A simple proof of Pitman-Yor’s Chinese restaurant process from its stick-breaking representation

Caroline Lawless, Julyan Arbel

To cite this version:

Caroline Lawless, Julyan Arbel. A simple proof of Pitman-Yor’s Chinese restaurant process from its stick-breaking representation. Dependence Modeling, De Gruyter, 2019. <hal-01950653>

HAL Id: hal-01950653
https://hal.archives-ouvertes.fr/hal-01950653
Submitted on 11 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A simple proof of Pitman–Yor’s Chinese restaurant process from its stick-breaking representation

Caroline Lawless and Julyan Arbel
Univ. Grenoble Alpes, Inria, CNRS, LJK, 38000 Grenoble, France

Abstract
For a long time, the Dirichlet process has been the gold standard discrete random measure in Bayesian nonparametrics. The Pitman–Yor process provides a simple and mathematically tractable generalization, allowing for a very flexible control of the clustering behaviour. Two commonly used representations of the Pitman–Yor process are the stick-breaking process and the Chinese restaurant process. The former is a constructive representation of the process which turns out very handy for practical implementation, while the latter describes the partition distribution induced. Obtaining one from the other is usually done indirectly with use of measure theory. In contrast, we provide here an elementary proof of Pitman–Yor’s Chinese Restaurant process from its stick-breaking representation.

1 Introduction
The Pitman–Yor process defines a rich and flexible class of random probability measures which was developed by Perman et al. (1992) and further investigated by Pitman (1995), Pitman and Yor (1997). It is a simple generalization of the Dirichlet process (Ferguson, 1973), whose mathematical tractability contributed to its popularity in machine learning theory (Caron et al., 2017), probabilistic models for linguistic applications (Teh, 2006, Wood et al., 2011), excursion theory (Perman et al., 1992, Pitman and Yor, 1997), measure-valued diffusions in population genetics (Petrov, 2009, Feng and Sun, 2010), combinatorics (Vershik et al., 2004, Kerov, 2006) and statistical physics (Derrida, 1981).

Its most prominent role is perhaps in Bayesian nonparametric statistics where it is used as a prior distribution, following the work of Ishwaran and James (2001). Applications in this setting embrace a variety of inferential problems, including species sampling (Favaro et al., 2009, Navarrete et al., 2008, Arbel et al., 2017), survival analysis and graphical models in genetics (Jara et al., 2010, Ni et al., 2018), image segmentation (Sudderth and Jordan, 2009), curve estimation (Canale et al., 2017), exchangeable feature allocations (Battiston et al., 2018) and time-series and econometrics (Caron et al., 2017, Bassetti et al., 2014).

Last but not least, the Pitman–Yor process is also employed in the context of nonparametric mixture modeling, thus generalizing the celebrated Dirichlet process mixture model of Lo (1984). Nonparametric mixture models based on the Pitman–Yor process are characterized by a more flexible parameterization than the Dirichlet process mixture model, thus
allowing for a better control of the clustering behaviour (De Blasi et al., 2015). In addition, see Ishwaran and James (2001), Favaro and Walker (2013), Arbel et al. (2018) for posterior sampling algorithms, Scricciolo et al. (2014), Miller and Harrison (2014) for asymptotic properties, and Scarpa and Dunson (2009), Canale et al. (2017) for spike-and-slab extensions.

The Pitman–Yor process has the following stick-breaking representation: if \( v_i \overset{\text{iid}}{\sim} \text{Beta}(1-d, \alpha + id) \) for \( i = 1, 2, \ldots \) with \( d \in (0, 1) \) and \( \alpha > -d \), if \( \pi_j = v_j \prod_{i=1}^{j-1} (1 - v_i) \) for \( j = 1, 2, \ldots \), and if \( \theta_1, \theta_2, \ldots \overset{\text{iid}}{\sim} H \), then the discrete random probability measure

\[
P = \sum_{j=1}^{\infty} \pi_j \delta_{\theta_j}
\]

is distributed according to the Pitman–Yor process, \( \text{PY}(\alpha, d, H) \), with concentration parameter \( \alpha \), discount parameter \( d \), and base distribution \( H \).

