# RENEWAL SEQUENCES AND RECORD CHAINS RELATED TO MULTIPLE ZETA SUMS

JEAN-JIL DUCHAMPS, JIM PITMAN, AND WENPIN TANG

ABSTRACT. For the random interval partition of $[0, 1]$ generated by the uniform stick-breaking scheme known as GEM(1), let $u_k$ be the probability that the first $k$ intervals created by the stick-breaking scheme are also the first $k$ intervals to be discovered in a process of uniform random sampling of points from $[0, 1]$. Then $u_k$ is a renewal sequence. We prove that $u_k$ is a rational linear combination of the real numbers $1, \zeta(2), \ldots, \zeta(k)$ where $\zeta$ is the Riemann zeta function, and show that $u_k$ has the limit $1/3$ as $k \to \infty$. Related results provide probabilistic interpretations of some multiple zeta values in terms of a Markov chain derived from the interval partition. This Markov chain has the structure of a weak record chain. Similar results are given for the GEM($\theta$) model, with beta$(1, \theta)$ instead of uniform stick-breaking factors, and for another more algebraic derivation of renewal sequences from the Riemann zeta function.

## CONTENTS

## 1. INTRODUCTION

Consider the following two sequences of random subsets of the set $[n] := \{1, \ldots, n\}$, generated by listing the cycles of a uniform random permutation $\pi_n$ of $[n]$ in two different orders: for a permutation $\pi_n$ with $K_n$ cycles,

- let $C_{1:n}, C_{2:n}, \ldots$, be the cycles of $\pi_n$ in *order of least elements*, so $C_{1:n}$ is the cycle of $\pi_n$ containing 1; if $C_{1:n} \neq [n]$, then $C_{2:n}$ is the cycle of $\pi_n$ containing the least $j \in [n]$ with $j \notin C_{1:n}$, and so on, with $C_{k:n} = \emptyset$ if $k > K_n$;

- let $R_{1:n}, R_{2:n}, \ldots, R_{K_n:n}$ be the same cycles in *order of greatest elements*, so $R_{1:n}$ is the cycle of $\pi_n$ containing $n$; if $R_{1:n} \neq [n]$, then $R_{2:n}$ is the cycle of $\pi_n$ containing the greatest $j \in [n]$ with $j \notin R_{1:n}$, and so on, with $R_{k:n} = \emptyset$ if $k > K_n$.

For $1 \leq k \leq n$ define the probability

$$(1.1) \qquad u_{k:n} := \mathbb{P}\left(\cup_{i=1}^k C_{i:n} = \cup_{i=1}^k R_{i:n}\right),$$

the probability that the same collection of $k$ cycles appears as the first $k$ cycles in both orders. It is elementary that the number of elements of $C_{1:n}$ has a discrete uniform distribution on $[n]$, and the same is true for $R_{1:n}$. Since $u_{1:n}$ is the probability that both 1 and $n$ fall in the same cycle of $\pi_n$, it follows easily that $u_{1:n} = 1/2$ for every $n \geq 2$. For $k \geq 2$ it is easy to give multiple summation formulas for $u_{k:n}$. Such formulas show that $u_{k:n}$ has some dependence on $n$ for $k \geq 2$, with limits

$$(1.2) \qquad \lim_{n \to \infty} u_{k:n} = u_k \text{ for each } k = 1, 2, \ldots,$$

which may be described as follows. It is known [10, page 25] that the asymptotic structure of sizes of cycles of $\pi_n$, when listed in either order, and normalized by $n$, is that of the sequence of lengths of subintervals of $[0,1]$ defined by the *uniform stick-breaking scheme*:

$$(1.3) \qquad P_1 := W_1, \qquad P_2 := (1 - W_1)W_2, \qquad P_3 := (1 - W_1)(1 - W_2)W_3, \ldots,$$

where the $W_i$ are i.i.d. uniform $[0,1]$ variables. In more detail,

$$(1.4) \qquad P_k := |I_k| \text{ where } I_k := [R_{k-1}, R_k) \text{ with } R_k = \sum_{i=1}^k P_k = 1 - \prod_{i=1}^k (1 - W_i),$$

with the convention $R_0 = 0$. The distribution of lengths $(P_1, P_2, \ldots)$ so obtained, with $\sum_i P_i = 1$ almost surely, is known as the GEM(1) model, after Griffiths, Engen, and McCloskey. The limit probabilities $u_k$ are easily evaluated directly in terms of the limit model, as follows. Let $(U_1, U_2, \ldots)$ be an i.i.d. uniform $[0,1]$ sequence of *sample points* independent of $(P_1, P_2, \ldots)$. Say that an interval $I_k$ has been discovered by time $i$ if it contains at least one of the sample points $U_1, U_2, \ldots, U_i$. The sampling process imposes a new order of discovery on the intervals $I_k$, which describes the large $n$ limit structure of the random permutation of cycles of $\pi_n$ described above. The limit $u_k$ in (1.2) is $u_k := \mathbb{P}(E_k)$, the probability of the event $E_k$ in the limit model that the union of the first $k$ intervals to be discovered in the sampling process equals $\cup_{i=1}^k I_i = [0, R_k)$ for $R_k = P_1 + \cdots + P_k$ as above. There are several different ways to express this event $E_k$. The most convenient for present purposes is to consider the stopping time $n(k,1)$ when the sampling process first discovers a point not in $[0, R_k)$. If at that time $n = n(k,1)$ (for $1 \leq i \leq k$ there is at least one sample point $U_j \in I_i$ with $1 \leq j < n$), then the event $E_k$ has occurred, and otherwise not. Thus for $k = 1, 2, \ldots$,

$$(1.5) \qquad u_k = \mathbb{P}\left(\cap_{i=1}^k (U_j \in I_i \text{ for some } 1 \leq j < n(k,1))\right),$$

and we adopt the convention that $u_0 := 1$. This sequence $(u_k, k \geq 0)$ is a renewal sequence appearing in the study of *regenerative permutations* in [51]. In that context it is easily shown that the limit $u_\infty := \lim_{k \to \infty} u_k$ exists, but it is difficult to evaluate the $u_k$ for general $k$. However, computation of $u_k$ for the first few $k = 1, 2, 3, \ldots$, by symbolic integration suggested a general formula for $u_k$ as a rational

linear combination of the real numbers $1, \zeta(2), \ldots, \zeta(k)$ where $\zeta$ is the Riemann zeta function,

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} \qquad \text{for } \operatorname{Re}(s) > 1.$$

This article establishes the following result, whose proof leads to some probabilistic interpretations of multiple zeta values and harmonic sums.

**Proposition 1.1.** *The renewal sequence $(u_k, \ k \geq 0)$ defined above by (1.5) in terms of uniform stick-breaking is characterized by any one of the following equivalent conditions:*

*(i). The sequence $(u_k, \ k \geq 0)$ is defined recursively by*

(1.6)  $$2u_k + 3u_{k-1} + u_{k-2} = 2\zeta(k) \quad \text{with } u_0 = 1, \ u_1 = 1/2.$$

*(ii). For all $k \geq 0$,*

(1.7)  $$u_k = (-1)^{k-1} \left( 2 - \frac{3}{2^k} \right) + \sum_{j=2}^{k} (-1)^{k-j} \left( 2 - \frac{1}{2^{k-j}} \right) \zeta(j).$$

*(iii). For all $k \geq 0$,*

(1.8)  $$u_k = \sum_{j=1}^{\infty} \frac{2}{j^k(j+1)(j+2)}.$$

*(iv). The generating function of $(u_k, \ k \geq 0)$ is*

(1.9)  $$U(z) := \sum_{k=0}^{\infty} u_k z^k = \frac{2}{(1+z)(2+z)} \left[ 1 + \left( 2 - \gamma - \psi(1-z) \right) z \right],$$

*for $|z| < 1$, where $\gamma := \lim_{n \to \infty} (\sum_{k=1}^{n} 1/k - \ln n) \approx 0.577$ is the Euler constant, and $\psi(z) := \Gamma'(z)/\Gamma(z)$ with $\Gamma(z) := \int_0^{\infty} t^{z-1} e^{-t} dt$, the digamma function.*

The proof is given in Section 3. We are also interested in $(u_k)$ when the partition $(I_k)$ follows a more general stick-breaking scheme, where the $(W_k)$ in (1.3) are i.i.d. with an arbitrary distribution on $(0, 1)$. We will develop in particular the case, for $\theta > 0$, where $W_k$ follows a beta$(1, \theta)$ distribution—with density $\theta(1-x)^{\theta-1}$ on $[0, 1]$. The sequence $u_k$ so defined is the limit (1.2) if the distribution of the random permutation $\pi_n$ is changed from the uniform distribution on permutations of $[n]$ to the *Ewens $(\theta)$ distribution* on permutations of $[n]$, in which the probability of any particular permutation of $[n]$ with $k$ cycles is $\theta^k/(\theta)_n$ instead of $1/(1)_n$, where $(\theta)_n := \theta(\theta+1) \cdots (\theta+n-1)$ is a rising factorial. In that case the limit distribution of interval lengths $(P_1, P_2, \ldots)$ is known as the GEM$(\theta)$ distribution [10, Section 5.4]. Our expressions for $u_k$ in this case are less explicit. In the following, the notation $\overset{\theta}{=}$ indicates evaluations for the GEM$(\theta)$ model. For instance, it was proved in [51, (7.16)] that

(1.10)  $$u_{\infty} \overset{\theta}{=} \frac{\Gamma(\theta+2)\Gamma(\theta+1)}{\Gamma(2\theta+2)} \overset{1}{=} \frac{1}{3}.$$

A consequence of Proposition 1.1 is that for each $k \geq 1$, the right-hand side of (1.7) is positive—in fact strictly greater than $u_{\infty} \overset{1}{=} 1/3$. Also, the probability $f_k$ of a first renewal at time $k$, which is determined by $u_1, \ldots, u_k$ by a well known

recursion recalled later in (4.7), is also strictly positive. These inequalities seem not at all obvious without the probabilistic interpretations offered here. The inequalities are reminiscent of Li's criterion [40] for the Riemann hypothesis, which has some probabilistic interpretations indicated in [14, Section 2.3]. The GEM(1) model also arises from the asymptotics of prime factorizations [23], but the results for sampling from GEM(1) described here do not seem easy to interpret in that setting.

The interpretation of $u_k$ sketched above and detailed in [51], that $u_k$ is the probability that the random order of discovery of intervals maps $[k]$ to $[k]$, yields the following corollary.

**Corollary 1.2.** *For* $\mathbf{w} := (w_1, w_2, \ldots) \in (0,1)^{\mathbb{N}_+}$, *let* $p_i(\mathbf{w}) := (1 - w_1) \cdots (1 - w_{i-1})w_i$. *Then for each* $k \geq 1$, *the expression*

$$(1.11) \qquad \sum_{\pi \in \mathfrak{S}_k} \int_{(0,1)^k} p_{\pi(1)}(\mathbf{w}) \prod_{i=2}^{k} \frac{p_{\pi(i)}(\mathbf{w})}{1 - \sum_{j=1}^{i-1} p_{\pi(j)}(\mathbf{w})} dw_1 \cdots dw_k$$

*is equal to* (1.7) *and to* (1.8), *where* $\mathfrak{S}_k$ *is the set of permutations of the finite set* $\{1, \ldots, k\}$.

