A conjugate prior for the Dirichlet distribution

Jean-Marc Andreoli
Naverlabs Europe

October 2014; last update October 2018

Abstract

This note investigates a conjugate class for the Dirichlet distribution in the exponential family.

1 Basic definitions

The exponential family of distributions is characterised by the existence of a systematic procedure to produce priors within that same family (see, e.g., Proposition 3.3.13 in [3]). For example, the class of Dirichlet distributions can be obtained by application of that procedure to the class of multinomial distributions, an essential class in the exponential family (see, e.g., Theorem 4.3 in [4], and the application of Dirichlet priors to document clustering in [1]). Now, applying the same procedure to the class of Dirichlet distributions itself, we obtain a new class of distributions, still in the exponential family, which we denote here as Boojum for lack of a better name.

Definition 1. The Boojum \((m, r)\) distribution, with parameter \(m \in \mathbb{R}\) (shape) and \(r \in \mathbb{R}^K\) (rate vector), is defined for \(x \in \mathbb{R}^K_+\) by

\[
\text{Boojum}(x; m, r) \triangleq \frac{1}{Z(m, r)} B(x)^{-m} \exp \sum_k r_k x_k
\]

where \(B\) denotes the multi-variate Beta function

\[
B(x) \triangleq \prod_k \Gamma(x_k) / \Gamma(\sum_k x_k),
\]

which is also the normalising constant of the Dirichlet distribution of parameter \(x\).

Alternatively, the distribution can be defined over \(\mathbb{R}^+ \times T_K\) where \(T_K \triangleq \{ t \in \mathbb{R}^K_+ | \sum_{k \in K} t_k = 1 \}\) is the \(\mathbb{R}^K_+\)-simplex. We define the homeomorphism from \(x \in \mathbb{R}^K_+\) into \(s, t \in \mathbb{R}^+ \times T_K\) as follows:

\[
direct: \quad s = \sum_k x_k \quad \forall k \quad t_k = \frac{x_k}{s}
\]

reverse: \(s = \sum_k x_k \quad \forall k \quad x_k = st_k\)

The computation of the Jacobian yields \(dx = s^{K-1} ds dt\), where \(dt\) denotes the measure on \(T_K\) obtained by projection of the simplex along any arbitrarily chosen axis. Hence the distribution in the alternate space \(\mathbb{R}^+ \times T_K\) is given by

\[
p(t|s) = \frac{1}{Z(s)} B(st)^{-m} \exp -s(r, t)
\]

\[
\bar{Z}(s) \triangleq \int_{t \in T} B(st)^{-m} \exp -s(r, t) dt
\]

\[
p(s) = \frac{1}{Z(m, r)} \bar{Z}(s) s^{K-1}
\]

\[
Z(m, r) = \int_s \bar{Z}(s) s^{K-1} ds
\]

Here, \(\langle ., . \rangle\) denotes the scalar product in \(\mathbb{R}^K\). To be strict, we should write \(\bar{Z}(s; m, r)\) for the normalising constant of the conditional, but we omit the dependency on \(m, r\) for the sake of clarity.

2 Study of convergence

We seek to determine the values of the parameter \(m, r\) which lead to a proper distribution, i.e. where the normalising constant \(Z(m, r)\) is finite. The main result, expressed by the following theorem, is informally summarised in Figure [1] when \(K=2\).
Theorem 1. Let \( m \in \mathbb{R} \) and \( r \in \mathbb{R}^K \). The distribution \( \text{Boojum}(m, r) \) is proper, i.e. \( Z(m, r) \) is finite, if and only if

\[
\forall k \in K \ r_k > 0 \quad \text{and} \quad m > -1 \quad \text{and} \quad \left( m \leq 0 \quad \text{or} \quad \sum_k \exp -\frac{r_k}{m} < 1 \right)
\]

The rest of this section is devoted to the proof of this result, which summarises Lemmas \( 2 \), \( 3 \), \( 5 \), \( 6 \), \( 7 \).

Lemma 1. For any given \( s > 0 \), the integral \( \bar{Z}(s) \) is finite if and only if \( m > -1 \).

