Stochastic Domination of Prime Powers of a Uniform Random Integer by Geometric Distributions

J. Squillace and C. George

Abstract - For each natural number $n \geq 2$ and for each prime $p \leq n$, we provide three proofs of the fact that the power, $C_p(n)$, of the prime p in the prime factorization of a uniformly chosen random integer from 1 to n is stochastically dominated by a nonnegative geometric random variable, Z_p, of parameter $1/p$. In one of these proofs, we construct a coupling of Z_p and $C_p(n)$ such that with probability one we have both $Z_p - 1 \leq C_p(n) \leq Z_p$ whenever $Z_p \leq \lfloor \log_p n \rfloor$ and $C_p(n) = \lfloor \log_p n \rfloor$ whenever $Z_p > \lfloor \log_p n \rfloor$; then we will show that any coupling of Z_p and $C_p(n)$ satisfying this constraint is a maximal coupling of Z_p and $C_p(n)$ if and only if $n = p^k$ for some positive integer k. We will also show how our couplings of the variables Z_p and $C_p(n)$ correspond to rigid bracings of an $\infty \times \lfloor \log_p(n) \rfloor$ rectangular grid if and only if n is not divisible by p.

Keywords : stochastic domination; maximal coupling; Strassen's theorem

Mathematics Subject Classification (2020) **:** 60C05; 11A41

1 Introduction

This paper uses the concept of stochastic domination to provide insights about prime numbers by comparing the distribution of the power of the prime p in the prime factorization of the uniform variable on the set $\{1, 2, \ldots, n\}$ with a nonnegative geometric distribution of parameter $1/p$. Given two random variables X and Y, not necessarily defined on the same probability space, what does it mean to say that Y is "larger" than X? The notion of stochastic dominance provides an answer to this question.

Definition 1.1 *Let X and Y be random variables with cumulative distribution functions (CDFs) $\mathbb{P}(X \leq a)$ and $\mathbb{P}(Y \leq a)$ for $a \in \mathbb{R}$. We say that Y has a* ***first-order stochastic dominance*** *over X, denoted by*

$$X \overset{D}{\leq} Y,$$

if $\mathbb{P}(X \leq a) \geq \mathbb{P}(Y \leq a)$ for all $a \in \mathbb{R}$.

Applications of stochastic domination have appeared in economics, statistics, and mathematical finance. The paper [5] shows how stochastic domination leads to constraints for problems in optimization.

The following definition appears on p. 193 of [7].

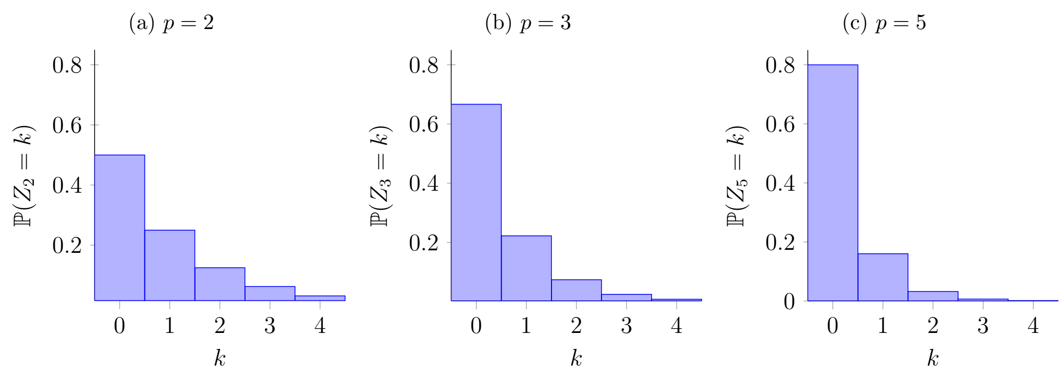
Definition 1.2 *If $X, X_1, X_2, \ldots$ is a sequence of random variables with respective distribution functions $F, F_1, F_2, \ldots$, we say that X_n* ***converges in distribution*** *to X, written $X_n \xrightarrow{D} X$, if $F_n \to F$ as $n \to \infty$.*

Fix a natural number $n \geq 2$, and consider a uniformly chosen random integer $N(n)$ on the set $\{1, 2, \ldots, n\}$. The fundamental theorem of arithmetic guarantees the existence of a unique prime factorization

$$N(n) = \prod_{p \leq n} p^{C_p(n)},$$

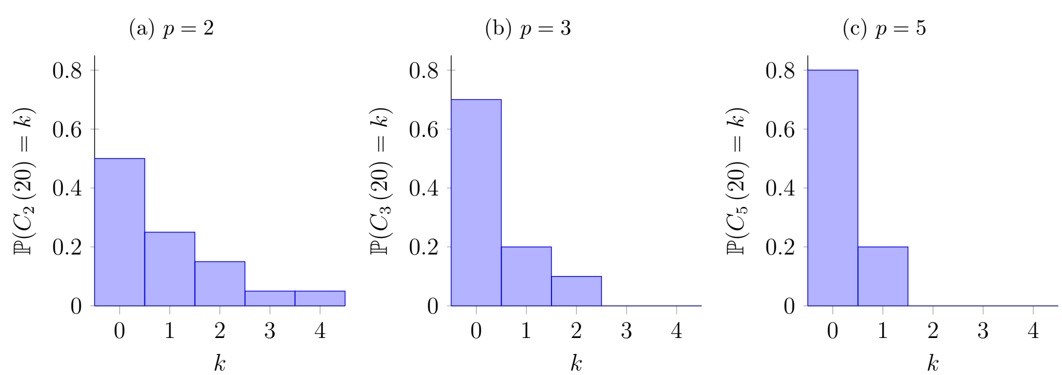
where the variable $C_p(n)$ is the power of the prime p. For any prime p, let Z_p denote a nonnegative geometric random variable of parameter $1/p$. Figures 1 and 2 contain the graphs of the probability mass functions (PMFs) of Z_p and $C_p(n)$ for $p = 2, 3, 5$, and $n = 20$. For any prime $p \leq n$, equations $(3, 4)$ imply that $C_p(n)$ converges to Z_p in distribution as $n \to \infty$.

