Generating random variates from a bicompositional Dirichlet distribution

Bergman, Jakob

2009

Generating random variates from a bicompositional Dirichlet distribution

Jakob Bergman

2009:5

LUND UNIVERSITY

DEPARTMENT OF STATISTICS

S-220 07 LUND
SWEDEN

Generating random variates from a bicompositional Dirichlet distribution

Jakob Bergman[*][†]

December 4, 2009

Abstract

A composition is a vector of positive components summing to a constant. The sample space of a composition is the simplex and the sample space of two compositions, a bicomposition, is a Cartesian product of two simplices. We present a way of generating random variates from a bicompositional Dirichlet distribution defined on the Cartesian product of two simplices using the rejection method. We derive a general solution for finding a dominating density function and a rejection constant, and also compare this solution to using a uniform dominating density function. Finally some examples of generated bicompositional random variates, with varying number of components.

Keywords: bicompositional Dirichlet distribution; composition; Dirichlet distribution; random variate generation; rejection method; simplex

1 Introduction

A composition is a vector of positive components summing to a constant. The components of a composition are what we usually think of as proportions (at least when the vector sums to 1). Compositions arise in many different areas; the geochemical compositions of different rock specimens, the proportion of expenditures on different commodity groups in household budgets, and the party preferences in a party preference survey are all examples of compositions from three different scientific areas. For more examples of compositions, see for instance Aitchison (2003).

[*]Department of Statistics, Lund University, Box 743, SE-220 07 Lund, Sweden
[†]e-mail: jakob.bergman@stat.lu.se

The sample space of a composition is the simplex. Without loss of generality we will always take the summing constant to be 1, and we define the D-dimensional simplex $\mathscr{S}^D$ as

$$\mathscr{S}^D = \left\{ \mathbf{x} = (x_1, \ldots, x_D)^{\mathrm{T}} \in \mathscr{R}_+^D : \sum_{j=1}^{D} x_j = 1 \right\}, \tag{1}$$

where $\mathscr{R}_+$ is the positive real space. The joint sample space of two compositions is the Cartesian product of two simplices $\mathscr{S}^D \times \mathscr{S}^D$. It should be noted that, unlike the case for real Cartesian product spaces, $\mathscr{S}^D \times \mathscr{S}^D \neq \mathscr{S}^{D+D}$ and that $\mathscr{S}^D \times \mathscr{S}^D$ is not even a simplex, but a manifold with two constraints.

2 The rejection method

Leydold (1998) notes that apart from the multinormal and Wishart distributions, papers on generating bivariate and multivariate random variates are rare and most suggested general methods have disadvantages. The only universal algorithm for generating multivariate random variates is the algorithm presented by Leydold and Hörmann (1998), which is a generalisation of algorithms for the univariate and bivariate case given in different versions by Gilks and Wild (1992) and Hörmann (1995). However, Leydold (1998) concludes that this algorithm is very slow and suggests an alternative algorithm which requires a function of the density to be concave. The class of distributions that will be utilized in this paper is very versatile and is therefore hard to find a function that fulfils the requirements. Hence we will use the rejection method to construct a specialized method for generating bicompositional random variates.

The following description of the rejection method of generating random variates is based on Devroye (1986, pp. 40–44).

Let f be the density from which we wish to generate random variates. Let $c \geq 1$ be a constant and g be a density such that

$$f(\mathbf{z}) \leq cg(\mathbf{z}) \tag{2}$$

for all $\mathbf{z}$. We now generate a random variate $\mathbf{Z}$ with density g and a random number U uniformly distributed on the unit interval. We let

$$T = c\frac{g(\mathbf{Z})}{f(\mathbf{Z})}. \tag{3}$$

The variate $\mathbf{Z}$ is accepted if $UT \leq 1$, otherwise we reject $\mathbf{Z}$ and generate new $\mathbf{Z}$ and U until acceptance.

We thus need to find a dominating density g and constant c, and preferably such choices that will have high probabilities of acceptance and hence make the random variate generation efficient.

