problems and local maxima (Robert and Casella 2013). Due to considerable developments in computational methods, Bayesian techniques have been applied to overcome above-mentioned drawbacks (Diebolt and Robert 1994), (Neal 1992). In this work, we apply Beta distribution as an alternative for Gaussian distribution and introduce a three parameters-based bivariate Beta mixture model to cluster two dimensional vectors with features defined between zero and one. After that, we discuss our mixture model and model learning, respectively. Then, the experimental results are provided to show the accuracy and performance of the proposed model. In this part first, we test the algorithm on synthetic data by comparing the real and estimated parameters and second the accuracy of the model is verified by applying the method on real datasets. For each real dataset, the performance of our model is compared with Gaussian mixture models in terms of accuracy. Finally, the last section concludes the work.

Bayesian Learning of a Finite Bivariate Beta Mixture Model

Bivariate Beta distribution with three shape parameters a, b, c and two real positive correlated random variables X and Y was introduced in (Olkin and Liu 2003) and its joint density function is given as:

$$f(X,Y) = \frac{X^{a-1}Y^{b-1}(1-X)^{b+c-1}(1-Y)^{a+c-1}}{B(a,b,c)(1-XY)^{(a+b+c)}} \quad (1)$$

$$B(a,b,c) = \frac{\Gamma(a)\Gamma(b)\Gamma(c)}{\Gamma(a+b+c)}$$

A finite bivariate Beta mixture model with M components is defined as $p(\vec{X} \mid \Theta) = \sum_{j=1}^{M} p(\vec{X} \mid \vec{\alpha}_j)P_j$ (Bishop 2006) where the P_j is the mixing probability and $p(\vec{X} \mid \Theta)$ is the bivariate Beta distribution. The symbol $\Theta = (P_j, \vec{\alpha}_j)$ refers to the set of weight and shape parameters of component j. $\vec{\alpha}_j = (a_j, b_j, c_j)$ is the set of parameters defining the j-th component, and $\vec{P}_j = (P_1, \ldots, P_M)$. The weight of each cluster ,P_j, is between zero and one and all the weights sum up to $= 1$ for $j = 1, \ldots, M$.

Bayesian Learning

One of the important problems in the model learning phase is the estimation of the mixture parameters. To do so, we have to initialize our values by applying k-means and method of moments (MOM) (Manouchehri and Bouguila 2018). Then, the new parameters are estimated by using Gibbs sampling within Metropolis-Hastings algorithm (Bouguila, Ziou, and Hammoud 2009).

Gibbs within Metropolis-Hastings

Let's assume that $\mathcal{X} = \{\vec{X}_1, \ldots, \vec{X}_N\}$ is a set of N 2-dimensional vectors where each vector is composed of M different finite clusters. Each observation, $\vec{X}_n = (X_n, Y_n)$, is generated from a finite but unknown bivariate Beta mixture model $p(\vec{X} \mid \Theta)$. We propose a M-dimensional membership $\vec{Z}_i = (Z_{i1}, \ldots, Z_{iM})$ to each observation, such that $Z_{ij} = 1$ if $\vec{X}_i$ belongs to component j and zero, otherwise. Thus, for $\mathcal{X}$ we have $\mathcal{Z} = \{Z_1, \ldots, Z_N\}$ and a complete form of data $(\mathcal{X}, \mathcal{Z})$ follows $p(\mathcal{X}, \mathcal{Z} \mid \Theta)$. In Bayesian inference, we assume a prior and posterior probability distribution for parameters Θ, which here we express them by $\pi(\Theta)$ and $\pi(\Theta \mid \mathcal{X}, \mathcal{Z})$, respectively (Robert and Casella 2013). This is based on Bayes formula defined as follows:

$$\pi(\Theta \mid \mathcal{X}, \mathcal{Z}) = \frac{p(\mathcal{X}, \mathcal{Z} \mid \Theta)\pi(\Theta)}{\int p(\mathcal{X}, \mathcal{Z} \mid \Theta)\pi(\Theta)} \propto p(\mathcal{X}, \mathcal{Z} \mid \Theta)\pi(\Theta) \tag{2}$$