Proof. For a given \( s \), we have, using the definition of \( \bar{Z} \) and the property \( \Gamma(x) = \frac{1}{x} \Gamma(1 + x) \)

\[
\bar{Z}(s) = \int_{t \in T} \left( \prod_k \frac{\Gamma(st_k)}{\Gamma(s)} \right)^{-m} \exp -s(r(t), t) \, dt = \int_{t \in T} \prod_k t_k^m \left( \prod_k \frac{\Gamma(1 + st_k)}{s^K \Gamma(s)} \right)^{-m} \exp -s(r(t), t) \, dt
\]

The underlined term is a continuous function of \( t \) over the compact \( T \), hence both lower and upper bounded. Hence, for some strictly positive constants \( u_s^+, u_s^- \) which depend on \( s \), we have

\[
u_s^+ \int_{t \in T} \prod_k t_k^m \, dt \leq \bar{Z}(s) \leq u_s^- \int_{t \in T} \prod_k t_k^m \, dt
\]

Hence, \( \bar{Z}(s) \) is finite if and only if so is \( \int_{t \in T} \prod_k t_k^m \, dt \). The latter is the normalising constant of the balanced Dirichlet distribution with parameter \( m + 1 \), which is proper if and only if \( m + 1 > 0 \).

\( \square \)

2.1 The case \( m \leq -1 \) or \( r \neq 0 \)

Lemma 2. If \( m \leq -1 \) or \( r \neq 0 \), then \( Z(m, r) \) is infinite.

Proof. When \( m \leq -1 \), consider \( Z(m, r) \) as an integral on space \( \mathbb{R}_+ \times T_K \). By Lemma 1, the integrand \( \bar{Z}(s) \) is infinite for all \( s \), hence, obviously, so is the integral \( Z(m, r) \). Now, let’s assume \( m > -1 \) and \( r \neq 0 \), i.e. \( r_k \leq 0 \) for some \( k \in K \). Consider \( Z(m, r) \) as an integral on space \( \mathbb{R}^K_+ \). We have, isolating \( x_k \) in the integrand,

\[
Z(m, r) = \int_{x \in \mathbb{R}^K_+ \setminus \{x\}} F(\sum_{h \neq k} x_h) \prod_{h \neq k} \Gamma(x_h)^{-m} \exp -r_h x_h \, dx_h \quad \text{where} \quad F(z) \triangleq \int_{x \in \mathbb{R}_+} \left( \frac{\Gamma(x)}{\Gamma(z + x)} \right)^{-m} \exp -r_x x \, dx
\]

Let \( \Delta \triangleq (0, 1] \) if \( m \leq 0 \) and \( \Delta \triangleq [1, \infty) \) if \( 0 < m \). Recall that \( \Gamma \) is increasing on \( [2, \infty) \). Hence, for any \( z \in \Delta \) and \( x \geq 2 \):

- If \( m \leq 0 \) then \( z \leq 1 \) hence \( \Gamma(z + x) \leq \Gamma(1 + x) = x \Gamma(x) \) hence \( \left( \frac{\Gamma(x)}{\Gamma(z + x)} \right)^{-m} \geq x^m \)

- If \( 0 < m \) then \( z \geq 1 \) hence \( \Gamma(z + x) \geq \Gamma(1 + x) = x \Gamma(x) \) hence \( \left( \frac{\Gamma(x)}{\Gamma(z + x)} \right)^{-m} \geq x^m \)

In both cases, by integration, we have for any \( z \in \Delta \)

\[
F(z) \geq \int_{x \geq 2} x^m \exp -r_x x \, dx \geq \int_{x \geq 2} x^m \, dx = \infty
\]

Hence \( F(\sum_{h \neq k} x_h) \) is infinite on a non null subset of \( \mathbb{R}^K_+ \setminus \{k\} \), hence \( Z(m, r) \) is infinite. \( \square \)
For a fixed $r$ let $m > 0$.

When $r > 0$, we get that $F$ and, since $Z$ is increasing, $F_{s} = Z_{s}$ is also increasing in $s$

Figure 2: Left: curve of the log $\Gamma$ function and its two approximations. Right: detail of the approximation gaps.

2.2 The case $r > 0$ and $-1 < m$

We now assume $r > 0$ and $-1 < m$. One case can easily be treated:

Lemma 3. If $r > 0$ and $m = 0$ then $\text{Boojum}(0,r)$ is a multivariate exponential distribution and $Z(0,r) = \prod_{k} r_{k}^{-1}$.

