

# A VIEW FROM THE BRIDGE SPANNING COMBINATORICS AND PROBABILITY

ROSS G. PINSKY

ABSTRACT. This paper presents an offering of some of the myriad connections between Combinatorics and Probability, directed in particular toward combinatorialists. The choice of material was dictated by the author's own interests, tastes and familiarity, as well as by a desire to present results with either complete proofs or well developed sketches of proofs, and to ensure that the arguments are rather accessible to combinatorialists. The first several sections collect some concepts and rudimentary results from probability theory that are needed to understand the rest of the paper.

## CONTENTS

1. Introduction	2
2. Some basic probabilistic concepts	4
3. The binomial distribution, the Poisson approximation and some other classical distributions	5
4. Some basic probabilistic tools and results	7
5. Moments of the Poisson distribution, Stirling numbers of the second kind and Touchard polynomials	11
6. Dobiński's formula for Bell numbers and an algorithm for selecting a random partition of $[n]$	14
7. Chinese restaurant and Feller constructions of random permutations with applications	16
8. Convergence in distribution of cycle counts of fixed length in random permutations to Poisson distributions	20
9. Limiting behavior of the lengths of the largest and smallest cycles in random permutations and the connection to the Dickman and Buchstab functions	22

---

2010 *Mathematics Subject Classification.* 05-02, 60C05.

*Key words and phrases.* Combinatorics, Probability .

10.	The statistical behavior of random partitions of large integers	27
11.	Threshold calculations for Erdős-Rényi random graphs using the first and second moment methods	31
12.	Arcsine law for random walks and a combinatorial lemma of Sparre-Andersen	39
13.	The probabilistic method	45
	References	47

## 1. INTRODUCTION

I was approached by the editors to write a survey article, accessible to combinatorialists, discussing the “bridge” between combinatorics and probability. The oldest lanes on that bridge are more than three hundred years old. Probability problems from that period and for quite some time afterwards as well, involving cards and dice, helped spur the development of basic combinatorial principles, which were then applied to solve the probability problems. The initial lanes constructed in the other direction might be said to exist almost as long; indeed, de Moivre’s seminal 1733 paper proving the Central Limit Theorem in the context of fair coin-tossing may be thought of as a way of estimating binomial coefficients through a probabilistic lens. A very different aspect of probability in the service of combinatorics dates to the 1940’s when the so-called “Probabilistic Method” was first popularized by Erdős. Around the same time, an interest in random permutations began to develop among certain probabilists. In this latter case, it is difficult to say which field is the giver and which the receiver. The rising tide of graph theory and computer algorithms over the past seventy years has certainly provided fertile ground for the use of probability in combinatorics. By today this bridge between combinatorics and probability has amassed a multitude of decks and lanes, far more than can be dealt with in a single survey of modest length. Thus, the choice of material was dictated by my own interests, tastes and familiarity. It was also guided by a desire to present results with either complete proofs or well developed sketches of proofs, and to ensure that the arguments are rather accessible to combinatorialists, including graduate students in the field. A large majority of the material herein

concerns enumerative combinatorics. For more on the historical view of the connection between combinatorics and probability, see [17, 26, 63].

Elementary probability based on equally likely outcomes is almost by definition co-extensive with enumerative combinatorics. Today, probabilists sometimes study combinatorial structures on which probability distributions other than the uniform ones are used. These other distributions are usually constructed as an exponential family with a parameter. For example, in the context of random permutations, a lot of work has been done in recent years in the case that the uniform measure is replaced by a measure from the family of Ewens sampling distributions or from the family of Mallows distributions. For the Ewens sampling distributions, see for example [5, 28]; for the Mallows distributions, see for example [9, 31, 32, 52, 58]. In this survey, we restrict to the uniform distribution on combinatorial objects.

I now note several from among many of the important and interesting directions that I do not touch upon at all. One such direction is Analytic Combinatorics, where probability and complex function theory unite to provide answers to problems concerning the probabilistic structure of large combinatorial objects. We refer the reader to the book [29] for a comprehensive treatment and many additional references. See also the book [54]. Another such direction is the problem of the length of the longest increasing subsequence in a random permutation. This very specific problem turns out to be packed with a great wealth and variety of mathematical ideas and directions. The original seedling for this direction was provided in [34]. Then fundamental contributions were provided in [73, 47], followed by the tour-de-force [7]. We refer the reader to the book [64] for a very readable treatment and comprehensive references. Another direction not touched upon here is random trees. We refer the reader to the book [21] and the survey article [45], along with references therein, to the classical articles [1, 2] on the Brownian continuum random tree, and for example to the papers [10, 50, 46]. Yet another direction not discussed in this article is the probabilistic aspects of pattern avoiding permutations. See for example [36, 37, 38, 39, 40] and references therein. For the combinatorial aspects of pattern avoiding permutations, see for example [14].

A nice set of lecture notes on combinatorics and probability is provided by [61]. The paper [60] aims to bridge several results in combinatorics and

probability. The book [67] concerns probabilistic methods in solving combinatorial problems. A combinatorics book with many probabilistic exercises is [16].

To access the bridge on the probability side, one must pass a toll booth, which consists of some basic probability. This is the material in sections 2, 3 and 4. An excellent reference for this material is [22]; see also [26]. Readers who so wish can skirt the toll booth and access the bridge at section 5, referring back if and when necessary .

## 2. SOME BASIC PROBABILISTIC CONCEPTS

Let  $\Omega$  be a set,  $\mathcal{F}$  a  $\sigma$ -algebra on  $\Omega$  and  $P$  a probability measure on the measure space  $(\Omega, \mathcal{F})$ . The triplet  $(\Omega, \mathcal{F}, P)$  is called a probability space. In the case that  $\Omega$  is finite or countably infinite, there is no reason not to choose  $\mathcal{F}$  to be the  $\sigma$ -algebra consisting of all subsets of  $\Omega$ . In such cases, we will refer to  $(\Omega, P)$  as a probability space. A random variable  $X$  on the probability space  $(\Omega, \mathcal{F}, P)$  is a measurable function from  $(\Omega, \mathcal{F})$  to  $(\mathbb{R}, \mathcal{B})$ , where  $\mathcal{B}$  is the Borel  $\sigma$ -algebra on  $\mathbb{R}$ . For  $A \in \mathcal{B}$ , we write  $P(X \in A)$  for  $P(X^{-1}(A))$ . The expected value of  $X$  is defined as

$$EX = \int_{\Omega} X dP,$$

if  $X$  is integrable (that is, if  $\int_{\Omega} |X| dP < \infty$ ) and if so, then the variance of  $X$  is defined as

$$\sigma^2(X) = E(X - EX)^2 = \int_{\Omega} (X - EX)^2 dP,$$

which can be written as

$$\sigma^2(X) = EX^2 - (EX)^2.$$

For  $k \in \mathbb{N}$ , the  $k$ th moment of the random variable is defined as

$$EX^k = \int_{\Omega} X^k dP,$$

if  $\int_{\Omega} |X|^k dP < \infty$ . If  $\int_{\Omega} |X|^k dP = \infty$ , we say that the  $k$ th moment does not exist. A subset  $D \subset \Omega$  is called an event. A collection  $\mathcal{D}$  of events is called independent if for every finite or countably infinite collection  $\{D_j\}_{j \in I} \subset \mathcal{D}$ , one has  $P(\cap_{j \in I} D_j) = \prod_{j \in I} P(D_j)$ . A sequence  $\{X_n\}_{n=1}^{\infty}$  of random variables is called independent if  $P(X_1 \in A_1, X_2 \in A_2, \dots) = \prod_{n=1}^{\infty} P(X_n \in A_n)$ , for all Borel sets  $A_n \subset \mathbb{R}$ ,  $n = 1, 2, \dots$ .

Given events  $A, B$  with  $P(A) > 0$ , the conditional probability of the event  $B$  given the event  $A$  is denoted by  $P(B|A)$  and defined by

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

The distribution  $\nu$  of a random variable  $X$  is the probability measure on  $\mathbb{R}$  defined  $\nu(A) = P(X \in A)$ , for Borel sets  $A \in \mathcal{B}$ . The distribution function  $F(x)$  of a random variable  $X$  is defined by

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

It follows that  $F$  is continuous at  $x$  if and only if  $x$  is not an atom for the distribution  $\nu$  of  $X$ ; that is if and only if  $\nu(\{x\}) = P(X = x) = 0$ . Given a probability measure  $\nu$  on  $\mathbb{R}$ , one can always construct a probability space and a random variable on it whose distribution is  $\nu$ .

A sequence  $\{\nu_n\}_{n=1}^{\infty}$  of probability measures on  $\mathbb{R}$  is said to converge weakly to the probability measure  $\nu$  on  $\mathbb{R}$  if  $\lim_{n \rightarrow \infty} \nu_n(A) = \nu(A)$ , for all Borel sets  $A \in \mathcal{B}$  which satisfy  $\nu(\partial A) = 0$ , where  $\partial A$  denotes the boundary of  $A$ . With the help of some nontrivial measure-theoretic machinery, weak convergence can be shown to be equivalent to the seemingly weaker requirement that replaces the class of all Borel sets above by the class of rays  $(-\infty, x]$ , with  $x \in \mathbb{R}$ .

A sequence  $\{X_n\}_{n=1}^{\infty}$  of random variables is said to converge in distribution to a random variable  $X$  if their distributions  $\{\nu_n\}_{n=1}^{\infty}$  converge weakly to the distribution  $\nu$  of  $X$ . In light of the above two paragraphs, this requirement can be written as

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \quad \text{for all } x \in \mathbb{R} \text{ at which } F \text{ is continuous.}$$

We write  $X_n \xrightarrow{\text{dist}} X$ . Note that if  $X$  is almost surely equal to the constant  $c$ , then  $X_n \xrightarrow{\text{dist}} X$  is equivalent to  $\lim_{n \rightarrow \infty} P(|X_n - c| \geq \epsilon) = 0$ , for every  $\epsilon > 0$ .