where $\int p(\mathcal{X}, \mathcal{Z} \mid \Theta)\pi(\Theta)$ is the marginal density of the complete data $(\mathcal{X}, \mathcal{Z})$. The posterior distribution assists to simulate Θ which follows $\pi(\Theta \mid \mathcal{X}, \mathcal{Z})$. This technique, called Gibbs sampling, is commonly applied in estimating of Bayesian mixtures. In fact, instead of computation, we consider membership vector $\vec{X}_i$ as a missing multinomial variable defined by $\vec{Z}_i \sim \mathcal{M}(1; \hat{Z}_{i1}, \ldots, \hat{Z}_{iM})$ where each of $\hat{Z}_{ij}$ can be expressed by $\hat{Z}_{ij} = \frac{p(\vec{X}_i|\xi_j)P_j}{\sum_{j=1}^{M} p(\vec{X}_i|\xi_j)P_j}$. Assuming $\pi(P \mid \mathcal{Z}, \mathcal{X})$ as a density function for weight and considering this fact that it is independent of $\mathcal{X}$, we can write $\pi(P \mid \mathcal{Z}, \mathcal{X}) = \pi(P \mid \mathcal{Z})$. As mentioned before, Θ includes weight and shape parameter. Thus we need to simulate missing value $\mathcal{Z}$, weight $\vec{P}$ and shape parameter $\vec{\alpha}$ sequentially by standard Gibbs sampler with following procedure (Diebolt and Robert 1994), (Marin, Mengersen, and Robert 2005), (Bouguila, Ziou, and Hammoud 2009):

1. Initialization with k-means and MOM
2. Step t: For $t = 1, \ldots$
   (a) Generate $\vec{Z}_i^{(t)} \sim \mathcal{M}(1; \hat{Z}_{i1}^{(t-1)}, \ldots, \hat{Z}_{iM}^{(t-1)})$
   (b) Generate *P* from $\pi(P \mid \mathcal{Z}^{(t)})$
   (c) Generate $\vec{\alpha}$ from $\pi(\vec{\alpha} \mid \mathcal{Z}^{(t)}, X)$

For missing parameters, as $\pi(P \mid \mathcal{Z}) \sim \pi(P)\pi(\mathcal{Z} \mid P)$, if we determine $\pi(P)$ and $\pi(\mathcal{Z} \mid P)$, we can find $\pi(P \mid \mathcal{Z})$. As mentioned before, in Bayesian inference we need to assign a prior to each parameter to simulate its posterior and in this part we find proper priors. To do this task, considering the nature of weight $\vec{P}$, as it sums to one and all its values $(P_1, \ldots, P_M)$ are positive, our natural choice is the Dirichlet distribution (Marin, Mengersen, and Robert 2005) which is defined by $\pi(P) = \frac{\Gamma(\sum_{j=1}^{M}\eta_j)}{\prod_{j=1}^{M}\Gamma(\eta_j)}\prod_{j=1}^{M} P_j^{\eta_j - 1}$ where $\eta = (\eta_1, \ldots, \eta_M)$ is the Dirichlet distribution's parameter vector. Thus, $\pi(\mathcal{Z} \mid P) = \prod_{i=1}^{N}\pi(Z_i \mid P) = \prod_{i=1}^{N}\prod_{j=1}^{M} P_j^{Z_{ij}} = \prod_{j=1}^{M} P_j^{n_j}$ where $\eta_j = \sum_{i=1}^{N} I_{Z_{ij}} =_j$. Then,

$$\pi(P \mid \mathcal{Z}) = \frac{\Gamma(\sum_{j=1}^{M}\eta_j)}{\prod_{j=1}^{M}\Gamma(\eta_j)}\prod_{j=1}^{M} P_j^{\eta_j + n_j - 1} \tag{3}$$
$$\propto \mathcal{D}(\eta_1 + n_1, \ldots, \eta_M + n_M)$$

where D is a Dirichlet distribution with parameters $D(\eta_1 + n_1, \ldots, \eta_M + n_M)$. To find $\pi(\alpha_j)$ as the prior of shape parameter, we consider the fact that bivariate Beta distribution belongs to the exponential family. In fact, if a S-parameter density p belongs to the exponential family and we assume θ as its distribution parameter, then we can write it as following (Bouguila, Ziou, and Hammoud 2009):

$$p(\vec{X} \mid \theta) = H(\vec{X}) exp\Big(\sum_{l=1}^{s} G_l(\theta)T_l(\vec{X}) + \Phi(\theta)\Big) \tag{4}$$

