

## On the dependence of the component counting process of a uniform random variable

Joseph Squillace\*

### Abstract

We are concerned with the general problem of proving the existence of joint distributions of two discrete random variables  $M$  and  $N$  subject to infinitely many constraints of the form  $\mathbb{P}(M = i, N = j) = 0$ . In particular, the variable  $M$  has a countably infinite range and the other variable  $N$  is uniformly distributed with finite range. The constraints placed on the joint distributions will require, for most elements  $j$  in the range of  $N$ ,  $\mathbb{P}(M = i, N = j) = 0$  for infinitely many values of  $i$  in the range of  $M$ , where the corresponding values of  $i$  depend on  $j$ . To prove the existence of such joint distributions, we apply a theorem proved by Strassen on the existence of joint distributions with prespecified marginal distributions.

We consider some combinatorial structures that can be decomposed into components. Given  $n \in \mathbb{N}$ , consider an assembly, multiset, or selection  $A_n$  among elements of  $\{1, 2, \dots, n\}$ , and consider a uniformly distributed random variable  $N(n)$  on  $A_n$ . For each  $i \leq n$ , denote by  $C_i(n)$  the number of components of  $N(n)$  of size  $i$  so that  $\sum_{i \leq n} iC_i(n) = n$ . In each of these combinatorial structures, there exists infinitely many processes  $((Z_i(n, x))_{i \leq n})_x$ , indexed by a real parameter  $x$ , consisting of non-negative independent variables  $(Z_i(n, x))_{i \leq n}$  such that the distribution of the vector  $(C_i(n))_{i \leq n}$  equals the distribution of the vector  $(Z_i(n, x))_{i \leq n}$  conditional on the event  $\{\sum_{i \leq n} iZ_i(n, x) = n\}$ . Let  $M(n, x)$  denote a random variable whose components are given by  $(Z_i(n, x))_{i \leq n}$ . We introduce the notion of pivot mass which is then combined with Strassen's work to provide couplings of  $M(n, x)$  and  $N(n)$  with desired properties. For each of these combinatorial structures, we prove that there exists a real number  $x(n)$  for which we can couple  $M(n, x)$  and  $N(n)$  with  $\sum_{i \leq n} (C_i(n) - Z_i(n, x))^+ \leq 1$  when  $x > x(n)$ . We are providing a partial answer to the question "how much dependence is there in the process  $(C_i(n))_{i \leq n}$ ?"

**Keywords:** component counting process; couplings; marriage theorem; pivot mass.

**MSC2020 subject classifications:** 60C05.

Submitted to ECP on May 10, 2020, final version accepted on May 2, 2021.

Supersedes arXiv:1910.12841.

## 1 Introduction

Our results regard the component counting process of a uniform random variable in a combinatorial structure, and these results are provided by establishing the existence of couplings of random variables.

---

\*University of Rhode Island, United States of America. E-mail: josephps@uri.edu

**Definition 1.1.** Let  $X$  and  $Y$  be random variables defined on probability spaces  $(\Omega_X, \mathcal{F}_X, \mathbb{P}_X)$  and  $(\Omega_Y, \mathcal{F}_Y, \mathbb{P}_Y)$ .<sup>1</sup> A **coupling** of  $X$  and  $Y$  is a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  in which there exists random variables  $X'$  and  $Y'$  such that  $X'$  has the same distribution as  $X$  and  $Y'$  has the same distribution as  $Y$ .<sup>2</sup>

Given a natural number  $n$ , this paper is concerned with combinatorial structures  $A_n$  on the set  $[n] := \{1, 2, \dots, n\}$  that can be decomposed into components (e.g., the components of a permutation are its cycles). The combinatorial structures considered in this paper are known as assemblies, multisets and selections, and these structures are outlined in §1.1. Given a uniform random variable  $N(n) \sim \text{Unif}(A_n)$ , let  $C_i(n)$  denote the number of components of  $N(n)$  of size  $i$ ,  $1 \leq i \leq n$ . The sequence  $(C_i(n))_{i \leq n}$  is called the **component counting process** of  $N(n)$ ; the variables  $C_i(n)$ ,  $1 \leq i \leq n$ , are dependent since  $\sum_{i \leq n} iC_i(n) = n$ . Our goal is to measure the amount of dependence in this sequence by showing that it is dominated by a related sequence  $(Z_i)_i$  of independent variables  $Z_i$  in the following sense.<sup>3</sup>

**Definition 1.2.** Let  $n \in \mathbb{N}$ . Given random variables  $M(n)$  and  $N(n)$  with components  $Z_i(n)$  and  $C_i(n)$ ,  $1 \leq i \leq n$ , respectively, the variable  $M(n)$   **$d$ -dominates**  $N(n)$  if there exists a coupling of  $M(n)$  and  $N(n)$  such that

$$\sum_{i \leq n} (C_i(n) - Z_i(n))^+ \leq d, \quad \text{always.} \tag{1.1}$$

Condition (1.1) is equivalent to the equation  $\mathbb{P}\left(\sum_{i \leq n} (C_i(n) - Z_i(n))^+ \leq d\right) = 1$ . The goal of this paper is to show that a uniform variable  $N$  is 1-dominated by a particular variable  $M$ .

There is a natural combinatorial setup [3], with an  $N(n)$  based on uniformly chosen instances, and related exponential families of independent variables  $Z_i(n, x)$ ,  $1 \leq i \leq n$ ; the free parameter  $x$  [3, 4] corresponds to Cramer tilting. The combinatorial  $N(n)$  and the independent variables  $Z_i(n, x)$  are related by the conditioning relation, given by equation (1.2) below. Let  $M(n, x)$  denote a random variable with  $i$  parts of size  $Z_i(n, x)$ . Now as  $x$  increases (to infinity for assemblies, to 1 for multisets and selections), the  $Z_i$  converge in distribution to the supremum of their support. Thus under any coupling, the terms  $(C_i(n) - Z_i(n, x))^+$  converge in distribution to 0 and  $\mathbb{P}\left(\sum_{i \leq n} (C_i(n) - Z_i(n, x))^+ \leq 1\right)$  tends to 1, where  $(\cdot)^+$  denotes the positive part. The main result of this paper, Theorem 1.4, strengthens this fact since it provides the existence of a coupling of  $N(n)$  and  $M(n, x)$  such that  $M(n, x)$  1-dominates  $N(n)$  for sufficiently large values of  $x$ . Moreover, it will be shown that  $M(n, x)$  does not 0-dominate  $N(n)$  (i.e., we do not have  $Z_i(n, x) \leq C_i(n)$  for all  $i \leq n$ ).