The expression (1.7) gives a *rational zeta series expansion* of the multiple integral (1.11). Similar expansions also appeared in Beukers' proof [13] of the irrationality of $\zeta(3)$. The expression (1.11) is a sum of $k!$ positive terms, while (1.7) is a linear combination of $1, \zeta(2), \ldots, \zeta(k)$ with alternating signs. By symbolic integration, we can identify each term of the sum in (1.11) for $k = 2, 3$, but some terms become difficult to evaluate for $k \geq 4$, and we have no general formula for these terms, and no direct algebraic explanation of why the terms in (1.11) should sum to a rational zeta series.

The Riemann zeta function plays an important role in analytic number theory [15, 24], and has applications in geometry [25, 56] and mathematical physics [12, 39]. Connections between the Riemann zeta function and probability theory have also been explored, for example:

- For each $s > 1$, the normalized terms of the Riemann zeta series define a discrete probability distribution of a random variable $Z_s$ with values on $\{1, 2, \ldots\}$, such that $\log Z_s$ has a compound Poisson distribution [5, 34, 44].
- The values $\zeta(2)$ and $\zeta(3)$ emerge in the limit of large random objects [4, 29].
- The values $1/\zeta(n)$ for $n = 2, 3, \ldots$, arise from the limit proportion of $n$-free numbers, that is, numbers not divisible by any $n$th power of a natural number; see [1, 27].
- The values $\zeta(1/2 - n)$ for $n \geq 0$ appear in the expected first ladder height of Gaussian random walks [21].
- The Riemann zeta function appears in the Mellin transforms of functionals of Brownian motion and Bessel processes [14, 55].
- Conjectured bounds for the zeta function on the critical line $\Re(s) = 1/2$ can be related to branching random walks [8].
- There are striking parallels between the behavior of zeros of the Riemann zeta function on the line $\Re(s) = 1/2$ and the structure of eigenvalues in the random matrix theory [38, 45, 47].

In the early 1990s, Hoffman [35] and Zagier [57] introduced the *multiple zeta value*

$$(1.12) \qquad \zeta(s_1, \ldots, s_k) := \sum_{0 < n_1 < \cdots < n_k} \frac{1}{n_1^{s_1} \cdots n_k^{s_k}},$$

and the *multiple zeta-star value*

$$(1.13) \qquad \zeta^*(s_1, \ldots, s_k) := \sum_{0 < n_1 \leq \cdots \leq n_k} \frac{1}{n_1^{s_1} \cdots n_k^{s_k}},$$

for each $k > 0$, and $s_i \in \mathbb{N}_+ := \{1, 2, \ldots\}$ with $s_1 > 1$ to ensure the convergence. Note that the multiple zeta-star value (1.13) can be written as the sum of multiple zeta values,

$$\zeta^*(s_1, \ldots, s_k) = \sum_{\mathbf{s}^*} \zeta(\mathbf{s}^*),$$

where the sum is over all $\mathbf{s}^* = (s_1 \square \cdots \square s_k)$, with each $\square$ filled by either a comma or a plus. To illustrate,

$$\zeta^*(s_1, s_2) = \zeta(s_1, s_2) + \zeta(s_1 + s_2),$$
$$\zeta^*(s_1, s_2, s_3) = \zeta(s_1, s_2, s_3) + \zeta(s_1, s_2 + s_3) + \zeta(s_1 + s_2, s_3) + \zeta(s_1 + s_2 + s_3).$$

See [6, 18, 36] for the algebraic structure and some evaluations of multiple zeta values. It was proved in [3, 58] that the multiple zeta functions (1.12) and (1.13) can also be continued meromorphically on the whole space $\mathbb{C}^k$.

These multiple zeta values appear in various contexts including algebraic geometry, knot theory, and quantum field theory; see [20]. But we are not aware of any previous probabilistic interpretation of these numbers. In this article we show how the zeta values $\zeta(2), \zeta(3), \ldots$, and (1.12) and (1.13) arise in the renewal sequence $(u_k)$ associated with the discovery of intervals for a GEM(1) partition of $[0, 1]$. Equivalently the same sequence $(u_k)$ can be expressed in terms of a GEM(1)-biased permutation of $\mathbb{N}_+$ [51], or of the *Bernoulli sieve* [32] driven by the GEM(1) distribution.

**Organization of the paper:** The rest of the paper is organized as follows:

- In Section 2 we introduce the main tool of our analysis, a Markov chain $(\widehat{Q}_k)$ derived from the discovery process of subintervals in the GEM($\theta$) stick-breaking model, and show its equality in distribution with a weak record chain.
- In Section 3 we give the proof of Proposition 1.1 for $\theta = 1$ and provide some partial results for general $\theta$.
- In Section 4 we define a number of renewal sequences satisfying a recursion involving the Riemann zeta function.
- In Section 5 we specialize again to $\theta = 1$ and examine further the distribution of the Markov chain $(\widehat{Q}_k)$, deriving expressions involving iterated harmonic sums and zeta values.
- In Section 6 we derive a formula for $u_{2:n}$ associated with random permutations, which provides evaluation of $u_2$ for general $\theta$ as the limit. Among those we identify the sequence $(u_k)$ defined by (1.5) in the GEM(1) model.

## 2. One-parameter Markov chains and record processes

Recall the definition (1.3) of the length $P_k = |I_k|$ of the $k$th interval in a stick-breaking partition and the uniform sequence $(U_i)$ of points that we use to discover intervals. Now define a random sequence of positive integers $(X_i)$ by setting

$$(2.1) \qquad X_i := k \iff U_i \in I_k.$$

So $X_i$ is the rank of the interval in which the $i$th sample point $U_i$ falls. Conditionally given the sequence of interval lengths $(P_1, P_2, \ldots)$, the $X_i$ are i.i.d. according to this distribution on $\mathbb{N}_+ := \{1, 2, \ldots\}$. Formula (1.5) can be recast as

$$(2.2) \qquad u_k = \mathbb{P}\left(\{X_1, X_2, \ldots, X_{n(k,1)-1}\} = \{1, 2, \ldots, k\}\right),$$

where $n(k, 1) = \inf\{i \geq 1, X_i \geq k + 1\}$.

The key to our analysis is the Markov chain $(\widehat{Q}_k)$ given by the following lemma from [51, Lemma 7.1]. This lemma is suggested by the work of Gnedin and coauthors on the Bernoulli sieve [30, 31], and subsequent work on extremes and gaps in sampling from a residual allocation model by Pitman and Yakubovich [49, 52].

**Lemma 2.1.** Let $X_1, X_2, \ldots,$ be as in (2.1) for a stick-breaking partition with i.i.d. factors $W_i \stackrel{(d)}{=} W$ as in (1.3) for some distribution of $W$ on $(0, 1)$. For $n \in \mathbb{N}_+$ and $k = 0, 1, \ldots,$ let

$$(2.3) \qquad Q_n^*(k) := \sum_{i=1}^n \mathbf{1}(X_i > k) = \sum_{i=1}^n \mathbf{1}(U_i \geq R_k)$$

represent the number of the first $n$ sample points which land outside the union $[0, R_k)$ of the first $k$ intervals. For $m = 1, 2, \ldots,$ let $n(k, m) := \min\{n : Q_n^*(k) = m\}$ be the first time $n$ that there are $m$ sample points outside the first $k$ intervals. Then

(i). For each $k$ and $m$ there is the equality of joint distributions

$$(2.4) \qquad \left(Q_{n(k,m)}^*(k - j), 0 \leq j \leq k\right) \stackrel{(d)}{=} \left(\widehat{Q}_j, 0 \leq j \leq k \,\Big|\, \widehat{Q}_0 = m\right),$$

where $(\widehat{Q}_0, \widehat{Q}_1, \ldots)$ with $1 \leq \widehat{Q}_0 \leq \widehat{Q}_1 \cdots$ is a Markov chain with state space $\mathbb{N}_+$ and stationary transition probability function

$$(2.5) \qquad \widehat{q}(m, n) := \binom{n-1}{m-1} \mathbb{E} W^{n-m}(1-W)^m \quad \text{for } m \leq n.$$

(ii). For each $k \geq 1$ the renewal probability $u_k$ defined by (2.2) is given by

$$(2.6) \qquad u_k = \mathbb{P}(\widehat{Q}_0 < \widehat{Q}_1 < \cdots < \widehat{Q}_k \,|\, \widehat{Q}_0 = 1).$$

(iii). The sequence $u_k$ is strictly decreasing, with limit $u_\infty \geq 0$ which is given by

$$(2.7) \qquad u_\infty = \mathbb{P}(\widehat{Q}_0 < \widehat{Q}_1 < \cdots \,|\, \widehat{Q}_0 = 1).$$

Here we study the Markov chain $(\widehat{Q}_k)$ for the GEM$(\theta)$ partition and show that it has an interpretation as a *weak record chain*. Let $X_1, X_2, \ldots,$ be a random sample from the GEM$(\theta)$ model with i.i.d. stick-breaking factors $W_i \stackrel{(d)}{=} W$ for $W$ following a beta$(1, \theta)$ distribution. Consider

$$(2.8) \qquad C_k^{\ell,\theta} := \sum_{j=1}^k \mathbf{1}\{Q_{n(k,\ell)}^*(j) = Q_{n(k,\ell)}^*(j-1)\} \quad \text{for } k \geq 1,$$

the number of empty intervals among the first $k$ intervals at the first time $n(k, \ell)$ there are $\ell$ points outside the first $k$ intervals. To study the random variables $C_k^{\ell,\theta}$, we introduce a family of one-parameter Markov chains $(\widehat{Q}_j^{\ell,\theta}, \ j \geq 0)$ with

- the initial value $\widehat{Q}_0^{\ell,\theta} = \ell \in \mathbb{N}_+$,
- the transition probability function $\widehat{q}^{\theta}(m, n)$ given by (2.5) for $W$ the beta $(1, \theta)$ distribution.

For $W$ the beta$(1, \theta)$ distribution,

$$\mathbb{E}W^{n-m}(1 - W)^m = \frac{(1)_{n-m} \, \theta}{(\theta + m)_{n-m-1}} \quad \text{for } m \leq n,$$

where

$$(x)_j := x(x + 1) \cdots (x + j - 1) = \frac{\Gamma(x + j)}{\Gamma(x)}.$$

So the transition probability $\widehat{q}^{\theta}$ of the $\widehat{Q}^{\ell,\theta}$ chain is given by

$$(2.9) \qquad \widehat{q}^{\theta}(m, n) = \frac{(m)_{n-m} \, \theta}{(\theta + m)_{n-m+1}} \quad \text{for } m \leq n.$$

Let

$$(2.10) \qquad G_i^{\ell,\theta}(k) := \sum_{j=1}^{k} 1\{\widehat{Q}_j^{\ell,\theta} = i\} \quad \text{for } i \geq \ell$$

be the occupation count of state $i$ for the Markov chain $(\widehat{Q}_j^{\ell,\theta}, \ 1 \leq j \leq k)$. According to Lemma 2.1 $(i)$, for each $k \geq 1$,

$$(2.11) \quad C_k^{\ell,\theta} \stackrel{(d)}{=} \widehat{C}_k^{\ell,\theta} := \sum_{j=1}^{k} 1\{\widehat{Q}_j^{\ell,\theta} = \widehat{Q}_{j-1}^{\ell,\theta}\}$$

$$= \sum_{j=1}^{k} 1\{\widehat{Q}_j^{\ell,\theta} = \widehat{Q}_{j-1}^{\ell,\theta} = \ell\} + \sum_{i=\ell+1}^{\infty} \sum_{j=1}^{k} 1\{\widehat{Q}_j^{\ell,\theta} = \widehat{Q}_{j-1}^{\ell,\theta} = i\}$$

$$(2.12) \qquad = G_\ell^{\ell,\theta}(k) + \sum_{i=\ell+1}^{\infty} (G_i^{\ell,\theta}(k) - 1)^+,$$

where the last equality follows from the fact that the process $(\widehat{Q}_j^{\ell,\theta}, \ j \geq 0)$ is weakly increasing starting at $\ell$.

