On matrix variate Dirichlet vectors

Konstencia Bobecka

Polytechnica, Warsaw, Poland

Richard Emilion

MAPMO, Université d'Orléans, 45100 Orléans, France

Jacek Wesolowski

Polytechnica, Warsaw, Poland

Abstract

A matrix variate Dirichlet vector is a random vector of independent Wishart matrices 'divided' by their sum. Many properties of Dirichlet vectors are usually established through integral computations assuming existence of density for the Wishart's. We propose a method to deal with the general case where densities need not exist. On the other hand, in dimension larger than 2, Dirichlet processes reduce to Dirichlet vectors and posterior of Dirichlet need not be Dirichlet.

Key words: Bayesian statistics, Definite positive matrix, Dirichlet vectors, random matrices, random distributions, Wishart distribution.

1 Introduction

Random matrix topic is known for its theoretical richness but also for its applications in a wide range of areas of both mathematics and theoretical physics. We are concerned here with Dirichlet distributions on matrix spaces. In dimension one, this interesting distribution is that of a vector of independent Gamma's real random variables divided by their sum. As Wishart distributions are usually considered as the generalization to the multivariate case of

Email addresses: bobecka@mini.pw.edu.pl (Konstencia Bobecka), richard.emilion@univ-orleans.fr (Richard Emilion), wesolo@mini.pw.edu.pl (Jacek Wesolowski).

Gamma distributions, multivariate Dirichlet distribution is defined as the distribution of a vector of independent Wishart matrices 'divided ' by their sum whenever a division algorithm is cautiously defined in the space of positive definite symmetric matrices (see e.g. [3]).
Several properties of dimension one Dirichlet's can be extended to higher dimension through elementary but rather nontrivial integral computations when assuming the existence of density for the Wishart's (see e.g. [3]). However such densities need not always exist and our purpose is to consider the general case. Our method, illustrated by two examples, hinges on a deep result due to Casalis-Letac [1] which states that the Dirichlet distribution doest not depend neither on the common second parameter of the Wishart's nor on the division algorithm.
The paper is structured as follows. In Section 2 we precise our notations and the notion of division algorithm. Section 3 contains the main results through two examples. In Section 4 it is shown that Dirichlet processes reduce to Dirichlet vectors and the last Section concerns the Bayesian approach.

2 Notations

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space on which are defined all the random variables (r.v.) mentioned in this paper. The probability distribution of a r.v. X will be denoted $\mathcal{P}_X$.
Let $n = 1, 2, \ldots$ and $m = 1, 2, \ldots$ be two positive integers and let $\mathcal{M}_{n\times m}$ (resp. $\mathcal{M}_n$) denote the vector space of real matrices with n lines and m (resp. n) columns. Let $\mathrm{Id}_n \in \mathcal{M}_n$ denote the identity matrix. The transpose of a matrix x will be denoted x^T.
Let $\mathcal{V}_n$ denote the euclidean space of symmetric $n \times n$ real matrices endowed with the inner product $(x, y) = \mathrm{tr}(xy)$. The Lebesgue measure on $\mathcal{V}_n$ is defined by imposing unit measure to the unit cube in this space. Let $\mathcal{V}_n^+$ be the cone of real symmetric positive definite $n \times n$ matrices and let $\overline{\mathcal{V}}_n^+$ denote its closure.
Let $\mathcal{T}_n$ denote the set of lower triangular real matrices and let $\mathcal{T}_n^+$ be the set of lower triangular real matrices with (strictly) positive diagonal terms.
Obviously, elements of $\mathcal{V}_n^+$ and that of $\mathcal{T}_n^+$ are invertible.
Let $\mathcal{G}_n$ be the set of mappings $\nu : \mathcal{V}_n \longrightarrow \mathcal{V}_n$ such that there exists an invertible $a \in \mathcal{M}_n$ with $\nu(x) = axa^T$ for any $x \in \mathcal{V}_n$, that is the subgroup of automorphisms which preserve $\mathcal{V}_n^+$.
Let $\mathcal{C}_n$ be the connected component of $\mathcal{G}_n$ containing Id_n.