$\pi(\theta) \propto exp\big(\sum_{l=1}^{s}\rho_l G_l(\theta) + \kappa\Phi(\alpha)\big)$ as a conjugate prior on θ is given where $\rho = (\rho_l, \ldots, \rho_s) \in Rs$ and $\kappa > 0$ are hyperparameter. Considering our density function and $\vec{X} = (X, Y)$, the prior distribution is as follows:

$$\begin{aligned}&\pi(\theta_j) \propto exp\big(\rho_1(a_j - 1) + \rho_2(b_j - 1) + \rho_3(b_j + c_j - 1)\\&+ \rho_4(a_j + c_j - 1) + \rho_5(a_j + b_j + c_j) + \kappa\times\\&(\log\Gamma(a_j + b_j + c_j) - \log\Gamma(a_j) - \log\Gamma(b_j) - \log\Gamma(c_j))\big)\end{aligned} \tag{5}$$

Thus, the posterior distribution is defined by following equation where $(\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \kappa)$ are prior hyperparameters:

$$\begin{aligned}&\pi(\theta_j \mid \mathcal{Z}, \mathcal{X}) \propto \pi(\theta_j)\prod_{Z_{ij}=1}\pi(X_i \mid \theta_j)\\&\propto exp((a_j - 1)(\rho_1 + \sum_{Z_{ij}=1}\log X_i) +\\&(b_j - 1)(\rho_2 + \sum_{Z_{ij}=1}\log Y_i) + (b_j + c_j - 1)\\&(\rho_3 + \sum_{Z_{ij}=1}\log(1 - X_i)) + (a_j + c_j - 1)\\&(\rho_4 + \sum_{Z_{ij}=1}\log(1 - Y_i)) + (a_j + b_j + c_j)\\&(\rho_5 + \sum_{Z_{ij}=1}\log\frac{1}{1 - X_iY_i}) + (\kappa + n_j)\\&((\log\Gamma(a_j + b_j + c_j) - \log\Gamma(a_j) - \log\Gamma(b_j)))\\&- \log\Gamma(c_j)))\end{aligned} \tag{6}$$

As the prior and the posterior have the same form, we conclude that $\pi(\theta_j)$ is a conjugate prior on θ_j. Thus, the posterior hyper parameters are as follows:
$\rho_1 + \sum_{Z_{ij}=1}\log X_i, \rho_2 + \sum_{Z_{ij}=1}\log Y_i$,
$\rho_3 + \sum_{Z_{ij}=1}\log(1 - X_i), \rho_4 + \sum_{Z_{ij}=1}\log(1 - Y_i)$,
$\rho_5 + \sum_{Z_{ij}=1}\log\frac{1}{1 - X_iY_i}, \kappa + n_j$
Considering (Kleiter 1992), (Castillo, Hadi, and Solares 1997), once the sample $\mathcal{X}$ is known, it can be used to get the prior hyperparameters (Bensmail et al. 1997). Then we held $(\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \kappa)$ fixed at following values:

$$\rho_1 = \sum_{Z_i=1}^{N}\log X_i, \rho_2 = \sum_{Z_i=1}^{N}\log Y_i, \rho_3 = \sum_{Z_i=1}^{N}\log(1 - X_i),$$
$$\rho_4 = \sum_{Z_i=1}^{N}\log(1 - Y_i), \rho_5 = \sum_{Z_i=1}^{N}\log\frac{1}{1 - X_iY_i}, \kappa = n_j = 1$$