Figure 1: Partial plots of the probability histograms of Z_p for $p = 2, 3, 5$.



Note that the support of Z_p is the set of nonnegative integers. Since $p^{C_p(n)} \leq n$, the support of $C_p(n)$ is $\{0, 1, \ldots, \lfloor \log_p n \rfloor\}$, where $\lfloor \cdot \rfloor$ denotes the floor function.

Figure 2: Plots of the probability histograms of $C_p(20)$ for $p = 2, 3, 5$.



All of our results will involve couplings, which we now define.

Definition 1.3 *A* ***coupling*** *of two random variables X and Y is a random vector (X', Y') such that X' is equal to X in distribution and Y' is equal to Y in distribution. A coupling (X', Y') of X and Y is a* ***maximal coupling*** *if $\mathbb{P}(X' = Y')$ is maximal among all couplings of X and Y.*

In a coupling, the component variables X and Y may be random vectors or stochastic processes. For example, the paper [6] uses maximal couplings to provide necessary and sufficient conditions for the existence of a coupling of two stochastic processes $X_n^{(1)}$ and $X_n^{(2)}$ such that the two processes eventually agree.

The goal of this paper is to prove the following theorem.

Theorem 1.4 *Fix a natural number $n \geq 2$, and let p denote a prime $\leq n$. Let $N(n) = \prod_{p \leq n} p^{C_p(n)}$ denote a uniform random variable on the set $\{1, 2, \ldots, n\}$, and let Z_p denote a nonnegative geometric random variable of parameter $1/p$.*

a) There exists a coupling $\left(Z_p', C_p'(n)\right)$ of Z_p and $C_p(n)$ satisfying

$$C_p'(n) = \begin{cases} Z_p' \text{ or } Z_p' - 1 & \text{if } Z_p' \leq \lfloor \log_p n \rfloor, \\ \lfloor \log_p n \rfloor & \text{if } Z_p' > \lfloor \log_p n \rfloor. \end{cases} \tag{1}$$

Consequently,

$$C_p(n) \overset{D}{\leq} Z_p. \tag{2}$$

b) Any coupling of Z_p and $C_p(n)$ satisfying (1) is a maximal coupling of Z_p and $C_p(n)$ if and only if $n = p^k$ for some natural number k.

c) When $n = p^k$, there exists a unique $\infty \times \lfloor \log_p n \rfloor$ joint probability table representing all maximal couplings of Z_p and $C_p(n)$. Furthermore, when $n = p^k$, the maximal coupling probability is $1 - 1/np$.

d) A coupling $\left(Z_p', C_p'(n)\right)$ of Z_p and $C_p(n)$ satisfying (1) corresponds to a bracing of a rigid rectangular $\infty \times \lfloor \log_p(n) \rfloor$ grid if and only if n is not divisible by p.

An open conjecture by Richard Arratia (p. 49 of [3]) asks, for each $n \geq 1$, whether there exists a coupling of the variables $\prod_{p \leq n} p^{C_p(n)}$ and $\prod_{p \leq n} p^{Z_p}$, where the $Z_p, p \leq n$, are independent variables, such that $\sum_{p \leq n} (C_p(n) - Z_p)^+ \leq 1$, where $(\cdot)^+$ denotes the positive part.

We will provide three proofs of (2). In Section 2, inequality (2) is proved via the definition of first-order stochastic dominance by comparing the CDFs of Z_p and $C_p(n)$. In Section 3, we prove inequality (2) by establishing a joint probability table with marginals Z_p' and $C_p'(n)$ corresponding to Z_p and $C_p(n)$, respectively, such that Z_p' and $C_p'(n)$ satisfy constraint (1). We will then refer to a theorem, which states that a variable Y has a stochastic domination over the variable X if and only if there exists a coupling (X', Y')

of X and Y such that X' is pointwise dominated by Y', to show that (2) follows from (1). In Section 3.1, we prove Theorem 1.4b; and we prove Theorem 1.4c by showing that there is a unique $\infty \times \lfloor \log_p n \rfloor$ joint probability table corresponding to all maximal couplings of Z_p and $C_p(n)$ when $n = p^k$.

In Section 4, we prove inequality (2) by using a variant of the marriage theorem due to Volker Strassen. In Section 5, we prove Theorem 1.4d by constructing a bipartite graph corresponding to the coupling constructed in Section 3; and then we show that the bipartite graph corresponds to a rigid bracing of a rectangular $\infty \times \lfloor \log_p(n) \rfloor$ grid if and only if n is not divisible by p.

2 Stochastic Domination via Direct Computation of Cumulative Distribution Functions

Our first proof of inequality (2) is based on the fact that we have formulas for the CDFs of $C_p(n)$ and Z_p.

Proof. [**Proof of Inequality** (2)**.**] Given a nonnegative integer j, we have the event equality

$$\{C_p(n) \geq j\} = \left\{p^j \text{ divides } N(n)\right\}.$$

There are $\lfloor n/p^j \rfloor$ multiples of p^j in the set $\{1, 2, \ldots, n\}$, and since $N(n)$ is uniformly distributed, the probability of the event $\{C_p(n) \geq j\}$ is

$$\mathbb{P}(C_p \geq j) = \frac{\lfloor n/p^j \rfloor}{n}. \tag{3}$$

Moreover, since Z_p is a nonnegative geometric random variable of parameter $1/p$, the probability of the event $\{Z_p \geq j\}$ is

$$\mathbb{P}(Z_p \geq j) = 1/p^j. \tag{4}$$

Thus,

$$\begin{aligned} \mathbb{P}(Z_p \leq j) &\stackrel{(4)}{=} 1 - 1/p^{j+1} \\ &\leq 1 - \frac{\lfloor n/p^{j+1} \rfloor}{n} \\ &\stackrel{(3)}{=} \mathbb{P}(C_p(n) \leq j). \end{aligned}$$