Definition 1 *A division algorithm is a measurable mapping* $g : \mathcal{V}_n^+ \longrightarrow \mathcal{C}_n$ *such that* $g(x)(x) = Id_n$ *for all* $x \in \mathcal{V}_n^+$

We will be concerned by the following two usual division algorithms

$$g(x)(y) = y^{-\frac{1}{2}} x y^{-\frac{1}{2}} \tag{1}$$

$$g(x)(y) = c(y)^{-1} x (c(y)^{-1})^T \tag{2}$$

where $c(y)$ is the matrix of $\mathcal{T}_n^+$ appearing in the Cholesky decomposition of $y \in \mathcal{V}_n^+$, that is $c(y)c(y)^T = y$. In this case we will take

$$y^{\frac{1}{2}} = c(y) \in \mathcal{T}_n^+, \text{ so that } y^{-\frac{1}{2}} \in \mathcal{T}_n^+ \text{ and } g(x)(y) = y^{-\frac{1}{2}} x (y^{-\frac{1}{2}})^T \tag{3}$$

For any positive integer $k = 1, 2, \ldots$, let

$$\mathcal{Q}_n(k) = \left\{(p_1, \ldots, p_k) : \ p_i \in \overline{\mathcal{V}}_n^+, \ i = 1, \ldots, k, \text{ and } \Sigma_{i=1}^k p_i = \mathrm{Id}_n\right\}$$

Let finally

$$\Lambda_n = \left\{0, \frac{1}{2}, \frac{2}{2}, \ldots, \frac{n-1}{2}\right\} \cup \left(\frac{n-1}{2}, \infty\right).$$

3 Wishart and Dirichlet distributions on matrix spaces

Definition 2 *For $\alpha \in \Lambda_n$ and $a \in \mathcal{V}_n^+$, the Wishart probability measure $\gamma_{\alpha,a}^{(n)}$ on $\overline{\mathcal{V}}_n^+$ is defined through its Laplace transform:*

$$L(\theta) = \int_{\mathcal{V}_n^+} e^{(\theta,x)} \, \gamma_{\alpha,a}^{(n)}(dx) = \left(\frac{\det(a)}{\det(a-\theta)}\right)^{\alpha}$$

for $a - \theta \in \mathcal{V}_n^+$.

If $\alpha > \frac{n-1}{2}$, then the probability measure $\gamma_{\alpha,a}^{(n)}$ is concentrated on the open cone $\mathcal{V}_n^+$ and has a density with respect to the Lebesgue measure on $\mathcal{V}_n$ of the form

$$\gamma_{\alpha,a}^{(n)}(dx) = \frac{[\det(a)]^{\alpha}}{\Gamma_n(\alpha)} [\det(x)]^{\alpha - \frac{n+1}{2}} e^{-(a,x)} I_{\mathcal{V}_n^+}(x) \, dx \, .$$

If $\alpha \in \{0, \frac{1}{2}, \ldots, \frac{n-1}{2}\}$, then $\gamma_{\alpha,a}^{(n)}$ is concentrated on the boundary $\partial \, \mathcal{V}_n^+$ of the cone, thus it does not have a density with respect to the Lebesgue measure on $\mathcal{V}_n$.

Now, consider a measure on the k-th cartesian power of $\overline{\mathcal{V}}_n^+$ which is a product of Wishart measures with same second parameter

$$\gamma^{(\otimes k)} = \gamma_{\alpha_1,a}^{(n)} \otimes \ldots \otimes \gamma_{\alpha_k,a}^{(n)} \, ,$$

such that $\sum_{j=1}^{k} \alpha_j > \frac{n-1}{2}$. That is its support is

$$\mathcal{H} = \{(x_1, \ldots, x_k) \in \left[\overline{\mathcal{V}}_n^+\right]^{\times k} : \ s = x_1 + \ldots + x_k \in \mathcal{V}_n^+\} \ .$$

Let $h : \mathcal{H} \to \mathcal{Q}_n(k)$ be defined by

$$h(x_1, \ldots, x_k) = s^{-1/2}(x_1, \ldots, x_k)s^{-1/2} \ , \quad \forall \, x_1, \ldots, x_k \in \mathcal{H}.$$