The notion of  $d$ -domination is motivated by a conjecture proposed by Richard Arratia in §2.2 of [1], which we now describe. Consider a uniformly distributed variable  $N(n) \in [n]$  with prime factorization  $N(n) = \prod_{p \leq n} p^{C_p(n)}$ . It can be shown that the prime power process  $(C_p(n))_{p \leq n}$  converges in distribution to a process  $(Z_p)_{p \leq n}$  of independent variables where  $Z_p$  is a geometric random variable of parameter  $\frac{1}{p}$  and range  $\mathbb{Z}_{\geq 0}$ , for each prime  $p \leq n$ . Defining  $M(n) = \prod_{p \leq n} p^{Z_p}$ , we state Arratia's conjecture.

**Conjecture 1.3** (Arratia). For all  $n \geq 1$ , it is possible to construct  $N(n)$  uniformly

<sup>1</sup>The probability measure  $\mathbb{P}_X$  is defined by  $\mathbb{P}_X(i) = \mathbb{P}(X = i)$ .

<sup>2</sup>For each of the random variables  $X$  considered in this paper,  $X$  and  $X'$  will share the same range. Thus, when describing a particular coupling of  $X$  and  $Y$ , we often write  $X$  and  $Y$  instead of  $X'$  and  $Y'$ , respectively.

<sup>3</sup>Note that only component counting processes  $(C_i(n))_{i \leq n}$  satisfy the condition  $\sum_{i \leq n} iC_i(n) = n$  for all  $n \in \mathbb{N}$ . In general, variables denoted by  $Z_i(n)$ ,  $i \in \mathbb{N}$ , will be independent, so we cannot assume  $\sum_{i \leq n} iZ_i(n) = n$  for any  $n \in \mathbb{N}$ .

distributed from 1 to  $n$ ,  $M(n)$  and a prime  $P(n)$  such that

$$\text{always } N(n) \text{ divides } M(n)P(n).$$

Equivalently, the conjecture states that there exists a coupling of  $M(n)$  and  $N(n)$  such that we always have  $\sum_{p \leq n} (C_p(n) - Z_p)^+ \leq 1$ , and this was the motivation for the definition of  $d$ -domination. However, Arratia's conjecture does not fall within the combinatorial settings considered in this paper due to the fact that the prime power process  $(C_p(n))_{p \leq n}$  does not satisfy the equation  $\sum_{p \leq n} pC_p(n) = n$ .

The combinatorial structures listed in §1.1 provide the frameworks in which we obtain our couplings. Theorem 1.4, the main result of this paper, is stated in §1.2. In §2, we describe how our constraints force a significant proportion of the entries of a prospective joint mass distribution of our variables to be 0. In §3, we introduce the notion of pivot mass, which depends on the constraints placed on the desired joint distribution. Some properties of the pivot mass are proved in §3 and §4. In §5, we apply results on the pivot mass and a theorem proved by Strassen to prove Theorem 1.4, thereby proving the existence of our couplings.

### 1.1 Three major combinatorial structures

All couplings constructed in this paper involve a uniform random variable in any one of the following three combinatorial classes. An **assembly**  $A_n$  is an example of a combinatorial structure in which the set  $[n]$  is partitioned into blocks and for each block of size  $i$  one of  $m_i$  possible structures is chosen. A **multiset**  $A_n$  is a pair  $([n], m)$ , where  $m : A \rightarrow \mathbb{N}$  is a function that gives the multiplicity  $m(a)$  of each element  $a \in [n]$ . Equivalently (see Meta-example 2.2 of §2.2 of [2]), the integer  $n$  is partitioned into parts, and for each part of size  $i$ , one of the  $m_i$  objects of weight  $i$  is chosen. **Selections** are similar to multisets, but now we require all parts to be distinct. To simplify the notation, let us define  $k_n := \#A_n$  for each of these structures.

### 1.2 Couplings of random variables

In each of the assembly, multiset, and selection settings, our methods of arriving at our desired couplings are similar. We start by considering  $N(n) \sim \text{Unif}(A_n)$ . Given  $i \leq n$ , if we denote by  $C_i(n)$  the number components of  $N(n)$  of size  $i$ , then  $0 \leq C_i(n) \leq n$  and  $\sum_{i \leq n} iC_i(n) = n$ . In these combinatorial settings, there exists an infinite family  $((Z_i(n, x))_{i \leq n})_x$ , parametrized by positive values of  $x$  (specifically,  $x > 0$  for assemblies,  $x \in (0, 1)$  for multisets, and  $x \in (0, \infty)$  for selections) of infinite sequences  $(Z_i(n, x))_{i \leq n}$  of nonnegative integer-valued independent random variables  $Z_i(n, x)$  for which

$$\mathcal{L}(C_1(n), \dots, C_n(n)) = \mathcal{L}\left(Z_1(n, x), \dots, Z_n(n, x) \mid \sum_{i \leq n} iZ_i(n, x) = n\right) \quad (1.2)$$

(§2.3 of [2]). Equation (1.2) states that the joint distribution of the vector  $(C_i(n))_{i \leq n}$  is equal to the joint distribution of the vector  $(Z_i(n, x))_{i \leq n}$  conditional on the event  $\left\{\sum_{i \leq n} iZ_i(n, x) = n\right\}$ . For a fixed  $x$ , we consider another random variable  $M(n, x)$  whose component counts<sup>4</sup> are given by  $(Z_i(n, x))_{i \leq n}$ , so the distribution of  $M(n, x)$  is determined by the independent process  $(Z_i(n, x))_{i \leq n}$ .

The main result of this paper is the following theorem which deals with the case  $d = 1$ ; this theorem asserts the existence of a finite  $x(n)$  such that, for all  $x > x(n)$ ,

<sup>4</sup>For fixed  $x$ , since the variables  $Z_i(n, x)$ ,  $i \leq n$ , are independent, it is not always true that  $\sum_{i \leq n} iZ_i(n, x) = n$ . Therefore, the variable  $M(n, x)$  does not always correspond to an element of  $A_n$ .

couplings exist in which  $M(n, x)$  1-dominates  $N(n)$ .<sup>5</sup>

**Theorem 1.4.** *Let  $n \in \mathbb{N}$  and suppose  $A_n$  denotes an assembly, multiset, or a selection among elements of  $[n]$ . Given  $N(n) \sim \text{Unif}(A_n)$  with component counting process  $(C_i(n))_{i \leq n}$ , there exists a positive real number  $x(n)$  for which, when  $x > x(n)$ , there exists a process  $(Z_i(n, x))_{i \leq n}$  of non-negative independent random variables satisfying (1.2) such that  $M(n, x)$  1-dominates  $N(n)$ .*

It is worth noting that  $M(n, x)$  does not 0-dominate  $N(n)$  due to the fact that the quantity  $\mathbb{P}(Z_i(n, x) = 0 \text{ for all } i \leq n)$  is positive for each of the combinatorial structures considered in this paper (see §4 and apply independence of the  $Z_i$ 's); however, the  $C_i(n)$ 's are never all 0 due to the constraint  $\sum_{i \leq n} iC_i(n) = n$ .