Thus, the steps of the Gibbs sampler are (Bouguila, Ziou, and Hammoud 2009) as follows:

1. Initialization
2. Step t: For $t = 1, \ldots$
   (a) Generate $\vec{Z}_i^{(t)} \sim M(1; \hat{Z}_{i1}^{(t-1)}, \ldots, \hat{Z}_{iM}^{(t-1)})$
   (b) Compute $\eta_j = \sum_{i=1}^{N} I_{Z_{ij}} =_j$
   (c) Generate $P^{(t)}$ by Equation (3) and $\alpha_j^{(t)}$ by Equation (6).

We are using a hybrid MH-within-Gibbs based Markov chain Monte Carlo (MCMC) algorithm based on both Gibbs and M-H sampling, known as Metropolis-within-Gibbs sampling. When one of the parameters is hard to sample, the M–H algorithm offers a solution which is simulating from the posterior distribution. Starting from point $\alpha_j^{(0)}$, the corresponding Markov chain explores the surface of the posterior distribution. At iteration t, the steps of the M-H algorithm can be described as follows (Bouguila, Ziou, and Hammoud 2009):

1. Generate $\tilde{\alpha}_j \propto q(\alpha_j \mid \alpha_j^{(t-1)})$ and U $\propto \mathcal{U}[0,1]$
2. Compute $r = \frac{\pi(\tilde{\alpha}_j|\mathcal{Z},\mathcal{X})q(\alpha_j^{(t-1)}|\tilde{\alpha}_j)}{\pi(\alpha_j^{(t-1)}|\mathcal{Z},\mathcal{X})q(\tilde{\alpha}_j|\alpha_j)^{(t-1)}}$
3. If $r < u$ then $\alpha_j^{(t)} = \tilde{\alpha}_j$ else $\alpha_j^{(t)} = \alpha_j^{(t-1)}$

To choose the proposal distribution q, we use the random walk M–H algorithm. As $\tilde{a}_{jl}, \tilde{b}_{jl}, \tilde{c}_{jl}$ are positive, we choose log-normal distributions for them and $\tilde{a}_{jl} \sim \mathcal{LN}(\log(a_{jl}^{(t-1)}), \sigma_1^2), \tilde{b}_{jl} \sim \mathcal{LN}(\log(b_{jl}^{(t-1)}), \sigma_2^2), \tilde{c}_{jl} \sim \mathcal{LN}(\log(c_{jl}^{(t-1)}), \sigma_3^2)$. These equations are equivalent to $\log(\tilde{a}_{jl}) = \log(a_{jl}^{(t-1)}) + \epsilon_1$, $\log(\tilde{b}_{jl}) = \log(b_{jl}^{(t-1)}) + \epsilon_2$, $\log(\tilde{c}_{jl}) = \log(c_{jl}^{(t-1)}) + \epsilon_3$ where $\epsilon_1 \sim \mathcal{N}(0, \sigma_1^2)$, $\epsilon_2 \sim \mathcal{N}(0, \sigma_2^2)$ and $\epsilon_3 \sim \mathcal{N}(0, \sigma_3^2)$. The random walk M–H algorithm is composed of the following steps:

1. Generate $\tilde{a}_{jl}, \tilde{b}_{jl}, \tilde{c}_{jl}$, $l = 1, \ldots, d$ and U $\propto \mathcal{U}[0,1]$
2. Compute:

$$r = \frac{\pi(\tilde{\alpha}_j \mid \mathcal{Z}, \mathcal{X}) \prod_{l=1}^{d} \mathcal{LN}((a_{jl}^{(t-1)}) \mid \log(\tilde{a}_{jl}), \sigma_1^2)}{\pi(\alpha_j^{(t-1)} \mid \mathcal{Z}, \mathcal{X}) \prod_{l=1}^{d} \mathcal{LN}(\tilde{a}_{jl} \mid \log(a_{jl}^{(t-1)}), \sigma_1^2)}$$
$$\frac{\mathcal{LN}((b_{jl}^{(t-1)}) \mid \log(\tilde{b}_{jl}), \sigma_2^2)\mathcal{LN}((c_{jl}^{(t-1)}) \mid \log(\tilde{c}_{jl}), \sigma_3^2)}{\mathcal{LN}(\tilde{b}_{jl} \mid \log(b_{jl}^{(t-1)}), \sigma_2^2)\mathcal{LN}(\tilde{c}_{jl} \mid \log(c_{jl}^{(t-1)}), \sigma_3^2)}$$