### 3. THE BINOMIAL DISTRIBUTION, THE POISSON APPROXIMATION AND SOME OTHER CLASSICAL DISTRIBUTIONS

For each  $n \in \mathbb{N}$  and each  $p \in [0, 1]$ , the *binomial distribution*  $\text{Binomial}(n, p)$  describes the random number  $X_{n,p}$  of “successes” in a sequence of independent trials, where for each trial the probability of success is  $p$ . Such trials are known as *Bernoulli trials*. Simple probabilistic and combinatoric reasoning

reveals that

$$P(X_{n,p} = j) = \binom{n}{j} p^j (1-p)^{n-j}, \quad j = 0, 1, \dots, n.$$

A random variable  $X_\lambda$  has the *Poisson distribution* with parameter  $\lambda > 0$  if

$$P(X_\lambda = j) = e^{-\lambda} \frac{\lambda^j}{j!}, \quad j = 0, 1, \dots.$$

We denote this distribution by  $\text{Poisson}(\lambda)$ . Unlike the case of the binomial distribution, at first glance one does not see any combinatorics here; yet as we shall see directly below and in later sections, quite a bit of combinatorics is lurking behind the scenes. A basic result appearing in a first course in probability theory is the ‘‘Poisson approximation to the binomial distribution,’’ which states that for large  $N$  and small  $p$ , the  $\text{Poisson}(Np)$  distribution approximates the  $\text{Binomial}(N, p)$  distribution. More precisely,

$$\lim_{N \rightarrow \infty, Np \rightarrow \lambda} P(X_{N,p} = j) = P(X_\lambda = j), \quad j = 0, 1, \dots;$$

that is,

$$(3.1) \quad \lim_{N \rightarrow \infty, Np \rightarrow \lambda} \binom{N}{j} p^j (1-p)^{N-j} = e^{-\lambda} \frac{\lambda^j}{j!}, \quad j = 0, 1, \dots.$$

The proof of (3.1) is an easy calculus exercise. We note that (3.1) is equivalent to the statement

$$X_{n,p_n} \xrightarrow{\text{dist}} X_\lambda \text{ as } n \rightarrow \infty \text{ and } np_n \rightarrow \lambda.$$

The Poisson distribution arises in a wealth of applications, the most canonical of which is probably with regard to call centers. Under many real-life situations, the number of calls that arrive at a call center during  $t$  units of time, for any  $t > 0$ , can be modeled by the Poisson distribution with parameter  $\lambda t$ , where  $\lambda$  is referred to as the intensity parameter. The intuition behind this is as follows. Fix  $t > 0$ , and let  $n \in \mathbb{N}$ . Divide the time interval  $[0, t]$  into subintervals  $\{J_i^{(n)}\}_{i=1}^n$  all of length  $\frac{t}{n}$ . One makes two assumptions: (i) the (random) number of calls in different subintervals are independent and identically distributed; (ii) in each subinterval  $J_i^{(n)}$ , the probability of at least one call is  $\frac{\lambda t}{n} + o(\frac{1}{n})$ , for some appropriate  $\lambda > 0$  that characterizes the call center, and the probability of more than one call is  $o(\frac{1}{n})$ . Under this scenario, it is easy to see that the distribution of the number of calls can be approximated by  $\text{Bin}(n, \frac{\lambda t}{n})$ . (Note that the probability that no interval has more than one call is given by  $(1 - o(\frac{1}{n}))^n$ ,

which approaches one as  $n \rightarrow \infty$ .) And by the Poisson approximation,  $\text{Binomial}(n, \frac{\lambda t}{n})$  is close to the  $\text{Poisson}(\lambda t)$  distribution. For basic applications of the Poisson approximation, see [26], and for a more extensive study, see [8].

A random variable  $X$  has the *geometric distribution* with parameter  $\rho \in (0, 1)$  if  $P(X = k) = \rho^k(1 - \rho)$ ,  $k = 0, 1, \dots$ .

So far all the distributions we've presented have been discrete; that is there exists a finite or countably infinite set  $\{x_j\}_{j \in I}$ , such that a random variable  $X$  with the distribution in question satisfies  $P(X = x_j) > 0$ , for all  $j \in I$ , and  $\sum_{j \in I} P(X = x_j) = 1$ . Now we recall three continuous distributions that will appear in this survey. A distribution is continuous if a random variable with the distribution in question satisfies  $P(X = x) = 0$ , for all  $x \in \mathbb{R}$ ; equivalently, the random variable's distribution function  $F(x)$  is a continuous function.

A random variable  $X$  has the *exponential distribution* with parameter  $\lambda > 0$  if its distribution function is given by  $F(x) = 1 - e^{-\lambda x}$ , for  $x \geq 0$ . Its density function is given by  $f(x) := F'(x) = \lambda e^{-\lambda x}$ ,  $x \geq 0$ . A random variable has the *standard normal* (or *Gaussian*) distribution, denoted by  $N(0, 1)$ , if its distribution function is given by  $F(x) = \int_{-\infty}^x \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy$ ,  $x \in \mathbb{R}$ . Its density function is  $f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ ,  $x \in \mathbb{R}$ . For  $X_{N(0,1)}$  a random variable with the standard normal distribution, one has  $EX_{N(0,1)} = 0$  and  $\sigma^2(X_{N(0,1)}) = 1$ . A random variable  $U$  has the *uniform distribution* on  $[0, 1]$  if its distribution function satisfies  $F(x) = x$ , for  $x \in [0, 1]$ . Its density function is  $f(x) = 1$ ,  $x \in [0, 1]$ .

#### 4. SOME BASIC PROBABILISTIC TOOLS AND RESULTS

Consider the probability space  $(\Omega, \mathcal{F}, P)$ . For events  $\{D_j\}_{j=1}^N$ , we have by the subadditivity of measures,

$$P(\cup_{j=1}^N D_j) \leq \sum_{j=1}^N P(D_j),$$

with equality if and only if  $P(D_i \cap D_j) = 0$ , for  $i \neq j$ .

Let  $X$  be a random variable, denote its expected value by  $EX$  and let  $\sigma^2(X) = E(X - EX)^2$  denote its variance. *Markov's inequality* states that

$$P(|X| \geq L) \leq \frac{1}{L} E|X|, \text{ for } L > 0.$$

Multiplying both sides of the above inequality by  $L$ , one sees that its proof is essentially immediate from the definition of the expectation. We recall *Chebyshev's inequality*, an extremely simple but often very useful formula. It states that

$$P(|X - EX| \geq \lambda) \leq \frac{\sigma^2(X)}{\lambda^2}, \quad \lambda > 0.$$

Here is a one-line proof using Markov's inequality (with  $L = 1$ ):

$$P(|X - EX| \geq \lambda) = P\left(\frac{(X - EX)^2}{\lambda^2} \geq 1\right) \leq E\frac{(X - EX)^2}{\lambda^2} = \frac{\sigma^2(X)}{\lambda^2}.$$

Let  $\{X_n\}_{n=1}^\infty$  be a sequence of random variables and let  $S_n = \sum_{j=1}^n X_j$ . We have

$$ES_n = \sum_{j=1}^n EX_j.$$

If the random variables are independent and have finite variances, then a rather straightforward calculation reveals that the variance of the sum equals the sum of the variances:

$$\sigma^2(S_n) = \sum_{j=1}^n \sigma^2(X_j).$$

Now assume that  $\{X_n\}_{n=1}^\infty$  are independent and identically distributed (henceforth IID). Assuming that the expected value exists for these random variables, and denoting it by  $\mu$ , the *Weak Law of Large Numbers* holds; namely,

$$\frac{S_n}{n} \xrightarrow{\text{dist}} \mu.$$

As noted earlier, the above result is equivalent to

$$(4.1) \quad \lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0, \quad \text{for all } \epsilon > 0.$$

Under the additional assumption of finite variance, (4.1) follows immediately by applying Chebyshev's inequality to the random variable  $\frac{S_n}{n}$ , and using the fact that  $\sigma^2(S_n) = n\sigma^2(X_1)$ .

Assuming that the IID random variables  $\{X_n\}_{n=1}^\infty$  have a finite variance  $\sigma^2$ , then the *Central Limit Theorem* holds. This result describes the probabilistic fluctuations of  $\frac{S_n}{n}$  from its expected value  $\mu$ :

$$\frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mu\right) = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{dist}} X_{N(0,1)},$$

where  $X_{N(0,1)}$  denotes a random variable with the standard normal distribution.

More generally, if  $\{W_n\}_{n=1}^\infty$  is an arbitrary sequence of random variables with finite expectations, we say that the weak law of large numbers holds for  $\{W_n\}_{n=1}^\infty$  if

$$\frac{W_n}{EW_n} \xrightarrow{\text{dist}} 1,$$

and if furthermore, the variances  $\sigma^2(W_n)$  are finite, we say that the central limit theorem holds for  $\{W_n\}_{n=1}^\infty$  if

$$\frac{W_n - EW_n}{\sigma(W_n)} \xrightarrow{\text{dist}} X_{N(0,1)}.$$

Let  $\{W_n\}_{n=1}^\infty$  be an arbitrary sequence of random variables. The *first moment method* states that:

$$(4.2) \quad \text{if } \lim_{n \rightarrow \infty} E|W_n| = 0, \text{ then } W_n \xrightarrow{\text{dist}} 0.$$

The proof is immediate from Markov's inequality; indeed,

$$P(|W_n| \geq \epsilon) \leq \frac{1}{\epsilon} E|W_n| \xrightarrow{n \rightarrow \infty} 0, \text{ for all } \epsilon > 0.$$

The *second moment method* states that

$$(4.3) \quad \text{if } \sigma^2(W_n) = o((EW_n)^2), \text{ then } \frac{W_n}{EW_n} \xrightarrow{\text{dist}} 1.$$

The proof is almost immediate from Chebyshev's inequality. Indeed, let  $Y_n = \frac{W_n}{EW_n}$ . Then  $EY_n = 1$  and  $\sigma^2(Y_n) = \frac{1}{(EW_n)^2} \sigma^2(W_n)$ , where we use the fact that  $\sigma^2(cX) = c^2 \sigma^2(X)$ , if  $X$  is a random variable and  $c$  is a constant. Thus, by Chebyshev's inequality, for any  $\epsilon > 0$ ,

$$P\left(\left|\frac{W_n}{EW_n} - 1\right| \geq \epsilon\right) = P(|Y_n - EY_n| \geq \epsilon) \leq \frac{\sigma^2(Y_n)}{\epsilon^2} = \frac{\sigma^2(W_n)}{\epsilon^2 (EW_n)^2} \xrightarrow{n \rightarrow \infty} 0.$$