Now we establish a connection between the one-parameter chain $\widehat{Q}^{\ell,\theta}$ and a record process. Fix $\ell \in \mathbb{N}_+$. For $X_1, X_2, \ldots$, i.i.d. with support $\{\ell, \ell + 1, \ldots\}$, let $(R_j, \ j \geq 0)$ be the *weak ascending record process* of $(X_j, \ j \geq 1)$. That is,

$$R_0 := \ell \quad \text{and} \quad R_j := X_{L_j} \text{ for } j \geq 1,$$

where $L_j$ is defined recursively by

$$L_1 := 1 \quad \text{and} \quad L_{j+1} := \min\{i > L_j : X_i \geq X_{L_j}\} \text{ for } j \geq 1.$$

The sequence $(R_j, \ j \geq 0)$ was first considered by Vervaat [54]; see also [9, Section 2.8] and [46, Lecture 15] for further discussion on records of discrete distributions.

It is known that $(R_j, \ j \geq 0)$ is a Markov chain with the transition probability function $r(m, n)$ given by

$$(2.13) \qquad r(m, n) = \frac{\mathbb{P}(X_1 = n)}{\mathbb{P}(X_1 \geq m)} \quad \text{for } m \leq n.$$

**Proposition 2.2.** *Let $(R_j, \ j \geq 0)$ be the weak ascending record process of the i.i.d. sequence $(X_j, \ j \geq 1)$ with $X_j \overset{(d)}{=} \widehat{Q}_1^{\ell, \theta}$; that is,*

$$\mathbb{P}(X_j = n) = \widehat{q}^\theta(\ell, n) \quad \text{for } n \geq \ell,$$

*where $\widehat{q}^\theta$ is defined by (2.9). Then there is the equality in joint distributions*

$$(2.14) \qquad (\widehat{Q}_j^{\ell, \theta}, j \geq 0) \overset{(d)}{=} (R_j, \ j \geq 0).$$

*Proof.* Observe that for $\ell \leq m \leq n$,

$$\widehat{q}^\theta(\ell, n) = \frac{(\ell)_{n-\ell}\, \theta}{(\theta + \ell)_{n-\ell+1}} = \frac{(\ell)_{m-\ell}}{(\theta + \ell)_{m-\ell}} \widehat{q}^\theta(m, n).$$

Sum this identity over $n \geq m$ to see that $\mathbb{P}(X_j \geq m) = (\ell)_{m-\ell}/(\theta + \ell)_{m-\ell}$, and hence that $\widehat{q}^\theta(m, \cdot)$ is the conditional distribution of $X_j$ given $X_j \geq m$, as required. $\quad\square$

It is known [52, Theorem 1.1] that the counts $G_i^{\ell, \theta}(\infty)$ of records at each possible value $i = \ell, \ell + 1, \ldots$, are independent and geometrically distributed on $\mathbb{N}_0 := \{0\} \cup \mathbb{N}_+$ with parameter $i/(i + \theta)$. Combined with Proposition 2.2, we get the following result which is a variant of [33, Proposition 5.1].

**Corollary 2.3.** *Let $C_k^{\ell, \theta}$ and $\widehat{C}_k^{\ell, \theta}$ be defined by (2.8) and (2.11). Then there is the increasing and almost sure convergence*

$$C_k^{\ell, \theta} \overset{(d)}{=} \widehat{C}_k^{\ell, \theta} \uparrow C_\infty^{\ell, \theta},$$

*along with convergence of all positive moments, where the probability generating function of $C_\infty^{\ell, \theta}$ is given by*

$$(2.15) \qquad F_{\ell, \theta}(z) := \mathbb{E} z^{C_\infty^{\ell, \theta}} = \frac{\Gamma(\ell + 1 + \theta)\Gamma(\ell + \theta - \theta z)}{\Gamma(\ell)\Gamma(\ell + 1 + 2\theta - \theta z)}.$$

*Consequently, the random variable $C_\infty^{\ell, \theta}$ has the mixed Poisson distribution with random parameter $-\theta \log H$, where $H$ has the beta$(\ell, \theta + 1)$ distribution.*

This result, combined with Lemma 2.1$(iii)$, leads to the formula (1.10):

$$u_\infty \overset{\theta}{=} \mathbb{P}(C_\infty^{1, \theta} = 0) = F_{1, \theta}(0) = \frac{\Gamma(\theta + 2)\Gamma(\theta + 1)}{\Gamma(2\theta + 2)}.$$

Also note that the random variable $C_\infty^{\ell, \theta}$ has a simple representation for $\theta \in \mathbb{N}_+$,

$$(2.16) \qquad C_\infty^{\ell, \theta} \overset{(d)}{=} \sum_{j=0}^{\theta} \mathcal{G}_j^{\ell, \theta},$$

where $\mathcal{G}_j^{\ell, \theta}$, $0 \leq j \leq \theta$ are independent and geometrically distributed on $\mathbb{N}_0$ with parameter $(\ell + j)/(\ell + j + \theta)$.

*Proof.* The identity (2.12) shows that

$$\widehat{C}_k^{\ell,\theta} \uparrow C_\infty^{\ell,\theta} := G_\ell^{\ell,\theta}(\infty) + \sum_{i=\ell+1}^{\infty} (G_i^{\ell,\theta}(\infty) - 1)^+ \quad a.s.,$$

where $G_i^{\ell,\theta}(\infty)$, $i \geq \ell$ are independent and geometrically distributed on $\mathbb{N}_0$ with parameter $p_{i,\theta} := i/(i+\theta)$. For $G$ geometrically distributed on $\mathbb{N}_0$ with parameter $p$,

$$\mathbb{E}z^G = \frac{p}{1 - (1-p)z} \quad \text{and} \quad \mathbb{E}z^{(G-1)^+} = p + \frac{(1-p)p}{1 - (1-p)z}.$$

As a result,

$$\begin{aligned}
\mathbb{E}z^{C_\infty^{\ell,\theta}} &= \frac{p_{\ell,\theta}}{1 - (1-p_{\ell,\theta})z} \prod_{i=\ell+1}^{\infty} \left( p_{i,\theta} + \frac{(1-p_{i,\theta})p_{i,\theta}}{1 - (1-p_{i,\theta})z} \right) \\
&= \frac{\ell}{\ell + \theta - \theta z} \prod_{i=\ell+1}^{\infty} \frac{i(i + 2\theta - \theta z)}{(i+\theta)(i+\theta-\theta z)} \\
&= \frac{\ell}{\ell + \theta - \theta z} \cdot \frac{\Gamma(\ell+1+\theta)\Gamma(\ell+1+\theta-\theta z)}{\Gamma(\ell+1)\Gamma(\ell+1+2\theta-\theta z)},
\end{aligned}$$

which leads to the formula (2.15). Recall that the generating function of the Poisson($u$) distribution is $e^{-u(1-z)}$, and that the Mellin transform of the beta($p,q$) variable $H_{p,q}$ is

$$\mathbb{E}H_{p,q}^\nu = \frac{\Gamma(\nu+p)\Gamma(p+q)}{\Gamma(p)\Gamma(\nu+p+q)} \quad \text{for } \nu > -p.$$

By taking $\nu = \theta(1-z)$, $p = \ell$, and $q = \theta+1$, we identify the distribution of $C_\infty^{\ell,\theta}$ with the stated mixed Poisson distribution. $\square$

Let $\psi(x) := \Gamma'(x)/\Gamma(x)$ be the digamma function, and $\psi^{(k)}(x)$ be the $k$th derivative of $\psi(x)$. For $k \geq 1$, define

$$(2.17) \qquad \Delta_{k,\ell,\theta}(z) := \psi^{(k-1)}(\ell + \theta - \theta z) - \psi^{(k-1)}(\ell + 1 + 2\theta - \theta z).$$

A simple calculation shows that $F_{\ell,\theta}'(z) = -\theta F_{\ell,\theta}(z)\Delta_{1,\ell,\theta}(z)$ and $\Delta_{k,\ell,\theta}'(z) = -\theta\Delta_{k+1,\ell,\theta}(z)$. By induction, the derivatives of $F_\theta$ can be written as

$$(2.18) \qquad F_{\ell,\theta}^{(k)}(z) = (-\theta)^k F_{\ell,\theta}(z) P_k(\Delta_{1,\ell,\theta}(z), \ldots, \Delta_{k,\ell,\theta}(z)),$$

where $P_k(x_1, \ldots, x_k)$ is the $k$th *complete Bell polynomial* [22, Section 3.3]. To illustrate,

$$\begin{aligned}
P_1(x_1) &= x_1, \\
P_2(x_1, x_2) &= x_1^2 + x_2, \\
P_3(x_1, x_2, x_3) &= x_1^3 + 3x_1 x_2 + x_3, \\
P_4(x_1, x_2, x_3, x_4) &= x_1^4 + 6x_1^2 x_2 + 4x_1 x_3 + 3x_2^2 + x_4, \\
P_5(x_1, x_2, x_3, x_4, x_5) &= x_1^5 + 10x_1^3 x_2 + 10x_1^2 x_3 + 15x_1 x_2^2 + 5x_1 x_4 + 10x_2 x_3 + x_5,
\end{aligned}$$

and so on. Now by expanding $F_{\ell,\theta}$ into power series at $z = 0$ and $z = 1$, we get

$$(2.19) \qquad \mathbb{P}(C_\infty^{\ell,\theta} = k) = \frac{(-\theta)^k}{k!} \frac{\Gamma(\ell+\theta)\Gamma(\ell+\theta+1)}{\Gamma(\ell)\Gamma(\ell+2\theta+1)} P_k(\Delta_{1,\ell,\theta}(0), \ldots, \Delta_{k,\ell,\theta}(0))$$

and

$$\mathbb{E}\binom{C_\infty^{\ell,\theta}}{k} = \frac{(-\theta)^k}{k!} P_k(\Delta_{1,\ell,\theta}(1), \dots, \Delta_{k,\ell,\theta}(1)), \tag{2.20}$$

where $\Delta_{k,\ell,\theta}(\cdot)$ is defined by (2.17). By taking $\theta = 1$ and $k = 1$ in (2.20), we get

$$\mathbb{E}C_\infty^{\ell,1} = \psi(\ell+2) - \psi(\ell) = \frac{1+2\ell}{\ell(\ell+1)}, \tag{2.21}$$

since $\psi(\ell) = \sum_{j=1}^{\ell-1} 1/j - \gamma$, with $\gamma$ the Euler constant.

## 3. Proof of Proposition 1.1

In this section we apply the results of Section 2 to evaluate the renewal sequence $(u_k)$ in the GEM(1) case, and extend to the general GEM($\theta$) case. The computation boils down to the study of the Markov chain $(\widehat{Q}_k^{\ell,\theta},\ k \geq 0)$ with $\ell = 1$. We start by proving Proposition 1.1, corresponding to the case where $\ell = 1$ and $\theta = 1$. To this end, we need the following duality formula due to Hoffman [35, Theorem 4.4] and Zagier [57, Section 9].