The Pitman–Yor process induces the following partition distribution: if \( P \sim \text{PY}(\alpha, d, H) \), for some nonatomic probability distribution \( H \), we observe data \( x_1, \ldots, x_n \mid P \overset{\text{iid}}{\sim} P \), and \( C \) is the partition of the first \( n \) integers \( \{1, \ldots, n\} \) induced by data, then

\[
P(C = C) = \frac{d^{|C|}}{(\alpha)_{n-1}} \left( \frac{\alpha}{d} \right)^{|C|} \prod_{c \in C} (1 - d)^{|c| - 1},
\]

where the multiplicative factor before the product in (2) is also commonly (and equivalently) written as \( (\prod_{i=1}^{C-1} (\alpha + id))/((\alpha)_{n-1}) \) in the literature. When the discount parameter \( d \) is set to zero, the Pitman–Yor process reduces to the Dirichlet process and the partition distribution (2) boils down to the celebrated Chinese Restaurant process (CRP, see Antoniak, 1974). By abuse of language, we call the partition distribution (2) the Pitman–Yor’s CRP. Under the latter partition distribution, the number of parts in a partition \( C \) of \( n \) elements, \( k_n = |C| \), grows to infinity as a power-law of the sample size, \( n^d \) (see Pitman, 2003, for details). This Pitman–Yor power-law growth is more in tune with most of empirical data (Clauset et al., 2009) than the logarithmic growth induced by the Dirichlet process CRP, \( \alpha \log n \).

The purpose of this note is to provide a simple proof of Pitman–Yor’s CRP (2) from its stick-breaking representation (1) (Theorem 2.1). This generalizes the derivation by Miller (2018) who obtained the Dirichlet process CRP (Antoniak, 1974) from Sethuraman’s stick-breaking representation (Sethuraman, 1994). In doing so, we also provide the marginal distribution of the allocation variables vector (3) in Proposition 2.2.

## 2 Partition distribution from stick-breaking

Suppose we make \( n \) observations, \( z_1, \ldots, z_n \). We denote the set \( \{1, \ldots, n\} \) by \([n]\). Our observations induce a partition of \([n]\), denoted \( C = \{c_1, \ldots, c_{k_n}\} \) where \( c_1, \ldots, c_{k_n} \) are disjoint sets and \( \bigcup_{i=1}^{k_n} c_i = [n] \), in such a way that \( z_i \) and \( z_j \) belong to the same partition if and only if \( z_i = z_j \). We denote the number of parts in the partition \( C \) by \( k_n = |C| \) and we denote the number of elements in partition \( j \) by \(|c_j|\). We use bold font to represent random variables.
We write \((x)_{(n)} = \prod_{j=0}^{n-1} (x + j)\) to denote the rising factorial.

**Theorem 2.1.** Suppose

\[
v_i \sim \text{Beta}(1 - d, \alpha + id) \text{ for } i = 1, 2, \ldots,
\]

\[
\pi_j = v_j \prod_{i=1}^{j-1} (1 - v_i) \text{ for } j = 1, 2, \ldots
\]

Let allocation variables be defined by

\[
z_1, \ldots, z_n | \pi \sim \pi \text{ iid, meaning, } \mathbb{P}(z_i = j | \pi) = \pi_j,
\]

and \(C\) denote the random partition of \([n]\) induced by \(z_1, \ldots, z_n\). Then

\[
\mathbb{P}(C = C) = \frac{d^{\lvert C \rvert}}{(\alpha)_{(n)}} \frac{(\alpha)}{(d - \lvert C \rvert)} \prod_{c \in C} (1 - d - (\lvert c \rvert - 1)).
\]

The proof of Theorem 2.1 follows the lines of Miller (2018)'s derivation. We need the next two technical results, which we will prove in Section 3. Let \(C_z\) denote the partition \([n]\) induced by \(z\) for any \(z \in \mathbb{N}^n\). Let \(k_n\) be the number of parts in the partition. We define \(m(z) = \max \{z_1, \ldots, z_n\}\), and \(g_j(z) = \#\{i : z_i \geq j\}\).