A frequently employed tool in these pages is the method of indicator random variables. The indicator random variable  $I_A$  for an event  $A \subset \Omega$  is defined by

$$1_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A; \\ 0, & \text{if } \omega \in \Omega - A. \end{cases}$$

We illustrate this tool first with the case of a binomial random variable. According to the definition, we have  $EX_{n,p} = \sum_{j=0}^n j \binom{n}{j} p^j (1-p)^{n-j}$ . A simpler way to calculate this expectation is to recall the probabilistic model that gave rise to  $X_{n,p}$ ; namely that  $X_{n,p}$  is the number of successes in a sequence of  $n$  independent Bernoulli trials, where for each trial the probability of success is  $p$ . Thus, we define the indicator random variable  $I_i$  to be equal to 1 if there was a success on the  $i$ th Bernoulli trial and 0 if there was a

failure. Then  $I_i$  has the following *Bernoulli distribution* with parameter  $p$ :  $P(I_i = 1) = 1 - P(I_i = 0) = p$ . Clearly  $EI_i = p$ . We can represent  $X_{n,p}$  as

$$X_{n,p} = \sum_{i=1}^n I_i,$$

and then by the linearity of the expectation, we immediately obtain

$$EX_{n,p} = np.$$

Note that this calculation does not use the fact that the  $\{I_i\}_{i=1}^n$  are independent. An easy calculation shows that the variance of  $I_i$  is equal to  $p(1-p)$ . Thus, we have

$$\sigma^2(X_{n,p}) = \sum_{i=1}^n \sigma^2(I_i) = np(1-p).$$

When we employ indicator random variables in this article, the typical situation will be

$$X = \sum_{i=1}^n 1_{A_i},$$

where the events  $\{A_i\}_{i=1}^n$  are not independent. Since  $E1_A = P(A)$ , we have

$$EX = \sum_{i=1}^n P(A_i).$$

Since the variance is given by  $\sigma^2(X) = EX^2 - (EX)^2$ , we will have the variance at our disposal as soon as we calculate the second moment. We have

$$(4.4) \quad EX^2 = E\left(\sum_{i=1}^n 1_{A_i}\right)^2 = \sum_{i=1}^n P(A_i) + 2 \sum_{1 \leq i < j \leq n} P(A_i \cap A_j).$$

We conclude this section with the method of moments for proving convergence in distribution [22]. Consider a random variable  $X$  which possesses all of its moments. Under the growth condition

$$\limsup_{k \rightarrow \infty} \frac{1}{k} (EX^{2k})^{\frac{1}{2k}} < \infty,$$

the convergence in distribution of a sequence of random variables  $\{X_n\}_{n=1}^{\infty}$  to  $X$  follows if

$$\lim_{n \rightarrow \infty} EX_n^k = EX^k, \text{ for all } k \in \mathbb{N}.$$

5. MOMENTS OF THE POISSON DISTRIBUTION, STIRLING NUMBERS OF THE SECOND KIND AND TOUCHARD POLYNOMIALS

It turns out that the Poisson distribution has an intimate connection with the *Stirling numbers of the second kind*  $S(n, k)$ ,  $n \in \mathbb{N}, 1 \leq k \leq n$ . Recall that  $S(n, k)$  counts the number of ways to partition a set of  $n$  labelled objects into  $k$  non-empty subsets [48, 71]. For convenience in formulas, one defines  $S(n, 0) = 0$ , for  $n \in \mathbb{N}$  and  $S(0, 0) = 1$ . The connection comes via the moments of the Poisson distribution. The  $n$ th moment, by definition, is given by

$$(5.1) \quad EX_\lambda^n = \sum_{k=0}^{\infty} (e^{-\lambda} \frac{\lambda^k}{k!}) k^n, \quad n = 0, 1, \dots;$$

however we will calculate the moments using a different tack.

It will be convenient to define the falling factorial:

$$(x)_j = x(x-1) \cdots (x-j+1) = j! \binom{x}{j}, \quad \text{for } j \in \mathbb{N} \text{ and } x \in \mathbb{R}.$$

Also, for convenience, one defines  $(x)_0 = 1$ . The falling factorials and the Stirling numbers of the second kind are connected by the formula

$$(5.2) \quad x^n = \sum_{j=0}^n S(n, j) (x)_j, \quad n = 0, 1, \dots.$$

We supply a quick proof of (5.2) [71]. We may assume that  $n \geq 1$ . It is enough to prove (5.2) for positive integers  $x$ , in which case  $x^n$  is the number of functions  $f : [n] \rightarrow A$ , where  $|A| = x$ . We now count such functions in another way. For each  $j \in [x]$ , consider all those functions whose range contains exactly  $j$  elements. The inverse images of these elements give a partition of  $[n]$  into  $j$  nonempty sets,  $\{B_i\}_{i=1}^j$ . We can choose the particular  $j$  elements in  $\binom{x}{j}$  ways, and we can order the sets  $\{B_i\}_{i=1}^j$  in  $j!$  ways. Thus there are  $S(n, j) \binom{x}{j} j! = S(n, j) (x)_j$  such functions. Therefore the number of functions  $f$  as above can also be represented as  $\sum_{j=0}^n S(n, j) (x)_j$ .

In the case of the Poisson distribution, the falling factorial moments,  $E(X_\lambda)_j$ , are very simple. We have

$$(5.3) \quad E(X_\lambda)_j = \sum_{k=0}^{\infty} (e^{-\lambda} \frac{\lambda^k}{k!}) (k)_j = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{(k-j)!} = \lambda^j, \quad j \geq 0.$$

Substituting  $X_\lambda$  for  $x$  in (5.2), taking expectations on both sides of the equation and using (5.3) gives

$$(5.4) \quad EX_\lambda^n = \sum_{j=0}^n S(n, j) E(X_\lambda)_j = \sum_{j=0}^n S(n, j) \lambda^j.$$

The *Touchard polynomials* (see, for example, [49]), are defined by

$$(5.5) \quad T_n(x) = \sum_{j=0}^n S(n, j) x^j, \quad n = 0, 1, \dots.$$

(For  $n \geq 1$ , we can begin the summation above from  $j = 1$ .) Thus, from (5.4), the moments of a Poisson-distributed random variable can be written in terms of the Touchard polynomials as

$$(5.6) \quad EX_\lambda^n = T_n(\lambda), \quad n = 0, 1, \dots.$$

Recall that for each  $n \in \mathbb{N}$ , the *Bell number*  $B_n$  counts the total number of partitions of the set  $[n]$ ; that is,

$$(5.7) \quad B_n = \sum_{k=0}^n S(n, k), \quad n = 1, 2, \dots.$$

It then follows from (5.6) that

$$(5.8) \quad EX_1^n = B_n, \quad n = 1, 2, \dots.$$

Thus, the  $n$ th moment of the Poisson distribution with parameter  $\lambda = 1$  is equal to the total number of partitions of the set  $[n]$ .

In each of the sections 6 and 8, we give an application of the connection between the Poisson distribution and Stirling numbers of the second kind as expressed in (5.6).

The calculations in (5.2)-(5.4) hardly lend intuition as to why the Poisson distribution should be connected to the Stirling numbers of the second kind via (5.6). The intuition (as well as an alternate proof of (5.6)) comes from the Poisson distribution's connection to combinatorics; namely, through the Poisson approximation to the binomial distribution.

*Alternate proof of (5.6).* We may assume that  $n \geq 1$ . The  $n$ th moment of a random variable distributed as  $\text{Bin}(N, \frac{\lambda}{N})$  is given by

$$(5.9) \quad EX_{N, \frac{\lambda}{N}}^n = \sum_{j=0}^N j^n \binom{N}{j} \left(\frac{\lambda}{N}\right)^j \left(1 - \frac{\lambda}{N}\right)^{N-j}.$$

Since

$$\sum_{j=M}^N j^n \binom{N}{j} \left(\frac{\lambda}{N}\right)^j \left(1 - \frac{\lambda}{N}\right)^{N-j} \leq \sum_{j=M}^{\infty} \frac{j^n}{j!} \lambda^j, \quad 0 \leq M \leq N,$$

we have  $\lim_{M \rightarrow \infty} \sup_{N \geq M} \sum_{j=M}^N j^n \binom{N}{j} \left(\frac{\lambda}{N}\right)^j \left(1 - \frac{\lambda}{N}\right)^{N-j} = 0$ . This in conjunction with (3.1) shows that

$$(5.10) \quad EX_{\lambda}^n = \lim_{N \rightarrow \infty} EX_{N, \frac{\lambda}{N}}^n, \quad n = 0, 1, \dots$$

Writing  $X_{N, \frac{\lambda}{N}}$  as a sum of indicator random variables,

$$X_{N, \frac{\lambda}{N}} = \sum_{j=1}^N I_j^{(N)},$$

where the  $\{I_j^{(N)}\}_{j=1}^N$  are IID and distributed according to the Bernoulli distribution with parameter  $\frac{\lambda}{N}$ , we have

$$(5.11) \quad X_{N, \frac{\lambda}{N}}^n = \sum_{j_1, j_2, \dots, j_n=1}^N \prod_{i=1}^n I_{j_i}^{(N)}, \quad n \in \mathbb{N}.$$

Note that

$$(5.12) \quad E \prod_{i=1}^n I_{j_i}^{(N)} = \left(\frac{\lambda}{N}\right)^k, \quad \text{where } k = |\{j_1, j_2, \dots, j_n\}|.$$