Proof. Simply observe that $\text{Boojum}(0,r)$ is the product of independent exponentials, each with rate $r_{k}$ for $k=1:K$. \hfill $\square$

When $m \neq 0$, the finiteness of $Z(m,r)$ depends on the behaviour of $\bar{Z}$ (which is finite by Lemma 1) near $0$ and near $\infty$. One side (depending on the sign of $m$) can easily be treated.

Lemma 4. If $r > 0$ and $-1 < m \neq 0$, then $Z_{\Delta}(m,r)$ is finite, where

$$Z_{\Delta}(m,r) \triangleq \int_{s \in \Delta} Z(s)s^{K-1}ds \quad \text{and} \quad \Delta \triangleq \begin{cases} [1, \infty) & \text{if } m < 0 \\ (0, 1] & \text{if } m > 0 \end{cases}$$

Hence, $Z(m,r)$ is finite if and only if so is $Z_{\mathbb{R}+\Delta}$, since $Z(m,r) = Z_{\Delta} + Z_{\mathbb{R}+\Delta}$.

Proof. For a fixed $t \in \mathcal{T}$, by derivation, we have

$$\frac{d\log \mathcal{B}(st)}{ds} = \frac{d}{ds} \left( \sum_{k} \log \Gamma(st_{k}) - \log \Gamma(s) \right) = \sum_{k} t_{k} \Psi(st_{k}) - \Psi(s) s \leq \sum_{k} t_{k} \Psi(s) - \Psi(s) = 0$$

Here $\Psi$ is the derivative of the log $\Gamma$ function (a.k.a. digamma function), which is increasing \cite{2}, hence $\Psi(st_{k}) < \Psi(s)$. Therefore, log $\mathcal{B}(st)$ is a decreasing function of $s$, hence so is $\mathcal{B}(st)$. Hence $\mathcal{B}(st)^{-m}$ is $m$-increasing in $s$ (meaning increasing when $m > 0$ and decreasing when $m < 0$).

Observe that $\exp -s(r-r^{*},t)$ is also $m$-increasing in $s$ (or constant if all the components of $r$ are equal), hence $\mathcal{B}(st)^{-m} \exp -s(r-r^{*},t)$ is $m$-increasing in $s$. By integration over $t$, we get that $F(s) \triangleq Z(s; m, r-r^{*})$, which is finite by Lemma 3 is $m$-increasing in $s$.

Observe that $\exp -s(r-r^{*},t) = \exp sr^{*} \exp -s(r,t)$, since $\sum_{k} t_{k} = 1$. Hence $F(s) = \bar{Z}(s) \exp sr^{*}$ and

$$Z_{\Delta} = \int_{s \in \Delta} F(s)s^{K-1} \exp -r^{*}sds$$

and, since $F$ is $m$-increasing, we have (the split value 1 is arbitrary)

- when $m < 0$ ($F$ is decreasing): $Z_{[1, \infty)} = \int_{s \geq 1}^{\infty} F(st)s^{K-1} \exp -r^{*}sds \leq F(1) \int_{s \geq 1} \exp -r^{*}sds \leq \infty$
- when $m > 0$ ($F$ is increasing): $Z_{(0, 1]} = \int_{s \leq 1} F(st)s^{K-1} \exp -r^{*}sds \leq F(1) \int_{s \leq 1} \exp -r^{*}sds \leq \infty$

Hence, $Z(m,r)$ is finite if and only if so is $Z_{\Delta}(m,r)$ where $\Delta = (0, 1]$ when $m < 0$ and $\Delta = [1, \infty)$ when $m > 0$. \hfill $\square$

Thus, we need only study the behaviour of $\bar{Z}$ near $0$ when $-1 < m < 0$ and near $\infty$ when $m > 0$. For that, we look for an approximation of $\log \mathcal{B}(st)$, which we obtain from an approximation of the log $\Gamma$ function, actually one near $0$ and a different one near $\infty$, as illustrated in Figure 3. Expanding the definition of $\bar{Z}$ in Equation 1 gives:

$$Z_{\Delta} = \int_{s \in \Delta} s^{K-1} \int_{t \in \mathcal{T}} \exp(-m \log \mathcal{B}(st) - s(r,t))dtds \quad \text{Equation 2}$$
2.3 The case \( r > 0 \) and \(-1 < m < 0\)