## 2 The joint mass distribution of $(M(n, x), N(n))$

For some fixed value of  $x$ , if we are to successively construct a joint probability mass function  $p(\cdot, \cdot)$  such that  $M(n, x)$  1-dominates  $N(n)$ , we must ensure that  $\mathbb{P}(M(n, x) = \cdot, N(n) = \cdot) = 0$  when  $\sum_{i \leq n} (C_i(n) - Z_i(n, x))^+ > 1$ . We can index the joint distribution by using the range of  $N(n)$  and the range of  $M(n, x)$  for the column labels and row labels, respectively. In particular, we can label the columns with the range of  $(C_i(n))_{i \leq n}$  in lexicographic order. Since we have infinitely many row labels, for each  $m \in \mathbb{Z}_{\geq 0}$ , we apply the lexicographic ordering on all elements  $(m_1, \dots, m_n) \in (\mathbb{Z}_{\geq 0})^n$  with  $\sum_{i \leq n} m_i = m$ , starting with  $m = 0$  (we start with  $m = 0$  since the  $Z_i(n, x)$ 's are non-negative). With respect to this ordering, we will often enumerate the columns by  $1, 2, \dots, k_n$  and the rows by  $1, 2, \dots$ .

In each of these three settings, there are additional constraints on any joint probability mass function of  $M(n, x)$  and  $N(n)$  since the marginal distributions are known. In particular, the sum along column  $N(n) = j, 1 \leq j \leq k_n$ , is  $\mathbb{P}(N(n) = j) = 1/k_n$ ; and the sum along the row  $M(n, x) = m, m \in \mathbb{N}$ , labeled  $(Z_i(n, x))_{i \leq n} = (m_i)_{i \leq n}$  is  $\mathbb{P}((Z_i(n, x))_{i \leq n} = (m_i)_{i \leq n}) = \prod_{i \leq n} \mathbb{P}(Z_i(n, x) = m_i)$ , where the latest equation is due to the independence of the process  $(Z_i(n, x))_{i \leq n}$ .

## 3 Pivot mass

Given columns  $j$  and  $k$ , with corresponding components  $(C_i(n))_{i \leq n}$  and  $(C'_i(n))_{i \leq n}$ , we seek a way to compare the corresponding sets of row labels in which column  $j$  or  $k$  must be 0. Any of our desired couplings has the property that column  $(C_i(n))_{i \leq n}$  has a zero in row  $(Z_i(n, x))_{i \leq n}$  when the constraint  $\sum_{i \leq n} (C_i(n) - Z_i(n, x))^+ \leq 1$  is violated, so we compare the probability measures of the sets  $\{(Z_i(n, x))_{i \leq n} : \sum_{i \leq n} (C_i(n) - Z_i(n, x))^+ > 1\}$  and  $\{(Z_i(n, x))_{i \leq n} : \sum_{i \leq n} (C'_i(n) - Z_i(n, x))^+ > 1\}$ . I.e., since we seek couplings of  $M(n, x)$  and  $N(n)$  for which  $M(n, x)$  1-dominates  $N(n)$ , we measure the probability that  $M(n, x)$  takes on a value  $i$  for which column  $j$  or  $k$  has a required 0 in row  $i$ . This motivates the following definition.

**Definition 3.1.** *We call the pair  $(i, j)$ , corresponding to the  $i$ th row label  $(Z_i(x))_{i \leq n}$  and the  $j$ th column label  $(C_i(n))_{i \leq n}$ , a **pivot** if  $\sum_{i \leq n} (C_i(n) - Z_i(x))^+ > 1$ . Denote the set of all pivots by  $P$ . The **pivot mass** in column  $N(n) = j$  is defined as*

$$\mathcal{PM}(j) = \mathcal{PM}_{(n,x)}(j) := \sum_{i:(i,j) \in P} \mathbb{P}(M(n, x) = i).$$

<sup>5</sup>To simplify the notation, we will sometimes (Figure 2, Theorem 3.3, and §§4–5) replace  $Z_i(n, x)$  with  $Z_i$ , replace  $C_i(n)$  with  $C_i$ , replace  $N(n)$  with  $N$ , and replace  $M(n, x)$  with  $M$ .

## The component counting process of a uniform random variable

Given a subset  $L(n)$  of column labels of  $[n]$ , the pivot mass in  $L(n)$  is defined as

$$\mathcal{PM}(L(n)) = \mathcal{PM}_{(n,x)}(L(n)) := \sum_{\substack{i:(i,j) \in P \\ \forall j \in L(n)}} \mathbb{P}(M(n,x) = i).$$

Theorem 3.3 gives a formula for  $\mathcal{PM}(j)$ . Fortunately, due to the role of the parameter  $x$ , it is not necessary to derive a formula for  $\mathcal{PM}(L(n))$  in order to prove Theorem 1.4. The fact that  $\mathcal{PM}(L(n)) \leq \mathcal{PM}(j)$  for any  $j \in L(n)$  will be sufficient. Figure 1 shows some of the features of our desired couplings.

		$N(n)$								
		1	2	...	$j$	...	$j_0$	...	$k_n$	
$M(n,x)$	1									$\mathbb{P}(M(n,x) = 1)$
	2									$\mathbb{P}(M(n,x) = 2)$
	...									...
	$i_0$						0			$\mathbb{P}(M(n,x) = i_0)$
	...									...
	$i$				$\mathbb{P}(M(n,x) = i, N(n) = j)$					$\mathbb{P}(M(n,x) = i)$
	...									...
	...									...
Column sum		$\frac{1}{k_n}$	$\frac{1}{k_n}$	...	$\frac{1}{k_n}$	...	$\frac{1}{k_n}$	...	$\frac{1}{k_n}$	

Figure 1: If  $(i_0, j_0)$  is a pivot, then our desired joint distribution table should have a 0 in the  $(i_0, j_0)$  entry.