From here on, we assume that  $N \geq n$ . (Recall from (5.10) that we will be letting  $N \rightarrow \infty$  with  $n$  fixed.) How many  $n$ -tuples  $(j_1, \dots, j_n) \in [N]^n$  satisfy  $|\{j_1, j_2, \dots, j_n\}| = k$ ? We first need to choose  $k$  integers from  $[N]$ ; there are  $\binom{N}{k}$  ways to make this choice. Having fixed  $k$  such integers, we then need to set  $j_i$  equal to one of these integers, for each  $i = 1, \dots, n$ , and we need to do it in such a way as to ensure that  $|\{j_1, \dots, j_n\}| = k$ . The number of ways to do this is  $k!$  times the number of ways to partition a set of size  $n$  into  $k$  non-empty sets; namely,  $k!S(n, k)$  ways. Using this observation along with (5.11) and (5.12), we have

$$(5.13) \quad EX_{N, \frac{\lambda}{N}}^n = \sum_{k=1}^n \binom{N}{k} k! S(n, k) \left(\frac{\lambda}{N}\right)^k, \quad n \in \mathbb{N}.$$

From (5.13), we conclude that

$$(5.14) \quad \lim_{N \rightarrow \infty} EX_{N, \frac{\lambda}{N}}^n = \sum_{k=0}^n S(n, k) \lambda^k = T_n(\lambda), \quad n = 0, 1, \dots$$

From (5.14) and (5.10), we obtain (5.6).  $\square$

6. DOBIŃSKI'S FORMULA FOR BELL NUMBERS AND AN ALGORITHM FOR  
SELECTING A RANDOM PARTITION OF  $[n]$

Considering (5.1) along with (5.8) yields the beautiful formula

$$(6.1) \quad B_n = e^{-1} \sum_{k=0}^{\infty} \frac{k^n}{k!}, \quad n \in \mathbb{N}.$$

It is quite striking that the sequence  $\{B_n\}_{n=1}^{\infty}$  of combinatorial numbers can be represented as such an infinite series with a parameter. This result is known as *Dobiński's formula* [19, 48, 49, 59, 65]. As an interesting application of Dobiński's formula, we define a constructive algorithm that chooses uniformly at random a partition of  $[n]$ ; that is, it chooses at random one of the  $B_n$  partitions of  $[n]$  in such a way that each partition has probability  $\frac{1}{B_n}$  of being chosen [70, 59].

In order to choose a partition of  $[n]$ , we must choose a  $k \in [n]$  that determines how many sets will be in the partition, and then given  $k$ , we must choose one of the  $S(n, k)$  partitions of  $[n]$  into  $k$  non-empty subsets. We utilize a balls-in-bins type of construction. For each  $n$ , let  $M_n$  be a positive, integer-valued random variable with distribution

$$(6.2) \quad P(M_n = m) = \frac{1}{eB_n} \frac{m^n}{m!}, \quad m = 1, 2, \dots.$$

Note that this is indeed a probability distribution in light of Dobiński's formula. Now take  $M_n$  bins, and sequentially place  $n$  balls, numbered from 1 to  $n$ , into these bins, uniformly at random. Let  $K_n$  denote the number of non-empty bins. (Of course,  $K_n$  is random.) We have now constructed a random set partition of  $[n]$  into  $K_n$  non-empty sets. Denote it by  $\Psi_n$ .

*Proof that the random partition  $\Psi_n$  is uniformly distributed.* We need to show that  $P(\Psi_n = \psi) = \frac{1}{B_n}$ , for every partition  $\psi$  of  $[n]$ . Fix a partition  $\psi$  of  $[n]$ . Let  $k$  denote the number of sets in the partition  $\psi$ . If  $M_n = m < k$ , then it is clear from the construction that it is not possible to have  $\Psi_n = \psi$ . That is, the conditional probability that  $\Psi_n = \psi$  given that  $M_n = m$  is equal to zero:

$$(6.3) \quad P(\Psi_n = \psi | M_n = m) = 0, \quad \text{if } m < k.$$

On the other hand, for  $m \geq k$ , we claim that

$$(6.4) \quad P(\Psi_n = \psi | M_n = m) = \frac{(m)_k}{m^n} = \frac{m(m-1) \cdots (m-k+1)}{m^n}, \quad m \geq k.$$

To see this, note first that conditioned on  $M_n = m$ , there are  $m^n$  possible ways to place the balls into the bins. Thus, to prove (6.4), we need to show that of these  $m^n$  ways, exactly  $m(m-1)\cdots(m-k+1)$  of them result in constructing the partition  $\psi$ . In the paragraph after the next one, we will prove this fact. In order to make the proof more reader friendly, in the next paragraph, we consider what must happen to the first three balls in two specific examples.

Let  $3 \leq k \leq n$  and of course, as in (6.4), let  $m \geq k$ . Assume first, for example, that the numbers 1 and 3 appear in the same subset of the partition  $\psi$ , and that the number 2 appears in a different subset of  $\psi$ . Then in order to construct the partition  $\psi$ , there are  $m$  bins to choose from for ball number 1, there are  $m-1$  bins to choose from for ball number 2, namely any bin except for the one containing ball number 1, and there is one bin to choose from for ball number 3, namely the bin containing ball number 1. Now assume alternatively that the numbers 1,2 and 3 all appear in different subsets in the partition  $\psi$ . Then in order to construct the partition  $\psi$ , there are  $m$  bins to choose from for ball number 1, there are  $m-1$  bins to choose from for ball number 2, and there are  $m-2$  bins to choose from for ball number 3.

We now give the full proof. Let  $l_1 = 0$  and for  $j = 2, \dots, n$ , let  $l_j$  equal the number of subsets in the partition  $\psi$  that contain at least one of the numbers in  $[j-1]$ . For each  $j \in [n]$ , consider the situation where balls numbered 1 through  $j-1$  have been placed in bins in such a way that it is still possible to end up with the partition  $\psi$ , and then let  $\gamma_j$  denote the number of bins in which ball number  $j$  can be placed in order to continue to preserve the possibility of ending up with the partition  $\psi$ . Of course  $\gamma_1 = m$ . For  $j \geq 2$ ,  $\gamma_j$  is equal to 1 if in  $\psi$ , the subset containing  $j$  also contains an element from  $[j-1]$ , while otherwise,  $\gamma_j = m - l_j$ . Furthermore, for  $j \geq 1$ , if  $\gamma_j = 1$ , then  $l_{j+1} = l_j$ , while if  $\gamma_j = m - l_j$ , then  $l_{j+1} = l_j + 1$ . From this analysis, it follows that the number of ways to place the  $n$  balls into the  $m$  bins in such a way as to result in constructing  $\psi$  is  $\prod_{j=1}^n \gamma_j = m(m-1)\cdots(m-j+1)$ .

From (6.2)-(6.4) we conclude that

$$P(\Psi_n = \psi) = \sum_{m=1}^{\infty} P(M_n = m)P(\Psi_n = \psi|M_n = m) =$$

$$\sum_{m=k}^{\infty} \frac{1}{eB_n} \frac{m^n}{m!} \frac{(m)_k}{m^n} = \frac{1}{eB_n} \sum_{m=k}^{\infty} \frac{1}{(m-k)!} = \frac{1}{B_n},$$

which completes the proof that the algorithm yields a uniformly random partition of  $[n]$ .  $\square$

We note that in [70] it is also proved that the number of empty bins appearing in the construction is independent of the particular partition produced, and is distributed according to the Poisson distribution with parameter 1.

For some more on probabilistic aspects of set partitions, see [48, 59]. For some results concerning the asymptotic behavior of certain partition statistics under the uniform distribution on partitions of  $[n]$ , see [35, 66].

## 7. CHINESE RESTAURANT AND FELLER CONSTRUCTIONS OF RANDOM PERMUTATIONS WITH APPLICATIONS

Consider the space  $S_n$  of permutations of  $[n]$  with the uniform probability measure  $P_n$  which gives probability  $\frac{1}{n!}$  to each permutation  $\sigma \in S_n$ . We present the so-called *Chinese restaurant construction* which simultaneously yields a uniformly random permutation  $\Sigma_n$  in  $S_n$ , for all  $n$  [61]. Furthermore, this construction is consistent in the sense that if one writes out the permutation  $\Sigma_n$  as the product of its cycles and deletes the number  $n$  from the cycle in which it appears, then the resulting random permutation of  $S_{n-1}$  is equal to  $\Sigma_{n-1}$ .

The construction works as follows. Consider a restaurant with an unlimited number of circular tables, each of which has an unlimited number of seats. Person number 1 sits at a table. Now for  $n \geq 1$ , suppose that persons number 1 through  $n$  have already been seated. Then person number  $n+1$  chooses a seat as follows. For each  $j \in [n]$ , with probability  $\frac{1}{n+1}$ , person number  $n+1$  chooses to sit to the left of person number  $j$ . Also, with probability  $\frac{1}{n+1}$ , person number  $n+1$  chooses to sit at an unoccupied table. Now for each  $n \in \mathbb{N}$ , the random permutation  $\Sigma_n \in S_n$  is defined by  $\Sigma_n(i) = j$ , if after the first  $n$  persons have taken seats, person number  $j$  is seated to the left of person number  $i$ .

To see that  $\Sigma_n$  is a uniformly random permutation in  $S_n$ , proceed by induction. It is true for  $n = 1$ . Now assume it is true for some  $n \geq 1$ . Let  $\sigma \in S_{n+1}$ , and let  $\sigma' \in S_n$  be the permutation obtained by writing  $\sigma$  as a product of cycles and deleting  $n + 1$ . By the inductive assumption,  $P(\Sigma_n = \sigma') = \frac{1}{n!}$ . By the construction,  $P(\Sigma_{n+1} = \sigma | \Sigma_n = \sigma') = \frac{1}{n+1}$  and  $P(\Sigma_{n+1} = \sigma | \Sigma_n \neq \sigma') = 0$ . Thus,

$$P(\Sigma_{n+1} = \sigma) = P(\Sigma_n = \sigma')P(\Sigma_{n+1} = \sigma | \Sigma_n = \sigma') + P(\Sigma_n \neq \sigma')P(\Sigma_{n+1} = \sigma | \Sigma_n \neq \sigma') = \frac{1}{n!} \frac{1}{n+1} = \frac{1}{(n+1)!}.$$

The above-noted consistency is also clear.