3 The bicompositional Dirichlet distribution

Bergman (2009) proposed a distribution, called the bicompositional Dirichlet distribution, for modelling random vectors on $\mathscr{S}^D \times \mathscr{S}^D$. The proposed distribution has the probability density function

$$f(\mathbf{x}, \mathbf{y}) = A \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) \left(\mathbf{x}^{\mathrm{T}} \mathbf{y} \right)^{\gamma}, \tag{4}$$

where $\mathbf{x}, \mathbf{y} \in \mathscr{S}^D$, $\alpha_j, \beta_j \in \mathscr{R}_+ (j = 1, \ldots, D)$ and $\gamma \in \mathscr{R}$. Expressions for the normalisation constant A are given in Bergman (2009). If $\gamma = 0$, the probability density function (4) is equal to the product of two Dirichlet probability density functions with parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)^{\mathrm{T}}$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_D)^{\mathrm{T}}$ respectively, and hence $\mathbf{X}$ and $\mathbf{Y}$ are independent.

When $\mathbf{X}, \mathbf{Y} \in \mathscr{S}^2$ we shall refer to this as the *bicomponent* case, and similarly to $\mathscr{S}^3$ as the *tricomponent* case, and to $\mathscr{S}^D (D > 2)$ as the *multicomponent* case.

4 Generating random bicompositions

Here, the bicomposition $(\mathbf{X}, \mathbf{Y})$ will take the role of $\mathbf{Z}$ in Section 2.

When $\gamma \geq 0$, we may use the product of two Dirichlet distributions as a dominating density, since $0 < \mathbf{x}^{\mathrm{T}} \mathbf{y} < 1$ and

$$A \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) (\mathbf{x}^{\mathrm{T}} \mathbf{y})^{\gamma} \leq A \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right).$$

Defining $B_{\boldsymbol{\alpha}} = \Gamma(\alpha_1 + \cdots + \alpha_D) / \prod_{j=1}^{D} \Gamma(\alpha_j)$ and analogously for $B_{\boldsymbol{\beta}}$, the inequality (2) becomes

$$A \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) (\mathbf{x}^{\mathrm{T}} \mathbf{y})^{\gamma} \leq c B_{\boldsymbol{\alpha}} \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} \right) B_{\boldsymbol{\beta}} \left(\prod_{j=1}^{D} y_j^{\alpha_j - 1} \right), \tag{5}$$

where the constant c is

$$c = \frac{A}{B_{\boldsymbol{\alpha}} B_{\boldsymbol{\beta}}}. \tag{6}$$

Generating a Dirichlet distributed random variate is easily done based on Gamma distributed variates. (Devroye, 1986, pp. 593–596)

Using a product of two Dirichlet distributions as dominating density is however not always very efficient, as $(\mathbf{x}^{\mathrm{T}} \mathbf{y})^{\gamma}$ will be close to 0 when γ is large. When $\gamma \geq 0$, and $\alpha_j, \beta_j > 1$ $(j = 1, \ldots, D)$, it is easily seen that the density (4) will have an upper bound. We may therefore use an uniform density as g, with $c = \max f(\mathbf{x}, \mathbf{y})$.

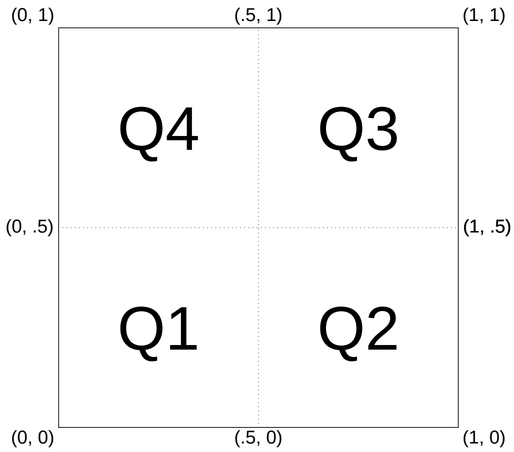

Figure 1. The four quadrants Q1-Q4 of the sample space $\mathscr{S}^2 \times \mathscr{S}^2$; the horizontal axis represents x and the vertical axis represents y.