**Lemma 3.1.** *Let* $\zeta(s_1, \dots, s_k)$ *be the multiple zeta value defined by* (1.12). *Then*

$$\zeta(\underbrace{1, \dots, 1}_{k-1}, h+1) = \zeta(\underbrace{1, \dots, 1}_{h-1}, k+1) \quad \text{for all } h, k \in \mathbb{N}_+. \tag{3.1}$$

*In particular,*

$$\zeta(\underbrace{1, \dots, 1}_{k-2}, 2) = \zeta(k) \quad \text{for all } k \geq 2. \tag{3.2}$$

*Proof of Proposition* 1.1. By Lemma 2.1$(ii)$, for $k \geq 2$,

$$
\begin{aligned}
u_k &\overset{1}{=} \sum_{1 < n_1 < \cdots < n_k} \mathbb{P}(\widehat{Q}_1^{1,1} = n_1, \dots, \widehat{Q}_k^{1,1} = n_k) \\
&= \sum_{1 < n_1 < \cdots < n_k} \widehat{q}^1(1, n_1)\, \widehat{q}^1(n_1, n_2) \cdots \widehat{q}^1(n_{k-1}, n_k) \\
&= \sum_{1 < n_1 < \cdots < n_{k-1}} \frac{1}{(n_1+1) \cdots (n_{k-2}+1)(n_{k-1}+1)^2} \\
&= \sum_{0 < n_1 < \cdots < n_{k-1}} \frac{1}{(n_1+2) \cdots (n_{k-2}+2)(n_{k-1}+2)^2}.
\end{aligned}
$$

For $k \geq 2$ and $x \geq 0$, let
$$
\zeta(\nu_1, \dots, \nu_{k-1}; x) := \sum_{0 < n_1 < \cdots < n_{k-1}} \frac{1}{(n_1+x)^{\nu_1} \cdots (n_{k-2}+x)^{\nu_{k-2}}(n_{k-1}+x)^{\nu_{k-1}}} \tag{3.3}
$$

be the *multiple Hurwitz zeta function* [19], and $h_k(x) := \zeta(\underbrace{1, \dots, 1}_{k-2}, 2; x)$. Therefore

$$u_k \overset{1}{=} h_k(2) \text{ for } k \geq 2. \tag{3.4}$$

We claim that for $k \geq 3$,

$$\zeta(\nu_1, \dots, \nu_{k-1}; x-1) = \zeta(\nu_1, \dots, \nu_{k-1}; x) + x^{-\nu_1}\zeta(\nu_2, \dots, \nu_{k-1}; x). \tag{3.5}$$

In fact,

$$\zeta(\nu_1, \ldots, \nu_{k-1}; x-1)$$

$$= \sum_{0 < n_1 < \cdots < n_{k-1}} \frac{1}{(n_1 + x - 1)^{\nu_1} \cdots (n_{k-2} + x - 1)^{\nu_{k-2}}(n_{k-1} + x - 1)^{\nu_{k-1}}}$$

$$= \sum_{0 \leq n_1 < \cdots < n_{k-1}} \frac{1}{(n_1 + x)^{\nu_1} \cdots (n_{k-2} + x)^{\nu_{k-2}}(n_{k-1} + x)^{\nu_{k-1}}},$$

and writing this expression as two sums over the distinct sets $\{0 = n_1 < n_2 < \cdots < n_{k-1}\}$ and $\{0 < n_1 < n_2 < \cdots < n_{k-1}\}$ yields the formula (3.5). Consequently,

$$(3.6) \qquad h_k(x-1) = h_k(x) + x^{-1} h_{k-1}(x) \quad \text{for } k \geq 3.$$

By taking $x = 2$ and $x = 1$ in (3.6), we get for $k \geq 3$,

$$h_k(1) = h_k(2) + \frac{1}{2} h_{k-1}(2) \quad \text{and} \quad h_k(0) = h_k(1) + h_{k-1}(1),$$

which implies that for $k \geq 4$,

$$(3.7) \qquad h_k(0) = h_k(2) + \frac{3}{2} h_{k-1}(2) + \frac{1}{2} h_{k-2}(2).$$

According to the formula (3.2),

$$(3.8) \qquad h_k(0) = \zeta(\underbrace{1, \ldots, 1}_{k-2}, 2) = \zeta(k).$$

By (3.4), (3.7), and (3.8), we derive the recursion (1.6) for $k \geq 4$. Recall that by definition, we have $u_0 = 1$, and it is easy to check that $u_1 \overset{1}{=} 1/2$. By symbolic integration, we get

$$u_2 \overset{1}{=} -\frac{5}{4} + \zeta(2) \quad \text{and} \quad u_3 \overset{1}{=} \frac{13}{8} - \frac{3}{2}\zeta(2) + \zeta(3),$$

which satisfies the recursion for $k = 2, 3$. So the part $(i)$ of the proposition is proved. The equivalences $(i) \Leftrightarrow (ii) \Leftrightarrow (iv)$ are straightforward, and $(ii) \Leftrightarrow (iii)$ follow by partial fraction decomposition. We will see in Section 4 that the parts $(i)$ and $(iv)$ in Proposition 1.1 are valid for general recursions of the form $au_{k-2} + bu_{k-1} + cu_k = \zeta(k)$. $\qquad \square$

In what follows, we aim to extend the above calculation to general $\theta > 0$. It is easily seen that

$$u_k \overset{\theta}{=} \sum_{1 < n_1 < \cdots < n_k} \mathbb{P}(\widehat{Q}_1^{1,\theta} = n_1, \ldots, \widehat{Q}_k^{1,\theta} = n_k)$$

$$= \sum_{1 < n_1 < \cdots < n_k} \frac{\theta^k (n_k - 1)!}{(\theta + n_1) \cdots (\theta + n_{k-1})(\theta + 1)_{n_k}}$$

$$= \theta^k \sum_{0 < n_1 < \cdots < n_{k-1}} \frac{1}{(\theta + n_1 + 1) \cdots (\theta + n_{k-1} + 1)} \sum_{n_k > n_{k-1}} \frac{n_k!}{(\theta + 1)_{n_k + 1}}.$$

Note that for all $k \geq 0$,

$$
\begin{aligned}
\sum_{n \geq k} \frac{n!}{(\theta + 1)_{n+1}} &= \sum_{n \geq k} \frac{1}{\theta} \frac{(\theta + n + 1)n! - (n + 1)!}{(\theta + 1)_{n+1}} \\
&= \sum_{n \geq k} \frac{1}{\theta} \left( \frac{n!}{(\theta + 1)_n} - \frac{(n + 1)!}{(\theta + 1)_{n+1}} \right) \\
&= \frac{k!}{\theta (\theta + 1)_k} = \frac{\Gamma(\theta) \, \Gamma(k + 1)}{\Gamma(\theta + k + 1)}.
\end{aligned}
$$

Therefore,

$$(3.9) \quad u_k \overset{\theta}{=} \theta^k \Gamma(\theta) \sum_{0 < n_1 < \cdots < n_{k-1}} \frac{1}{(\theta + n_1 + 1) \cdots (\theta + n_{k-1} + 1)} \frac{\Gamma(n_{k-1} + 2)}{\Gamma(n_{k-1} + \theta + 2)}.$$

It seems to be difficult to simplify the expression (3.9) for general $\theta$. We focus on the case where $\theta \in \mathbb{N}_+$. Let

$$h_{k,\theta}(x) := \theta^k \Gamma(\theta) \sum_{0 < n_1 < \cdots < n_{k-1}} \frac{\Gamma(n_{k-1} + 1 - \theta + x)}{(n_1 + x) \cdots (n_{k-1} + x) \Gamma(n_{k-1} + 1 + x)},$$

so $u_k \overset{\theta}{=} h_{k,\theta}(\theta + 1)$. Again it is elementary to show that

$$h_{k,\theta}(x - 1) = h_{k,\theta}(x) + \frac{\theta}{x} h_{k-1,\theta}(x).$$

Consequently, the sequence $(u_k, \ k \geq 0)$ satisfies a $(\theta + 1)$-order recursion,

$$(3.10) \quad u_k + a_{1,\theta} u_{k-1} + \cdots + a_{\theta+1,\theta} u_{k-\theta-1} = h_{k,\theta}(0),$$

where

$$a_{i,\theta} := \sum_{0 < n_1 < \cdots < n_i \leq \theta + 1} \frac{\theta^i}{n_1 \cdots n_i} \quad \text{for } 1 \leq i \leq \theta + 1,$$

and $h_{k,\theta}(0)$ is a variant of the multiple Hurwitz zeta function.

## 4. Renewal sequences derived from the zeta function

Look at the sequence

$$(4.1) \quad u_k := \sum_{n=1}^{\infty} \frac{n^{-k}}{q(n)} \quad (k = 0, 1, 2, \ldots),$$

where

$$(4.2) \quad q(n) := an^2 + bn + c$$

is a generic quadratic function of $n$. We are interested in conditions on $q$ which allow the sequence $(u_k, k = 0, 1, \ldots)$ to be interpreted as a renewal sequence [28]. Basic requirements are that $q(n) > 0$ for all $n = 1, 2, \ldots$, so at least $a > 0$, and that $u_0 = 1$, which is a matter of normalization of coefficients of $q$. The sequence $1/q(n), \ n = 1, 2, \ldots$, then defines a probability distribution on the positive integers. If $X$ denotes a random variable with this distribution, so $\mathbb{P}(X = n) = 1/q(n)$, $n = 1, 2, \ldots$, then (4.1) becomes

$$(4.3) \quad u_k = \mathbb{E}(1/X)^k \quad (k = 0, 1, 2, \ldots).$$

That is to say, $u_k$ is the $k$th moment of the probability distribution of $1/X$ on $[0, 1]$. Obviously, $0 \leq u_k \leq 1$, and by the Cauchy-Schwartz inequality applied to

$$(1/X)^k = (1/X)^{(k-1)/2}(1/X)^{(k+1)/2},$$

$$(4.4) \qquad u_k^2 \leq u_{k-1}u_{k+1} \qquad (k = 1, 2, \ldots).$$

A sequence $(u_k)$ bounded between 0 and 1 with $u_0 = 1$ and subject to (4.4) is called a *Kaluza sequence* [37]. By a classical theorem of Kaluza, every such sequence is a *renewal sequence* [37]. See [51] for an elementary proof and further references. In view of Proposition 1.2, we are motivated to study such renewal sequences $(u_k)$ and the associated distribution $(f_k)$ of the time until first renewal, whose generating functions

$$(4.5) \qquad U(z) := \sum_{k=0}^{\infty} u_k z^k \text{ and } F(z) := \sum_{k=1}^{\infty} f_k z^k \qquad (|z| < 1)$$

are known [28] to be related by

$$(4.6) \qquad U(z) = (1 - F(z))^{-1} \text{ and } F(z) = 1 - U(z)^{-1}.$$

This identity of generating functions corresponds to the basic relation

$$(4.7) \qquad u_k = f_k + f_{k-1}u_1 + \cdots + f_1 u_{k-1} \qquad (k = 1, 2, \ldots),$$

which allows either of the sequences $(u_k)$ and $(f_k)$ to be derived from the other. Observe that the definition $q(n) = an^2 + bn + c$ gives

$$(4.8) \qquad \frac{cn^{-k}}{q(n)} + \frac{bn^{-(k-1)}}{q(n)} + \frac{an^{-(k-2)}}{q(n)} = n^{-k}$$

and hence, for $k \geq 2$,

$$(4.9) \qquad cu_k + bu_{k-1} + au_{k-2} = \sum_{n=1}^{\infty} n^{-k} = \zeta(k).$$