Definition 3 *The Dirichlet probability measure* $Dir_n\,(\alpha_1, \ldots, \alpha_k)$ *on* $\mathcal{Q}_n(k)$ *is defined as transformation of* $\gamma^{(\otimes k)}$ *through the mapping* h*, that is for any* A*, a Borel set in* $\mathcal{Q}_n(k)$*,*

$$Dir_n\,(\alpha_1, \ldots, \alpha_k)\,(A) = \gamma^{(\otimes k)}(h^{-1}(A)) \ .$$

If $\alpha_i \in (\frac{n-1}{2}, \infty)$, for all $i = 1, \ldots, k$, then the projection of $Dir_n\,(\alpha_1, \ldots, \alpha_k)$ on the first $k-1$ matrix coordinates is supported on the open set

$$\mathcal{U}_k^{(n)} = \{(x_1, \ldots, x_{k-1}) \in \left[\mathcal{V}_n^+\right]^{\times(k-1)} : \ \mathrm{Id}_n - \sum_{j=1}^{k-1} x_j \in \mathcal{V}_n^+\}$$

and has the density with respect to the Lebesgue measure on $[\mathcal{V}_n^+]^{\times(k-1)}$ of the form

$$f(x) = \frac{\Gamma_n\left(\sum\limits_{j=1}^{k} \alpha_j\right)}{\prod\limits_{j=1}^{k} \Gamma_n(\alpha_j)} \prod_{j=1}^{k-1} (\det(x_j))^{\alpha_j - \frac{n+1}{2}} \left(\det\left(\mathrm{Id}_n - \sum_{j=1}^{k-1} x_j\right)\right)^{\alpha_k - \frac{n+1}{2}} I_{\mathcal{U}_k^{(n)}}(x). \tag{4}$$

Remark 1 *Alternatively, if* $W_1, \ldots, W_k$ *are independent Wishart random matrices,* $W_i \sim \gamma_{\alpha_i, a}^{(n)}$*,* $i = 1, \ldots, k$*, such that* $W = W_1 + \ldots + W_k$ *is positive definite a.s. then the random vector*

$$\mathbf{P} = (P_1, \ldots, P_k) = W^{-1/2}(W_1, \ldots, W_k)W^{-1/2} \sim Dir_n\,(\alpha_1, \ldots, \alpha_k) \ .$$

If $\alpha_i \in (\frac{n-1}{2}, \infty)$*, for all* $i = 1, \ldots, k$*, then* $(P_1, \ldots, P_{k-1})$ *has the density given by (4) otherwise there is no density.*

Actually the Dirichlet distribution does not depend neither on the second parameter a of the Wishart's nor on the division algorithm. This is precised by the following nice theorem of Casalis-Letac ([1], Theorem 3.1) which will be crucial in our proofs.

Theorem 2 *Let* $W_i \overset{ind}{\sim} \gamma_{\alpha_i, a}^{(n)}$*,* $i = 1, \ldots, k$ *be independent Wishart random matrices such that* $W = W_1 + \ldots + W_k$ *is positive definite and let* g *be any division* *algorithm then the random vector $(g(W)W_1,\dots,g(W)W_k)$ is independent of W and its distribution does not depend neither on a nor on g.*

4 Main results

4.1 Diagonal blocks

The following theorem shows that the vector of diagonal blocks of a Dirichlet is still a Dirichlet.

Theorem 3 *Let $\mathbf{P} = (P_1,\dots,P_k)$ be a random vector assuming values in $\mathcal{Q}_n(k)$ having a Dirichlet distribution $Dir_n(\alpha_1,\dots,\alpha_k)$. Let y be a $n\times m$, $m \le n$, non-random matrix such that $y^T y = I_m$. Then*

$$y^T\mathbf{P}y \sim Dir_m(\alpha_1,\dots,\alpha_k)\,. \tag{5}$$

The same holds if y is random and independent of $\mathbf{P}$.

We divide the proof into two lemmas.

Lemma 4 *Let $b \in \mathcal{M}_{n\times m}$ be a non-random matrix such that $x \in \mathcal{M}_{m\times 1}$ and $bx = 0$ implies $x = 0$. Let $W_i \overset{ind}{\sim} \gamma^{(n)}_{\alpha_i,a}$, $i = 1,\dots,k$ be independent Wishart random matrices such that $\sum_{i=1}^k \alpha_i \in (\frac{n-1}{2},\infty)$, then the vector of random matrices*

$$\Phi(b) = (b^TWb)^{-\frac{1}{2}}b^T(W_1,\dots,W_k)b(b^TWb)^{-\frac{1}{2}}$$

and $W = \sum_{i=1}^k W_i$ are independent and $\Phi(b) \sim Dir_m(\alpha_1,\dots,\alpha_k)$.