4.1 The bicomponent case

The bicomponent case is treated separately as $\mathbf{x} = (x, 1-x)^{\mathrm{T}}$ and $\mathbf{y} = (y, 1-y)^{\mathrm{T}}$, and the density hence may be viewed as a function of x and y.

Bergman (2009) showed that the bicomponent bicompositional Dirichlet density exists if and only if $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$. If $\gamma < 0$, the factor $(\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma}$ will tend to infinity when x is close to 0 and y is close to 1, or when x is close to 1 and y is close to 0. We therefore divide the sample space $\mathscr{S}^2 \times \mathscr{S}^2$ into four quadrants, denoted Q1-Q4 counter-clockwise from the origin. Figure 1 shows the $\mathscr{S}^2 \times \mathscr{S}^2$ with the four quadrants.

To generate a random variate from a bicomponent bicompositional Dirichlet distribution with parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and

$$-\min(\alpha_2, \beta_2) < \gamma < 0,$$

we first randomly choose a quadrant with probability

$$p_k = \iint_{\mathrm{Q}k} f(x, y)\,dx\,dy \qquad (k = 1, 2, 3, 4), \tag{7}$$

where $f(x, y)$ is the bicomponent bicompositional Dirichlet probability density function (4) viewed as a function of x and y. Expressions for the cumulative distribution function has been given by Bergman (2009), which may be used in calculating p_k. Depending on which quadrant is chosen, we then choose a dominating density g and a constant c in the following manner.

Q1 & Q3 In quadrants Q1 and Q3, $\mathbf{x}^\mathrm{T}\mathbf{y} > 1/2$ and we may hence use a product of two Dirichlet (or equivalently Beta) distributions with parameters $\boldsymbol{\alpha}$ respectively $\boldsymbol{\beta}$ as g and a constant

$$c = \frac{A}{B_{\boldsymbol{\alpha}} B_{\boldsymbol{\beta}} 2^{\gamma}}. \tag{8}$$

Q2 In quadrant Q2, $\mathbf{x}^\mathrm{T}\mathbf{y}$ is bounded from below by $(1-x)/2$, and hence

$$(\mathbf{x}^\mathrm{T}\mathbf{y})^{\gamma} \le 2^{-\gamma}(1-x)^{\gamma}$$

as $\gamma < 0$. We therefore use a product of two Dirichlet distributions with parameters $(\alpha_1, \alpha_2 + \gamma)$ respectively $\boldsymbol{\beta}$ as the density g and the constant c given by

$$c = \frac{A}{B_{(\alpha_1,\alpha_2+\gamma)} B_{\boldsymbol{\beta}} 2^{\gamma}}. \tag{9}$$

Q4 Analogously, in quadrant Q4, $\mathbf{x}^\mathrm{T}\mathbf{y} > (1-y)/2$ and we hence use a product of two Dirichlet distributions with parameters $\boldsymbol{\alpha}$ respectively $(\beta_1, \beta_2 + \gamma)$ as g and c given in (10).

$$c = \frac{A}{B_{\boldsymbol{\alpha}} B_{(\beta_1,\beta_2+\gamma)} 2^{\gamma}} \tag{10}$$

We must though assure that the generated variates with density g are limited to that particular quadrant.

5 Comparison of the two dominating densities

The efficiency of the generation process will usually depend on the choice of dominating density. In most cases we have a possibility to choose between two different dominating densities: a product of two independent Dirichlet densities or a uniform density. In general, the product of two Dirichlet distributions will often be more efficient when γ is close to 0, but may however be highly inefficient when γ is large.

To compare the efficiency of the two dominating densities we generated 25,000 random variates for each of the dominating densities from a number of different bicomponent bicompositional Dirichlet distributions, and calculated the average number of trials to generate one random variate. Table 1 shows the results presented as the estimated probability of acceptance (the reciprocal of the average number of trials) as well as the results for a distribution where only a Dirichlet product is available as dominating density as the distribution density function does not have an upper bound. We note that the probability of

Table 1. Comparisons of the estimated acceptance probabilities depending on choice of dominating density. We clearly see that the product of two Dirichlet densities can be very inefficient for large values of γ, but also that it may be much more efficient than a uniform density for some distributions.