□

3 Constructing a Coupling of Z_p and $C_p(n)$ with $C_p(n) \leq Z_p$ Pointwise

In this section, Theorem 1.4a is proved by constructing a coupling $\left(Z'_p, C'_p(n)\right)$ of Z_p and $C_p(n)$ such that $\left(Z'_p, C'_p(n)\right)$ satisfies constraint (1); constraint (1) implies that

$C'_p(n)$ is pointwise dominated by Z'_p. The construction is achieved by describing an $\infty \times \lfloor \log_p(n) \rfloor$ joint probability table, consisting of nonnegative numbers, such that row i sums to $\mathbb{P}(Z_p = i)$ for each integer $i \geq 0$, column j sums to $\mathbb{P}(C_p(n) = j)$ for each integer j with $0 \leq j \leq \lfloor \log_p(n) \rfloor$, and we impose the constraint that the entry in row i and column j is 0 if both of the statements

(i) $i - 1 \leq j \leq i < \lfloor \log_p n \rfloor$,

(ii) $i \geq \lfloor \log_p n \rfloor$ and $j = \log_p n$

are false. Then we apply the following theorem (Theorem 3.1 in Chapter 1 of [10]) to prove inequality (2).

Theorem 3.1 *Let X and Y be random variables. Then $X \overset{D}{\leq} Y$ if and only if there is a coupling (X', Y') of X and Y such that (pointwise) $X' \leq Y'$.*

Proof. [**Proof of Theorem 1.4a.**] Consider a prime $p \leq n$. Any coupling of Z_p and $C_p(n)$ has row and column sums determined by the PMFs of Z_p and $C_p(n)$, given by

$$\mathbb{P}(Z_p = i) \overset{(4)}{=} \frac{1 - 1/p}{p^i}, \;\; i \in \mathbb{Z}_{\geq 0} \tag{5}$$

and

$$\mathbb{P}(C_p(n) = j) \overset{(3)}{=} \frac{\lfloor n/p^j \rfloor - \lfloor n/p^{j+1} \rfloor}{n}, \;\; 0 \leq j \leq \left\lfloor \log_p n \right\rfloor. \tag{6}$$

Table 1 illustrates the desired row and column sums of any joint probability distribution corresponding to a coupling of Z_p and $C_p(n)$. To simplify the notation in the following tables, define $p(i,j) := \mathbb{P}\left(Z'_p = i, C'_p(n) = j\right)$ and $k := \lfloor \log_p(n) \rfloor$.

Table 1: Visual features of a coupling $\left(Z'_p, C'_p(n)\right)$ of Z_p and $C_p(n)$.

Z'_p \ C'_p	0	1	2	$\cdots$	k	Row sum
0	$p(0,0)$	$p(0,1)$	$p(0,2)$	$\cdots$	$p(0,k)$	$1-1/p$
1	$p(1,0)$	$p(1,1)$	$p(1,2)$	$\cdots$	$p(1,k)$	$(1-1/p)/p$
2	$p(2,0)$	$p(2,1)$	$p(2,2)$	$\cdots$	$p(2,k)$	$(1-1/p)/p^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
k	$p(k,0)$	$p(k,1)$	$p(k,2)$	$\cdots$	$p(k,k)$	$(1-1/p)/p^k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Column sum	$1 - \left\lfloor \frac{n}{p} \right\rfloor /n$	$\left(\left\lfloor \frac{n}{p} \right\rfloor - \left\lfloor \frac{n}{p^2} \right\rfloor\right)/n$	$\left(\left\lfloor \frac{n}{p^2} \right\rfloor - \left\lfloor \frac{n}{p^3} \right\rfloor\right)/n$	$\cdots$	$\left(\left\lfloor \frac{n}{p^k} \right\rfloor - \left\lfloor \frac{n}{p^{k+1}} \right\rfloor\right)/n$	

We will construct a coupling corresponding to equation (1) by placing at most two nonzero entries each row and each column except the final column. The resulting joint probability PMF is provided by Table 2.

Table 2: Visualizing a joint probability table corresponding to equation (1).

Z_p' \ C_p'	0	1	2	$\cdots$	$k-1$	k	Row sum
0	$p(0,0)$	0	0	$\cdots$	0	0	$1-1/p$
1	$p(1,0)$	$p(1,1)$	0	$\cdots$	0	0	$(1-1/p)/p$
2	0	$p(2,1)$	$p(2,2)$	$\cdots$	0	0	$(1-1/p)/p^2$
3	0	0	$p(3,2)$	$\cdots$	0	0	$(1-1/p)/p^3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
k	0	0	0	0	$p(k,k-1)$	$p(k,k)$	$(1-1/p)/p^k$
$k+1$	0	0	0	0	0	$p(k+1,k)$	$(1-1/p)/p^{k+1}$
$k+2$	0	0	0	0	0	$p(k+2,k)$	$(1-1/p)/p^{k+2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
Column sum	$1-\left\lfloor\frac{n}{p}\right\rfloor/n$	$\left(\left\lfloor\frac{n}{p}\right\rfloor-\left\lfloor\frac{n}{p^2}\right\rfloor\right)/n$	$\left(\left\lfloor\frac{n}{p^2}\right\rfloor-\left\lfloor\frac{n}{p^3}\right\rfloor\right)/n$	$\cdots$	$\left(\left\lfloor\frac{n}{p^{k-1}}\right\rfloor-\left\lfloor\frac{n}{p^k}\right\rfloor\right)/n$	$\left(\left\lfloor\frac{n}{p^k}\right\rfloor-\left\lfloor\frac{n}{p^{k+1}}\right\rfloor\right)/n$	