**Lemma 5.** If \( r > 0 \) and \(-1 < m < 0\), then \( Z(m, r) \) is finite.

**Proof.** Define \( h(z) \triangleq \log \Gamma(z) + \log z \). Thus we have

\[
\log \Gamma(s) = -s + s + h(s) \\
\sum_k \log \Gamma(st_k) = \sum_k -s + st_k + h(st_k)
\] \( \Rightarrow \) \( -\log B(st) = (K - 1) \log s + \sum_k \log t_k + h(s) - \sum_k h(st_k) \)

Hence, reporting in Equation (2), we get

\[
Z_{[0,1]} = \int_{s \leq 1} s^{K-1} \int_{t \in T} \exp \left( m((K-1) \log s + \sum_k \log t_k + H(s, t)) - s<r,t> \right) dt ds \\
= \int_{s \leq 1} s^{(1+m)(K-1)-1} \int_{t \in T} \prod_k t_k^m \exp(-mH(s, t) - s<r,t>) dt ds
\]

By construction, \( h \) is bounded on \((0, 1]\), hence \( H(s, t) \) is bounded for \( s, t \in (0, 1] \times T \), and so is \( s<r,t> \). Hence, we have for some strictly positive constant \( U \)

\[
Z_{[0,1]} \leq U \int_{s \leq 1} s^{(1+m)(K-1)-1} \int_{t \in T} \prod_k t_k^m dt ds = UB(1 + m) \int_{s \leq 1} s^{(1+m)(K-1)} ds < \infty
\]

Hence \( Z_{[0,1]} \) is finite, hence so is \( Z(m, r) \) by Lemma 4. \( \square \)

2.4 The case \( r > 0 \) and \( 0 < m \)

**Lemma 6.** If \( r > 0 \) and \( 0 < m \), then \( Z(m, r) \) is finite if \( T < 1 \) and infinite if \( T > 1 \), where \( T \triangleq \sum_k \exp -\frac{t_k}{m} \).

**Proof.** Define \( h(z) \triangleq \log \Gamma(z) + z - (z - \frac{1}{2}) \log z \). Thus, \( h \) denotes the difference between \( \log \Gamma \) and its Stirling’s approximation (up to a constant). We have, for any \( t \in T \), using \( \sum_k t_k = 1 \),

\[
\log \Gamma(s) = -s + s + h(s) \\
\sum_k \log \Gamma(st_k) = \sum_k -st_k + (st_k - \frac{1}{2}) \log st_k + h(st_k)
\] \( \Rightarrow \)

\[
-\log B(st) = \frac{K-1}{2} \log s - s \sum_k t_k \log t_k + h(s) - \sum_k h(st_k) + \frac{1}{2} \sum_k \log t_k
\]

Observe that the terms in \( s \log s \) cancelled out. Hence, reporting in Equation (2), we get

\[
Z_{[1,\infty)} = \int_{s \geq 1} s^{K-1} \int_{t \in T} \exp \left( m((K-1) \log s - s \log J(t)) + H(s, t) - s<r,t> \right) dt ds \\
= \int_{s \geq 1} s^{(1+m)(K-1)-1} \int_{t \in T} \exp m(-s(J(t) + \frac{r,t}{m}) + H(s, t)) dt ds
\]

We first seek bounds for the term \( H(s, t) \). Function \( h \) is decreasing and lower bounded by \( \log \sqrt{2\pi} \), hence positive. Therefore, for any \( s \geq 1 \):

- For any \( t \in T \), since \( h(st_k) > 0 \) and \( t_k \leq 1 \), hence \( \log t_k \leq 0 \), we have \( H(s, t) < h(s) \leq h(1) \).
- On the other hand, \( h \) is not upper bounded near \( 0 \), hence for any given \( s \), the function \( H(s, \cdot) \) is not lower bounded near the border of \( T \). However, choose \( \tau \in T \) and for any \( 0 \leq \alpha \leq 1 \) consider the sub-domain of \( T \) defined by

\[
T_\alpha \triangleq \{ t \in T | \forall k \in K \ t_k \geq \alpha \tau_k \}
\]