<u>Proof of lemma 4</u>
The hypothesis on b implies $b^TW_ib \in \mathcal{V}_m^+$ so that $\Phi(b)$ is a random vector with k components $\in \mathcal{V}_m^+$.
As $b^TW_ib \overset{ind}{\sim} \gamma^{(m)}_{(\alpha_i,b^Tab)}$, it is seen by Theorem 2 that $\Phi(b) \sim Dir_m(\alpha_1,\dots,\alpha_k)$ and that for any Borel subset A of $(\mathcal{V}_m^+)^k$

$$\mathcal{P}(\Phi(b)\in A) \text{ does not depend on parameter } a. \tag{6}$$

Let

$$D = W^{-\frac{1}{2}}(W_1,\dots,W_k)W^{-\frac{1}{2}}.$$

Observing that

$$\Phi(b) = (b^TWb)^{-\frac{1}{2}}b^TW^{\frac{1}{2}}DW^{\frac{1}{2}}b(b^TWb)^{-\frac{1}{2}},$$

write $\Phi(b)$ as a function of D and W, say $\Phi(b) = f(D,W)$. Then,

$$\mathcal{P}(\Phi(b) \in A|W = w) = \mathcal{P}(f(D,W) \in A|W = w) = \int_{d\in\mathcal{Q}_n(k)} I_{\{d:f(d,w)\in A\}}\mathcal{P}_{D|W=w}(d).$$

As D and W are independent by Theorem 2, we have

$$\mathcal{P}(\Phi(b) \in A|W = w) = \int_{d\in\mathcal{Q}_n(k)} I_{\{d:f(d,w)\in A\}}\mathcal{P}_D(d). \tag{7}$$

Again due to Theorem 2, $\mathcal{P}_D$ does not depend on a, so that (7) yields

$$\mathcal{P}(\Phi(b) \in A|W = w) \text{ does not depend on parameter } a. \tag{8}$$

Now, writing equality

$$\mathcal{P}(\Phi(b) \in A) = \int_{w\in\mathcal{V}_n^+} \mathcal{P}(\Phi(b) \in A|W = w)\mathcal{P}_W(w)$$

as

$$\int_{w\in\mathcal{V}_n^+} [\mathcal{P}(\Phi(b) \in A) - \mathcal{P}(\Phi(b) \in A|W = w)]\mathcal{P}_W(w) = 0$$

and observing that W has a Wishart density proportional to $e^{-(w,a)}dw$ because of the condition$\sum_{i=1}^k \alpha_i \in (\frac{n-1}{2}, \infty)$, we arrive at

$$\int_{w\in\mathcal{V}_n^+} [\mathcal{P}(\Phi(b) \in A) - \mathcal{P}(\Phi(b) \in A|W = w)]e^{-(w,a)}dw = 0. \tag{9}$$

By (6) and (7), it is seen that the expression within brackets in (9) does not depend on a. Further, if we replace all the Wishart's $W_i(\alpha_i, a)$ with independent $W_i(\alpha_i, a')$, the same arguments used above show that equality (9) still holds for a' instead of a. Thus (9) holds for any $a \in \mathcal{V}_n^+$. As for the Laplace transform, this implies that

$$\mathcal{P}(\Phi(b) \in A) = \mathcal{P}(\Phi(b) \in A|W = w) \text{ for a.a. } w.$$

In other words $\Phi(b)$ and W are independent. □

Lemma 5 *Let $B : \mathcal{V}_n^+ \longrightarrow \mathcal{M}_{n\times m}$ be a measurable mapping such that the conditions $v \in \mathcal{V}_n^+$, $x \in \mathcal{M}_{m\times 1}$ and $B(v)x = 0$ imply $x = 0$.*
Let $W_i, \alpha_i\ (i = 1, \ldots, k), W$ and Φ be as in lemma 4.
Then, the vector of random matrices $\Phi(B(W))$ is $Dir_m(\alpha_1, \ldots, \alpha_k)$.