**Proposition 2.2.** For any \(z \in \mathbb{N}^n\), the marginal distribution of the allocation variables vector \(z = (z_1, \ldots, z_n)\) is given by

\[
\mathbb{P}(z = z) = \frac{1}{(\alpha)_{(n)}} \prod_{c \in C_z} \frac{\Gamma(\lvert c \rvert + 1 - d)}{\Gamma(1 - d)} \prod_{j=1}^{m(z)} \frac{(\alpha + (j - 1)d)}{g_j(z) + \alpha + (j - 1)d}.
\]

**Lemma 2.3.** For any partition \(C\) of \([n]\),

\[
\sum_{z \in \mathbb{N}^n} 1(C_z = C) \prod_{j=1}^{m(z)} \frac{(\alpha + (j - 1)d)}{g_j(z) + \alpha + (j - 1)d} = \frac{d^{\lvert C \rvert}}{\prod_{c \in C} (\lvert c \rvert - d)} \left( \frac{\alpha}{d} \right)^{\lvert C \rvert}.
\]
Proof of Theorem 2.1.

\[ \mathbb{P}(C = C) = \sum_{z \in \mathbb{N}^n} \mathbb{P}(C = C|z = z) \mathbb{P}(z = z) \]

\[= \sum_{z \in \mathbb{N}^n} 1(C_z = C) \frac{1}{(\alpha)_n} \prod_{c \in C_z} \frac{\Gamma(|c| + 1 - d)}{\Gamma(1 - d)} \prod_{j=1}^{m(z)} \frac{\alpha + (j - 1)d}{g_j(z) + \alpha + (j - 1)d} \]

\[= \frac{1}{(\alpha)_n} \prod_{c \in C} \frac{\Gamma(|c| + 1 - d)}{\Gamma(1 - d)} \sum_{z \in \mathbb{N}^n} 1(C_z = C) \prod_{j=1}^{m(z)} \frac{\alpha + (j - 1)d}{g_j(z) + \alpha + (j - 1)d} \]

\[= \frac{1}{(\alpha)_n} \prod_{c \in C} \frac{\Gamma(|c| + 1 - d)}{\Gamma(1 - d)} \frac{d^{|C|}}{\prod_{c \in C} (|c| - d)} \left( \frac{\alpha}{d} \right)_{(|C|)} \]

\[= \frac{d^{|C|}}{(\alpha)_n} \left( \frac{\alpha}{d} \right)_{(|C|)} \prod_{c \in C} (1 - d)_{(|c| - 1)}, \]

where (a) is by Proposition 2.2, (b) is by Lemma 2.3, and (c) is since \( \Gamma(|c| + 1 - d) = (|c| - d) \Gamma(|c - d|) \).

\[\square\]

3 Proofs of the technical results

3.1 Additional lemmas

We require the following additional lemmas.

Lemma 3.1. For \( a + c > 0 \), and \( b + d > 0 \), if \( y \sim \text{Beta}(a, b) \), then \( \mathbb{E}[y^c (1 - y)^d] = \frac{B(a + c, b + d)}{B(a, b)} \) where \( B \) denotes the beta function.