Since $p(0,0)$ is the only nonzero entry in the zeroth row, we necessarily have

$$p(0,0)=1-1/p. \tag{7}$$

Since $p(0,0)$ and $p(1,0)$ are the only possible nonzero entries in the zeroth column, we necessarily have

$$p(0,0)+p(1,0)=1-\frac{\lfloor n/p\rfloor}{n}.$$

Thus,

$$p(1,0)=1-\frac{\lfloor n/p\rfloor}{n}-p(0,0)\overset{(7)}{=}1/p-\frac{\lfloor n/p\rfloor}{n}. \tag{8}$$

Since $p(1,0)$ and $p(1,1)$ are the only entries in the first row, it follows that

$$p(1,0)+p(1,1)=\frac{1-1/p}{p},$$

so

$$p(1,1)=\frac{1-1/p}{p}-p(1,0)\overset{(8)}{=}\frac{\lfloor n/p\rfloor}{n}-1/p^2. \tag{9}$$

Since $p(1,1)$ and $p(2,1)$ are the only entries in the first column,

$$p(1,1)+p(2,1)=\frac{\lfloor n/p\rfloor-\lfloor n/p^2\rfloor}{n}.$$

Thus,

$$p(2,1)=\frac{\lfloor n/p\rfloor-\lfloor n/p^2\rfloor}{n}-p(1,1)\overset{(9)}{=}1/p^2-\frac{\lfloor n/p^2\rfloor}{n}. \tag{10}$$

Similarly, it follows that

$$\begin{aligned}
p(2,2) &= \frac{1-1/p}{p^2}-p(2,1)\\
&\overset{(10)}{=} \left(1/p^2-1/p^3\right)-\left(1/p^2-\frac{\lfloor n/p^2\rfloor}{n}\right)\\
&= \frac{\lfloor n/p^2\rfloor}{n}-1/p^3.
\end{aligned}$$

Continuing in this fashion, we can obtain formulas for the entries $p(3,2), p(3,3), \ldots, p(k,k-1), p(k,k)$; each row and column will have at most two nonzero entries except the last column (since we have yet to assign values to the entries $p(i,k)$ for $i>k$). By construction,

$$p(i,i) = \frac{\lfloor n/p^i \rfloor}{n} - \frac{1}{p^{i+1}}, \qquad 0 \leq i \leq k, \tag{11}$$

and

$$p(i+1,i) = 1/p^{i+1} - \frac{\lfloor n/p^{i+1} \rfloor}{n}, \qquad 0 \leq i \leq k-1. \tag{12}$$

Before we deal with the final column, let us show that the terms we have constructed in equation (11) are positive and the terms given by (12) are nonnegative. The case $i=0$ has been resolved by equations (7, 8), so consider $1 \leq i \leq \lfloor \log_p n \rfloor$. By the division algorithm (the use of the division algorithm was suggested by [8]), we have

$$n = qp^i + r, \quad 0 \leq r < p^i, \quad q \in \mathbb{N}. \tag{13}$$

Therefore,

$$\begin{aligned} p(i,i) &\overset{(11)}{=} \frac{\lfloor n/p^i \rfloor}{n} - \frac{1}{p^{i+1}} \\ &\overset{(13)}{=} q/n - 1/p^{i+1} \\ &= \frac{qp^{i+1} - n}{np^{i+1}} \\ &= \frac{qp^i(p-1) - r}{np^{i+1}} \\ &\overset{(13)}{>} 0, \end{aligned}$$

and

$$\begin{aligned} p(i+1,i) &\overset{(12)}{=} \frac{1}{p^{i+1}} - \frac{\lfloor n/p^{i+1} \rfloor}{n} \\ &\geq 0. \end{aligned}$$

At the moment, we have only assigned one entry in the final column, namely, $p\left(\lfloor \log_p n \rfloor, \lfloor \log_p n \rfloor\right)$. By equation (11), we have

$$p\left(\lfloor \log_p n \rfloor, \lfloor \log_p n \rfloor\right) = \frac{\left\lfloor \frac{n}{p^{\lfloor \log_p n \rfloor}} \right\rfloor}{n} - \frac{1}{p^{\lfloor \log_p n \rfloor + 1}}.$$

Moreover, since in rows $i > \lfloor \log_p n \rfloor$ all entries in Table 2 are zero except in the final column, we necessarily have

$$p\left(i, \lfloor \log_p n \rfloor\right) = \text{ row } i \text{ sum } \overset{(5)}{=} \frac{1-1/p}{p^i}, \qquad i > \lfloor \log_p n \rfloor. \tag{14}$$

Summing over the entries in the final column, leads to

$$\begin{aligned} p\left(\left\lfloor \log_p n\right\rfloor, \left\lfloor \log_p n\right\rfloor\right) + \sum_{i>\left\lfloor \log_p n\right\rfloor} \frac{1-1/p}{p^i} &= \text{column } \left\lfloor \log_p n\right\rfloor \text{ sum} \\ &\overset{(6)}{=} \frac{\left\lfloor \frac{n}{p^{\left\lfloor \log_p n\right\rfloor}}\right\rfloor - \left\lfloor \frac{n}{p^{\left\lfloor \log_p n\right\rfloor+1}}\right\rfloor}{n} \\ &= \frac{\left\lfloor \frac{n}{p^{\left\lfloor \log_p n\right\rfloor}}\right\rfloor}{n}. \end{aligned}$$