$$r = \frac{\pi(\tilde{\alpha}_j \mid \mathcal{Z}, \mathcal{X}) \prod_{l=1}^{d} \tilde{a}_{jl}\tilde{b}_{jl}\tilde{c}_{jl}}{\pi(\alpha_j^{(t-1)} \mid \mathcal{Z}, \mathcal{X}) \prod_{l=1}^{d} a_{jl}^{(t-1)} b_{jl}^{(t-1)} c_{jl}^{(t-1)}}$$

3. If $r < u$ then $\alpha_j^{(t)} = \tilde{\alpha}_j$ else $\alpha_j^{(t)} = \alpha_j^{(t-1)}$

Experimental Results

In this section, we verify the performance of our proposed algorithm. To do so, first we test it on three synthetic datasets generated from bivariate Beta mixtures with different parameters. Table 1 demonstrates the real and estimated parameters. Regarding the results, our approach estimates parameters successfully in all three cases. Then, we evaluate our model accuracy on three real datasets which are normalized first. To assess the accuracy of the algorithm, we compare the accuracy of our model with Gaussian mixture models. We hereby introduce each datasets shortly describing their bivariate attributes, classes. Then, we present the

Three clusters		
Parameter	Real Values	Estimated Values
a_1	1.41	1.4336
b_1	4.14	4.0779
c_1	7	7.0291
p_1	0.5	0.4927
a_2	5.154	5.1706
b_2	3.923	3.9326
c_2	2.511	2.5176
p_2	0.1571	0.1534
a_3	5.33	5.3448
b_3	1.42	1.4584
c_3	8.631	8.6294
Four clusters		
a_1	1.515	1.5023
b_1	4.177	4.1026
c_1	7.6491	7.6583
p_1	0.48	0.4734
a_2	5.1149	5.1895
b_2	3.1595	3.1875
c_2	2.599	2.6023
p_2	0.15	0.1557
a_3	5.2137	5.1854
b_3	1.5444	1.4965
c_3	5.1469	5.1646
p_3	0.16	0.1584
a_4	1.112	1.0937
b_4	9.192	9.0896
c_4	11.2	11.1893
Five clusters		
a_1	3.8	3.7937
b_1	4.1	4.064
c_1	11.58	11.5478
p_1	0.2	0.2074
a_2	8.3	8.3490
b_2	1	1.0023
c_2	8.3	8.3469
p_2	0.15	0.1563
a_3	7.111	7.0737
b_3	7	6.9753
c_3	1.1	1.2046
p_3	0.25	0.2495
a_4	1.26	1.2947
b_4	11.1330	11.0978
c_4	8.8233	8.8641
p_4	0.25	0.2434
a_5	11.133	11.1402
b_5	3.133	3.1238
c_5	2.483	2.4376

Table 1: Real and estimated values of parameters for 3, 4, 5-component mixture

Dataset	BBMM	GMM
Elephants Classification	93.48%	81.88%
PET Film	97.73%	84.09%
Cane Age Estimation	79.7%	37.91%

Table 2: Accuracy of model performance in real datasets

results of our evaluation in Table 2. As it is obvious, our propose model bivariate Beta mixture models (BBMM) outperforms Gaussian mixture models (GMM).

Male African Elephants Classification: The first real dataset (Lee et al. 2013a) is based on a research about 138 male African elephants that lived through droughts in the first two years of their life (Lee et al. 2013b). Each observation is described by two features, namely, age and height. The animals are labeled one if they are firstborn and zero, otherwise.