**Example 3.2.** Fix  $n = 3$  and consider the assembly  $A_3 = S_3$  of permutations of  $\{1, 2, 3\}$ . The elements of  $S_3$  are  $1, (12), (13), (23), (123), (132)$ , and their respective component counts are  $(3, 0, 0), (1, 1, 0), (1, 1, 0), (1, 1, 0), (0, 0, 1), (0, 0, 1)$ . Some key features of a desired joint mass distribution of  $(M(3, x), N(3))$  are given in Figure 2.

		$N$						
		$(0, 0, 1)$	$(0, 0, 1)$	$(1, 1, 0)$	$(1, 1, 0)$	$(1, 1, 0)$	$(3, 0, 0)$	
$M$	$(0, 0, 0)$			0	0	0	0	$\mathbb{P}((Z_i)_{i \leq 3} = (0, 0, 0))$
	$(0, 0, 1)$			0	0	0	0	$\mathbb{P}((Z_i)_{i \leq 3} = (0, 0, 1))$
	$(0, 1, 0)$						0	$\mathbb{P}((Z_i)_{i \leq 3} = (0, 1, 0))$
	$(1, 0, 0)$						0	$\mathbb{P}((Z_i)_{i \leq 3} = (1, 0, 0))$
	$(0, 0, 2)$			0	0	0	0	$\mathbb{P}((Z_i)_{i \leq 3} = (0, 0, 2))$
	$(0, 1, 1)$						0	$\mathbb{P}((Z_i)_{i \leq 3} = (0, 1, 1))$
	$(0, 2, 0)$						0	$\mathbb{P}((Z_i)_{i \leq 3} = (0, 2, 0))$
	$(1, 0, 1)$						0	$\mathbb{P}((Z_i)_{i \leq 3} = (1, 0, 1))$
	$(1, 1, 0)$						0	$\mathbb{P}((Z_i)_{i \leq 3} = (1, 1, 0))$
	$(2, 0, 0)$						0	$\mathbb{P}((Z_i)_{i \leq 3} = (2, 0, 0))$
	...							...
	...							...
Column sum		$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	
$\mathcal{PM}(N)$		0	0	$\mathbb{P}(Z_1 = Z_2 = 0)$	$\mathbb{P}(Z_1 = Z_2 = 0)$	$\mathbb{P}(Z_1 = Z_2 = 0)$	$\mathbb{P}(Z_1 \leq 1)$	

Figure 2: A desired coupling of  $M(3, x)$  and  $N(3)$  should have a zero at any location  $((Z_i(3, x))_{i \leq 3}, (C_i(3))_{i \leq 3})$  satisfying  $\sum_{i \leq 3} (C_i(3) - Z_i(3, x))^+ > 1$ .

Each column with a pivot contains infinitely many pivots. E.g., in Figure 2, column  $(3, 0, 0)$  has a pivot in any row of the form  $(a, b, c)$  with  $a \in \{0, 1\}, b, c \geq 0$ . Columns labeled  $(1, 1, 0)$  have a pivot in any row of the form  $(0, 0, l)$  for any  $l \in \mathbb{Z}_{\geq 0}$ . Moreover,

note that 0-domination would require each entry in the first row to be 0; however, each row mass is positive for all of the combinatorial structures considered in this paper.

The following theorem plays a key role in the proof of Theorem 1.4.<sup>6</sup> For convenience, in the proof of the following theorem, we simplify the notation by writing  $C_i(n) = C_i$ ,  $Z_i(n, x) = Z_i$ ,  $M(n, x) = M$  and  $N(n) = N$ . Moreover, the notion of pivot mass introduced in this section may be generalized; in a particular setting, one should define pivot mass based on the constraints required of their desired coupling. It is both a combinatorial and probabilistic object since it is a sum of probability masses indexed by the counting constraint  $\sum_{i \leq n} (C_i(n) - Z_i(n, x))^+ \leq 1$ .

**Theorem 3.3** (Pivot Mass Formula for 1 Column). *Consider a fixed column label  $N(n) \in A_n$  and denote its component counting process by  $(C_i(n))_{i \leq n}$ . Its pivot mass is*

$$\begin{aligned} \mathcal{PM}(N(n)) &= 1 - \sum_{j \leq n} \left( 1_{\{C_j > 0\}} (1 - \mathbb{P}(Z_j \leq C_j - 2)) \prod_{\substack{i \neq j, \\ i \leq n}} (1 - \mathbb{P}(Z_i \leq C_i - 1)) \right) \\ &\quad + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \prod_{i \leq n} (1 - \mathbb{P}(Z_i \leq C_i - 1)). \end{aligned}$$

*Proof.* Given  $1 \leq j \leq n$ , let  $\vec{e}_j$  denote the row vector of length  $n$  whose  $j$ th entry is 1 and whose other entries are 0. Given two vectors  $(a_i)_{i \leq n}, (b_i)_{i \leq n}$  in  $\mathbb{R}^n$ , we write  $(a_i)_{i \leq n} \leq (b_i)_{i \leq n}$  if  $a_i \leq b_i$  for each  $i \leq n$ . Since  $\sum_{k=1}^{\infty} \mathbb{P}(M(n, x) = k) = 1$ , we have

$$\mathcal{PM}(N) = 1 - \sum_{k: (k, N) \notin P} \mathbb{P}(M = k). \tag{3.1}$$

We have the event equality

$$\{(M, N) \notin P\} = \left\{ \exists j \leq n : (Z_i)_{i \leq n} \geq (C_i)_{i \leq n} - \vec{e}_j \cdot 1_{\{C_j > 0\}} \right\}$$

since the pair  $(M, N)$  is a not pivot if and only if  $Z_i \geq C_i$  for all  $i$  except possibly one value  $j$  with  $Z_j = C_j - 1$ . Since each  $Z_i, 1 \leq i \leq n$ , is nonnegative, we can only have  $Z_j = C_j - 1$  when  $C_j > 0$ . Note that if  $Z_i \geq C_i$  for all  $i$ , then any  $j$  satisfies  $(Z_i)_{i \leq n} \geq (C_i)_{i \leq n} - \vec{e}_j \cdot 1_{\{C_j > 0\}}$ . On the other hand, if there exists a value  $j$  for which  $Z_j = C_j - 1$  and  $Z_i \geq C_i$  for all  $i \neq j$ , then  $(Z_i)_{i \leq n} \geq (C_i)_{i \leq n} - \vec{e}_j \cdot 1_{\{C_j > 0\}}$ . Therefore, the right hand side of equation (3.1) is

$$1 - \sum_{k: (k, N) \notin P} \mathbb{P}(M = k) = 1 - \mathbb{P}\left(\exists j \leq n : (Z_i)_{i \leq n} \geq (C_i)_{i \leq n} - \vec{e}_j \cdot 1_{\{C_j > 0\}}\right). \tag{3.2}$$