The construction above allows for easy analysis of certain properties of random permutations. We illustrate this with regard to the number of cycles in a random permutation. Let  $N^{(n)}(\sigma)$  denote the number of cycles in the permutation  $\sigma \in S_n$ . Then  $N^{(n)}$  is a random variable on the probability space  $(S_n, P_n)$ . Fix  $n \in \mathbb{N}$ . The number of cycles in the random permutation  $\Sigma_n$  is the number of persons from among persons 1 through  $n$  who chose to sit at an unoccupied table. The probability that person  $j$  chose to sit at an unoccupied table is  $\frac{1}{j}$ . It is clear that whether or not person  $j$  chose to sit at an empty table is independent of whether or not any of the other persons chose to do so. In light of this, it follows that the distribution of the random variable  $N_n$  is equal to the distribution of  $\sum_{i=1}^n I_i$ , where  $\{I_i\}_{i=1}^n$  are independent and for each  $i$ ,  $I_i$  has the Bernoulli distribution with parameter  $\frac{1}{i}$ . From this it is immediate that

$$EN^{(n)} = \sum_{i=1}^n \frac{1}{i} \sim \log n \text{ as } n \rightarrow \infty.$$

Recalling that the variance of a sum of independent random variables is equal to the sum of their variances, a simple calculation gives

$$\text{Var}(N^{(n)}) = \sum_{i=1}^n \frac{i-1}{i^2} \sim \log n \text{ as } n \rightarrow \infty.$$

Now the second moment method (4.3) yields the weak law of large numbers for  $N^{(n)}$ .

**Theorem 1.** *The total number of cycles  $N^{(n)}$  in a uniformly random permutation from  $S_n$  satisfies*

$$(7.1) \quad \frac{N^{(n)}}{\log n} \xrightarrow{\text{dist}} 1.$$

And some rather basic probabilistic machinery yields the central limit theorem for  $N^{(n)}$  [22].

**Theorem 2.** *The total number of cycles  $N^{(n)}$  in a uniformly random permutation from  $S_n$  satisfies*

$$(7.2) \quad \frac{N^{(n)} - \log n}{\sqrt{\log n}} \xrightarrow{\text{dist}} N(0, 1).$$

We now present an alternative construction of a random permutation in  $S_n$ . The construction, called the Feller construction [25, 5], builds the permutation's cycles one by one. Begin with the number 1. Choose uniformly at random a number  $j$  from  $[n]$  and set it equal to  $\sigma_1$ :  $\sigma_1 = j$ . If  $j = 1$ , then we have constructed a complete cycle (of length one) in the permutation. If  $j \neq 1$ , choose uniformly at random a number  $k$  from  $[n] - \{j\}$  and set it equal to  $\sigma_j$ :  $\sigma_j = k$ . If  $k = 1$ , then we have constructed a complete cycle (of length two). If  $k \neq 1$ , choose uniformly at random a number  $l$  from  $[n] - \{j, k\}$  and set it equal to  $\sigma_k$ :  $\sigma_k = l$ . If  $l = 1$ , then we have constructed a complete cycle (of length three). If  $l \neq 1$ , choose uniformly at random a number  $m$  from  $[n] - \{j, k, l\}$  and set it equal to  $\sigma_l$ :  $\sigma_l = m$ . Continue like this until the number 1 is finally chosen and a cycle is completed. Once a cycle is completed, start the process over again beginning, say, with the smallest number that has not been used in the completed cycle. This number now takes on the role that the number 1 had above. After  $n$  selections of a number uniformly at random, each time from a set whose size has shrunk by one from the previous selection, the process ends and the construction of the permutation is completed. Denote the permutation constructed here by  $\Sigma'_n$ . It is clear from the construction that  $\Sigma'_n$  is uniformly distributed on  $S_n$ . As an example, let  $n = 6$ . If we first select uniformly at random from  $[6]$  the number 4, and then select uniformly at random from  $[6] - \{4\}$  the number 1, then we have  $\sigma_1 = 4$  and  $\sigma_4 = 1$ , completing a cycle. Now we start with the smallest number not yet used, which is 2. If we select uniformly at random from  $[6] - \{1, 4\}$  the number 2, then  $\sigma_2 = 2$ , and we have completed another cycle. Now we start again with the smallest number not yet used, which is

3. If we select uniformly at random from  $[6] - \{1, 2, 4\}$  the number 6, then we have  $\sigma_3 = 6$ . If then we select uniformly at random from  $[6] - \{1, 2, 4, 6\}$  the number 5, then  $\sigma_6 = 5$ . The last step of course is deterministic; we must choose  $\sigma_5 = 3$ . We have constructed the permutation  $\Sigma'_6 = 426135$  (where the permutation has been written in one-line notation).

From the above description, it is clear that the probability of completing a cycle at the  $j$ th selection of a number is  $\frac{1}{n-j+1}$ ; indeed, at the  $j$ th selection, there are  $n-j+1$  numbers to choose from, and only one of them completes a cycle. Define the indicator random variable  $I_{n-j+1}$  to be one if a cycle was completed at the  $j$ th selection, and zero otherwise. Then  $I_{n-j+1}$  has the following Bernoulli distribution with parameter  $\frac{1}{n-j+1}$ :  $P(I_{n-j+1} = 1) = 1 - P(I_{n-j+1} = 0) = \frac{1}{n-j+1}$ . A little thought also shows that the  $\{I_i\}_{i=1}^n$  are mutually independent; knowledge of the values of certain of the  $I_i$ 's has no influence on the probabilities concerning the other  $I_i$ 's. Thus, as with the Chinese restaurant construction, the Feller construction leads immediately to the fact that  $N_n$  is distributed as  $\sum_{i=1}^n I_i$ .

For  $j \in [n]$  and  $\sigma \in S_n$ , let  $L^{(n);j}(\sigma)$  denote the length of the cycle in  $\sigma$  that contains  $j$ . Then  $L^{(n);j}$  is a random variable on the probability space  $(S_n, P_n)$ . Since  $P_n$  is the uniform distribution, it is clear that the distribution of  $L^{(n);j}$  is the same for all  $j \in [n]$ . Using the Feller construction, we easily obtain the following result.

**Proposition 1.** *Under  $P_n$ , for each  $j \in [n]$ , the random variable  $L^{(n);j}$  is uniformly distributed on  $[n]$ .*

*Proof.* As noted above, it suffices to consider  $j = 1$ . From the Feller construction, the probability that for the uniformly random permutation  $\Sigma'_n$  in  $S_n$ , the cycle containing 1 is of size  $j$  is given by  $\frac{1}{n}$ , for  $j = 1$ , and is given by  $\frac{n-1}{n} \frac{n-2}{n-1} \cdots \frac{n-j+1}{n-j+2} \frac{1}{n-j+1} = \frac{1}{n}$ , for  $j = 2, \dots, n$ .  $\square$

Now for each  $n \in \mathbb{N}$  and  $j \in [n]$ , let  $A_j^{(n)} = A_j^{(n)}(\Sigma'_n)$  denote the length of the  $j$ th cycle constructed via the Feller construction, with  $A_j^{(n)} = 0$  if fewer than  $j$  cycles were constructed. We refer to  $\{A_j^{(n)}\}_{j=1}^n$  as the *ordered cycles*. Note that whereas  $L^{(n);j}$  was defined on  $(S_n, P_n)$ , independent of any particular construction,  $A_j^{(n)}$  is defined in terms of the Feller construction. Let  $\{U_n\}_{n=1}^\infty$  be a sequence of IID random variables with the uniform distribution on  $[0, 1]$ . Let  $X_1 = U_1$  and  $X_n = (1 - U_1) \cdots (1 - U_{n-1})U_n$ , for

$n \geq 2$ . The  $\{X_n\}_{n=1}^\infty$  can be understood in terms of the uniform *stick breaking* model: take a stick of length one and break it at a uniformly random point. Let the length of the left hand piece be  $X_1$ . Take the remaining piece of stick, of length  $1 - X_1$ , and break it at a uniformly random point. Let the length of the left hand piece be  $X_2$ , etc. In light of Proposition 1 and the Feller construction, the following theorem is almost immediate.

**Theorem 3.** *For any  $k \in \mathbb{N}$ , the random vector  $\frac{1}{n}(A_1^{(n)}, \dots, A_k^{(n)})$  converges in distribution to the random vector  $(X_1, \dots, X_k)$ , where*

$$X_j = (1 - U_1) \cdots (1 - U_{j-1})U_j, \quad j \in \mathbb{N},$$

and  $\{U_n\}_{n=1}^\infty$  are IID random variables distributed according to the uniform distribution on  $[0, 1]$ .

The distribution of  $\{X_n\}_{n=1}^\infty$  is known as the GEM distribution [5]. It will be mentioned again in section 9.

## 8. CONVERGENCE IN DISTRIBUTION OF CYCLE COUNTS OF FIXED LENGTH IN RANDOM PERMUTATIONS TO POISSON DISTRIBUTIONS

In this section we present a second application of the connection between the Poisson distribution and Stirling numbers of the second kind as expressed in (5.6). As in the previous section, we consider the space  $S_n$  of permutations of  $[n]$  with the uniform probability measure  $P_n$ . For each  $m \in [n]$ , let  $C_m^{(n)} : S_n \rightarrow \mathbb{N}$  count the number of  $m$ -cycles in a permutation; that is,  $C_m^{(n)}(\sigma)$  equals the number of  $m$ -cycles in  $\sigma \in S_n$ . Then  $C_m^{(n)}$  is a random variable on the probability space  $(S_n, P_n)$ . We use the standard probability notation  $P_n(C_m^{(n)} = k)$ :

$$P_n(C_m^{(n)} = k) = P_n(\{\sigma \in S_n : C_m^{(n)}(\sigma) = k\}) = \frac{|\{\sigma \in S_n : C_m^{(n)}(\sigma) = k\}|}{n!}.$$

For fixed  $n$ , the distribution of  $C_m^{(n)}$  is complicated, but we will prove that as  $n \rightarrow \infty$ , the distribution of  $C_m^{(n)}$  converges in distribution to the Poisson( $\frac{1}{m}$ ) distribution. That is,

$$(8.1) \quad \lim_{n \rightarrow \infty} P_n(C_m^{(n)} = k) = e^{-\frac{1}{m}} \frac{(\frac{1}{m})^k}{k!}, \quad k = 0, 1, \dots; \quad m = 1, 2, \dots.$$

In fact, a stronger result holds; namely that for any  $m \in \mathbb{N}$ , the distribution of the random vector  $(C_1^{(n)}, \dots, C_m^{(n)})$  converges in distribution to a

random vector, call it  $(X_1, \dots, X_m)$ , with independent components, where  $X_i$  is distributed according to the Poisson distribution with parameter  $\frac{1}{i}$ . In the theorem below, the product on the right hand side expresses the independence of the components  $\{X_i\}_{i=1}^m$ .