Thus, upon summing a geometric series, we have

$$\begin{aligned} p\left(\left\lfloor \log_p n\right\rfloor, \left\lfloor \log_p n\right\rfloor\right) &= \frac{\left\lfloor \frac{n}{p^{\left\lfloor \log_p n\right\rfloor}}\right\rfloor}{n} - \sum_{i>\left\lfloor \log_p n\right\rfloor} \frac{p-1}{p^{i+1}} \\ &= \frac{\left\lfloor \frac{n}{p^{\left\lfloor \log_p n\right\rfloor}}\right\rfloor}{n} - \frac{1}{p^{\left\lfloor \log_p n\right\rfloor+1}}, \end{aligned}$$

which is consistent with the value of $p\left(\left\lfloor \log_p n\right\rfloor, \left\lfloor \log_p n\right\rfloor\right)$ determined by equation (11). This completes the construction of the coupling $\left(Z_p', C_p'(n)\right)$ of Z_p and $C_p(n)$ satisfying the constraint (1). Moreover, constraint (1) implies that Z_p' pointwise dominates $C_p'(n)$; thus, by Theorem 3.1, inequality (2) is true. $\square$

3.1 Maximal Couplings of $C_p(n)$ and Z_p Satisfying Constraint (1)

In this section, we will prove Theorem 1.4b-c. First we describe the notion of a maximal coupling, and then we show that any coupling of $C_p(n)$ and Z_p satisfying (1) (e.g., any coupling whose joint probability table is of the form given by Table 2) is a maximal coupling of $C_p(n)$ and Z_p if and only if $n = p^k$ for some positive integer k (Theorem 1.4b). Then we give a short proof of Theorem 1.4c, which is basically a corollary of Theorem 1.4b. The following definition appears in Section 4.1 of [10].

Definition 3.2 *Let $\mathbb{I}$ be an arbitrary index set. Suppose $(X_i')_{i\in\mathbb{I}}$ is a coupling of the collection of random variables $X_i, i \in \mathbb{I}$. A **coupling event** is an event C such that*

$$C \subseteq \left\{X_i' = X_j' \text{ for all } i, j \in \mathbb{I}\right\}. \tag{15}$$

*Suppose that the variables $X_i, i \in \mathbb{I}$ are discrete and take values in a finite or countable set E, and denote the PMF of X_i by p_i. A coupling is a **maximal coupling** if the event inequality (15) holds with equality. The sum $\sum_{x\in E} \inf_{i\in\mathbb{I}} p_i(x)$ is the **maximal coupling probability**.*

Visually, Theorem 1.4b says that the coupling in Table 2 is maximal if and only if the entries below the main diagonal are 0 for columns labeled $0, 1, \ldots, \lfloor \log_p n \rfloor - 1$; and this is due to the fact that, by equation (12), $p(i+1, i) = 0$ when $n = p^k$. The following theorem (see Section 4.2 of [10]) gives a formula for the maximal coupling probability.

Theorem 3.3 *Let $\mathbb{I}$ denote an index set. Suppose $X_i, i \in \mathbb{I}$, are discrete random variables taking values in a finite or countable set E. Then there exists a maximal coupling; i.e., there exists a coupling with coupling event C such that $\mathbb{P}(C) = \sum_{x \in E} \inf_{i \in \mathbb{I}} p_i(x)$.*

By Theorem 3.3 and equations (5) and (6) the maximal coupling probability of $C_p(n)$ and Z_p is

$$\sum_{l=0}^{\lfloor \log_p n \rfloor} \min \left\{ \frac{\lfloor n/p^l \rfloor - \lfloor n/p^{l+1} \rfloor}{n}, \frac{1 - 1/p}{p^l} \right\}.$$

Proof. [**Proof of Theorem 1.4b.**] The proof of theorem 1.4a provided in Section 3 proved the existence of a coupling of Z_p and $C_p(n)$ satisfying equation (1); it suffices to show that equation (1) corresponds to a maximal coupling of Z_p and $C_p(n)$ if and only if $n = p^k$.

Suppose $n = p^k$. Then

$$\frac{\lfloor n/p^l \rfloor - \lfloor n/p^{l+1} \rfloor}{n} = \begin{cases} \frac{1-1/p}{p^l} & l \neq k, \\ 1/n & l = k \end{cases}$$

which implies

$$\min \left\{ \frac{\lfloor n/p^l \rfloor - \lfloor n/p^{l+1} \rfloor}{n}, \frac{1 - 1/p}{p^l} \right\} = \frac{1 - 1/p}{p^l} \tag{16}$$

for each $l = 0, 1, \ldots, k$. Therefore,

$$\begin{aligned} \mathbb{P}\left(C_p'(n) = Z_p'\right) &= \sum_{l=0}^{k} p(l, l) \\ &\overset{(11)}{=} \sum_{l=0}^{k} \left(\frac{\lfloor n/p^l \rfloor}{n} - \frac{1}{p^{l+1}} \right) \\ &= \sum_{l=0}^{k} \left(1/p^l - 1/p^{l+1}\right) \\ &\overset{(16)}{=} \sum_{l=0}^{k} \min \left\{ \frac{\lfloor n/p^k \rfloor - \lfloor n/p^{k+1} \rfloor}{n}, \frac{1 - 1/p}{p^l} \right\}. \end{aligned}$$

By Theorem 3.3, this shows that the coupling in Table 2 is indeed a maximal coupling of $C_p(n)$ and Z_p when $n = p^k$.

Conversely, suppose that $n \neq p^k$ for every nonnegative integer k. Then at least one of the terms $p(i+1,i)$, $i = 0,1,\ldots,k-1$, in Table 2 is positive. To see this, note that n is of the form $n = p^l b$, where $0 \leq l \leq k-1$, $b > 1$, and p does not divide b; thus,

$$p(l+1,l) \overset{(12)}{=} \frac{1}{p^{l+1}} - \frac{\lfloor n/p^{l+1} \rfloor}{n} > 0.$$