We rewrite the probability  $\mathbb{P}\left(\exists j \leq n : (Z_i)_{i \leq n} \geq (C_i)_{i \leq n} - \vec{e}_j \cdot 1_{\{C_j > 0\}}\right)$  by applying an inclusion-exclusion argument. Corresponding to any  $j \leq n$  with  $C_j > 0, Z_j \geq C_j - 1$ , and  $Z_i \geq C_i$  for  $i \neq j$ , we add the term  $\mathbb{P}(Z_j \geq C_j - 1, \text{ and } Z_i \geq C_i \text{ for all } i \neq j)$ . As a result, we have added those elements with  $Z_i \geq C_i$  for all  $i$  a total of  $\sum_{i=1}^n 1_{\{C_i > 0\}}$  many times. Therefore, we compensate by subtracting the term  $(\sum_{i=1}^n 1_{\{C_i > 0\}} - 1) \mathbb{P}((Z_i)_{i \leq n} \geq (C_i)_{i \leq n})$ . Further, applying independence of the process  $(Z_i)_{i \leq n}$ , we have

$$\begin{aligned} \mathbb{P}(Z_j \geq C_j - 1 \text{ and } Z_i \geq C_i \text{ for all } i \neq j) &= \mathbb{P}(Z_j \geq C_j - 1) \mathbb{P}(Z_i \geq C_i \text{ for all } i \neq j) \\ &= \mathbb{P}(Z_j \geq C_j - 1) \prod_{\substack{i \neq j, \\ i \leq n}} \mathbb{P}(Z_i \geq C_i) \end{aligned}$$

<sup>6</sup>When Theorem 3.3 is applied in §4, additional indicator functions will be included to remind us that  $\mathbb{P}(Z_i(n, x) \leq k) = 0$  if  $k < 0$ .

and  $\mathbb{P}\left(\left(Z_i\right)_{i \leq n} \geq \left(C_i\right)_{i \leq n}\right) = \prod_{i \leq n} \mathbb{P}\left(Z_i \geq C_i\right)$ . Thus, the right hand side of equation (3.2) becomes

$$1 - \sum_{j \leq n} \left( 1_{\{C_j > 0\}} \mathbb{P}\left(Z_j \geq C_j - 1\right) \prod_{\substack{i \neq j, \\ i \leq n}} \mathbb{P}\left(Z_i \geq C_i\right) \right) + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \prod_{i \leq n} \mathbb{P}\left(Z_i \geq C_i\right). \tag{3.3}$$

Using the fact that  $\mathbb{P}\left(Z_i \geq a\right) = 1 - \mathbb{P}\left(\left(Z_i \leq a - 1\right)\right)$ , expression (3.3) becomes

$$1 - \sum_{j \leq n} \left( 1_{\{C_j > 0\}} \left(1 - \mathbb{P}\left(Z_j \leq C_j - 2\right)\right) \prod_{\substack{i \neq j, \\ i \leq n}} \left(1 - \mathbb{P}\left(Z_i \leq C_i - 1\right)\right) \right) + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \prod_{i \leq n} \left(1 - \mathbb{P}\left(Z_i \leq C_i - 1\right)\right). \quad \square$$

The following result shows that only columns with label  $\left(C_i(n)\right)_{i \leq n} = \vec{e}_n$  have zero pivot mass. In this paper, we will only apply the  $(\Leftarrow)$  part of the statement.<sup>7</sup>

**Theorem 3.4.** *For any nonempty collection  $L(n)$  of column labels,  $\mathcal{PM}(L(n)) = 0$  if and only if a column with label  $\left(C_i(n)\right)_{i \leq n} = \vec{e}_n$  belongs to  $L(n)$ .*

*Proof.*  $(\Leftarrow)$  Given any row label  $\left(Z_i(n, x)\right)_{i \leq n}$ , the vector  $\left(C_i(n)\right)_{i \leq n} = \vec{e}_n$  satisfies

$$\begin{aligned} \sum_{i \leq n} \left(C_i(n) - Z_i(n, x)\right)^+ &= \left(C_n(n) - Z_n(n, x)\right)^+ \\ &= \left(1 - Z_n(n, x)\right)^+ \\ &\leq 1. \end{aligned}$$

Thus,  $\mathcal{PM}(\vec{e}_n) = 0$ . Therefore, given  $\vec{e}_n \in L(n)$ , we have

$$\mathcal{PM}(L(n)) \leq \mathcal{PM}(\vec{e}_n) = 0.$$

$(\Rightarrow)$  Now suppose  $\vec{e}_n \notin L(n)$ . Recall that any column label  $\left(C_i(n)\right)_{i \leq n}$  satisfies  $\sum_{i \leq n} i C_i(n) = n$ . Since  $\vec{e}_n$  is the only column label with  $\sum_{i \leq n} C_i(n) = 1$ , this gives us one of two cases for each column label in  $L(n)$ . Either (a) there exists some  $j$  with  $C_j(n) \geq 2$  or (b) there exists distinct  $j, k$  with  $C_j(n) \geq 1, C_k(n) \geq 1$ . In case (a), using any row label  $\left(Z_i(n, x)\right)_{i \leq n}$  with  $Z_j(n, x) = 0$ , we have

$$\sum_{i \leq n} \left(C_i(n) - Z_i(n, x)\right)^+ \geq C_j(n) - Z_j(n, x) \geq 2.$$

In case (b), we can take any  $\left(Z_i(n, x)\right)_{i \leq n}$  with  $Z_j(n, x) = Z_k(n, x) = 0$  to ensure that

$$\sum_{i \leq n} \left(C_i(n) - Z_i(n, x)\right)^+ \geq \left(C_j(n) - Z_j(n, x)\right) + \left(C_k(n) - Z_k(n, x)\right) \geq 2.$$

Since we have just showed that each column label other than  $\vec{e}_n$  has a pivot, we use the fact that each of these columns has a pivot in the first row (labeled  $\left(Z_i(n, x)\right)_{i \leq n} = (0, 0, \dots, 0)$ ). Note that  $\mathbb{P}(M(n, x) = i) > 0$  for all distributions in this paper (see §4), so we have

$$\mathcal{PM}(L(n)) \geq \mathbb{P}\left(Z_i(n, x) = 0, \forall i \leq n\right) = \mathbb{P}\left(M(n, x) = 1\right) > 0. \quad \square$$

<sup>7</sup>Note that  $(\Rightarrow)$  implies that each column label other than  $\vec{e}_n$  has pivots. Using equations (2.2)–(2.4) in §2.2 of [2] (which give the number of columns with label  $\vec{e}_n$  in each of these combinatorial settings), we can always determine the number of columns that contain pivots.