It follows that $U(z)$, and hence $F(z)$ can always be expressed in terms of the well known (see [2, formula 6.3.14]) generating function of $\zeta$ values

$$(4.10) \qquad G(z) := \sum_{n=2}^{\infty} \zeta(n)z^n = -z(\gamma + \psi(1 - z)) \qquad (|z| < 1),$$

where $\gamma$ is Euler's constant and $\psi(x) := \Gamma'(x)/\Gamma(x)$ is the digamma function, as

$$c(U(z) - u_0 - u_1 z) + bz(U(z) - u_0) + az^2 U(z) = G(z)$$

or

$$q(z)U(z) - c(u_0 + u_1 z) - u_0 bz = G(z),$$

which rearranges as

$$(4.11) \qquad U(z) = \frac{cu_0 + (bu_0 + cu_1)z + G(z)}{q(z)}.$$

Defining $r_1, r_2 \in \mathbb{C}$ as the two roots of $q$, we have

$$q(z) = a(z - r_1)(z - r_2),$$
$$b = -a(r_1 + r_2) \text{ and } c = ar_1 r_2.$$

Note that our assumption that $q(n) > 0$ for all $n = 1, 2, \ldots$, implies that the roots $r_1$ and $r_2$ are not positive integers. A straightforward computation shows that the condition $u_0 = 1$ implies

$$a = \begin{cases} \dfrac{\psi(1 - r_2) - \psi(1 - r_1)}{r_1 - r_2} & \text{if } r_1 \neq r_2, \\ \psi'(1 - r_1) & \text{if } r_1 = r_2, \end{cases}$$

and that we have

$$u_1 = \frac{1}{2c}\left(-b + 2\gamma + \psi(1 - r_1) + \psi(1 - r_2)\right).$$

Finally, obtaining $F(z)$ from (4.6) and (4.11) and taking derivatives gives us

(4.12)     $F'(1) = q(1),$

$F''(1) + q(1)(1 - q(1))$

(4.13)
$$= (4 - q(1))a - q(-1) + q(1)(c(1 + 2u_1) + b)$$
$$= (4 - q(1))a - q(-1) + q(1)(c + 2\gamma + \psi(1 - r_1) + \psi(1 - r_2)).$$

To summarize, and combine with some standard renewal theory, see the following proposition.

**Proposition 4.1.** *Let $q(n)$ be any quadratic function of $n = 1, 2, \ldots$, with $q(n) > 0$ for all $n$, normalized so that $u_0 := \sum_{n=1}^{\infty} 1/q(n) = 1$, and let $u_k := \sum_{n=1}^{\infty} n^{-k}/q(n)$ for $k \geq 1$. Then $(u_k)$ is a decreasing, positive recurrent renewal sequence, with*

$$\lim_{k \to \infty} u_k = 1/q(1).$$

*The corresponding distribution of an i.i.d. sequence of positive integer valued random variables $Y_1, Y_2, \ldots$, with $\mathbb{P}(Y_1 + \cdots + Y_m = k$ for some $m) = u_k$ has distribution with mean and variance*

$$\mathbb{E}(Y_1) = q(1),$$
$$Var(Y_1) = (4 - q(1))a - q(-1) + q(1)(c + 2\gamma + \psi(1 - r_1) + \psi(1 - r_2)),$$

*and probability generating function $F(z) := \mathbb{E}(z^{Y_1})$ given by (4.6) for $U(z)$ as in (4.11).*

**Example.** Take $a = 1/2, b = 3/2, c = 1$. Then $q(n) = (n+1)(n+2)/2$ makes $u_0 = 1$, $1/u_\infty = \mathbb{E}(Y_1) = q(1) = 3$, and $Var(Y_1) = 11$. From Proposition 1.1, we know that this sequence $(u_k)$ is the renewal sequence associated with a GEM(1) random partition of $[0, 1]$. Equivalently in the terminology of random permutations [51], $(u_k)$ is the renewal sequence of the *splitting times* of a GEM(1)-biased permutation. Therefore $Y_1$ is distributed as $T_1$, the first splitting time of $\Pi$ a GEM(1)-biased permutation. In particular, we have $\mathbb{E}(T_1) = 3$, $Var(T_1) = 11$.

## 5. DEVELOPMENT OF THE GEM(1) CASE

The Markov chain $(\widehat{Q}_k)$ described in Lemma 2.1, with uniform stick-breaking factors—i.e., in the GEM(1) case—was first studied by Erdős, Rényi, and Szüsz [26], where it appears as Engel's series derived from $U$ a uniform random variable on $(0, 1)$. More precisely, if $2 \leq q_1 \leq q_2 \leq \cdots$ is the unique random sequence of integers such that

$$U = \frac{1}{q_1} + \frac{1}{q_1 q_2} + \cdots + \frac{1}{q_1 q_2 \cdots q_n} + \cdots,$$

then we have

$$(q_i - 1, \, i \geq 1) \overset{(d)}{=} (\widehat{Q}_i, \, i \geq 1),$$

where we condition on $\widehat{Q}_0 = 1$.

Here we give explicit formulas for the distribution of $\widehat{Q}_k$ in terms of iterated harmonic sums and the Riemann zeta function. The transition probabilities $\widehat{q}(m, n) := \mathbb{P}(\widehat{Q}_{k+1} = n \mid \widehat{Q}_k = m)$ are given by

$$\widehat{q}(m, n) = \frac{m}{n(n+1)}.$$

Then the joint probability distribution of $\widehat{Q}_1, \ldots, \widehat{Q}_k$ is given by the formula

(5.1)
$$\mathbb{P}(\widehat{Q}_1 = n_1, \ldots, \widehat{Q}_{k-1} = n_{k-1}, \widehat{Q}_k = n_k) = \frac{1(1 \leq n_1 \leq \cdots \leq n_{k-1} \leq n_k)}{(n_1 + 1) \cdots (n_{k-1} + 1)(n_k + 1)n_k}.$$

It follows that for $k = 1, 2, \ldots$,

$$\mathbb{P}(\widehat{Q}_k = n) = \frac{1}{n(n+1)} \sum_{1 \leq n_1 \leq \cdots \leq n_{k-1} \leq n} \frac{1}{(n_1 + 1) \cdots (n_{k-1} + 1)}$$

(5.2)
$$= \frac{H^*_{k-1}(n+1) - H^*_{k-2}(n+1)}{n(n+1)},$$

where $H^*_{-1}(n) = 0, H^*_0(n) = 1$, and $H^*_k(n)$ for $k = 1, 2, \ldots$, is the $k$th *iterated harmonic sum* defined by

(5.3)
$$H^*_k(n) := \sum_{m=1}^{n} \frac{H^*_{k-1}(m)}{m} = \sum_{1 \leq n_1 \leq n_2 \leq \cdots \leq n_k \leq n} \frac{1}{n_1 n_2 \cdots n_k}.$$

In particular, $H^*_1(n) = H(n)$ is the $n$th harmonic number, and $H^*_2(n) = \sum_{m=1}^{n} H(m)/m$. Such *iterated* or *multiple* harmonic sums have attracted the attention of a number of authors [6, 7, 17]. Since (5.2) describes a probability distribution over $n \in \{1, 2, 3, \ldots\}$, we deduce by induction that

(5.4)
$$\sum_{n=1}^{\infty} \frac{H^*_k(n+1)}{n(n+1)} = k + 1 \quad (k \geq 0).$$

This identity has the probabilistic interpretation that for each $k = 1, 2, \ldots$, $n = 1, 2, \ldots$,

(5.5)
$$\mathbb{E}\left[ \sum_{j=1}^{k} 1(\widehat{Q}_j = n) \right] = \sum_{j=1}^{k} \mathbb{P}(\widehat{Q}_j = n) = \frac{H^*_{k-1}(n+1)}{n(n+1)},$$

where $\sum_{j=1}^{k} 1(\widehat{Q}_j = n)$ is the number of times $j$ with $1 \leq j \leq k$ that $\widehat{Q}_j$ has value $n$. It is easily seen that for each fixed $n$ the sequence $H^*_{k-1}(n)$ is increasing with limit $n$ as $k \to \infty$. So the limit version of (5.5) is for $n = 1, 2, \ldots$,

(5.6)
$$\mathbb{E}\left[ \sum_{j=1}^{\infty} 1(\widehat{Q}_j = n) \right] = \sum_{j=1}^{\infty} \mathbb{P}(\widehat{Q}_j = n) = \frac{1}{n}.$$

As observed by Rényi [53], the random variables $S_1, S_2, \ldots$, with $S_n := \sum_{j=1}^{\infty} 1(\widehat{Q}_j = n)$, are independent with the geometric distribution on $\{0, 1, 2, \ldots\}$ with parameter $1 - 1/(n+1)$:

$$(5.7) \qquad \mathbb{P}(S_n = s) = \frac{1}{(n+1)^s} \left( 1 - \frac{1}{n+1} \right) \quad (n = 1, 2, \ldots, s = 0, 1, \ldots)$$

which also implies (5.6).

Some known results about the iterated harmonic sums $H_k^*(n)$ can now be interpreted as features of the distributions of $\widehat{Q}_k$. Arakawa and Kaneko [7] defined the function

$$(5.8) \qquad \xi_m(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} t^{s-1} e^{-t} \frac{\mathrm{Li}_m(1 - e^{-t})}{(1 - e^{-t})} dt,$$

where $\mathrm{Li}_m$ is the polylogarithm function

$$\mathrm{Li}_m(z) := \sum_{n=1}^{\infty} z^n / n^m.$$

The integral converges for $\Re(s) > 0$, and the function $\xi_m(s)$ continues to an entire function of $s$. They showed that values of $\xi_m(s)$ for $s$ can be expressed in terms of multiple zeta values, and observed in particular that

$$\xi_1(s) = s\zeta(s+1),$$

which can readily be derived from (5.8) using the identity $\mathrm{Li}_1(1 - e^{-t}) = t$. Ohno [48] then showed that for positive integer $m$ and $k$,

$$(5.9) \qquad \xi_m(k) = \sum_{1 \leq n_1 \leq \cdots \leq n_k} \frac{1}{n_1 n_2 \cdots n_{k-1} n_k^{m+1}} = \sum_{n=1}^{\infty} \frac{H_{k-1}^*(n)}{n^{m+1}}.$$

That is, with the replacement $k \to k+1$ and taking $m = 1$,

$$(5.10) \qquad \sum_{n=1}^{\infty} \frac{H_k^*(n)}{n^2} = (k+1)\zeta(k+2) \quad (k = 0, 1, 2, \ldots).$$

Subtracting 1 from both sides of this identity gives a corresponding formula with summation from $n = 2$ to $\infty$ on the left. Comparing with the more elementary formula (5.4), and using

$$\frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)},$$

it follows that

$$(5.11) \qquad \sum_{n=1}^{\infty} \frac{H_k^*(n+1)}{n(n+1)^2} = k + 2 - (k+1)\zeta(k+2) \quad (k = 0, 1, 2, \ldots).$$