Proof.

\[ \mathbb{E}[y^c (1 - y)^d] = \int_0^1 y^c (1 - y)^d \frac{1}{B(a, b)} y^{a-1} (1 - y)^{b-1} dy \]

\[= \frac{1}{B(a, b)} \int_0^1 y^{a+c-1} (1 - y)^{b+d-1} dy \]

\[= \frac{B(a + c, b + d)}{B(a, b)}. \]

\[\square\]

Let \( S_{k_n} \) denote the set of \( k_n! \) permutations of \([k_n]\). The following lemma is key for proving Lemma 2.3.
Lemma 3.2. For any \( n_1, \ldots, n_{k_n} \in \mathbb{N} \),
\[
\sum_{\sigma \in S_{k_n}} \prod_{i=1}^{k_n} \frac{1}{a_i(\sigma) - (k_n - i + 1)d} = \frac{1}{\prod_{i=1}^{k_n} (n_i - d)}
\]
where \( a_i(\sigma) = n_{\sigma_i} + n_{\sigma_i+1} + \cdots + n_{\sigma_{k_n}} \).

Proof. Consider the process of sampling without replacement \( k_n \) times from an urn containing \( k_n \) balls. The balls have sizes \( n_1 - d, \ldots, n_{k_n} - d \), and the probability of drawing ball \( i \) is proportional to its size \( n_i - d \). Thus for any permutation \( \sigma \in S_{k_n} \) we have that
\[
p(\sigma_1) = \frac{n_{\sigma_1} - d}{n - td} = \frac{n_{\sigma_1} - d}{a_1(\sigma) - td},
p(\sigma_2|\sigma_1) = \frac{n_{\sigma_2} - d}{n - n_{\sigma_1} - (k_n - 1)d} = \frac{n_{\sigma_2} - d}{a_2(\sigma) - (k_n - 1)d},
p(\sigma_i|\sigma_1, \ldots, \sigma_{i-1}) = \frac{n_{\sigma_i} - d}{n - n_{\sigma_1} - \cdots - n_{\sigma_{i-1}} - (k_n - i + 1)d} = \frac{n_{\sigma_i} - d}{a_i(\sigma) - (k_n - i + 1)d}.
\]
Therefore,
\[
p(\sigma) = p(\sigma_1)p(\sigma_2|\sigma_1) \cdots p(\sigma_{k_n}|\sigma_1, \ldots, \sigma_{k_n-1}) = \prod_{i=1}^{k_n} \frac{n_{\sigma_i} - d}{a_i(\sigma) - (k_n - i + 1)d}.
\]
This way, we construct a distribution on \( S_{k_n} \). We know that \( \sum_{\sigma \in S_{k_n}} p(\sigma) = 1 \). Applying this to Equation (4) and dividing both sides by \( (n_{\sigma_1} - d) \cdots (n_{\sigma_{k_n}} - d) = (n_1 - d) \cdots (n_{k_n} - d) \) gives the result.

Lemma 3.3. Let \( b_i \in \mathbb{N} \) for \( i \in \{1, \ldots, k_n\} \) and let \( b_0 = 0 \). We define \( \bar{b}_i = b_0 + b_1 + \cdots + b_i \). Then
\[
\prod_{i=1}^{k_n} \sum_{b_i \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{i-1})(b_i)}{(\frac{\alpha}{d} + \bar{b}_i)(b_i)} = \frac{(\frac{\alpha}{d})(k_n)}{\prod_{i=1}^{k_n} (\frac{\alpha}{d} - (k_n + 1 - i))}.
\]
Proof. Let \( A_j \) denote the intermediate sum \( A_j = \prod_{i=j}^{k_n} \sum_{b_i \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{i-1})(b_i)}{(\frac{\alpha}{d} + \bar{b}_i)(b_i)} \). We show by induction decreasing from \( j = k_n \) to \( j = 0 \) that
\[
A_j = \frac{(\frac{\alpha}{d} + \bar{b}_{j-1})(k_n + 1)}{\prod_{i=j}^{k_n} (\frac{\alpha}{d} - (k_n + 1 - i))}.
\]
When \( j = k_n \) we have
\[
A_{k_n} = \sum_{b_{k_n} \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{k_n-1})(b_{k_n})}{(\frac{\alpha}{d} + \bar{b}_{k_n})(b_{k_n})} = \sum_{b_{k_n} \in \mathbb{N}} \mathbb{E}[X^{b_{k_n}}]
\]
where \( X \sim \text{Beta}(\frac{\alpha}{a} + \bar{b}_{k_n - 1}, \frac{a_{k_n}}{d}) \). We have that