#### 4 Pivot mass can be made arbitrarily small for assemblies, multi-sets, and selections

The following condition on  $\mathcal{PM}$  will be verified for our three combinatorial structures:

$$\forall n \in \mathbb{N} \quad \forall \varepsilon > 0 \quad \exists x(n) : x > x(n) \implies \text{equation (1.2) holds and } \mathcal{PM}_{(n,x)}(\cdot) < \varepsilon. \quad (4.1)$$

##### 4.1 Assemblies

In the assembly setting, we can take<sup>8</sup>  $Z_i(n, x) \sim \text{Po}\left(\frac{m_i x^i}{i!}\right)$  for any  $x > 0$  to obtain equation (1.2) (§2.3 of [2]). Recall that the CDF of a random variable  $Z \sim \text{Po}(\lambda)$  is given by  $\mathbb{P}(Z \leq k) = \frac{\Gamma([k+1], \lambda)}{[k]!}$  for  $k \in \mathbb{Z}_{\geq 0}$ , where  $\Gamma(a, b)$  is the **upper incomplete gamma function** - i.e.,  $\Gamma(a, b) = \int_b^\infty t^{a-1} e^{-t} dt$ .

**Lemma 4.1.** For a fixed  $a > 0$ , we have  $\lim_{b \rightarrow \infty} \Gamma(a, b) = 0$ .

*Proof.* Since  $\Gamma(a, 0) = \Gamma(a)$  is convergent for  $a > 0$ , we have

$$\begin{aligned} \Gamma(a, b) &= \Gamma(a) - \int_0^b t^{a-1} e^{-t} dt \\ &\rightarrow \Gamma(a) - \Gamma(a) \text{ as } b \rightarrow \infty \\ &= 0. \quad \square \end{aligned}$$

We can apply Lemma 4.1 and take  $x \rightarrow \infty$  to obtain

$$\frac{\Gamma\left(C_i, \frac{m_i x^i}{i!}\right)}{(C_i - 1)!} \rightarrow 0 \text{ when } C_i > 0, \quad (4.2)$$

$$\frac{\Gamma\left(C_j - 1, \frac{m_j x^j}{j!}\right)}{(C_j - 2)!} \rightarrow 0 \text{ when } C_j > 1. \quad (4.3)$$

Therefore, Theorem 3.3 implies that  $\mathcal{PM}(N(n))$  equals

$$\begin{aligned} &1 - \sum_{j \leq n} \left( 1_{\{C_j > 0\}} \left( 1 - 1_{\{C_j > 1\}} \frac{\Gamma\left(C_j - 1, \frac{m_j x^j}{j!}\right)}{(C_j - 2)!} \right) \prod_{\substack{i \neq j, \\ i \leq n}} \left( 1 - 1_{\{C_i > 0\}} \frac{\Gamma\left(C_i, \frac{m_i x^i}{i!}\right)}{(C_i - 1)!} \right) \right) \\ &+ \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \prod_{i \leq n} \left( 1 - 1_{\{C_i > 0\}} \frac{\Gamma\left(C_i, \frac{m_i x^i}{i!}\right)}{(C_i - 1)!} \right). \end{aligned}$$

If we let  $x \rightarrow \infty$ , we can apply (4.2) and (4.3) to deduce that

$$\mathcal{PM}(N(n)) \rightarrow 1 - \sum_{j \leq n} \left( 1_{\{C_j > 0\}} (1 - 0) \prod_{\substack{i \neq j, \\ i \leq n}} (1 - 0) \right)$$

<sup>8</sup>Although the distribution of  $Z_i(n, x)$  does not depend on  $n$ , the choice the process  $(Z_i(n, x))_{i \leq n}$  satisfying (1.2) does depend on  $n$ . I.e., if  $(Z_i(n, x))_{i \leq n}$  and  $(Z_i(n+1, x))_{i \leq n+1}$  equal  $(C_i(n))_{i \leq n}$  and  $(C_i(n+1))_{i \leq n+1}$ , respectively, conditional on the events  $\{\sum_{i \leq n} i Z_i = n\}$  and  $\{\sum_{i \leq n+1} i Z_i = n+1\}$ , respectively, then we need not have  $(Z_i(n, x))_{i \leq n} = (Z_i(n+1, x))_{i \leq n}$ .



$$\begin{aligned}
 & + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \prod_{i \leq n} (1 - 0) \\
 & = 1 - \sum_{j \leq n} 1_{\{C_j > 0\}} + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \\
 & = 0.
 \end{aligned}$$

This verifies condition (4.1) for assemblies.

### 4.2 Multisets

In the multiset setting, we can take  $Z_i(n, x) \sim \text{NB}(m_i, x^i)$ , for any  $x \in (0, 1)$ , to obtain equation (1.2) (§2.3 of [2]). Recall that the CDF of  $Z \sim \text{NB}(r, p)$  is given by  $\mathbb{P}(Z \leq k) = 1 - I_p(k + 1, r)$ , where  $I_p$  is the **regularized incomplete beta function**. That is,  $I_x(a, b) = \frac{B(x; a, b)}{B(a, b)}$ , where  $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ , defined for  $\text{Re}(a) > 0$  and  $\text{Re}(b) > 0$ , is the **beta function** and  $B(x; a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$  is the **incomplete beta function**.

**Lemma 4.2.** Given  $a > 0$ ,  $\lim_{x \rightarrow 1} I_x(a, b) = 1$ .

*Proof.* We have

$$\begin{aligned}
 \lim_{x \rightarrow 1} I_x(a, b) &= \lim_{x \rightarrow 1} \frac{B(x; a, b)}{B(a, b)} \\
 &= \lim_{x \rightarrow 1} \frac{\int_0^x t^{a-1} (1-t)^{b-1} dt}{\int_0^1 t^{a-1} (1-t)^{b-1} dt} \\
 &= \frac{\int_0^1 t^{a-1} (1-t)^{b-1} dt}{\int_0^1 t^{a-1} (1-t)^{b-1} dt} \\
 &= 1. \quad \square
 \end{aligned}$$

Using Theorem 3.3,  $\mathcal{PM}(N(n))$  equals

$$\begin{aligned}
 & 1 - \left( \sum_{j \leq n} \left( 1_{\{C_j > 0\}} (1 - 1_{\{C_j > 1\}} (1 - I_{x^j}(C_j - 1, m_j))) \prod_{\substack{i \neq j, \\ i \leq n}} (1 - 1_{\{C_i > 0\}} (1 - I_{x^i}(C_i, m_i))) \right) \right) \\
 & + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \prod_{i \leq n} (1 - 1_{\{C_i > 0\}} (1 - I_{x^i}(C_i, m_i))).
 \end{aligned}$$