Plugged into formula (5.2) for the distribution of $\widehat{Q}_k$, this gives a formula for the first inverse moment of $\widehat{Q}_k + 1$,

$$(5.12) \qquad \mathbb{E}\left( \frac{1}{\widehat{Q}_k + 1} \right) = 1 - k\zeta(k+1) + (k-1)\zeta(k).$$

By this stage, we have reached some identities for multiple zeta values which cannot easily be verified symbolically using *Mathematica*, though they are readily

checked for modest values of $k$ to limits of numerical precision. The case $k = 1$ of (5.10) reduces to

$$\sum_{n=1}^{\infty} \frac{H(n)}{n^2} = 2\zeta(3),$$

which Borwein *et al.* [17] attribute to Euler, and which can be confirmed symbolically with *Mathematica*. The case $k = 2$ of (5.10) expands to

(5.13)
$$\sum_{n=1}^{\infty} \frac{H_2^*(n)}{n^2} = \frac{1}{2}\sum_{n=1}^{\infty} \frac{H^2(n)}{n^2} + \frac{1}{2}\sum_{n=1}^{\infty} \frac{H_2(n)}{n^2} = 3\zeta(4),$$

with $H_k(n) = \sum_{m=1}^{n} 1/m^k$, which can also be confirmed symbolically with *Mathematica*. The decomposition into power sums is found to be

(5.14)
$$\frac{1}{2}\sum_{n=1}^{\infty} \frac{H^2(n)}{n^2} = \frac{17}{8}\zeta(4),$$

(5.15)
$$\frac{1}{2}\sum_{n=1}^{\infty} \frac{H_2(n)}{n^2} = \frac{7}{8}\zeta(4),$$

and summing these two identities yields (5.13) as required. According to Borwein and Borwein [16], formula (5.14) was first discovered numerically by Enrico Au-Yeun, and provided impetus to the surge of effort at simplification of multiple Euler sums by Borwein and coauthors [11]. Borwein and Borwein's first rigorous proof of (5.14) was based on the integral identity

$$\frac{1}{\pi}\int_0^{\pi} \theta^2 \log^2(2\cos\theta/2)d\theta = \frac{11}{2}\zeta(4),$$

which they derived with Fourier analysis using Parseval's formula. Later [17], they gave a systematic account of evaluations of multiple harmonic sums, including (5.14) as an exemplar case. In particular, they made a systematic study of the Euler sums

(5.16)
$$s_h(k,s) := \sum_{n=1}^{\infty} \frac{H^k(n)}{(n+1)^s}$$

and

(5.17)
$$\sigma_h(k,s) := \sum_{n=1}^{\infty} \frac{H_k(n)}{(n+1)^s},$$

and a number of other similar sums. They proved a number of exact reductions of such sums to evaluations of the zeta function at integer arguments and established other such reductions beyond a reasonable doubt by numerical computation.

With this development, we can prove the following result for $C_k^{1,1}$, the number of empty intervals among the first $k$ intervals at the stopping time $n(k,1)$ when the first sample point falls outside the union $[0, R_k)$ of these intervals.

**Proposition 5.1.** *Let $C_k^{1,1}$ be defined by (2.8) for $\ell = 1$ and $\theta = 1$. Then*

(5.18)
$$\mathbb{E}C_k^{1,1} = \begin{cases} 1/2 + k - (k-1)\zeta(k) & \text{for } k \geq 2, \\ 1/2 & \text{for } k = 1. \end{cases}$$

*Proof.* According to (2.11),

$$\mathbb{E}C_k^{1,1} = \sum_{j=1}^{k} \mathbb{P}(\widehat{Q}_j^{1,1} = \widehat{Q}_{j-1}^{1,1}),$$

where for $j \geq 2$ we use (5.2) to evaluate $\mathbb{P}(\widehat{Q}_j^{1,1} = n)$. So for $j \geq 2$,

$$\mathbb{P}(\widehat{Q}_j^{1,1} = \widehat{Q}_{j-1}^{1,1}) = \sum_{n \geq 1} \mathbb{P}(\widehat{Q}_j^{1,1} = \widehat{Q}_{j-1}^{1,1} = n)$$

$$= \sum_{n \geq 1} \mathbb{P}(\widehat{Q}_{j-1}^{1,1} = n)\widehat{q}^1(n,n)$$

$$= \sum_{n \geq 1} \frac{H_{j-2}^*(n+1) - H_{j-3}^*(n+1)}{n(n+1)^2}$$

$$= \begin{cases} 1 - (j-1)\zeta(j) + (j-2)\zeta(j-1) & \text{for } j \geq 3, \\ 2 - \zeta(2) & \text{for } j = 2, \end{cases}$$

where the last equality follows from (5.11). Also note that

$$\mathbb{P}(\widehat{Q}_1^{1,1} = \widehat{Q}_0^{1,1}) = \widehat{q}^1(1,1) = 1/2.$$

The formulas (5.18) follow from the above computations. $\qquad\square$

Recall the interpretation of the binomial moments $\mathbb{E}\binom{C_k^{1,1}}{j}$ from [51, (6.2)]. The case $j = 1$ has been evaluated in (5.18). For $j = k$,

(5.19) $$\mathbb{E}\binom{C_k^{1,1}}{k} = \mathbb{P}(C_k^{1,1} = k) = \frac{1}{2^k},$$

and for $j = k - 1$,

$$\mathbb{E}\binom{C_k^{1,1}}{k-1} = \frac{k}{2^k} + \mathbb{P}(C_k^{1,1} = k - 1).$$

Note that

$$\mathbb{P}(C_k^{1,1} = k - 1) = \sum_{m=0}^{k-1} \sum_{n>1} \widehat{q}^1(1,1)^m \widehat{q}^1(1,n)\widehat{q}^1(n,n)^{k-1-m}$$

$$= \frac{1}{2^{k-1}} - 2\sum_{n \geq 1} \frac{1}{n(n+1)(n+2)^k}.$$

By partial fraction decomposition,

$$\frac{1}{n(n+1)(n+2)^k} = \frac{1}{2^k n} - \frac{1}{n+1} + \frac{2^k - 1}{2^k(n+2)} + \sum_{j=2}^{k} \frac{2^{k+1-j} - 1}{2^{k+1-j}(n+2)^j},$$

which leads to

$$\sum_{n \geq 1} \frac{1}{n(n+1)(n+2)^k} = \frac{k}{2^{k+1}} - k + 1 + \sum_{j=2}^{k} \frac{2^{k+1-j} - 1}{2^{k+1-j}}\zeta(j).$$

Therefore,

(5.20) $$\mathbb{E}\binom{C_k^{1,1}}{k-1} = 2k - 2 + \frac{1}{2^{k-1}} - \sum_{j=2}^{k} \frac{2^{k+1-j} - 1}{2^{k-j}}\zeta(j).$$

We have proved in Corollary 2.3 that the random variables $C_k^{1,1}$ converge in distribution as $k \to \infty$. Consequently,

$$\mathbb{E}\binom{C_k^{1,1}}{k-1} \longrightarrow 0 \quad \text{as } k \to \infty.$$

It is easily seen from the expression (5.20) that this is equivalent to the well known formula

$$\sum_{n=2}^{\infty} (\zeta(n) - 1) = 1.$$

But the formulas for other binomial moments seem to be difficult, even for $j = 2$. Generally, we are interested in the exact distribution of $C_k^{1,1}$ on $\{0, 1, \ldots, k\}$ for $k = 1, 2, \ldots$. Simple formulas are found for $k = 1, 2, 3, 4$ as displayed in the following table.

Table of $\mathbb{P}(C_k^{1,1} = j)$ with $0 \le j \le k$ for $k = 1, 2, 3, 4$.

| $k$ | $0$ | $1$ | $2$ | $3$ | $4$ |
|---|---|---|---|---|---|
| 1 | $\frac{1}{2}$ | $\frac{1}{2}$ | | | |
| 2 | $-\frac{5}{4} + \zeta(2)$ | $2 - \zeta(2)$ | $\frac{1}{4}$ | | |
| 3 | $\frac{13}{8} - \frac{3}{2}\zeta(2) + \zeta(3)$ | $-\frac{37}{8} + 3\zeta(2)$ | $\frac{31}{8} - \frac{3}{2}\zeta(2) - \zeta(3)$ | $\frac{1}{8}$ | |
| 4 | $-\frac{29}{16} + \frac{7}{4}\zeta(2) - \frac{3}{2}\zeta(3) + \zeta(4)$ | $\frac{57}{8} - \frac{21}{4}\zeta(2) + \frac{3}{2}\zeta(3)$ | $-\frac{41}{4} + \frac{21}{4}\zeta(2) + \frac{3}{2}\zeta(3)$ | $\frac{47}{8} - \frac{7}{4}\zeta(2) - \frac{3}{2}\zeta(3) - \zeta(4)$ | $\frac{1}{16}$ |

The details are left to the reader. Observe that up to $k = 4$, all of the point probabilities in the distribution of $C_k^{1,1}$ are rational linear combinations of zeta values. This is also true with $j = 0, k - 1, k$ for all $k$. We leave open the problem of finding an explicit formula for $\mathbb{P}(C_k^{1,1} = j)$ for general $j$ and $k$, but make the following conjecture.

**Conjecture 5.2.** *For each $k \ge 1$ and $0 \le j \le k$,*

$$\mathbb{P}(C_k^{1,1} = j) = q_{k,1} + \sum_{j=2}^{k} q_{k,j}\zeta(j),$$

*with $q_{k,j}$ rational numbers.*

## 6. EVALUATION OF $u_{2:n}$ AND ITS LIMIT

In this section we derive an explicit formula for $u_2$ in the general $\mathrm{GEM}(\theta)$ case, which is based on evaluation of the combinatorial expressions of $u_{2:n}$ given later in (6.2). In principle, the analysis of $u_{2:n}$ can be extended to $u_{k:n}$ for $k \ge 3$, but there will be an annoying proliferation of cases. Already for $k = 2$, it requires considerable care not to overcount or undercount the cases.

Let $\Pi_n$ be the partition of $[n]$ generated by a random permutation of $[n]$, or, more generally, by any consistent sequence of exchangeable random partitions of $[n]$, with *exchangeable partition probability function (EPPF) p*. See [50] for background. The function $p$ is a function of compositions $(n_1, \ldots, n_k)$ of $n$, which gives for every $m \ge n$ the probability $p(n_1, \ldots, n_k)$ that for each particular listing of elements of $[n]$ by a permutation, say $(x_1, \ldots, x_n)$, that the first $n_1$ elements $\{x_1, \ldots, x_{n_1}\}$ fall in one block of $\Pi_m$; and if $n_1 < n$, the next $n_2$ elements $\{x_{n_1+1}, \ldots, x_{n_1+n_2}\}$ fall in another block of $\Pi_m$; and if $n_1 + n_2 < n$, the next

$n_3$ elements $\{x_{n_1+n_2+1}, \ldots, x_{n_1+n_2+n_3}\}$ fall in a third block of $\Pi_m$, and so on. In other words, $p(n_1, \ldots, n_k)$ is the common probability, for every $m \geq n$, that the restriction of $\Pi_m$ to $[n]$ equals any particular partition of $[n]$ whose blocks are of sizes $n_1, \ldots, n_k$. For the Ewens $(\theta)$ model, there is the well known formula

$$p_\theta(n_1, \ldots, n_k) = \frac{\theta^{k-1}}{(1+\theta)_{n-1}} \prod_{i=1}^{k} (1)_{n_i-1}, \text{ where } n = n_1 + \cdots + n_k.$$

This formula for $\theta = 1$,

$$p_1(n_1, \ldots, n_k) = \frac{\prod_{i=1}^{k}(n_i - 1)!}{n!}, \text{ where } n = n_1 + \cdots + n_k,$$

corresponds to the case when $\Pi_n$ is the partition of $[n]$ generated by the cycles of a uniformly distributed random permutation of $[n]$. Then the denominator $n!$ is the number of permutations of $[n]$, while the product in the numerator is the obvious enumeration of the number of permutations of $n$ in which $[n_1]$ forms one cycle, $[n_1 + n_2] \setminus [n_1]$ forms a second cycle, and so on. Essential for the following arguments is the less obvious *consistency property* of uniform random permutations that $p_1(n_1, \ldots, n_k)$ is also, for every $m \geq n$, the probability that $[n_1]$ is the restriction to $[n]$ of one cycle of $\Pi_m$, $[n_1 + n_2] \setminus [n_1]$ is the restriction to $[n]$ of a second cycle of $\Pi_m$, and so on. This basic consistency property of random permutations allows the sequence of random partitions $\Pi_n$ of $[n]$ to be constructed according to the *Chinese Restaurant Process*, so the restriction of $\Pi_m$ to $[n]$ is $\Pi_n$ for every $n < m$.