Proof of lemma 5
For any Borel subset A of $(\mathcal{V}_m^+)^k$, we have

$$\mathcal{P}(\Phi(B(W)) \in A) = \int_{w\in\mathcal{V}_n^+} \mathcal{P}(\Phi(B(w)) \in A|W = w)\mathcal{P}_W(w)$$

Applying lemma 4 with $b = B(w)$, it is then seen that

$$\begin{aligned}\mathcal{P}(\Phi(B(W)) \in A) &= \int_{w\in\mathcal{V}_n^+} \mathcal{P}(\Phi(B(w)) \in A))\mathcal{P}_W(w) \\ &= \int_{w\in\mathcal{V}_n^+} Dir_m(\alpha_1,\ldots,\alpha_k)(A)\mathcal{P}_W(w) \\ &= Dir_m(\alpha_1,\ldots,\alpha_k)(A). \qquad \square\end{aligned}$$

Proof of Theorem 3

Let y be non-random as in the statement of theorem 3. Take $B(v) = v^{-\frac{1}{2}}y$ for any $v \in \mathcal{V}_n^+$. Then $x \in \mathcal{M}_{m\times 1}$ and $B(v)x = 0$ means that $v^{-\frac{1}{2}}yx = 0$. This straightforward implies that $v^{\frac{1}{2}}v^{-\frac{1}{2}}yx = 0$, $yx = 0$, $y^T yx = 0$ and $x = 0$ so that B satifies the requirement of lemma 4.

Then $B(W) = W^{-\frac{1}{2}}y$ implies that

$$(B(W)^T W B(W))^{-\frac{1}{2}} = (y^T W^{-\frac{1}{2}} W W^{-\frac{1}{2}} y)^{-\frac{1}{2}} = 1.$$

Therefore

$$\Phi(B(W)) = (B(W)^T W B(W))^{-\frac{1}{2}} B(W)^T (W_1,\ldots,W_k) B(W) (B(W)^T W B(W))^{-\frac{1}{2}}$$

reduces to

$$\Phi(B(W)) = B(W)^T (W_1,\ldots,W_k) B(W) = y^T W^{-\frac{1}{2}} (W_1,\ldots,W_k) W^{-\frac{1}{2}} y.$$

Since $\Phi(B(W))$ is a Dirichlet vector by lemma 4 and the same for $W^{-\frac{1}{2}}(W_1,\ldots,W_k)W^{-\frac{1}{2}} \stackrel{d}{=} \mathbf{P}$ by definition, we get for any deterministic y such that $y^T y = \mathrm{Id}_m$

$$y^T \mathbf{P} y \sim Dir_m(\alpha_1,\ldots,\alpha_k). \tag{10}$$

Next, suppose that y is a random matrix such that $y^T y = \mathrm{Id}_m$ and suppose that y and P are independent. Let $W_i \stackrel{ind}{\sim} \gamma^{(m)}_{(\alpha_i,a)}, i = 1,\ldots,k$ be independent Wishart's independent of y. Then (10) implies that the conditional distribution of $y^T W^{-\frac{1}{2}}(W_1,\ldots,W_k)W^{-\frac{1}{2}}y$ given y is $Dir_m(\alpha_1,\ldots,\alpha_k)$ because $W^{-\frac{1}{2}}(W_1,\ldots,W_k)W^{-\frac{1}{2}}$ and y are independent. Integrating out w.r.t. the distribution of y we find that

$$y^T W^{-\frac{1}{2}}(W_1,\ldots,W_k)W^{-\frac{1}{2}}y \sim Dir_m(\alpha_1,\ldots,\alpha_k). \tag{11}$$

Morevoer, as y and $W^{-\frac{1}{2}}(W_1,\ldots,W_k)W^{-\frac{1}{2}}$ are independent and as y and $\mathbf{P}$ are independent by hypothesis and finally as $W^{-\frac{1}{2}}(W_1,\ldots,W_k)W^{-\frac{1}{2}} \stackrel{d}{=} \mathbf{P}$ by definition, we have $(y, W^{-\frac{1}{2}}(W_1,\ldots,W_k)W^{-\frac{1}{2}}) \stackrel{d}{=} (y,\mathbf{P})$. Thus $y^T W^{-\frac{1}{2}}(W_1,\ldots,W_k)W^{-\frac{1}{2}}y \stackrel{d}{=} y^T\mathbf{P}y$ and (11) implies $y^T\mathbf{P}y \sim Dir_m(\alpha_1,\ldots,\alpha_k)$.