Taking  $x \rightarrow 1$  and applying Lemma 4.2, we have

$$\begin{aligned}
 \mathcal{PM}(N(n)) &\rightarrow 1 - \left( \sum_{j \leq n} \left( 1_{\{C_j > 0\}} (1 - 1_{\{C_j > 1\}} (1 - 1)) \prod_{\substack{i \neq j, \\ i \leq n}} (1 - 1_{\{C_i > 0\}} (1 - 1)) \right) \right) \\
 & + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \prod_{i \leq n} (1 - 1_{\{C_i > 0\}} (1 - 1)) \\
 & = 1 - \sum_{j \leq n} 1_{\{C_j > 0\}} + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \\
 & = 0,
 \end{aligned}$$

which verifies condition (4.1) for multisets.

### 4.3 Selections

In the selection setting, we can take  $Z_i(n, x) \sim \text{Bin}\left(m_i, \frac{x^i}{1+x^i}\right)$ ,  $0 < x < \infty$ , in order to obtain equation (1.2) (§2.3 of [2]). In our case, we are taking  $p = \frac{x^i}{1+x^i}$ , so  $p \rightarrow 1$  if and only if  $x \rightarrow \infty$ . Recall that the CDF of  $Z \sim \text{Bin}(n, p)$  is given by  $\mathbb{P}(Z \leq k) = I_{1-p}(n - k, 1 + k)$ . Using Theorem 3.3, we can express  $\mathcal{PM}(N(n))$  as

$$1 - \sum_{j \leq n} \left( 1_{\{C_j > 0\}} \left( 1 - 1_{\{C_j > 1\}} I_{1-p}(m_j - C_j + 2, C_j - 1) \prod_{\substack{i \neq j, \\ i \leq n}} (1 - 1_{\{C_i > 0\}} I_{1-p}(m_i - C_i + 1, C_i)) \right) \right) \\ + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) \prod_{i \leq n} (1 - 1_{\{C_i > 0\}} I_{1-p}(m_i - C_i + 1, C_i)).$$

**Lemma 4.3.** We have  $\lim_{p \rightarrow 1} I_{1-p}(n - k, 1 + k) = 0$ .

*Proof.*

$$\begin{aligned} \lim_{p \rightarrow 1} I_{1-p}(n - k, 1 + k) &= \lim_{p \rightarrow 1} \frac{B(1 - p; n - k, 1 + k)}{B(n - k, 1 + k)} \\ &= \lim_{p \rightarrow 1} \frac{\int_0^{1-p} t^{n-k-1} (1 - t)^k}{\int_0^1 t^{n-k-1} (1 - t)^k} \\ &= 0. \end{aligned} \quad \square$$

Using Lemma 4.3, we see that

$$I_{1-p} = I_{1 - \frac{x^i}{1+x^i}} \rightarrow 0 \tag{4.4}$$

if  $x \rightarrow \infty$ . Thus, we apply Theorem 3.3 and Lemma 4.3 while taking  $x \rightarrow \infty$  to obtain

$$\mathcal{PM}(N(n)) \stackrel{(4.4)}{\rightarrow} 1 - \sum_{j \leq n} 1_{\{C_j > 0\}} + \left( \sum_{i \leq n} 1_{\{C_i > 0\}} - 1 \right) = 0,$$

which verifies condition (4.1) for selections.

## 5 Using pivot mass to provide couplings

Given complete separable metric spaces  $S$  and  $T$ , denote by  $p_S$  the projection of  $S \times T$  onto  $S$ . Let  $\omega$  be a nonempty closed subset of  $S \times T$  and  $\varepsilon \geq 0$ . The following result is Theorem 11 of [6].

**Theorem 5.1** (Strassen<sup>9</sup>). *There is a probability measure  $\lambda$  in  $S \times T$  with marginals  $\mu$  and  $\nu$  such that  $\lambda(\omega) \geq 1 - \varepsilon$ , if and only if for all closed sets  $L \subseteq T$*

$$\nu(L) \leq \mu(p_S(\omega \cap (S \times L))) + \varepsilon. \tag{5.1}$$

*Proof of Theorem 1.4.* Define  $S = (\mathbb{Z}_{\geq 0})^n$  and  $T = \{(a_i)_{i \leq n} \in (\mathbb{Z}_{\geq 0})^n : \sum_{i \leq n} i a_i = n\}$ , corresponding to the set of row labels and the set of column labels respectively, and endow both  $S$  and  $T$  with the metric  $d$  on  $\mathbb{Z}^n$  defined as  $d((x_i)_{i \leq n}, (y_i)_{i \leq n}) := \max_{i \leq n} |x_i - y_i|$ . Since  $S$  is finite and  $T$  is countably infinite, both  $S$  and  $T$  are separable. In both  $S$  and  $T$  we have

$$(x_i)_{i \leq n} \neq (y_i)_{i \leq n} \implies d((x_i)_{i \leq n}, (y_i)_{i \leq n}) \geq 1 \tag{5.2}$$

<sup>9</sup>Thanks to Anthony Quas for suggesting the use of Hall's Marriage Theorem. Strassen's Theorem is a variant of the marriage theorem.

since our  $n$ -tuples are integer-valued. Therefore, every Cauchy sequence in  $S$  (or in  $T$ ) converges in  $S$  (or in  $T$ ). Thus,  $S$  and  $T$  are complete. Our goal is to apply Theorem 5.1 with  $\varepsilon = 0$  and

$$\begin{aligned} \omega &= P^c, \\ L &= L(n), \\ \mu_{(n,x)}(i) &= \mathbb{P}(M(n, x) = i), i \in S, \\ \nu_n(j) &= \mathbb{P}(N(n) = j), j \in T, \\ \lambda &= p, \end{aligned}$$

where  $P = \{(i, j) \in S \times T : (i, j) \text{ is a pivot}\}$ ,  $L(n)$  denotes an arbitrary subset of  $T$ , and  $p$  is our desired joint PMF, with marginals corresponding to  $M(n, x)$  and  $N(n)$ , such that  $(i, j) \in P$  implies  $p(i, j) = 0$ . Let us endow  $\omega$  with the metric  $d_\omega$  obtained by restricting the metric