**Theorem 4.** *Let  $C_i^{(n)}$  denote the total number of cycles of length  $i$  in a uniformly random permutation from  $S_n$ . Then*

$$(8.2) \quad \lim_{n \rightarrow \infty} P_n(C_1^{(n)} = k_1, \dots, C_m^{(n)} = k_m) = \prod_{i=1}^m e^{-\frac{1}{i}} \frac{(\frac{1}{i})^{k_i}}{k_i!},$$

for  $k_i = 0, 1, \dots, 1 \leq i \leq m; m = 1, 2, \dots$ .

The proof we give of (8.1) readily generalizes to a proof Theorem 4, but the notation is more cumbersome. The combinatorial part of our proof is contained in the following proposition.

**Proposition 2.** *For  $n \geq mk$ ,*

$$E_n(C_m^{(n)})^k = T_k\left(\frac{1}{m}\right).$$

The proof of (8.1) follows from Proposition 2, (5.6) and the method of moments described in the final paragraph of section 4. To see that the condition required to employ the method of moments indeed holds in the case at hand, see, for example, the explanation in the first two paragraphs of the proof of Theorem C in [56].

We note that there are ways other than the method of moments to prove (8.1) and Theorem 4. See for example [5] which uses (9.1) in Chapter 9 and the method of inclusion-exclusion, or [53] for a completely different approach. A proof of (8.2) in [74] (and also in [55]) uses generating functions in a rather involved way. The first proof of (8.1) seems to be in [33] and the proof of (8.2) may go back to [43].

*Proof of Proposition 2.* Assume that  $n \geq mk$ . For  $D \subset [n]$  with  $|D| = m$ , let  $1_D(\sigma)$  be the indicator random variable equal to 1 or 0 according to whether or not  $\sigma \in S_n$  possesses an  $m$ -cycle consisting of the elements of  $D$ . Then we have

$$(8.3) \quad C_m^{(n)}(\sigma) = \sum_{\substack{D \subset [n] \\ |D|=m}} 1_D(\sigma),$$

and

$$(8.4) \quad E_n(C_m^{(n)})^k = \sum_{\{(D_1, \dots, D_k) \subset [n]^k : |D_j|=m, j \in [k]\}} E_n \prod_{j=1}^k 1_{D_j}.$$

Now  $E_n \prod_{j=1}^k 1_{D_j} \neq 0$  if and only if for some  $l \in [k]$ , there exist disjoint sets  $\{A_i\}_{i=1}^l$  such that  $\{D_j\}_{j=1}^k = \{A_i\}_{i=1}^l$ . If this is the case, then

$$(8.5) \quad E_n \prod_{j=1}^k 1_{D_j} = E_n \prod_{j=1}^l 1_{A_j} = \frac{(n-lm)!((m-1)!)^l}{n!}.$$

(Here we have used the assumption that  $n \geq mk$ .) The number of ways to construct a collection of  $l$  disjoint sets  $\{A_i\}_{i=1}^l$ , each of which consists of  $m$  elements from  $[n]$ , is  $\frac{n!}{(m!)^l (n-lm)! l!}$ . Given the  $\{A_i\}_{i=1}^l$ , the number of ways to choose the sets  $\{D_j\}_{j=1}^k$  so that  $\{D_j\}_{j=1}^k = \{A_i\}_{i=1}^l$  is equal to the Stirling number of the second kind  $S(k, l)$ , the number of ways to partition a set of size  $k$  into  $l$  nonempty parts, multiplied by  $l!$ , since the labeling must be taken into account. From these facts along with (8.4) and (8.5), we conclude that for  $n \geq mk$ ,

$$\begin{aligned} E_n(C_m^{(n)})^k &= \\ & \sum_{l=1}^k \left( \frac{(n-lm)!((m-1)!)^l}{n!} \right) \left( \frac{n!}{(m!)^l (n-lm)!} \right) S(k, l) = \\ & \sum_{l=1}^k \frac{1}{m^l} S(k, l) = T_k\left(\frac{1}{m}\right). \end{aligned}$$

□

## 9. LIMITING BEHAVIOR OF THE LENGTHS OF THE LARGEST AND SMALLEST CYCLES IN RANDOM PERMUTATIONS AND THE CONNECTION TO THE DICKMAN AND BUCHSTAB FUNCTIONS

Theorem 4 deals with the limiting distributions of the number of cycles of *fixed* lengths. We now consider the limiting behavior of the *lengths* of the largest cycles or of the smallest cycles. We begin by considering the largest cycles.

One way to analyze this begins with the formula of Cauchy which counts the number of permutations with given cycle numbers. Let  $\{c_j\}_{j=1}^n$  be nonnegative integers. Cauchy's formula [5] states that the number of permutations  $\sigma \in S_n$  that satisfy  $C_j^{(n)}(\sigma) = c_j$ , for all  $j \in [n]$ , is equal to

$n! \prod_{j=1}^n \left(\frac{1}{j}\right)^{c_j} \frac{1}{c_j!}$ , if  $\sum_{j=1}^n jc_j = n$ , and of course is equal to zero otherwise. Therefore, for a random permutation of  $S_n$ , we have

$$(9.1) \quad P_n(C_1^{(n)} = c_1, \dots, C_n^{(n)} = c_n) = \prod_{j=1}^n \left(\frac{1}{j}\right)^{c_j} \frac{1}{c_j!}, \text{ if } \sum_{j=1}^n jc_j = n.$$

The product on the right hand side above reminds one of Poisson distributions. Let  $\{Z_j\}_{j=1}^\infty$  be a sequence of independent random variables, with  $Z_j \sim \text{Poisson}\left(\frac{1}{j}\right)$ , and define  $T_{[n]} = \sum_{j=1}^n jZ_j$ , for  $n \in \mathbb{N}$ . We have

$$(9.2) \quad \begin{aligned} P(Z_1 = c_1, \dots, Z_n = c_n, T_{[n]} = n) &= \prod_{j=1}^n e^{-\frac{1}{j}} \left(\frac{1}{j}\right)^{c_j} \frac{1}{c_j!} = \\ e^{-\sum_{j=1}^n \frac{1}{j}} \prod_{j=1}^n \left(\frac{1}{j}\right)^{c_j} \frac{1}{c_j!}, &\text{ if } \sum_{j=1}^n jc_j = n, \end{aligned}$$

since the requirement  $T_{[n]} = n$  is automatically fulfilled if  $\sum_{j=1}^n jc_j = n$ . Summing (9.1) over all  $\{c_j\}_{j=1}^n$  which satisfy  $\sum_{j=1}^n jc_j = n$  gives

$$(9.3) \quad \sum_{\substack{(c_1, \dots, c_n): \\ \sum_{j=1}^n jc_j = n}} \prod_{j=1}^n \left(\frac{1}{j}\right)^{c_j} \frac{1}{c_j!} = 1.$$

Summing (9.2) over this same set of  $\{c_j\}_{j=1}^n$ , and using (9.3) gives

$$(9.4) \quad P(T_{[n]} = n) = e^{-\sum_{j=1}^n \frac{1}{j}}.$$

From (9.1), (9.2) and (9.4), we conclude that

$$(9.5) \quad P_n(C_1^{(n)} = c_1, \dots, C_n^{(n)} = c_n) = P(Z_1 = c_1, \dots, Z_n = c_n | T_{[n]} = n).$$

The representation in (9.5) of the distribution of the lengths of the cycles in terms of a conditioned distribution of independent random variables can be exploited. We demonstrate this by sketching the method in the case of the longest cycle [5]. Let  $L_j^{(n)}(\sigma)$  denote the length of the  $j$ th longest cycle in  $\sigma \in S_n$ ,  $j = 1, 2, \dots$ . We will consider only  $L_1^{(n)}$ , but one can work similarly with the random vector  $(L_1^{(n)}, \dots, L_j^{(n)})$ . The event  $\{L_1^{(n)} \leq m\}$

can be written as  $\{C_{m+1}^{(n)} = C_{m+2}^{(n)} = \dots = C_n^{(n)} = 0\}$ . Using (9.5) this gives

$$(9.6) \quad \begin{aligned} P_n(L_1^{(n)} \leq m) &= P(Z_{m+1} = \dots = Z_n = 0 | T_{[n]} = n) = \\ &= \frac{P(Z_{m+1} = \dots = Z_n = 0, T_{[n]} = n)}{P(T_{[n]} = n)} = \frac{P(Z_{m+1} = \dots = Z_n = 0, T_{[m]} = n)}{P(T_{[n]} = n)} = \\ &= \frac{P(Z_{m+1} = \dots = Z_n = 0)P(T_{[m]} = n)}{P(T_{[n]} = n)} = \frac{P(T_{[m]} = n)}{P(T_{[n]} = n)} \exp\left(-\sum_{j=m+1}^n \frac{1}{j}\right), \end{aligned}$$

where we have used the independence of  $T_{[m]}$  and  $\{Z_j\}_{j=m+1}^n$  in the next to the last equality.