$\square$

In dimension one ($n = m = 1$), it is well known that if $\mathbf{P}^{(j)} \overset{ind}{\sim} Dir_n\left(\alpha_1^{(j)}, \ldots, \alpha_k^{(j)}\right)$, $j = 1, 2$ are two independent Dirichlet vectors and if $Q \sim Beta(s^{(1)}, s^{(2)})$, where $s^{(j)} = \sum_{i=1}^k \alpha_i^{(j)}$, is independent of $(\mathbf{P}^{(1)}, \mathbf{P}^{(2)})$, then $Q\mathbf{P}^{(1)} + (1-Q)\mathbf{P}^{(2)} \sim Dir_n\left(s^{(1)}, s^{(1)}\right)$ (see e.g. [6], Section 7). This result which was used for example in the Sethuraman stick-breaking constructive definition of a Dirichlet process ([5] lemma 3.1), is generalized to the the matrix case by the following theorem where square root of a matrix $\in \mathcal{V}_n^+$ is taken in the sense of (3):

Theorem 6 *Let* $\mathbf{P}^{(j)} = (P_1^{(j)}, \ldots, P_k^{(j)}) \overset{ind}{\sim} Dir_n(\alpha_1^{(j)}, \ldots, \alpha_k^{(j)}), j = 1, \ldots, J$ *be independent matrix Dirichlet random vectors such that* $\alpha_i^{(j)} \in \Lambda_n$, $i = 1, \ldots, k$, *and* $s^{(j)} = \sum_{i=1}^k \alpha_i^{(j)} > \frac{n-1}{2}$.
Let $Q = (Q_1, \ldots, Q_J) \sim Dir_n(s^{(1)}, \ldots, s^{(J)})$ *be a real r.v. independent of the* $\mathbf{P}^{(j)}$*'s. Then*

$$\left(Q_1^{\frac{1}{2}}\mathbf{P}^{(1)}(Q_1^{\frac{1}{2}})^T, \ldots, Q_J^{\frac{1}{2}}\mathbf{P}^{(J)}(Q_J^{\frac{1}{2}})^T\right) \sim Dir_n(\alpha_1^{(1)}, \ldots, \alpha_k^{(1)}, \ldots, \alpha_1^{(J)}, \ldots, \alpha_k^{(J)})$$

and

$$Q_1^{\frac{1}{2}}\mathbf{P}^{(1)}(Q_1^{\frac{1}{2}})^T + \ldots + Q_J^{\frac{1}{2}}\mathbf{P}^{(J)}(Q_J^{\frac{1}{2}})^T \sim Dir_n(\alpha_1^{(1)} + \ldots + \alpha_1^{(J)}, \ldots, \alpha_k^{(1)} + \ldots + \alpha_k^{(J)}).$$

Proof. As the definition of Dirichlet distribution doest not depend on the division algorithm (again by Theorem 2), we will use here the division agorithm derived from Choleski decomposition as mentioned in (2) and (3).
First, using (2) and the independence hypothesis it is seen that there exists

$$W_i^{(j)} \overset{ind}{\sim} \gamma_{\alpha_i^{(j)}, a}^{(n)}, i = 1, \ldots, k \text{ and } j = 1, \ldots, J$$

such that

$$\mathbf{P}^{(j)} = W_{(j)}^{-\frac{1}{2}}(W_1^{(j)}, \ldots, W_k^{(j)})(W_{(j)}^{-\frac{1}{2}})^T \tag{12}$$

where $W_{(j)} = \sum_{i=1}^k W_i^{(j)}$.
Moreover, as the sum over i of independent Wishart's $\gamma_{\alpha_i^{(j)}, a}^{(n)}$ is a Wishart $\gamma_{s^{(j)}, a}^{(n)}$, it can be assumed that

$$Q_j = W^{-\frac{1}{2}} W_{(j)} (W^{-\frac{1}{2}})^T \tag{13}$$

where $W = \sum_{j=1}^K W_{(j)}$.
The point is that, since $W_{(j)} = W_{(j)}^{\frac{1}{2}}(W_{(j)}^{\frac{1}{2}})^T$ by Cholesky decomposition, (13)