Therefore,

$$\begin{aligned}
p(l+1,l+1) &= \mathbb{P}(Z_p = l+1) - p(l+1,i) \\
&= \frac{1-1/p}{p^{l+1}} - \left(\frac{1}{p^{l+1}} - \frac{\lfloor n/p^{l+1} \rfloor}{n}\right) \\
&= \frac{\lfloor n/p^{l+1} \rfloor}{n} - \frac{1}{p^{l+2}} \\
&< \frac{\lfloor n/p^{l+1} \rfloor}{n} - \frac{\lfloor n/p^{l+2} \rfloor}{n}. \\
&= \mathbb{P}(C_p(n) = l+1);
\end{aligned}$$

thus, $p(l+1,l+1)$ is less than $\min\{\mathbb{P}(Z_p(n) = l+1), \mathbb{P}(C_p(n) = l+1)\}$, which implies that $\sum_{l=0}^{k} p(l,l)$ is strictly less than the maximal coupling probability. Therefore, the coupling in Table 2 is not maximal in this case. □

Proof. [**Proof of Theorem 1.4c.**] If $n = p^k$ for some positive integer k, then the maximal coupling probability was expressed as $\sum_{l=0}^{k}\left(\frac{1-1/p}{p^l}\right)$ in the proof of Theorem 1.4b, and

$$\sum_{l=0}^{k}\left(\frac{1-1/p}{p^l}\right) = 1 - \frac{1}{np}.$$

Moreover, by equation (12) the entries $p(i+1,i)$, where $0 \leq i < k$, are 0, and this gives a unique solution to Table 2.

Now we will show that if $n = p^k$, then any maximal coupling of Z_p and $C_p(n)$ satisfies the constraint

$$C'_p(n) = \begin{cases} Z'_p & \text{if } Z'_p \leq k, \\ k & \text{if } Z'_p > k, \end{cases}$$

which is a special case of equation (1) (and therefore corresponds to a maximal coupling by Theorem 1.4b). Note that in the case $n = p^k$, by equations $(5,6)$, the row i sum equals the column i sum in Table 2 if $0 \leq i \leq k-1$. Thus, if any of the entries not on the main diagonal of the sub-table obtained by removing column k of Table 2 are nonzero, this will ensure that at least one of the entries $p(i,i)$ will be less than the row i sum – as a result, the coupling will not be maximal. □

4 Combinatorial Optimization and Stochastic Domination

In this section, we will use a theorem proved by Volker Strassen to prove inequality (2). Strassen's theorem provides a necessary and sufficient criterion for the existence of a joint probability distribution given a set of marginal distributions and constraints. As a consequence, it can be applied to model the dependence between random variables.

Let S and T be complete metric spaces. Denote by p_S the projection of the Cartesian product $S \times T$ onto S. Let ω be a nonempty closed subset of $S \times T$ and $\varepsilon \geq 0$. The following result is Theorem 11 of [9].

Theorem 4.1 *(Strassen) There is a probability measure λ in $S \times T$ with marginals μ and ν such that $\lambda(\omega) \geq 1 - \varepsilon$ if and only if for all closed sets $L \subseteq T$*

$$\nu(L) \leq \mu\left(p_S\left(\omega \cap (S \times L)\right)\right) + \varepsilon. \tag{17}$$

Proof. [**Proof of Inequality** (2)**.**] Let us define

$$\begin{aligned} S &:= \mathbb{Z}_{\geq 0}, \\ T &:= \left\{0, 1, \ldots, \left\lfloor \log_p n \right\rfloor\right\}, \end{aligned}$$

corresponding to the support of Z_p and the support of $C_p(n)$, respectively. Let us endow S and T with metrics, denoted by d_S and d_T, by restricting the Euclidean metric d on $\mathbb{R}$ given by

$$d(x, y) = |x - y|,$$

to the sets S and T, respectively. Since S is countably infinite and T is finite, both S and T are separable. In S and T, we have

$$i_1 \neq i_2 \implies d_S(i_1, i_2) \geq 1$$

and

$$j_1 \neq j_2 \implies d_T(j_1, j_2) \geq 1.$$

Therefore, every Cauchy sequence in S converges in S, and every Cauchy sequence in T converges in T. Thus, S and T are complete separable metric spaces. Let us apply Strassen's theorem with

$$\begin{aligned} \omega &= \{i, j : i \in S, j \in T, j \leq i\}, \\ \mu_p(i) &= \mathbb{P}(Z_p = i), i \in S, \\ \nu_{(n,p)}(j) &= \mathbb{P}(C_p(n) = j), j \in T, \\ \lambda &= p(\cdot, \cdot), \\ L &\subseteq \left\{0, 1, \ldots, \left\lfloor \log_p n \right\rfloor\right\}, \\ \varepsilon &= 0, \end{aligned}$$

where $p(\cdot, \cdot)$ is our desired joint PMF with marginals corresponding to Z_p and $C_p(n)$ satisfying

$$C_p(n) = j > i = Z_p \implies p(i, j) = 0.$$

Endow ω with the metric d_ω obtained by restricting the metric

$$d_{S \times T}\left((i_1, j_1), (i_2, j_2)\right) := \max\left\{d_S(i_1, i_2), d_T(i_1, i_2)\right\}$$

on $S \times T$ to ω. The set ω is closed since both S and T are closed. Further, $\omega \neq \varnothing$ since, e.g., $(0,0) \in \omega$. Moreover,

$$\nu_{(n,p)}(L) = \sum_{j \in L} \mathbb{P}(C_p(n) = j)$$

and

$$\begin{aligned} \mu_p(p_S(\omega \cap (S \times L))) &= \mathbb{P}(Z_p \in p_S(\omega \cap (S \times L))) \\ &= \mathbb{P}(Z_p \geq \min L) \\ &\overset{(4)}{=} 1/p^{\min L}. \end{aligned}$$

Therefore, inequality (17) reduces to

$$\sum_{j \in L} \mathbb{P}(C_p(n) = j) \leq \frac{1}{p^{\min L}} + \varepsilon. \tag{18}$$