This compact convex set, illustrated in Figure 3, contains \( \tau \) and has a non null measure in \( T \), whenever \( \alpha < 1 \) (for \( \alpha = 1 \) it degenerates into the singleton \( \{ \tau \} \)). For \( \alpha = 0 \) we have \( T_0 = T \) over which \( H(s, \cdot) \) is not lower bounded, but whenever \( \alpha > 0 \) we get a lower bound over \( T_\alpha \), since \(-h \) and \( \log \) are increasing and \( \forall t \in T_\alpha , t \geq \alpha \tau \):

\[
H(s, t) \geq h(s) - \sum_k h(s\alpha \tau_k) + \frac{1}{2} \sum_k \log(\alpha \tau_k) \geq - \sum_k h(\alpha \tau_k) + \frac{1}{2} \sum_k \log(\alpha \tau_k)
\]
If \( \text{Lemma 7.} \)

Figure 3: An illustration of the sub-domain \( T_\alpha \) of the simplex \( T \) used for lower bounding \( Z(m, r) \) when \( m > 0 \). The idea is to avoid the border of \( T \) (when \( \alpha > 0 \)) while still being non-null (when \( \alpha < 1 \)).

Neither the upper bound nor the lower bound of \( H(s, t) \) are dependent on \( s, t \), hence, we have for some strictly positive constants \( U_\alpha^+, U_T \), for all \( s, t \in [1, \infty) \times T_\alpha \):

\[
U_\alpha^+ \leq \exp(mH(s, t)) \leq U_T
\]

Let’s introduce the short-hands \( M \triangleq (1 + \frac{m}{2})(K - 1) \) and

\[
F_\alpha(s) \triangleq \int_{t \in T_\alpha} \exp(-ms(J(t) + \frac{(r, t)}{m})dt)
\]

Observe that \( M > 0 \). Using the bounds on \( \exp(mH) \), we get:

\[
U_\alpha^+ \int_{s \geq 1} s^M F_\alpha(s)ds \leq Z_{[1, \infty)} \leq U_T \int_{s \geq 1} s^M F_\alpha(s)ds
\]

We now seek bounds for \( F_\alpha \). We use a specific choice of \( \tau \):

\[
\tau_k \triangleq \frac{1}{T} \exp(-\frac{r_k}{m}) \quad \text{where} \quad T \triangleq \sum_k \exp(-\frac{r_k}{m})
\]

It is then easy to show that, for all \( t \in T \),

\[
J(t) + \frac{(r, t)}{m} = D(t) - \log T \quad \text{where} \quad D(t) \triangleq \sum_k t_k \log \frac{t_k}{\tau_k}
\]

Observe that \( D(t) \) is the information divergence between \( t \) and \( \tau \) viewed as discrete distributions over \( K \). It is a convex function of \( t \) which has minimum 0 reached at \( \tau \) and maximum \( \mu_\alpha \) over \( T_\alpha \), reached at one of its corners. Hence, reporting in Equation \( 3 \) we get, for any \( s \geq 1 \)

\[
A_\alpha \exp ms(\log T - \mu_\alpha) \leq F_\alpha(s) \leq A_\alpha \exp ms \log T \quad \text{where} \quad A_\alpha \triangleq \int_{t \in T_\alpha} dt
\]

Reporting in Equation \( 4 \), we get

\[
A_\alpha U_\alpha^+ \int_{s \geq 1} s^M \exp ms(\log T - \mu_\alpha)ds \leq Z_{[1, \infty)} \leq A_\alpha U_T \int_{s \geq 1} s^M \exp ms \log Tds
\]

Hence \( Z_{[1, \infty)} \) is finite if \( \log T < 0 \), and infinite if \( \log T - \mu_\alpha \geq 0 \) for at least some \( \alpha \), i.e. \( \log T > \inf_\alpha \mu_\alpha \). By Lemma\( 4 \) the same holds for \( Z(m, r) \). Now recall that \( \mu_\alpha = D(t) \) where \( t \) is one of the corners of \( T_\alpha \). When \( \alpha \) tends to 1, all the corners of \( T_\alpha \) tend to \( \tau \), hence \( \mu_\alpha \) tends to \( D(\tau) = 0 \). Hence \( \inf_\alpha \mu_\alpha = 0 \).

**Lemma 7.** If \( r > 0 \) and \( 0 < m \) and \( T = 1 \), then \( Z(m, r) \) is infinite (where \( T \) is defined as in the previous lemma).