Parameter values					Dominating density	
α_1	α_2	β_1	β_2	γ	Dirichlet	Uniform
2.1	3.1	5.5	2.3	0.3	0.769	0.222
2.1	3.1	5.5	2.3	3.2	0.103	0.200
2.1	3.1	5.5	2.3	7.7	0.007	0.110
2.1	3.1	5.5	2.3	-1.2	0.208	0.208
2.1	3.1	0.7	2.3	3.2	0.185	NA
7.1	4.2	6.3	8.5	0.3	0.769	0.119
7.1	4.2	6.3	8.5	3.2	0.100	0.125
7.1	4.2	6.3	8.5	7.7	0.005	0.135
7.1	1.2	12.5	3.1	3.2	0.357	0.031

acceptance with a uniform density can be much (almost 30 times) larger than the probability of acceptance with a with a Dirichlet density. On the other hand we also see that there are distributions for which the probability of acceptance with a with a Dirichlet density is more than 10 times the probability of acceptance with a uniform density. As an graphical illustration of the differences between the distributions, 150 generated random variates from four of the distributions in Table 1 are plotted for each of the two dominating densities in Fig. 2 together with contour curves of the density.

The differences in efficiency between the two dominating densities is even more obvious for the multicomponent bicompositional Dirichlet distribution examples presented in Table 2. Here again, we generated 25,000 random vari-

Table 2. Comparisons of the estimated acceptance probabilities for some multicomponent bicompositional Dirichlet distributions.

Parameter values			Dominating density	
$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	γ	Dirichlet	Uniform
(2, 2, 2)	(2, 2, 2)	1	0.333	0.145
(2, 2, 2)	(2, 2, 2)	7	0.001	0.085
(2.1, 1.2, 3.2, 4.1, 2.8)	(3.2, 2.2, 5.3, 1.8, 2.9)	1	0.204	< 0.001
(2.1, 1.2, 3.2, 4.1, 2.8)	(3.2, 2.2, 5.3, 1.8, 2.9)	3	0.009	< 0.001