Furthermore,

$$\begin{aligned} \sum_{j \in L} \mathbb{P}(C_p(n) = j) &\overset{(6)}{=} \sum_{j \in L} \frac{\lfloor n/p^j \rfloor - \lfloor n/p^{j+1} \rfloor}{n} \\ &\leq \sum_{j=\min L}^{\max L} \frac{\lfloor n/p^j \rfloor - \lfloor n/p^{j+1} \rfloor}{n} \\ &= \frac{\lfloor n/p^{\min L} \rfloor}{n} - \frac{\lfloor n/p^{1+\max L} \rfloor}{n} \\ &\leq 1/p^{\min L}. \end{aligned}$$

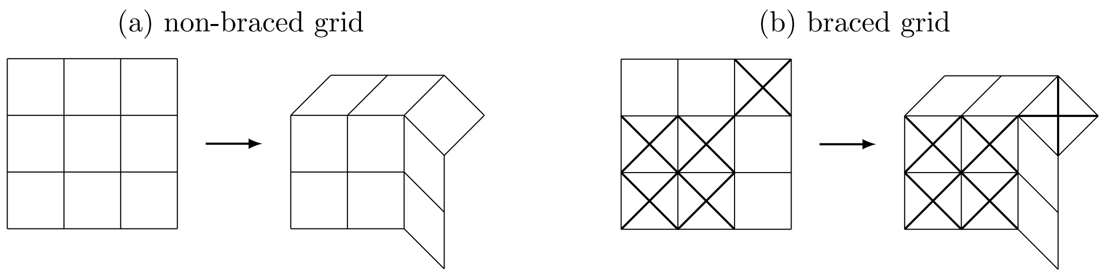
Thus, inequality (18) is true when $\varepsilon = 0$. Therefore, $p(\omega) \geq 1 - \varepsilon = 1$; thus, by Strassen's theorem, there exists a coupling $\left(Z'_p, C'_p(n)\right)$ of Z_p and $C_p(n)$ such that Z'_p pointwise dominates $C'_p(n)$. By Theorem 3.1, this proves inequality (2). □

5 Bipartite Graphs Corresponding to $C_p(n)$ and Z_p

In this section, we will construct a bipartite graph corresponding to our coupling in Section 3, and then we will prove Theorem 1.4d by showing that the bracing of the $\infty \times \lfloor \log_p(n) \rfloor$ grid corresponding to our bipartite graph is rigid if and only if n is not divisible by p.

Consider the following bracing problem described in pp. 74-75 of [4]: "A rectangular grid constructed of rigid struts can flex in many ways if no diagonal braces are included. Even if some braces are present, flexing may be possible." Figure 3 provides two examples of 3×3 rectangular grids in which flexing occurs.

Figure 3: Flexing in a 3×3 grid: (A) without bracing and (B) with bracing.



For each braced grid, we can identify its corresponding bipartite graph by drawing an edge from vertex i to vertex j whenever there is a brace in cell (i, j) of the grid.

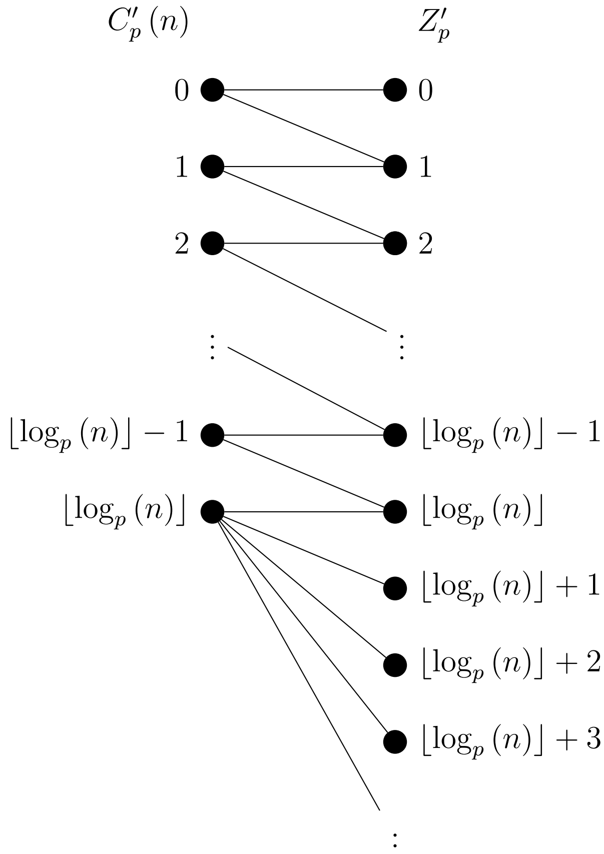
One variation of rigid bracing, known as tension bracing, is discussed on page 85 of [1]. In this setting, the cells are braced by wires, instead of rods, which may bend to shorter lengths among flexing.

Proof. [**Proof of Theorem 1.4d.**] When n is not divisible by p, the terms $p\,(i+1, i)$, $0 \le i \le k-1$, defined by equation (12) are positive. In Section 3, it was shown that $p\,(i, i) > 0$ when $0 \le i \le k$. Moreover, when $i \ge k$, we have