Consider first for $n \geq 2$ the probability that the same block of $\Pi_n$ is discovered first in examining elements of $[n]$ from left to right as in examining elements of $[n]$ from right to left. This is the probability that $1$ and $n$ fall in the same block of $\Pi_n$. By exchangeability, this is the same as the probability that $1$ and $2$ fall in the same block, that is,

$$(6.1) \qquad u_{1:n} := \mathbb{P}(1 \text{ and } n \text{ in the same block}) = p(2) \overset{\theta}{=} p_\theta(2) = \frac{1}{1+\theta}.$$

Next, consider for $n \geq 3$ the probability $u_{2:n}$ that the union of the first two blocks found in sampling left to right equals the union of the first two blocks found in sampling right to left.

**Proposition 6.1.** *For each $n \geq 3$, and each exchangeable random partition $\Pi_n$ of $[n]$ with EPPF $p$,*

$$u_{2:n} = p(n) + (n-2)p(n-1,1) + \sum_{1<j<k<n} p(j-1+n-k, 2)$$

$$(6.2) \qquad + \sum_{j=1}^{n-1} p(j, n-j) + \sum_{1<j<k<n} p(j, n-k+1).$$

*Proof.* The terms in (6.2) are accounted for as follows:

- $p(n)$ is the probability that there is only one block, with $1$ and $n$ in this block.
- $(n-2)p(n-1,1)$ is the probability that there are only two blocks, with both $1$ and $n$ in this block, while the second block is a singleton, which may be $\{j\}$ for any one of the $n-2$ elements $j \in \{2, \ldots, n-1\}$.

- $\sum_{1<j<k<n} p(j-1+n-k, 2)$ is the sum of the probabilities that there are two or more blocks, with both 1 and $n$ in this block, with $j$ the first element and $k$ the last element of some second block, which is both the second block to appear from left to right and the second block to appear from right to left. The probability of this event determined by $1 < j < k < n$ is the probability that the set $[j] \cup ([n] \setminus [k-1])$ of $j+1+n-k$ elements is split by the partition into the two particular subsets $[j-1] \cup ([n] \setminus [k])$ and $\{j\} \cup \{k\}$ of $j-1+n-k$ and 2, respectively. Hence the $p(j-1+n-k, 2)$, by the exchangeability and consistency properties of the random partitions of various subsets of $[n]$.
- $\sum_{j=1}^{n-1} p(j, n-j)$ is the probability of the event that there are exactly two blocks $[j]$ and $[n] \setminus [j]$ for some $1 \le j < n$.
- $\sum_{1<j<k<n} p(j, n-k+1)$ is the sum of the probabilities that there are two or more blocks, with 1 and $n$ in different blocks, with $j$ the first element of the block containing $n$, which is the second block to appear from left to right, and $k$ the last element of the block containing 1, which is the second block to appear from right to left. The probability of this event determined by $1 < j < k < n$ is the probability that the set $[j] \cup ([n] \setminus [k-1])$ of $j+1+n-k$ elements is split by the partition into the two particular subsets $[j-1] \cup \{k\}$ and $\{j\} \cup ([n] - [k])$ of sizes $j$ and $n - k + 1$, respectively, and hence the $p(j, n-k+1)$, again by the exchangeability and consistency properties of the random partitions of various subsets of $[n]$. $\qquad \square$

In this classification of five kinds of terms contributing to the probability $u_{2:n}$, the first three kinds account for all cases in which 1 and $n$ fall in the same block, while the last two kinds account for all cases in which 1 and $n$ fall in different blocks. The double sum for the third kind of term is 0 unless $n \ge 4$, in which case it always simplifies to a single sum by grouping terms according to the value $h$ of $j - 1 + n - k$ :

$$(6.3) \qquad \sum_{1<j<k<n} p(j-1+n-k, 2) = \sum_{h=2}^{n-2} (h-1) p(h, 2)$$

$$(6.4) \qquad \overset{\theta}{=} [p_\theta(2) - p_\theta(n) - (n-2) p_\theta(n-1, 1)] \, p_\theta(2)$$

with further simplification as indicated by $\overset{\theta}{=}$ for the Ewens $(\theta)$ model. For the Ewens $(\theta)$ model, the $p(h, 2)$ in (6.3) becomes

$$p_\theta(h, 2) = \frac{\theta(1)_{h-1}}{(1+\theta)_{h+1}},$$

while the expression in (6.4) features

$$p_\theta(2) = \frac{1}{1+\theta}; \qquad p_\theta(n) = \frac{(1)_{n-1}}{(1+\theta)_{n-1}}; \qquad p_\theta(n-1, 1) = \frac{\theta(1)_{n-2}}{(1+\theta)_{n-1}}.$$

The evaluation $\overset{\theta}{=}$ in (6.4) is easily checked algebraically, by first checking it for $n = 4$, and then checking the equality of differences as $n$ is incremented. This evaluation (6.4) is an expression of the well known characteristic property of *noninterference* in the Ewens $(\theta)$ model, according to which, given that 1 and $n$ fall in the same block of some size $b$ with $2 \le b < n$, the remaining $n - b$ elements are partitioned according to the Ewens $(\theta)$ model for $n - b$ elements. The sums in (6.3) evaluate

the probability that 1 and $n$ fall in the same block, whose size $b$ is at most $n-2$, and that the partition of the remaining $n-b \geq 2$ elements puts the least of these elements in the same block as the greatest of these elements. For a general EPPF the probability that 1 and $n$ fall in the same block, whose size $b$ is at most $n-2$, is $p(2) - p(n) - (n-2)p(n-1,1)$, as in the first factor of (6.4) for $p = p_\theta$. For the Ewens $(\theta)$ model, given this event and the size $b \leq n-2$ of the block containing 1 and $n$, the probability that the remaining $n-b \geq 2$ elements have their least and greatest elements in the same block is just $p_\theta(2)$, regardless of the value of $b$, and hence the factorization in the expression of (6.4) for the Ewens $(\theta)$ model.

While the sum of the first three kinds of terms in (6.2) can be simplified as above in the Ewens $(\theta)$ model, even for $\theta = 1$ there is no comparable simplification for the sum of the last two kinds of terms in (6.2), representing the probability of the event that 1 and $n$ fall in different blocks, while the same union of the first two blocks is found by examining elements from left to right as in examining elements from right to left. Asymptotics as $n \to \infty$ are easy for the sum of the first three kinds of terms in (6.2). The limit of the contribution of these three terms is $p_\theta(2)^2 = (1+\theta)^{-2} \overset{1}{=} 1/4$. As for the remaining two kinds of terms, it is obvious that $\sum_{j=1}^{n-1} p(j, n-j) \to 0$ for any partition structure, since this is the probability that there are only two classes, and all elements of one class appear in a sample of size $n$ before all members of the other class. So for the Ewens $(\theta)$ model this gives

$$\lim_{n\to\infty} u_{2:n} \overset{\theta}{=} (1+\theta)^{-2} + \lim_{n\to\infty} \sum_{1<j<k<n} \frac{\theta(1)_{j-1}(1)_{n-k}}{(1+\theta)_{j+n-k}}$$

$$(6.5) \qquad = (1+\theta)^{-2} + \sum_{j=2}^{\infty} \frac{\theta(1)_{j-1}}{(\theta+j-1)(\theta+1)_j},$$

where the limit can also be evaluated as an integral with respect to the joint distribution of $P_1 = W_1$ and $P_2 = (1-W_1)W_2$ for $W_i$ independent beta $(1,\theta)$ variables. We can write

$$(6.6) \qquad u_2 \overset{\theta}{=} \frac{1}{(\theta+1)^2} + \frac{1}{\theta+1}\left( {}_3F_2\left(\begin{matrix}1,1,\theta\\\theta+1,\theta+2\end{matrix}\bigg|1\right) - 1\right)$$

in terms of the generalized hypergeometric function ${}_3F_2$. Lima [43, Lemma 1] gives the following formula for *Catalan's constant* $G$:

$$(6.7) \qquad \frac{1}{2}{}_3F_2\left(\begin{matrix}1,1,\frac{1}{2}\\\frac{3}{2},\frac{3}{2}\end{matrix}\bigg|1\right) = G = \beta(2),$$

where $\beta(s) := \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n+1)^s}$ for $s > 0$. By manipulating the hypergeometric ${}_3F_2$ function, one can see that for $\theta = n + 1/2$ where $n$ is an integer, $u_2$ is of the form $q + rG$, where $q$ and $r$ are rational numbers. Lima's articles [41, 42] contain many related formulas, as well as references to zeta and beta values.

## Acknowledgment

We thank David Aldous for various pointers to the literature.

## References

[1] Sergei Abramovich and Yakov Yu. Nikitin, On the Probability of Co-primality of two Natural Numbers Chosen at Random: From Euler identity to Haar Measure on the Ring of Adeles. *Bernoulli News*, 24(1):7–13, 2017.

[2] *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*, Edited by Milton Abramowitz and Irene A. Stegun, Dover Publications, Inc., New York, 1966. MR0208797

[3] Shigeki Akiyama, Shigeki Egami, and Yoshio Tanigawa, *Analytic continuation of multiple zeta-functions and their values at non-positive integers*, Acta Arith. **98** (2001), no. 2, 107–116, DOI 10.4064/aa98-2-1. MR1831604

[4] David Aldous and J. Michael Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 1–72, DOI 10.1007/978-3-662-09444-0_1. MR2023650

[5] Kenneth S. Alexander, Kenneth Baclawski, and Gian-Carlo Rota, *A stochastic interpretation of the Riemann zeta function*, Proc. Natl. Acad. Sci. U.S.A. **90** (1993), no. 2, 697–699, DOI 10.1073/pnas.90.2.697. MR1199075

[6] Takashi Aoki, Yasuhiro Kombu, and Yasuo Ohno, *A generating function for sums of multiple zeta values and its applications*, Proc. Amer. Math. Soc. **136** (2008), no. 2, 387–395, DOI 10.1090/S0002-9939-07-09175-7. MR2358475

[7] Tsuneo Arakawa and Masanobu Kaneko, *Multiple zeta values, poly-Bernoulli numbers, and related zeta functions*, Nagoya Math. J. **153** (1999), 189–209, DOI 10.1017/S0027763000006954. MR1684557

[8] Louis-Pierre Arguin, David Belius, and Adam J. Harper, *Maxima of a randomized Riemann zeta function, and branching random walks*, Ann. Appl. Probab. **27** (2017), no. 1, 178–215, DOI 10.1214/16-AAP1201. MR3619786

[9] Barry C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *Records*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1998. A Wiley-Interscience Publication. MR1628157

[10] Richard Arratia, A. D. Barbour, and Simon Tavaré, *Logarithmic combinatorial structures: a probabilistic approach*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2003. MR2032426

[11] David H. Bailey, Jonathan M. Borwein, and Roland Girgensohn, *Experimental evaluation of Euler sums*, Experiment. Math. **3** (1994), no. 1, 17–30. MR1302815

[12] M Berry, Riemann's Zeta function: A model for quantum chaos? In *Quantum chaos and statistical nuclear physics*, pages 1–17. Springer, 1986.