$$\begin{aligned} d_{S \times T} \left( \left( (s_i)_{i \leq n}, (t_i)_{i \leq n} \right), \left( (s'_i)_{i \leq n}, (t'_i)_{i \leq n} \right) \right) \\ := \max \left( d \left( (s_i)_{i \leq n}, (s'_i)_{i \leq n} \right), d \left( (t_i)_{i \leq n}, (t'_i)_{i \leq n} \right) \right) \end{aligned} \tag{5.3}$$

on  $S \times T$  to  $\omega$ . To show that  $\omega$  is closed, we first show that  $S$  and  $T$  are closed. The set  $T$  is closed since it is finite. Suppose that  $\left( (s_i(k))_{i \leq n} \right)_{k \in \mathbb{N}}$  is a sequence of  $n$ -tuples  $(s_i(k))_{i \leq n} \in S$  with  $\lim_{k \rightarrow \infty} (s_i(k))_{i \leq n} = l_1$  for some  $n$ -tuple  $l_1 \in \mathbb{Z}^n$ . To show that  $S$  is closed, it suffices to show that  $l_1 \in S$ . For all  $\varepsilon' \in (0, 1)$  there exists a constant  $K \in \mathbb{N}$  such that if  $k > K$  then  $d \left( (s_i(k))_{i \leq n}, l_1 \right) < \varepsilon'$ . Since  $\varepsilon' < 1$ , (5.2) implies  $l_1 = (s_i(K + 1))_{i \leq n}$ , so  $l_1 \in S$ . Therefore,  $S$  is closed. Now to show that  $\omega$  is closed in  $S \times T$ , suppose that  $\left( (s_i(k))_{i \leq n}, (t_i(k))_{i \leq n} \right)_{k \in \mathbb{N}}$  is a sequence of pairs  $\left( (s_i(k))_{i \leq n}, (t_i(k))_{i \leq n} \right) \in \omega$  of  $n$ -tuples  $s_i(k)_{i \leq n} \in S, (t_i(k))_{i \leq n} \in T$  with  $\lim_{k \rightarrow \infty} \left( (s_i(k))_{i \leq n}, (t_i(k))_{i \leq n} \right) = (l_1, l_2)$ , for some  $n$ -tuples  $l_1, l_2 \in \mathbb{Z}^n$ . Since  $S$  and  $T$  are closed, we have  $l_1 \in S$  and  $l_2 \in T$ . For all  $\varepsilon' \in (0, 1)$  there exists a constant  $K \in \mathbb{N}$  such that  $k > K \implies d_{S \times T} \left( (s_i(k))_{i \leq n}, (l_1, l_2) \right) < \varepsilon'$ . Therefore,

$$k > K \xrightarrow{(5.3)} d \left( (s_i(k))_{i \leq n}, l_1 \right), d \left( (t_i(k))_{i \leq n}, l_2 \right) < \varepsilon'.$$

Since  $\varepsilon' < 1$ ,

$$k > K \xrightarrow{(5.2)} d \left( (s_i(k))_{i \leq n}, l_1 \right) = d \left( (t_i(k))_{i \leq n}, l_2 \right) = 0.$$

Therefore, applying (5.2) twice, we obtain  $(l_1, l_2) = \left( (s_i(K + 1))_{i \leq n}, (t_i(K + 1))_{i \leq n} \right) \in \omega$ , so  $\omega$  is a closed subset of  $S \times T$ . Further,  $\omega \neq \emptyset$  since given any column label  $j$ , the pair  $(j, j)$  belongs to  $\omega$ . Note that the set  $L(n)$  is a closed subset of the column labels since  $L(n)$  is a finite set. Moreover,  $\nu_n(L(n))$  is equal to  $\mathbb{P}(N(n) \in L(n)) = \frac{\#L(n)}{k_n}$ , and

$$\begin{aligned} \mu_{(n,x)}(p_S(\omega \cap (S \times L))) &= \mathbb{P}(M \in p_S(\omega \cap (S \times L(n)))) \\ &= \mathbb{P}(M \in p_S(P^c \cap (S \times L(n)))) \\ &= \mathbb{P}(\exists j \in L(n) : (M, j) \notin P) \\ &= 1 - \mathcal{PM}_{(n,x)}(L(n)). \end{aligned}$$

Therefore, inequality (5.1) is equivalent to  $\frac{\#L(n)}{k_n} \leq 1 - \mathcal{PM}_{(n,x)}(L(n))$ . The latest inequality is equivalent to

$$\mathcal{PM}_{(n,x)}(L(n)) \leq 1 - \frac{\#L(n)}{k_n}. \tag{5.4}$$

By (4.1), the left hand side can be made arbitrarily small, so (5.4) holds when  $1 - \frac{\#L(n)}{k_n} > 0$ . When  $1 - \frac{\#L(n)}{k_n} = 0$ , we must have  $L(n) = A_n$ , so that  $\mathcal{PM}_{(n,x)}(L(n)) = 0$  by Theorem 3.4. Therefore, by the conclusion of Strassen's Theorem, there exists a joint probability measure  $p$ , with marginals  $\mathbb{P}(M(n, x) = \cdot)$  and  $\mathbb{P}(N(n) = \cdot)$ , such that  $p(\omega) = 1$ . I.e, the probability of having no pivot in this joint distribution is 1. Hence, the proof of Theorem 1.4 is complete.  $\square$

## References

- [1] Arratia, R. On the amount of dependence in the prime factorization of a uniform random integer. In B. Bollobás, editor, *Contemporary Combinatorics*, pp 29–91. Bolyai Society Mathematical Studies, Volume **10**, 2002.
- [2] Arratia, R., Barbour, A.D., and Tavaré, S. *Logarithmic Combinatorial Structures: A Probabilistic Approach*. EMS Monographs in Mathematics. European Mathematical Society (EMS), 2003.
- [3] Arratia, R. and Tavaré, S. Independent process approximations for random combinatorial structures. *Advances in mathematics* **104** (1), pp 90–154, 1994.
- [4] Duchon, P., Flajolet, P., Louchard, G., and Schaeffer, G. Boltzmann samplers for the random generation of combinatorial structures. *Combinatorics, Probability and Computing*, **13** (4-5): pp 577–625, 2004.
- [5] Quas, A. Reference Request for Couplings with Conditions, URL (version: 2017-10-27): <https://mathoverflow.net/q/284525>.
- [6] Strassen, V. The existence of probability spaces with given marginals. *Ann. Math. Statist.*, Volume **36**, Number 2, pp 423–439, 1965.

**Acknowledgments.** Thanks to Prof. Michael Cranston and Prof. Nathan Kaplan for their support in this project. Thanks to Prof. Richard Arratia and the referee for the helpful suggestions.