$$p\,(i, k) \overset{(14)}{=} \frac{1 - 1/p}{p^i} > 0.$$

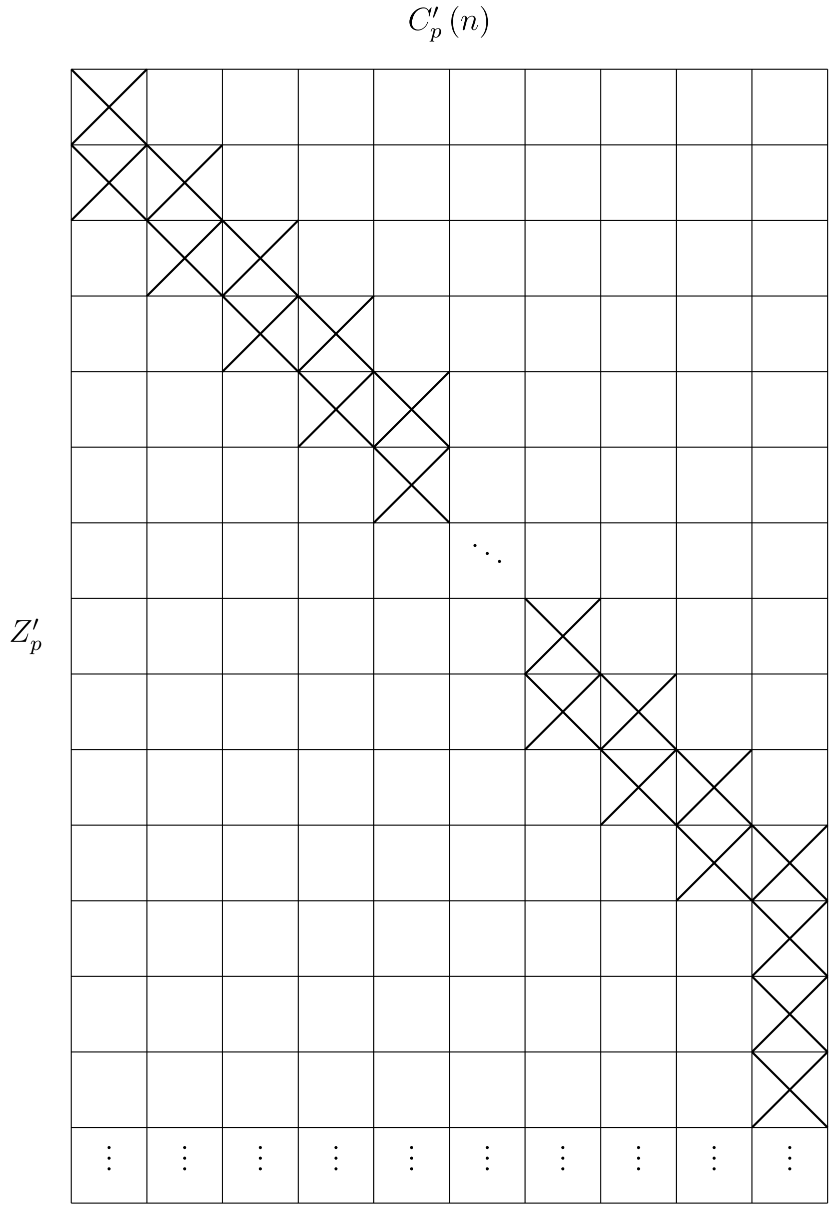
Thus, when n is not divisible by p, Table 2 corresponds to the bipartite graph shown in Figure 4 - an edge is drawn between the vertex $Z_p' = i$ and the vertex $C_p'\,(n) = j$ whenever $p\,(i, j) > 0$.

Figure 4: A bipartite graph corresponding to the coupling of Z_p and $C_p(n)$ given by Table 2 when n is not divisible by p.



Moreover, by [2], it is known that a finite braced grid is rigid (i.e., no flexing occurs) if and only if the corresponding bipartite graph is connected. The bipartite graph in Figure 4 is connected since there is a path between any two vertices, regardless of which set they are in. Therefore, the grid in Figure 5 is rigid.

Figure 5: An $\infty \times \lfloor \log_p(n) \rfloor$ rigid braced grid corresponding to the bipartite graph in Figure 4.



For if the grid in Figure 5 was not rigid, then there would exist a smallest row label l such that a cell in the lth row bends or flexes. To see that this is not possible, note the corresponding bipartite subgraph obtained by removing the vertices $Z'_p > l$ in Figure 4 is connected.

On the other hand, if n is divisible by p, then the by (8) the term $p(1,0)$ is equal to zero; as a result, there is no edge connecting the vertices $Z'_p = 1$ and $C'_p(n) = 0$ in the bipartite graph. This proves that the corresponding bipartite graph is not connected; thus, the grid is not rigid when n is divisible by p. $\square$

References

[1] J. Baglivo, J. Graver, *Incidence and Symmetry in Design and Architecture*, Cambridge University Press, 1983.

[2] E.D. Bolker, H. Crapo, How to brace a one-story building, *Environment and Planning B: Planning and Design*, **4** (1977), 125–152.

[3] B. Bollobás (Ed.), *Contemporary Combinatorics: 10 (Bolyai Society Mathematical Studies)*, Springer-Verlag, 2002.

[4] R.C. Bose, B. Manvel, *Combinatorial Theory*, John Wiley & Sons, 1984.

[5] D. Dentcheva, A. Ruszczyński, Optimization with stochastic dominance constraints, *SIAM Journal on Optimization*, **14** (2003), 548–566.

[6] S. Goldstein, Maximal coupling, *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **46** (1979), 193–204.

[7] G. Grimmett, D. Stirzaker, *Probability and Random Processes*, Oxford University Press, 2001.

[8] Is $\left\lfloor \frac{n}{p^i} \right\rfloor > \frac{n}{p^{i+1}}$, for $n \geq 1$, $p \leq n$ any prime, and $0 \leq i \leq \lfloor \log_p n \rfloor$, available online at the URL: `https://math.stackexchange.com/questions/2452461/is-left-lfloor-fracnpi-right-rfloor-fracnpi1-for-n-ge1-p`

[9] V. Strassen, The existence of probability spaces with given marginals, *Ann. Math. Stat.*, **36** (1965), 423–439.

[10] H. Thorisson, *Coupling, Stationarity, and Regeneration*, Springer, 2000.

Joseph Squillace
University of the Virgin Islands
10000 Castle Burke
Kingshill, USVI 00850-9960
E-mail: `joseph.squillace@uvi.edu`

Carissa George
University of the Virgin Islands
10000 Castle Burke
Kingshill, USVI 00850-9960
E-mail: `carissa.george@students.uvi.edu`

Received: December 19, 2024 **Accepted**: May 16, 2025
Communicated by Steven J. Miller