**Proof.** Now, \( \log T = 0 \), and, with the notations of the previous lemma, we have, using the change of variable \( z = m\mu_\alpha s \):

\[
Z_{[1, \infty)} \geq A_\alpha U_\alpha^+ \int_{s \geq 1} s^M \exp -m\mu_\alpha sds = A_\alpha U_\alpha^+ \int_{z \geq m\mu_\alpha} z^M \exp -zdzd
\]

\[
(\mu_\alpha)^{M+1}
\]

5
When $\alpha$ tends to 1, $\mu_\alpha$ tends to 0 and the integral in the numerator tends to $\Gamma(M + 1)$, while $U_{\alpha}^r$ tends to $U_1^r > 0$. Hence for some constant $R > 0$ we have for $\alpha$ in a neighbourhood of 1

$$Z_{[1, \infty]} \geq R \frac{A_{\alpha}}{\mu_\alpha^M + 1}.$$  

The $k$-th corner of $T_\alpha$ is given by $t_k = \alpha \tau_k$ for $k \neq k$ and $t_k = \alpha \tau_k + 1 - \alpha$. It is easy to show, using the fact that $\mu_\alpha$ is the maximum of $D$ on the corners of $T_\alpha$, that

$$\mu_\alpha = \max_k \left( \log \alpha + (\alpha \tau_k + 1 - \alpha) \log(1 + \frac{1-\alpha}{\tau_k \alpha}) \right) \sim (1-\alpha)^2 \max_k \frac{1 - \tau_k}{2\tau_k}.$$  

Furthermore, using the definition of $A_\alpha$ and the homeomorphism $t \mapsto \tau + \frac{1-\tau}{\alpha}$ between $T_\alpha$ and $T$, we get $A_\alpha = (1-\alpha)^K - 1 A_0$. Hence for some constant $R' > 0$ we have for $\alpha$ in a neighbourhood of 1

$$Z_{[1, \infty]} \geq R'(1-\alpha)^{(K-1)-2(M+1)} = R'(1-\alpha)^{-(K-1)(1+m)-2}.$$  

When $\alpha$ tends to 1, the right-hand side tends to $\infty$, hence $Z_{[1, \infty]}$ is infinite, hence so is $Z(m, r)$ by Lemma 4.

\[ \square \]

### 3 Characterisation of the distribution

**Theorem 2.** Let $y_{1:N}$ be a finite family of random variables over $T$ and $x$ a random variable over $\mathbb{R}^K$. We have

\[
\begin{align*}
\text{Prior:} & \quad x \sim \text{Boojum}(m, r) \\
\text{Observations:} & \quad \{ y_n | x \sim \text{Dirichlet}(x) \} \overset{N}{=} 1 \\
\text{Independence:} & \quad \{ y_{1:N} | x \} \Rightarrow \text{Posterior:} \quad x | y_{1:N} \sim \text{Boojum}(m + N, r - \sum_{n=1}^{N} \log y_n)
\end{align*}
\]

This holds whenever the prior is proper, in which case so is the posterior.

**Proof.** Simple application of the definitions. Distribution Boojum has been explicitly constructed to be a conjugate prior to the Dirichlet distribution, which is exactly what Theorem 2 states. Obviously, the properness of the prior implies that of the posterior. As a consistency check, we give here a redundant proof of this result, in the case of a single observation $y \in T$, i.e. $N = 1$; the general result ($N$ finite) is simply obtained by reiterating the argument.

Assume the prior Boojum$(m, r)$ is proper. By Theorem 2 we must first have $r > 0$ and $-1 < m$. Hence the posterior Boojum$(m + 1, r - \log y)$ satisfies $r - \log y > 0$ (because $y \in T$ hence $\forall k\ y_k \leq 1$) and $0 < m + 1$. By Theorem 1 we must also have $m \leq 0$ or $\sum_k \exp -\frac{r_k}{m} < 1$:

- If $-1 < m \leq 0$, then $\frac{1}{m + 1} \geq 1$, hence $u^{\frac{1}{m + 1}} \leq u$ for any $0 < u \leq 1$ and we have

$$\sum_k \exp -\frac{r_k - \log y_k}{m + 1} = \sum_k \left( \sum_k \exp -\frac{r_k}{m + 1} \right) \leq \sum_k y_k = 1.$$  

- If $0 < m$ and $\sum_k \exp -\frac{r_k}{m} < 1$, using Hölder’s inequality with $p = m + 1$ and $q = \frac{m + 1}{m}$ we get

$$\sum_k \exp -\frac{r_k - \log y_k}{m + 1} = \sum_k \left( \sum_k \exp -\frac{r_k}{m + 1} \right) \leq \left( \sum_k y_k \right)^{\frac{1}{m+1}} \left( \sum_k \exp -\frac{r_k}{m} \right)^{\frac{1}{m+1}} < 1.$$  