It is known that the random variable  $\frac{1}{n}T_{[n]} = \frac{1}{n}\sum_{j=1}^n jZ_j$  converges in distribution as  $n \rightarrow \infty$  to the so-called *Dickman distribution* [5, 57]. This distribution, supported on  $[0, \infty)$ , has density  $e^{-\gamma}\rho(x)$ , where  $\gamma$  is Euler's constant and  $\rho$  is the *Dickman function*, defined as the unique continuous function satisfying the differential-delay equation

$$(9.7) \quad \begin{aligned} \rho(x) &= 1, \quad x \in (0, 1]; \\ x\rho'(x) + \rho(x-1) &= 0, \quad x > 1. \end{aligned}$$

(The proof that  $\int_0^\infty \rho(x)dx = e^\gamma$  is not immediate [51], [72].) For use below, we note that

$$(9.8) \quad \int_{x-1}^x \rho(y)dy = x\rho(x), \quad x \geq 1.$$

This identity follows by defining  $H(x) = \int_{x-1}^x \rho(y)dy$  and using (9.7) to obtain  $H'(x) = (x\rho(x))'$  and  $H(1) = 1$ . For more on the interesting topic of convergence in distribution to the Dickman distribution, see [57] and references therein. In (9.6), replace  $m$  by  $m_n$  and assume that  $m_n \sim nx$  as  $n \rightarrow \infty$ , where  $x \in (0, 1)$ . Then the quotient on the right hand side of (9.6) can be written as

$$(9.9) \quad \frac{P(T_{[m_n]} = n)}{P(T_{[n]} = n)} = \frac{P(\frac{1}{m_n}T_{[m_n]} = \frac{n}{m_n})}{P(\frac{1}{n}T_{[n]} = 1)}.$$

On the right hand side above, both  $\frac{1}{m_n}T_{[m_n]}$  and  $\frac{1}{n}T_{[n]}$  converge in distribution to the Dickman distribution, and  $\lim_{n \rightarrow \infty} \frac{n}{m_n} = \frac{1}{x}$ . Convergence in distribution to the Dickman distribution yields

$$(9.10) \quad \lim_{N \rightarrow \infty} P\left(\frac{1}{N}T_{[N]} \in [a, b]\right) = \int_a^b e^{-\gamma}\rho(x)dx, \quad 0 \leq a < b < \infty,$$

but it gives no information about the point probabilities on the right hand side of (9.9). However, applying a technique called *size-biasing*, which has particularly nice properties in the case of sums of independent Poisson random variables, one can express the point probabilities of  $T_{[N]}$  in terms of interval probabilities:

$$(9.11) \quad P(T_{[N]} = k) = \frac{1}{k} \sum_{l=1}^N P(T_{[N]} = k - l), \quad k = 1, 2, \dots .$$

Using (9.10) and (9.11) with  $N = n$  and with  $N = m_n$ , along with (9.9) and (9.8), we obtain

$$(9.12) \quad \lim_{n \rightarrow \infty} \frac{P(T_{[m_n]} = n)}{P(T_{[n]} = n)} = \int_{\frac{1}{x}-1}^{\frac{1}{x}} \rho(y) dy = \frac{1}{x} \rho\left(\frac{1}{x}\right), \quad \text{if } \lim_{n \rightarrow \infty} \frac{m_n}{n} = x \in (0, 1).$$

Substituting  $m_m \sim nx$  in (9.6), noting that  $\lim_{n \rightarrow \infty} \exp(-\sum_{j=m_n+1}^n \frac{1}{j}) = x$ , and using (9.12), we arrive at the following result.

**Theorem 5.** *Let  $L_1^{(n)}$  denote the length of the largest cycle in a uniformly random permutation from  $S_n$ . Then  $\frac{1}{n}L_1^{(n)}$  converges in distribution to the distribution whose distribution function is given by  $\rho(\frac{1}{x})$ ,  $x \in [0, 1]$ . That is,*

$$\lim_{n \rightarrow \infty} P_n\left(\frac{1}{n}L_1^{(n)} \leq x\right) = \rho\left(\frac{1}{x}\right), \quad x \in [0, 1].$$

As noted, we have followed [5] for the proof of Theorem 5; the original proof appears in [68].

The distribution arising above, whose distribution function is  $\rho(\frac{1}{x})$ , for  $x \in [0, 1]$ , is the first component of the *Poisson-Dirichlet distribution*, which can be defined as follows. Recall the GEM distribution, that is, the stick-breaking model  $\{X_n\}_{n=1}^\infty$  introduced at the end of section 7 and appearing in Theorem 3. Let  $\{\hat{X}_n\}_{n=1}^\infty$  denote the decreasing rearrangement of  $\{X_n\}_{n=1}^\infty$ . The Poisson-Dirichlet distribution can be defined as the distribution of  $\{\hat{X}_n\}_{n=1}^\infty$ . See [5] and references therein. In particular, the distribution of  $\hat{X}_1$  is the distribution arising in Theorem 5. It can be shown that for any  $j \in \mathbb{N}$ , the random vector  $(L_1^{(n)}, \dots, L_j^{(n)})$  converges in distribution to the distribution of  $(\hat{X}_1, \dots, \hat{X}_j)$  [5].

In number theory, the Poisson-Dirichlet distribution comes up in a parallel fashion in relation to *smooth numbers*, which are integers with no large prime divisors. Let  $p_1^+(k)$  denote the largest prime factor of  $k \in \mathbb{N}$ . It was proved

by Dickman [18] that

$$\lim_{n \rightarrow \infty} \frac{1}{n} |\{k \in [n] : p_1^+(k) \leq n^x\}| = \rho\left(\frac{1}{x}\right), \quad x \in [0, 1].$$

See [51, 72]. Similar to the extension from  $L_1^{(n)}$  to the vector  $(L_1^{(n)}, \dots, L_j^{(n)})$ , the above result can be extended to the vector  $(p_1^+(\cdot), \dots, p_j^+(\cdot))$ , where  $p_i^+(k)$  denotes the  $i$ th largest prime factor (counted with multiplicities) of  $k$ , and  $p_i^+(k) = 0$  if  $k$  has fewer than  $i$  prime factors [11].

We now turn to the smallest cycles. Let  $S_j^{(n)}(\sigma)$  denote the length of the  $j$ th shortest cycle of  $\sigma \in S_n$ . We have

$$P_n(S_j^{(n)} > m) = P_n\left(\sum_{i=1}^m C_i^{(n)} < j\right).$$

Thus, by (8.2),

$$(9.13) \quad \lim_{n \rightarrow \infty} P_n(S_j^{(n)} > m) = P\left(\sum_{i=1}^m Z_i < j\right),$$

where  $\{Z_i\}_{i=1}^{\infty}$  are independent random variables and  $Z_i$  has the Poisson( $\frac{1}{i}$ ) distribution. A straight forward calculation shows that if  $X$  and  $Y$  are independent and distributed according to Poisson( $\lambda_1$ ) and Poisson( $\lambda_2$ ) respectively, then

$$P(X + Y = m) = \sum_{i=0}^m P(X = i)P(Y = m - i) = \frac{(\lambda_1 + \lambda_2)^m}{m!} e^{-(\lambda_1 + \lambda_2)};$$

that is  $X + Y$  has the Poisson( $\lambda_1 + \lambda_2$ ) distribution. Thus, from (9.13) we have

$$(9.14) \quad \lim_{n \rightarrow \infty} P_n(S_j^{(n)} > m) = \sum_{i=0}^{j-1} e^{-h(m)} \frac{(h(m))^i}{i!}, \quad \text{where } h(m) = \sum_{i=1}^m \frac{1}{i}.$$

A theory similar to that discussed above for the longest cycles, that also uses (9.5), can be developed to obtain a result concerning the so-called *large deviations* of  $S_j^{(n)}$  [5]. This result involves the *Buchstab function*  $\omega(x)$ , which is defined for  $x \geq 1$  as the unique continuous function satisfying

$$\omega(x) = \frac{1}{x}, \quad 1 \leq x \leq 2,$$

and satisfying the differential-delay equation

$$(x\omega(x))' = \omega(x - 1), \quad x > 2.$$

In particular, for the shortest cycle, the result is as follows [5].

**Theorem 6.**

$$P_n(S_1^{(n)} > \frac{n}{x}) \sim \frac{x\omega(x)}{n}, \quad x > 1, \text{ as } n \rightarrow \infty.$$

In number theory, the Buchstab function comes up in a parallel fashion in relation to *rough* numbers, which are integers with no small prime divisors. Let  $p_1^-(k)$  denote the smallest prime factor of  $k \in \mathbb{N}$ . It was proved by Buchstab [15] that

$$(9.15) \quad \frac{1}{n} |\{k \in [n] : p_1^-(k) \geq n^{\frac{1}{x}}\}| \sim x\omega(x) \frac{1}{\log n}, \quad x > 1, \text{ as } n \rightarrow \infty.$$

See [51, 72]. (Note that (9.15) for  $x = 1$  is the Prime Number Theorem. Buchstab assumed the Prime Number Theorem when proving his result.)

10. THE STATISTICAL BEHAVIOR OF RANDOM PARTITIONS OF LARGE INTEGERS

A partition  $\lambda$  of a positive integer  $n$  is a  $j$ -tuple  $(a_1, \dots, a_j)$  where  $j \in [n]$ ,  $\{a_i\}_{i=1}^j$  are positive integers satisfying  $a_1 \geq \dots \geq a_j$  and  $\sum_{i=1}^j a_i = n$ . Let  $\Lambda_n$  denote the set of partitions of  $n$ . There is a natural map, call it  $\mathcal{M}_n$ , from the set of permutations  $S_n$  of  $[n]$  onto the set of partitions of  $[n]$ , via the cycle decomposition of the permutation. We could write out a formal definition, but a simple example is probably clearer and should suffice. Let  $\sigma \in S_9$  be given in terms of its cycle decomposition by  $\sigma = (154)(2)(386)(79)$ . Then  $\mathcal{M}_9(\sigma)$  is the partition  $(3,3,2,1)$  of the integer 9. The map  $\mathcal{M}_n$  along with the uniform probability measure  $P_n$  on  $S_n$  induces a (non-uniform) probability measure  $\mathcal{P}_n$  on  $\Lambda_n$ ; namely,

$$\mathcal{P}_n(\lambda) = P_n(\{\sigma : \mathcal{M}_n(\sigma) = \lambda\}), \text{ for } \lambda \in \Lambda_n.$$

For  $k \in \mathbb{N}$ , let  $X_k(\lambda)$  denote the number of parts of the partition  $\lambda \in \Lambda_n$  that are equal to  $k$ . For  $s \in \mathbb{N}$ , let  $Y_s(\lambda)$  denote the  $s$ -th largest part of  $\lambda$ , with  $Y_s(\lambda) = 0$  if  $\lambda$  has fewer than  $s$  parts. (For convenience later on, we consider  $X_k$  and  $Y_s$  to be defined simultaneously on  $\Lambda_n$ , for all  $n$ .) For example, consider the partition of 13,  $\lambda = (4, 4, 2, 1, 1, 1)$ . Then  $X_1(\lambda) = 3$ ,  $X_2(\lambda) = 1$ ,  $X_4(\lambda) = 2$  and  $X_k(\lambda) = 0$  for  $k \notin \{1, 2, 4\}$ ;  $Y_1(\lambda) = Y_2(\lambda) = 4$ ,  $Y_3(\lambda) = 2$ ,  $Y_4(\lambda) = Y_5(\lambda) = Y_6(\lambda) = 1$  and  $Y_k(\lambda) = 0$ , for  $k \geq 7$ .