\[
\sum_{b_{k_n} \in \mathbb{N}} \mathbb{E}[X^{b_{k_n}}] = \mathbb{E} \left[ \sum_{b_{k_n} \in \mathbb{N}} X^{b_{k_n}} \right] = \mathbb{E} \left[ \frac{X}{1 - X} \right] = \frac{\alpha + d\bar{b}_{k_n - 1}}{a_{k_n} - d},
\]
due to Lemma 3.1, which proves the initialization for (5).

We now consider the case of an arbitrary \( j \), greater than 0 and less than \( k_n \). By the induction hypothesis, we have that Equation (5) holds for \( j + 1 \), that is

\[
A_{j+1} = \frac{(\frac{\alpha}{a} + \bar{b}_{j})_{(k_n-j)}}{\prod_{i=j+1}^{k_n}(\frac{\alpha}{a} - (k_n + 1 - i))}.
\]

Therefore,

\[
A_j = \sum_{b_j \in \mathbb{N}} \frac{(\frac{\alpha}{a} + \bar{b}_{j-1})(b_j)}{(\frac{\alpha}{a} + \bar{b}_{j})_{(k_n-j)}} \prod_{i=j+1}^{k_n} \sum_{b_i \in \mathbb{N}} \frac{(\frac{\alpha}{a} + \bar{b}_{i-1})(b_i)}{(\frac{\alpha}{a} + \bar{b}_{i})_{(k_n-i)}}
= \sum_{b_j \in \mathbb{N}} \frac{(\frac{\alpha}{a} + \bar{b}_{j-1})(b_j)}{(\frac{\alpha}{a} + \bar{b}_{j})_{(k_n-j)}} \frac{(\frac{\alpha}{a} + \bar{b}_{j})_{(k_n-j)}}{\prod_{i=j+1}^{k_n}(\frac{\alpha}{a} - (k_n + 1 - i))}
\]

Rearranging the rising factorials in the numerator, we can write

\[
\left(\frac{\alpha}{d} + \bar{b}_{j-1}\right)_{(b_j)} \left(\frac{\alpha}{d} + \bar{b}_{j}\right)_{(k_n-j)} = \left(\frac{\alpha}{d} + \bar{b}_{j-1}\right)_{(b_j)} \left(\frac{\alpha}{d} + \bar{b}_{j-1} + b_j\right)_{(k_n-j)}
= \left(\frac{\alpha}{d} + \bar{b}_{j-1}\right)_{(b_j+k_n-j)}
= \left(\frac{\alpha}{d} + \bar{b}_{j-1}\right)_{(k_n-j)} \left(\frac{\alpha}{d} + \bar{b}_{j-1} + k_n - j\right)_{(b_j)}
\]

and thus factorize the terms independent of \( b_j \) in order to obtain

\[
A_j = \frac{(\frac{\alpha}{a} + \bar{b}_{j-1})_{(k_n-j)}}{\prod_{i=j+1}^{k_n}(\frac{\alpha}{a} - (k_n + 1 - i))} \sum_{b_j \in \mathbb{N}} \frac{(\frac{\alpha}{a} + \bar{b}_{j-1} + k_n - j)_{(b_j)}}{\prod_{i=j+1}^{k_n}(\frac{\alpha}{a} + \bar{b}_{j-1})_{(b_j)}}
\]

The sum above can be rewritten, using \( X \sim \text{Beta}(\frac{\alpha}{a} + \bar{b}_{j-1} + (k_n - j), \frac{\alpha}{a} - (k_n - j)) \), as

\[
\sum_{b_j \in \mathbb{N}} \mathbb{E}[X^{b_j}] = \mathbb{E} \left[ \frac{X}{1 - X} \right] = \frac{\alpha + \bar{b}_{j-1} + (k_n - j)}{\frac{\alpha}{a} - (k_n + 1 - j)}.
\]