Hence, by Theorem 1 the posterior Boojum$(m + 1, r - \log y)$ is proper.

\[ \square \]

**Theorem 3.** Whenever Boojum$(m, r)$ is a proper distribution, its moment generating function $\phi(v)$ for $v \in \mathbb{R}^K$, and its $n$-th moment $M_n$ for $n \in \mathbb{N}^K$ are given by

$$\phi(v) = \frac{Z(m, r - v)}{Z(m, r)} \quad M_n = (-1)^n \sum_k \frac{\partial^n \phi}{\partial v_k^n} Z(m, r).$$

**Proof.** By definition of the moment generating function, we have

$$\phi(v) = \mathbb{E}_{x \sim \text{Boojum}(m, r)}[\exp(v, x)] = \frac{1}{Z(m, r)} \int B(x)^{-m} \exp(-\langle r, x \rangle + \langle v, x \rangle) dx = \frac{Z(m, r - v)}{Z(m, r)}.$$  

And we have the general result $M_n = \frac{\partial^n \phi}{\partial v_k^n} (0).$  

\[ \square \]

In particular, the expectation of Boojum$(m, r)$ is given by $(M_{|h=k| \in K})_{k \in K}$, hence

$$\mathbb{E}[\text{Boojum}(m, r)] = -\frac{1}{Z(m, r)} \left( \frac{\partial Z}{\partial r_k} (m, r) \right)_{k \in K} = -\nabla_r \log Z(m, r).$$
Now, consider the dilatation of ratio \( \frac{1}{\lambda} \)

For any \( \lambda \),\n
**Lemma 8.**\n
\[
\lambda_K = 1 \text{ for all } K \geq 1.
\]

**Proof.** For any \( N \geq 0 \) and \( K \geq 1 \), let \( q_{NK} \equiv |\mathcal{P}_{NK}| \). The sequences of \( \mathbb{N}^{K+1} \) summing to \( N \) are in bijection with the sequences of \( \mathbb{N}^K \) summing to a number between 0 and \( N \). Hence \( q_{N(K+1)} = \sum_{m=0}^{N} q_{mK} \). And, by definition, \( q_{N1} = 1 \).
Let \( F = \sum_{N \geq 0, K \geq 1} q_{N,K} X^N Y^K \) be the characteristic function of that sequence. We have, using the recurrence,\footnote{Many thanks to Zsolt Papp for pointing out several typos.}

\[
F = \sum_N X^N Y + \sum_{N,K \geq 1} q_{m,K} X^N Y^{K+1} = \sum_N X^N Y + \sum_{m \geq m} q_{m,K} (\sum_{N \geq m} X^N) Y^{K+1}
\]

\[
F = \frac{1}{1 - X} Y + \sum_{m \geq m} q_{m,K} (1 + F) \text{ hence } F = \frac{1}{1 - X} Y + \sum_{m} (N + K - 1) X^N Y^{K-1} = \sum_{N} (N + K - 1) X^N Y^K
\]

Hence \( q_{N,K} = \frac{(N+K-1)}{(K-1)!} \) and it is easy to see that \( q_{N,K} \sim \frac{N^K}{(K-1)!} \) when \( N \to \infty \). Applying Equation (5) to the indicator function of the simplex \( T_K \), we get

\[
\int_T dt = \lim_{N \to \infty} \frac{\lambda}{(K-1)!} = \frac{\lambda}{(K-1)!} \text{ hence } \lambda = 1.
\]

On the other hand, we have \( \int_T dt = B(1) = \frac{\Gamma(1)^K}{\Gamma(K)} = \frac{1}{(K-1)!} \), hence \( \lambda = 1 \).