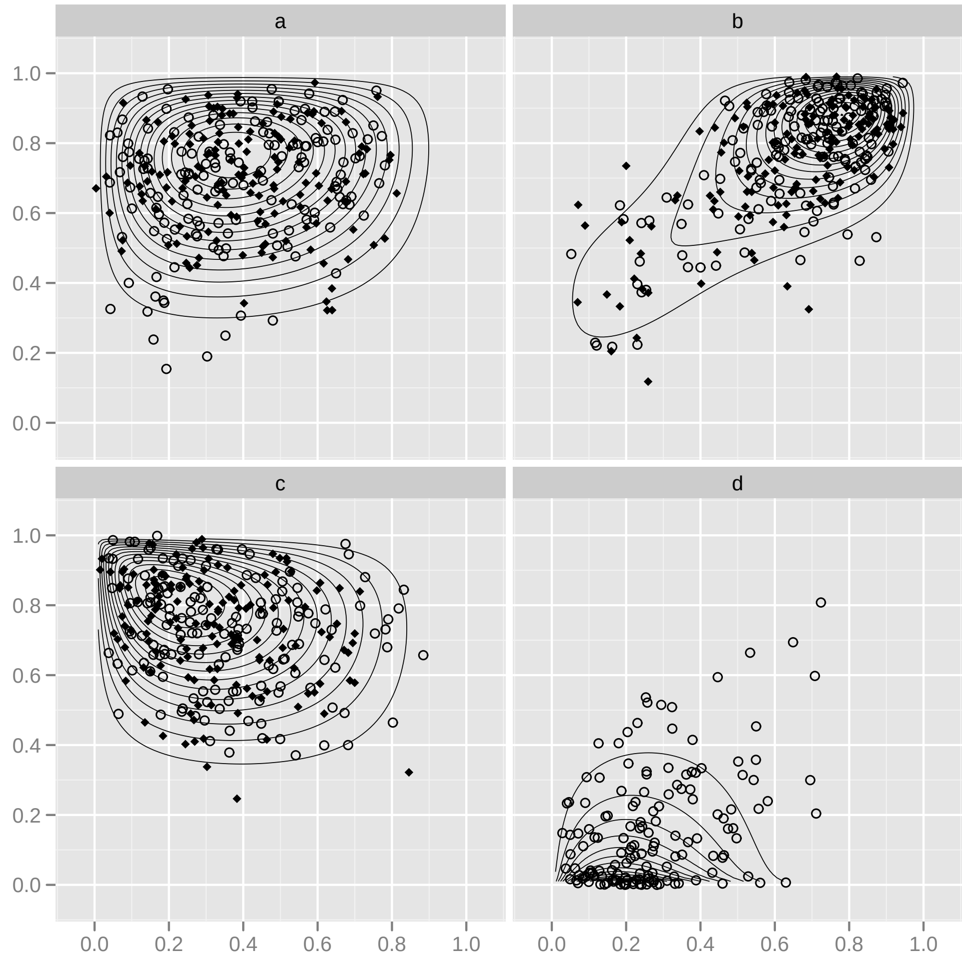

Figure 2. 150 random variates generated from four different bi-component bicompositional Dirichlet distributions with $(\alpha; \beta; \gamma)$ parameters $(2.1, 3.1;\ 5.5, 2.3;\ 0.3)$ (a), $(2.1, 3.1;\ 5.5, 2.3;\ 7.7)$ (b), $(2.1, 3.1;\ 5.5, 2.3;\ -1.2)$ (c), and $(2.1, 3.1;\ 0.7, 2.3;\ 3.2)$ (d), using the product of two Dirichlet densities (○) and a uniform density (◆) as dominating density. Since the distribution in (d) does not have an upper bound, a uniform density may not be utilized. As a reference, the contour curves of the true densities are also drawn.

ates, this time from four different multicomponent bicompositional Dirichlet distributions using both of the two dominating densities. For the tricomponent distributions, when $\gamma = 1$, the Dirichlet density has a probability of acceptance of more than twice that of the uniform density, but when $\gamma = 7$ the probability of acceptance of the uniform density is more than 80 times that of the the Dirichlet density. (Illustrations of random variates from the above tricomponent distributions are available as Online Resources 1 and 2.) For the two distributions with five components, we see that the Dirichlet density is much more effective for both cases. This is in accordance with Devroye (1986, p. 557), who notes that as the dimension D increases the rejection constant often deteriorates quickly when using an uniform density.

6 Conclusions

The choice of the dominating density is evidently crucial to the efficiency of this random variate generation. When γ is close to 0 or the number of components is large, a product of two Dirichlet density functions seems the most efficient, otherwise a uniform density function (if possible) is recommended. What is meant by close is however dependent of the other parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, so when in doubt, the recommendation would be to generate a small number of variates with each dominating density and see which is the most efficient for the particular parameter values in question. We note that the efficiency of the method seems to degrade as the dimension (i.e. the number of components) increases, and that further research is needed to find more efficient dominating densities for distributions with a large number of components and for large *gamma* values.

It remains yet to find a way of generating random numbers for the bicomponent case when $-\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1) < \gamma < -\min(\alpha_2, \beta_2)$ and the density function does not have an upper bound.

The random variate generation might further be made more efficient for at least the bicomponent case, by adopting the quadrant scheme also for positive γ; especially when the probability mass is concentrated in one or two of the quadrants, which is often the case for large γ, this might speed up the generation process considerably.

References

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Caldwell, NJ: The Blackburn Press.

Bergman, J. (2009, December). A bicompositional Dirichlet distribution. Technical Report 3, Department of Statistics, Lund University.

Devroye, L. (1986). *Non-uniform random variate generation*. New York: Springer-Verlag.

Gilks, W. R. and P. Wild (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics 41*(2), 337–348.

Hörmann, W. (1995). A rejection technique for sampling from T-concave distributions. *ACM Trans. Math. Softw. 21*(2), 182–193.

Leydold, J. (1998). A rejection technique for sampling from log-concave multivariate distributions. *ACM Trans. Model. Comput. Simul. 8*(3), 254–280.

Leydold, J. and W. Hörmann (1998). A sweep-plane algorithm for generating random tuples in simple polytopes. *Math. Comput. 67*(224), 1617–1635.