can be written as

$$Q_j = W^{-\frac{1}{2}} W_{(j)}^{\frac{1}{2}} (W_{(j)}^{\frac{1}{2}})^T (W^{-\frac{1}{2}})^T = W^{-\frac{1}{2}} W_{(j)}^{\frac{1}{2}} (W^{-\frac{1}{2}} W_{(j)}^{\frac{1}{2}})^T \tag{14}$$

which is the Cholesky decomposition of Q_j, so that we get

$$Q_j^{\frac{1}{2}} = W^{-\frac{1}{2}} W_{(j)}^{\frac{1}{2}}. \tag{15}$$

Then (12) and (15) yield

$$Q_j^{\frac{1}{2}} \mathbf{P}^{(j)} (Q_j^{\frac{1}{2}})^T = W^{-\frac{1}{2}} W_{(j)}^{\frac{1}{2}} W_{(j)}^{-\frac{1}{2}} (W_1^{(j)}, \ldots, W_k^{(j)}) (W_{(j)}^{-\frac{1}{2}})^T (W^{-\frac{1}{2}} W_{(j)}^{\frac{1}{2}})^T$$

and

$$Q_j^{\frac{1}{2}} \mathbf{P}^{(j)} (Q_j^{\frac{1}{2}})^T = W^{-\frac{1}{2}} (W_1^{(j)}, \ldots, W_k^{(j)}) (W^{-\frac{1}{2}})^T. \tag{16}$$

Writing (16) for $j = 1, \ldots, J$, it is seen that $\left(Q_1^{\frac{1}{2}} \mathbf{P}^{(1)} (Q_1^{\frac{1}{2}})^T, \ldots, Q_J^{\frac{1}{2}} \mathbf{P}^{(J)} (Q_J^{\frac{1}{2}})^T \right)$

$$= W^{-\frac{1}{2}} \left(W_1^{(1)}, \ldots, W_k^{(1)}, \ldots, W_1^{(J)}, \ldots, W_k^{(J)} \right) (W^{-\frac{1}{2}})^T \tag{17}$$

and thus has, when using division algorithm (3), the announced Dirichlet distribution since $W = \sum_{j=1}^{J} \sum_{i=1}^{K} W_i^{(j)}$.
Finally, representation (17) straightforward yields

$$Q_1^{\frac{1}{2}} \mathbf{P}^{(1)} (Q_1^{\frac{1}{2}})^T + \ldots + Q_J^{\frac{1}{2}} \mathbf{P}^{(J)} (Q_J^{\frac{1}{2}})^T \sim Dir_n(\alpha_1^{(1)} + \ldots + \alpha_1^{(J)}, \ldots, \alpha_k^{(1)} + \ldots + \alpha_k^{(J)})$$

since the sum over j of independent Wishart's $\gamma^{(n)}_{\alpha_i^{(j)}, a}$ is Wishart $\gamma^{(n)}_{\alpha_i^{(1)} + \ldots + \alpha_i^{(J)}, a}$.

□

5 Dirichlet processes, for $n \geq 2$, reduce to Dirichlet vectors

Let $(S, \mathcal{S})$ be a measurable space. We define a set of '**definite positive matrix-variate probability measures**' on S as

$$\begin{aligned} \mathcal{Q}(S) \;=\; &\{p : \mathcal{S} \to \overline{\mathcal{V}}_n^+, \text{ such that} \\ p(S) = \mathrm{Id}_n \text{ and } &p(\cup_j S_j) = \Sigma_j\, p(S_j) \text{ for any pair-wise disjoint } S_j \in \mathcal{S}\} \,. \end{aligned}$$

Let $\widetilde{\mathcal{Q}}(S)$ be a Borel σ-field of subsets of $\mathcal{Q}(S)$.
Let $\mathcal{D}$ denote the family of all finite measurable partitions of S.
A generalization to the matrix variate case of Dirichlet processes as defined by Ferguson [2] could then be defined as follows:
Let α be a finite nonegative measure defined on $\widetilde{\mathcal{Q}}(S)$. A Dirichlet process $\mathbf{P}$ with basis α is a r.v. from Ω to $\mathcal{Q}(S)$ such that for any partition