We now consider Equation (5) in the case where \( f \) is null outside \( T_K \), and is factorised, i.e. there exists a family \( f_{1:K} \) of scalar functions \( f_k : \mathbb{R}_+ \to \mathbb{R} \) such that

\[
f(t) = \prod_{k=1}^K f_k(t_k)
\]

In that case, for a given \( N \), the sum in the right-hand side of Equation (5) becomes

\[
S_{N,K}(f_{1:K}) \triangleq \sum_{x \in \mathbb{R}_+^K} \prod_{k=1}^K f_k(x_k/N)
\]

By construction, if \( \otimes \) denotes the convolution operator on vectors indexed by \( 0:N \), then

\[
S_{N,K}(f_{1:K}) = \left( \bigotimes_{k=1}^K \left( f_k(n) \right) \right)_{n=0:N}
\]

This convolution can be computed efficiently without exhaustively enumerating the grid \( \mathcal{P}_N \). Depending on the magnitude of \( N \), working with Fast Fourier Transforms may even be more efficient. In the log domain, a simple method makes use of the \( \log \otimes \exp \) operator, which is associative commutative like \( \otimes \), and which can be computed efficiently for any \( 0:N \)-dimensional vectors \( x, y \) by:

\[
\log \otimes \exp(x, y) = \log \sum \exp(T(x)+y)
\]

where \( T(x) \) is the \( 0:N \times 0:N \) Toeplitz matrix where \( T(x)_{nm} \) is equal to \( x_{n-m} \) if \( n \geq m \) and \( -\infty \) otherwise, operator \( \dagger \) adds a vector to a matrix row-wise and returns a matrix, and operator \( \log \sum \exp \) applies operator \( \log \sum \exp \) to each row of a matrix and returns a vector. Operator \( \log \sum \exp \) must be efficiently implemented to avoid numerical instability.

### 4.2 Approximate computation of \( Z(m, r) \)

For a given \( s \in \mathbb{R}_+ \), the integrand in the definition of \( Z(s) \) is factorised (up to a multiplicative constant):

\[
\tilde{Z}(s) = \frac{1}{\Gamma(s)^m} \int_T \prod_{k=1}^K \exp(-m \log \Gamma(st_k + r_k st_k)) dt
\]

Thus, it can be approximated using Equations (5) and (6) where \( f_k(u, s) = \exp(-m \log \Gamma(su + r_k su)) \). For \( N \) sufficiently large,

\[
Z(s) \approx \frac{1}{N^{K-1}} S_{N,K}(f_{1:K}(s))
\]

Now, choose any pivot value \( \rho \in \mathbb{R}_+ \). Let \( \Gamma(K, \rho) \) be the Gamma distribution with shape \( K \) and rate \( \rho \), we have:

\[
Z(m, r) = \int Z(s) s^{K-1} ds = \int Z(s) \exp(s \rho) s^{K-1} \exp(-s \rho) ds = \frac{\Gamma(K)}{\rho^K} E_{s \sim \Gamma(K, \rho)} [Z(s) \exp(s \rho)]
\]
The expectation can be approximated by the average of a sample \((s_p)_{p=1:P}\) from the distribution \(\text{Gam}(K, \rho)\) for \(P\) sufficiently large. Combining with Equation (7), we get:

\[
Z(m, r) \approx \frac{\Gamma(K)}{e^P \rho^K N^K - 1} \sum_{p=1}^{P} S_{NK}(f_{1:K}(:, s_p)) \Gamma(s_p)^m \exp s_p \rho
\]  

(8)

In log scale, this becomes

\[
\log Z(m, r) \approx \log \Gamma(K) - \log P - K \log \rho - (K - 1) \log N + \log \sum_{p} \exp(\log S_{NK}(f_{1:K}(:, s_p)) + m \log \Gamma(s_p) + \rho s_p)
\]

Moreover, for \(p \in 1:P\), let \(D_p\) be the \(0:N\) vector \((s_p \frac{n}{N})_{n=0:N}\). Then

\[
\log S_{NK}(f_{1:K}(:, s_p)) = \left(\log \bigotimes_{k=1}^{K} \exp(-m \log \Gamma(D_p) + r_k D_p)\right)_N
\]

References

[1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. “Latent dirichlet allocation”. In: the Journal of machine Learning research 3 (2003), pp. 993–1022.

[2] Cristinel Mortici and Chao-Ping Chen. “New Sharp Double Inequalities for Bounding the Gamma and Digamma Function”. In: Analele Universitatii de Vest din Timisoara 49.2 (2011), pp. 69–75.

[3] Christian P. Robert. The Bayesian choice: from decision-theoretic foundations to computational implementation. 2nd ed. Springer texts in statistics. New York: Springer, 2007.

[4] Jayaram Sethuraman. “A Constructive Definition of Dirichlet Priors”. In: Statistica Sinica 4 (1994), pp. 639–650.