[13] F. Beukers, *A note on the irrationality of $\zeta(2)$ and $\zeta(3)$*, Bull. London Math. Soc. **11** (1979), no. 3, 268–272, DOI 10.1112/blms/11.3.268. MR554391

[14] Philippe Biane, Jim Pitman, and Marc Yor, *Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions*, Bull. Amer. Math. Soc. (N.S.) **38** (2001), no. 4, 435–465, DOI 10.1090/S0273-0979-01-00912-0. MR1848256

[15] Enrico Bombieri, *The Riemann hypothesis*, The millennium prize problems, Clay Math. Inst., Cambridge, MA, 2006, pp. 107–124. MR2238277

[16] David Borwein and Jonathan M. Borwein, *On an intriguing integral and some series related to $\zeta(4)$*, Proc. Amer. Math. Soc. **123** (1995), no. 4, 1191–1198, DOI 10.2307/2160718. MR1231029

[17] David Borwein, Jonathan M. Borwein, and Roland Girgensohn, *Explicit evaluation of Euler sums*, Proc. Edinburgh Math. Soc. (2) **38** (1995), no. 2, 277–294, DOI 10.1017/S0013091500019088. MR1335874

[18] Jonathan M. Borwein, David M. Bradley, David J. Broadhurst, and Petr Lisoněk, *Combinatorial aspects of multiple zeta values*, Electron. J. Combin. **5** (1998), Research Paper 38, 12. MR1637378

[19] Olivier Bouillot, *On Hurwitz multizeta functions*, Adv. in Appl. Math. **71** (2015), 68–124, DOI 10.1016/j.aam.2015.06.003. MR3406959

[20] José Ignacio Burgos Gil and Javier Fresán, Multiple zeta values: from numbers to motives. Available at `http://javier.fresan.perso.math.cnrs.fr/mzv.pdf`.

[21] Joseph T. Chang and Yuval Peres, *Ladder heights, Gaussian random walks and the Riemann zeta function*, Ann. Probab. **25** (1997), no. 2, 787–802, DOI 10.1214/aop/1024404419. MR1434126

[22] Louis Comtet, *Advanced combinatorics*, Revised and enlarged edition, D. Reidel Publishing Co., Dordrecht, 1974. The art of finite and infinite expansions. MR0460128

[23] Peter Donnelly and Geoffrey Grimmett, *On the asymptotic distribution of large prime factors*, J. London Math. Soc. (2) **47** (1993), no. 3, 395–404, DOI 10.1112/jlms/s2-47.3.395. MR1214904

[24] H. M. Edwards, *Riemann's zeta function*, Dover Publications, Inc., Mineola, NY, 2001. Reprint of the 1974 original [Academic Press, New York; MR0466039 (57 #5922)]. MR1854455

[25] Torsten Ekedahl and Gerard van der Geer, *Cycles representing the top Chern class of the Hodge bundle on the moduli space of abelian varieties*, Duke Math. J. **129** (2005), no. 1, 187–199, DOI 10.1215/S0012-7094-04-12917-3. MR2155061

[26] P. Erdős, A. Rényi, and P. Szüsz, *On Engel's and Sylvester's series*, Ann. Univ. Sci. Budapest. Eötvös. Sect. Math. **1** (1958), 7–32. MR0102496

[27] C. J. A. Evelyn and E. H. Linfoot, *On a problem in the additive theory of numbers*, Ann. of Math. (2) **32** (1931), no. 2, 261–270, DOI 10.2307/1968190. MR1502996

[28] William Feller, *An introduction to probability theory and its applications. Vol. I*, Third edition, John Wiley & Sons, Inc., New York-London-Sydney, 1968. MR0228020

[29] A. M. Frieze, *On the value of a random minimum spanning tree problem*, Discrete Appl. Math. **10** (1985), no. 1, 47–56, DOI 10.1016/0166-218X(85)90058-7. MR770868

[30] Alexander Gnedin, Alex Iksanov, and Uwe Roesler, *Small parts in the Bernoulli sieve*, Fifth Colloquium on Mathematics and Computer Science, Discrete Math. Theor. Comput. Sci. Proc., AI, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008, pp. 235–242. MR2508790

[31] Alexander Gnedin, Alexander Iksanov, and Alexander Marynych, *The Bernoulli sieve: an overview*, 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10), Discrete Math. Theor. Comput. Sci. Proc., AM, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2010, pp. 329–341. MR2735350

[32] Alexander V. Gnedin, *The Bernoulli sieve*, Bernoulli **10** (2004), no. 1, 79–96, DOI 10.3150/bj/1077544604. MR2044594

[33] Alexander V. Gnedin, Alexander M. Iksanov, Pavlo Negadajlov, and Uwe Rösler, *The Bernoulli sieve revisited*, Ann. Appl. Probab. **19** (2009), no. 4, 1634–1655, DOI 10.1214/08-AAP592. MR2538083

[34] Allan Gut, *Some remarks on the Riemann zeta distribution*, Rev. Roumaine Math. Pures Appl. **51** (2006), no. 2, 205–217. MR2275304

[35] Michael E. Hoffman, *Multiple harmonic series*, Pacific J. Math. **152** (1992), no. 2, 275–290. MR1141796

[36] Michael E. Hoffman, *Algebraic aspects of multiple zeta values*, Zeta functions, topology and quantum physics, Dev. Math., vol. 14, Springer, New York, 2005, pp. 51–73, DOI 10.1007/0-387-24981-8_4. MR2179272

[37] Th. Kaluza, *Über die Koeffizienten reziproker Potenzreihen* (German), Math. Z. **28** (1928), no. 1, 161–170, DOI 10.1007/BF01181155. MR1544949

[38] Nicholas M. Katz and Peter Sarnak, *Zeroes of zeta functions and symmetry*, Bull. Amer. Math. Soc. (N.S.) **36** (1999), no. 1, 1–26, DOI 10.1090/S0273-0979-99-00766-1. MR1640151

[39] Klaus Kirsten, *Basic zeta functions and some applications in physics*, A window into zeta and modular physics, Math. Sci. Res. Inst. Publ., vol. 57, Cambridge Univ. Press, Cambridge, 2010, pp. 101–143. MR2648362

[40] Xian-Jin Li, *The positivity of a sequence of numbers and the Riemann hypothesis*, J. Number Theory **65** (1997), no. 2, 325–333, DOI 10.1006/jnth.1997.2137. MR1462847

[41] F. M. S. Lima, *An Euler-type formula for $\beta(2n)$ and closed-form expressions for a class of zeta series*, Integral Transforms Spec. Funct. **23** (2012), no. 9, 649–657, DOI 10.1080/10652469.2011.622274. MR2968884

[42] F. M. S. Lima, *A simpler proof of a Katsurada's theorem and rapidly converging series for $\zeta(2n + 1)$ and $\beta(2n)$*, Ann. Mat. Pura Appl. (4) **194** (2015), no. 4, 1015–1024, DOI 10.1007/s10231-014-0409-3. MR3357692

[43] F. M. S. Lima, *A rapidly converging ramanujan-type series for catalan's constant*. arXiv:1207.3139 (2012).

[44] Gwo Dong Lin and Chin-Yuan Hu, *The Riemann zeta distribution*, Bernoulli **7** (2001), no. 5, 817–828, DOI 10.2307/3318543. MR1867083

[45] H. L. Montgomery, *The pair correlation of zeros of the zeta function*, Analytic number theory (Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., St. Louis, Mo., 1972), Amer. Math. Soc., Providence, R.I., 1973, pp. 181–193. MR0337821

[46] Valery B. Nevzorov, *Records: mathematical theory*, Translations of Mathematical Monographs, vol. 194, American Mathematical Society, Providence, RI, 2001. Translated from the Russian manuscript by D. M. Chibisov. MR1791071

[47] A. M. Odlyzko, *The $10^{22}nd$ zero of the Riemann zeta function*, Dynamical, spectral, and arithmetic zeta functions (San Antonio, TX, 1999), Contemp. Math., vol. 290, Amer. Math. Soc., Providence, RI, 2001, pp. 139–144, DOI 10.1090/conm/290/04578. MR1868473

[48] Yasuo Ohno, *A generalization of the duality and sum formulas on the multiple zeta values*, J. Number Theory **74** (1999), no. 1, 39–43, DOI 10.1006/jnth.1998.2314. MR1670544

[49] Jim Pitman. Extremes and gaps in sampling from a residual allocation model (to be published).

[50] J. Pitman, *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002; With a foreword by Jean Picard. MR2245368

[51] Jim Pitman and Wenpin Tang, *Regenerative random permutations of integers*. arXiv:1704.01166 (2017).

[52] Jim Pitman and Yuri Yakubovich, *Extremes and gaps in sampling from a GEM random discrete distribution*, Electron. J. Probab. **22** (2017), Paper No. 44, 26, DOI 10.1214/17-EJP59. MR3646070

[53] A. Rényi, *A new approach to the theory of Engel's series*, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. **5** (1962), 25–32. MR0150123

[54] Wim Vervaat, *Limit theorems for records from discrete distributions*, Stochastic Processes Appl. **1** (1973), 317–334. MR0362457

[55] David Williams, *Brownian motion and the Riemann zeta-function*, Disorder in physical systems, Oxford Sci. Publ., Oxford Univ. Press, New York, 1990, pp. 361–372. MR1064572

[56] Edward Witten, *On quantum gauge theories in two dimensions*, Comm. Math. Phys. **141** (1991), no. 1, 153–209. MR1133264

[57] Don Zagier, *Values of zeta functions and their applications*, First European Congress of Mathematics, Vol. II (Paris, 1992), Progr. Math., vol. 120, Birkhäuser, Basel, 1994, pp. 497–512. MR1341859

[58] Jianqiang Zhao, *Analytic continuation of multiple zeta functions*, Proc. Amer. Math. Soc. **128** (2000), no. 5, 1275–1283, DOI 10.1090/S0002-9939-99-05398-8. MR1670846

Collége de France, 11 place Marcelin Berthelot, 75231 Paris
*Email address*: `jean-jil.duchamps@normalesup.org`

Statistics Department, University of California, 367 Evans Hall, Berkeley, California 94720
*Email address*: `pitman@stat.berkeley.edu`

Statistics Department, University of California, 367 Evans Hall, Berkeley, California 94720
*Email address*: `wenpintang@stat.berkeley.edu`