Failure Time of Polyethylene Terephthalate (PET) Film: The second dataset (Hirose 1993) is about the failure time of PET film which is used in electrical insulation. This experiment tests the failure times for 44 samples in gas-insulated transformers in which four different voltage levels were used. This dataset contains two features: the voltage (in kv) and the failure or censoring time in hours. The target is the censoring indicator. 1 and 0 means right or left-censored data, respectively.

Age Estimation of Bramble Canes: In the last case of experimenting on real data, we used brambles dataset (Diggle and Milne 1983) which has 823 observations. In this research, the location of living bramble canes in a nine-meter square plot was recorded. We take nine meters to be the unit of distance so that the plot can be thought of as a unit square. As the target of the experiment, the bramble canes were classified by their age. Therefore; the dataset consists of the following features: the x coordinate of the position of the cane in the plot as well as the y coordinate. And the target value is the age classification of the canes; 0 indicates a newly emerged cane, 1 indicates a one-year old cane and 2 indicates a two-year old cane.

Conclusion

In this paper, we introduced first a new mixture model based on a bivariate Beta distribution with three parameters. Secondly, relying on the missing data structure of the mixture model, we presented an MCMC method to evaluate the posterior distribution and Bayes estimators by Gibbs sampling. After presenting our parameter estimation algorithm, we evaluated the statistical mixture model using synthetic and real datasets. From the results, we can conclude that the bivariate Beta mixture has a better accuracy in comparison with the performance of Gaussian mixture model.

References

Bensmail, H.; Celeux, G.; Raftery, A. E.; and Robert, C. P. 1997. Inference in model-based cluster analysis. *statistics and Computing* 7(1):1–10.

Bishop, C. M. 2006. *Pattern recognition and machine learning*. springer.

Bouguila, N.; Ziou, D.; and Hammoud, R. I. 2009. On bayesian analysis of a finite generalized dirichlet mixture via a metropolis-within-gibbs sampling. *Pattern Analysis and Applications* 12(2):151–166.

Castillo, E.; Hadi, A. S.; and Solares, C. 1997. Learning and updating of uncertainty in dirichlet models. *Machine Learning* 26(1):43–63.

Diebolt, J., and Robert, C. P. 1994. Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society: Series B (Methodological)* 56(2):363–375.

Diggle, P. J., and Milne, R. K. 1983. Bivariate cox processes: some models for bivariate spatial point patterns. *Journal of the Royal Statistical Society: Series B (Methodological)* 45(1):11–21.

Hirose, H. 1993. Estimation of threshold stress in accelerated life-testing. *IEEE Transactions on Reliability* 42(4):650–657.

Kleiter, G. D. 1992. Bayesian diagnosis in expert systems. *Artificial Intelligence* 54(1-2):1–32.

Lee, P. C.; Bussière, L. F.; Webber, C. E.; Poole, J. H.; and Moss, C. J. 2013a. Enduring consequences of early experiences: 40 year effects on survival and success among african elephants (loxodonta africana). *Biology letters* 9(2):20130011.

Lee, P. C.; Bussière, L. F.; Webber, C. E.; Poole, J. H.; and Moss, C. J. 2013b. Enduring consequences of early experiences: 40 year effects on survival and success among african elephants (loxodonta africana). *Biology letters* 9(2):20130011.

Manouchehri, N., and Bouguila, N. 2018. Learning of finite two-dimensional beta mixture models. In *2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC)*, 227–232. IEEE.

Marin, J.-M.; Mengersen, K.; and Robert, C. P. 2005. Bayesian modelling and inference on mixtures of distributions. *Handbook of statistics* 25:459–507.

McLachlan, G. J., and Krishnan, T. 1997. The em algorithm and extensions.

Neal, R. M. 1992. Bayesian mixture modeling. In *Maximum Entropy and Bayesian Methods*. Springer. 197–211.

Olkin, I., and Liu, R. 2003. A bivariate beta distribution. *Statistics & Probability Letters* 62(4):407–412.

Robert, C., and Casella, G. 2013. *Monte Carlo statistical methods*. Springer Science & Business Media.