$(B_1, \ldots, B_k) \in \mathcal{D}$, the matrix-variate random vector $(\mathbf{P}(B_1), \ldots, \mathbf{P}(B_k)) \sim Dir_n(\alpha(B_1), \ldots, \alpha(B_k))$.
However to be consistent with Dirichlet vectors definition the preceding definition requires that for any measurable $B \in \widetilde{\mathcal{Q}}(S)$

$$\alpha(B) \in \Lambda_n = \left\{0, \frac{1}{2}, \frac{2}{2}, \ldots, \frac{n-1}{2}\right\} \cup \left(\frac{n-1}{2}, \infty\right). \tag{18}$$

In particular if $n \geq 2$, since $\alpha(\mathcal{Q}(S)) < \infty$, the measure α has a finite number of (disjoint) atoms, say $A_1, \ldots, A_k$, having positive measure $\alpha_1, \ldots, \alpha_k \in \{\frac{1}{2}, \frac{2}{2}, \ldots, \frac{n-1}{2}\} \cup \left(\frac{n-1}{2}, \infty\right)$, respectively.
Hence, if $\mathbf{P}_i = \mathbf{P}(A_i)$, $\mathbf{P}$ can be identified to a matrix variate Dirichlet vector $(\mathbf{P}_1, \ldots, \mathbf{P}_k) \sim Dir_n(\alpha_1, \ldots, \alpha_k)$ with

$$\mathbf{P}(B) = \sum_{i=1,\ldots,k: A_i \subseteq B} \mathbf{P}_i.$$

Actually, when examining Kingman's construction [4], it is seen that the deep reason for which Dirichlet process construction cannot be extended in dimension $n \geq 2$ is that the Whishart distributions are not indifinitely divisible while Gamma ones are.

6 Matrix variate Bayesian setting

Let $P = (P_1, \ldots, P_k) \sim Dir_n(\alpha_1, \ldots, \alpha_k)$ be a matrix variate Dirichlet vector. Applying Theorem 3 with $y^T = (0, \ldots, 1, \ldots, 0)$ where 1 is at the i-th position, we see that the vector of diagonal terms $P(i,i) = (P_1(i,i), \ldots, P_k(i,i)) \sim Dir_1(\alpha_1, \ldots, \alpha_k)$ is a one-dimensional Dirichlet, that is a random (classical) probability vector. Let X_i be a random variable assuming values in $\{1, \ldots, k\}$ such that

$$\mathcal{P}(X_i = j|P) = P_j(i,i).$$

Then, Bayes formula can be stated as follows:

Proposition 7 *Let $P = (P_1, \ldots, P_k) \sim Dir_n(\alpha_1, \ldots, \alpha_k)$ be a Dirichlet vector of random matrices, let X_i be a r.v. taking its values in $\{1, \ldots, k\}$ and such that $\mathcal{P}(X_i = j|P) = P_j(i,i)$, then $(P|X_i = j) \propto P_j(i,i) \cdot Dir(\alpha_1, \ldots, \alpha_k)$.*

In dimension one, since $P_j(i,i) = det(P_j)$ is a (normalization) constant, this statement is nothing but the well-known property that the posterior of a Dirichlet is a Dirichlet. In dimension $n > 1$ we see that this posterior need not be a Dirichlet.
Notice finally that we have a similar Bayes formula using all the diagonal

terms for a r.v. X such that

$$\mathcal{P}(X = j|P) = \frac{1}{n}\sum_{i=1}^{n} P_j(i,i).$$

References

[1] Casalis, M. and Letac, G. (1996). The Lukas-Olkin-Rubin characterization of Wishart distributions on symmetric cones. *Ann. Statist.*, Vol. 24, No. 2, 763-786.

[2] Ferguson, T. S (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

[3] Gupta, A. K. and Nagar, D. K. (1973). *Matrix variate distributions.* Chapman Hall/CRC, London.

[4] Kingman, J. F. C. (1975). Random discrete distributions. *J. Roy. Statist. Soc. B*, **37,** 1-22.

[5] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4** 639-650

[6] Wilks, S. S. (1962). *Mathematical Statistics.* John wiley, New York.