Putting this all together,

\[
A_j = \frac{(\frac{\alpha}{a} + \bar{b}_{j-1} + (k_n - j))}{\frac{\alpha}{a} - (k_n + 1 - j)} \frac{(\frac{\alpha}{a} + \bar{b}_{j-1})_{(k_n-j)}}{\prod_{i=j+1}^{k_n}(\frac{\alpha}{a} - (k_n + 1 - i))}
\]
which proves the desired result for $j$. By induction, this result is true for all $j \in \{1, \ldots, k_n\}$. Letting $j = 1$ gives the result stated in the lemma, since $\bar{b}_0 = b_0 = 0$. 

3.2 Proof of Proposition 2.2 and Lemma 2.3

Proof of Proposition 2.2. For simplicity, we fix the allocation variable vector to a value $z$ and denote $m(z)$ by $m$ and $g_j(z)$ by $g_j$. We have

$$
\Pr(z = z|\pi_1, \ldots, \pi_m) = \prod_{i=1}^{n} \pi_{z_i} = \prod_{j=1}^{m} \pi_j^{e_j}
$$

where $e_j = \#\{i : z_i = j\}$. Thus,

$$
\Pr(z = z|v_1, \ldots, v_m) = \prod_{j=1}^{m} \left(v_j \prod_{i=1}^{j-1}(1 - v_i)^{e_j}\right) = \prod_{j=1}^{m} v_j^{e_j}(1 - v_j)^{f_j}
$$

where $f_j = \#\{i : z_i > j\}$. Therefore,

$$
\Pr(z = z) = \int \Pr(z = z|v_1, \ldots, v_m)p(v_1, \ldots, v_m)dv_1 \cdots dv_m
$$

$$
= \int \left(\prod_{j=1}^{m} v_j^{e_j}(1 - v_j)^{f_j}\right)p(v_1) \cdots p_m(v_m)dv_1 \cdots dv_m
$$

$$
= \prod_{j=1}^{m} \int v_j^{e_j}(1 - v_j)^{f_j}p_j(v_j)dv_j
$$

$$
\overset{(a)}{=} \prod_{j=1}^{m} \frac{B(e_j + 1 - d, f_j + \alpha + jd)}{B(1 - d, \alpha + jd)}
$$

$$
= \prod_{j=1}^{m} \frac{\Gamma(e_j + 1 - d)\Gamma(f_j + \alpha + jd)\Gamma(\alpha + (j - 1)d) + 1}{\Gamma(e_j + f_j + \alpha + (j - 1)d) + 1}\Gamma(1 - d)\Gamma(\alpha + jd)
$$

$$
\overset{(b)}{=} \prod_{j=1}^{m} \frac{\Gamma(e_j + 1 - d)}{\Gamma(1 - d)} \prod_{j=1}^{m} \frac{\Gamma(g_j + \alpha + (j - 1)d) + 1}{\Gamma(g_j + \alpha + (j - 1)d + 1)} \prod_{j=1}^{m} \Gamma(\alpha + (j - 1)d + 1)
$$

$$
\overset{(c)}{=} \prod_{j=1}^{m} \frac{\Gamma(e_j + 1 - d)}{\Gamma(1 - d)} \prod_{j=1}^{m} \frac{\alpha + (j - 1)d}{g_j + \alpha + (j - 1)d} \Gamma(g_{m+1} + \alpha + md)\Gamma(\alpha)
$$

$$
\overset{(d)}{=} \frac{\Gamma(\alpha)}{\Gamma(n + \alpha)} \prod_{c \in C_s} \frac{\Gamma([c] + 1 - d)}{\Gamma(1 - d)} \prod_{j=1}^{m} \frac{\alpha + (j - 1)d}{g_j + \alpha + (j - 1)d}
$$
where step (a) follows from Lemma 3.1, step (b) since $f_j = g_{j+1}$ and $g_j = e_j + f_j$, step (c) since $\Gamma(x+1) = x\Gamma(x)$, and step (d) since $g_1 = n$ and $g_{m+1} = 0$.

**Proof of Lemma 2.3.** As before, we denote the parts of $C$ by $c_1, \ldots, c_{k_n}$, and we let $k_n = |C|$. We denote the distinct values taken on by $z_1, \ldots, z_n$ by $j_1 < \cdots < j_{k_n}$. We define $j_0 = b_0 = 0$, $b_i = j_i - j_{i-1}$, and $\bar{b}_i = b_0 + \cdots + b_i$ for $i \in \{1, \ldots, k_n\}$. We use the notation $a_i(\sigma) = n_{\sigma_i} + \cdots + n_{\sigma_{k_n}}$, where $\sigma$ is the permutation of $[k_n]$ such that $c_{\sigma_i} = \{\ell : z_\ell = j_i\}$.

Then for any $z \in \mathbb{N}^n$ such that $C_z = C$,

$$\prod_{j=1}^{m(z)} \frac{\alpha + (j - 1)d}{g_j(z) + \alpha + (j - 1)d} = \prod_{j=1}^{m(z)} \frac{\alpha d + j - 1}{g_j(z) + \alpha + (j - 1)d}$$

$$= \prod_{i=1}^{k_n} \frac{\bar{b}_i}{\prod_{j=b_{i-1}+1}^{b_i} g_j(z) + \alpha + (j - 1)d} = \prod_{i=1}^{k_n} \frac{\alpha d + \bar{b}_{i-1}(b_i)}{\alpha + a_i(\sigma) + \bar{b}_{i-1}(b_i)},$$

because $g_j(z) = a_i(\sigma)$ for $\bar{b}_{i-1} < j \leq \bar{b}_i$. It follows from the definition of $b = (b_1, \ldots, b_{k_n})$ and $\sigma$ that there is a one-to-one correspondence between $\{z \in \mathbb{N}^n : C_z = C\}$ and $\{(\sigma, b) : \sigma \in S_{k_n}, b \in \mathbb{N}^{k_n}\}$. Therefore,

$$\sum_{z \in \mathbb{N}^n} 1(C_z = C) \prod_{j=1}^{m(z)} \frac{\alpha + (j - 1)d}{g_j(z) + \alpha + (j - 1)d} = \sum_{\sigma \in S_{k_n}} \sum_{b \in \mathbb{N}^{k_n}} \prod_{i=1}^{k_n} \frac{\alpha d + \bar{b}_{i-1}(b_i)}{\alpha + a_i(\sigma) + \bar{b}_{i-1}(b_i)}$$

$$= \sum_{\sigma \in S_{k_n}} \sum_{b \in \mathbb{N}^{k_n}} \prod_{i=1}^{k_n} \frac{\alpha d + \bar{b}_{i-1}(b_i)}{\alpha + a_i(\sigma) + \bar{b}_{i-1}(b_i)}$$

$$= \sum_{\sigma \in S_{k_n}} \sum_{i=1}^{k_n} \frac{\alpha d}{a_i(\sigma) - (k_n - i + 1)}$$

$$= d^{k_n} \left(\frac{\alpha}{d}\right)^{(k_n)} \sum_{\sigma \in S_{k_n}} \frac{1}{\prod_{i=1}^{k_n} a_i(\sigma) - (k_n - i + 1)d}$$

$$= \frac{d^{k_n}}{\prod_{c \in C} (|c| - d)} \left(\frac{\alpha}{d}\right)^{(k_n)},$$

where step (a) follows from Lemma 3.3 and step (b) follows from Lemma 3.2.

**Acknowledgement**

The authors would like to thank Bernardo Nipoti for fruitful discussions that initiated this work.
References