Recalling from section 8 that  $C_k^{(n)}(\sigma)$  denotes the number of cycles of length  $k$  in the permutation  $\sigma \in S_n$ , it follows that

$$X_k(\lambda) = C_k^{(n)}(\sigma), \text{ where } \lambda = \mathcal{M}_n(\sigma).$$

Recalling from section 9 that  $L_j^{(n)}(\sigma)$  denotes the length of the  $j$ th longest cycle in  $\sigma \in S_n$ , it follows that

$$Y_s(\lambda) = L_s^{(n)}(\sigma), \text{ where } \lambda = \mathcal{M}_n(\sigma).$$

In light of (8.1), it follows that under the probability measures  $\mathcal{P}_n$ , the distribution of  $X_k$  converges to the Poisson( $\frac{1}{k}$ ) distribution:

$$(10.1) \quad \lim_{n \rightarrow \infty} \mathcal{P}_n(X_k = j) = e^{-\frac{1}{k}} \frac{(\frac{1}{k})^j}{j!}, \quad j = 0, 1, \dots$$

And in light of Theorem 5,

$$(10.2) \quad \lim_{n \rightarrow \infty} \mathcal{P}_n\left(\frac{Y_1}{n} \leq x\right) = \rho\left(\frac{1}{x}\right), \quad x \in [0, 1],$$

where we recall that  $\rho$  is the Dickman function. We recall from the discussion after Theorem 5 that the distribution on  $[0, 1]$  whose distribution function is given by  $\rho(\frac{1}{x})$  is the distribution of the first component  $\hat{X}_1$  of the random vector  $\{\hat{X}_n\}_{n=1}^\infty$  whose distribution is called the Poisson-Dirichlet distribution. From that discussion, it then follows more generally that

$$(10.3) \quad \lim_{n \rightarrow \infty} \mathcal{P}_n\left(\frac{Y_k}{n} \leq x\right) = P(\hat{X}_k \leq x), \quad x \in [0, 1], k \geq 1.$$

In this section, we consider the statistical behavior of  $\Lambda_n$  under the *uniform* measure that gives probability  $\frac{1}{|\Lambda_n|}$  to each partition  $\lambda \in \Lambda_n$ . The results are very different from the results in (10.1)-(10.3) in the case of the measures  $\{\mathcal{P}_n\}_{n=1}^\infty$ . The paper [30] makes beautiful use of generating functions to study the asymptotic behavior of partition statistics in the case of the uniform distribution. We state here three of the many results in that paper. Denote the uniform measure on  $\Lambda_n$  by  $P_n^{\text{part}}$ . Recall that  $X_k$  is the number of parts of size  $k$  in the random partition.

**Theorem 7.** *If  $\lim_{n \rightarrow \infty} \frac{k_n}{n^{\frac{1}{2}}} = 0$ , then*

$$\lim_{n \rightarrow \infty} P_n^{\text{part}}\left(\frac{\pi}{\sqrt{6n}} k_n X_{k_n} \leq x\right) = 1 - e^{-x}, \quad x \geq 0.$$

The above theorem states that for any fixed  $k$  ( $k_n = k$ ) as well as for  $k_n$  growing at a sufficiently slow rate, the rescaled quantity  $\frac{\pi}{\sqrt{6n}} k_n X_{k_n}$  converges in distribution as  $n \rightarrow \infty$  to an exponential random variable with parameter 1. Compare this with (10.1).

The next theorem concerns  $Y_s$ , the  $s$ -th largest part of the random partition. In the case  $s = 1$ , it was originally proved in [23].

**Theorem 8.**

$$\lim_{n \rightarrow \infty} P_n^{part} \left( \frac{\pi}{\sqrt{6n}} Y_s - \log \frac{\sqrt{6n}}{\pi} \leq x \right) = \int_{-\infty}^x \frac{\exp(-e^{-y} - sy)}{(t-1)!} dy, \quad x \in \mathbb{R}, s \in \mathbb{N}.$$

Compare Theorem 8 to (10.2) and (10.3).

If instead of looking at  $Y_s$ , one considers  $Y_{s_n}$  with  $s_n \rightarrow \infty$  at a sufficiently slow rate, then one gets convergence in distribution to a Gaussian limiting distribution.

**Theorem 9.** *If  $\lim_{n \rightarrow \infty} s_n = \infty$  and  $\lim_{n \rightarrow \infty} \frac{s_n}{n^{\frac{1}{4}}} = 0$ , then*

$$\lim_{n \rightarrow \infty} P_n^{part} \left( \pi \sqrt{\frac{s_n}{6n}} Y_{s_n} - \sqrt{s_n} \log \frac{\sqrt{6n}}{\pi s_n} \leq x \right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy, \quad x \in \mathbb{R}.$$

We now sketch the method used to prove Theorem 7. The proofs of Theorems 8 and 9 use the same method along with some additional ingredients. Let  $\Lambda$  denote the set of all partitions of all nonnegative integers, including the empty partition of 0. Let  $p(n) = |\Lambda_n|$  denote the number of partitions of  $n$ . As is well-known and easy to check, the generating function  $\sum_{n=0}^{\infty} p(n)q^n$  for  $\{p(n)\}_{n=0}^{\infty}$  is given by the formula

$$(10.4) \quad \sum_{n=0}^{\infty} p(n)q^n = \prod_{j=1}^{\infty} (1 - q^j)^{-1}.$$

Define  $|\lambda| = n$ , if  $\lambda \in \Lambda_n$ . For  $q \in (0, 1)$ , consider the following probability measure on the set  $\Lambda$  of all partitions of nonnegative integers:

$$(10.5) \quad Q_q(\lambda) = q^{|\lambda|} \prod_{j=1}^{\infty} (1 - q^j), \quad \lambda \in \Lambda.$$

To see that  $Q_q$  is a probability measure, note that

$$\sum_{\lambda \in \Lambda_n} Q_q(\lambda) = p(n)q^n \prod_{j=1}^{\infty} (1 - q^j),$$

and thus using (10.4),

$$\sum_{\lambda \in \Lambda} Q_q(\lambda) = \sum_{n=0}^{\infty} \sum_{\lambda \in \Lambda_n} Q_q(\lambda) = \prod_{j=1}^{\infty} (1 - q^j) \sum_{n=0}^{\infty} q^n p(n) = 1.$$

The key observation is that under  $Q_q$ , the random variables  $\{X_j\}_{j=1}^{\infty}$  are independent and the distribution of  $X_j$  is geometric with parameter  $q^j$ .

To verify the above observation, note that the values  $\{X_j(\lambda)\}_{j=1}^{\infty}$  uniquely determine  $\lambda$ , and note that  $|\lambda| = \sum_{j=1}^{\infty} jX_j(\lambda)$ . Thus, from (10.5),

$$Q_q(X_j = x_j, j = 1, 2, \dots) = q^{\sum_{j=1}^{\infty} jx_j} \prod_{j=1}^{\infty} (1-q^j) = \prod_{j=1}^{\infty} (q^j)^{x_j} (1-q^j), n = 0, 1, \dots.$$

The product on the right hand side above indicates the independence of  $\{X_j\}_{j=1}^{\infty}$  under  $Q_q$ .

Let  $N = |\lambda|$ . Of course,  $P_n^{\text{part}}(N = n) = 1$ . However, under  $Q_q$ , the number  $N$  being partitioned is random. From (10.5), we have

$$Q_q(N = n) = p(n)q^n \prod_{j=1}^{\infty} (1 - q^j), \quad n = 0, 1, \dots.$$

Note from the definition of the uniform measure  $P_n^{\text{part}}$  on  $\Lambda_n$  and from the definition of the measure  $Q_q$  that if  $\lambda_1$  and  $\lambda_2$  are partitions for which  $N(\lambda_1) = N(\lambda_2) = n$ , then  $P_n^{\text{part}}(\lambda_1) = P_n^{\text{part}}(\lambda_2)$  and  $Q_q(\lambda_1) = Q_q(\lambda_2)$ . From these fact it follows that

$$(10.6) \quad P_n^{\text{part}}(A) = Q_q(A|N = n), \quad A \subset \Lambda_n.$$

Let  $E^{Q_q}$  denote expectation with respect to  $Q_q$ . Consider the probability generating function of  $N$  under  $Q_q$ , defined by  $\Phi(r) = E^{Q_q} r^N$ . We have

$$(10.7) \quad \begin{aligned} \Phi(r) &= E^{Q_q} r^N = \sum_{n=0}^{\infty} Q_q(N = n) r^n = \\ &= \prod_{j=1}^{\infty} (1 - q^j) \sum_{n=0}^{\infty} p(n) q^n r^n = \prod_{j=1}^{\infty} \frac{1 - q^j}{1 - (qr)^j}, \end{aligned}$$

where we have used (10.4) in the final equality. Using the fact that  $\Phi'(1) = E^{Q_q} N$  and  $\Phi''(1) = E^{Q_q} N(N-1)$ , one can calculate the expected value and variance of  $N$ :

$$\begin{aligned} E^{Q_q} N &= \sum_{j=1}^{\infty} \frac{j q^j}{1 - q^j}; \\ \text{Var}_{Q_q}(N) &= \sum_{j=1}^{\infty} \frac{j^2 q^j}{(1 - q^j)^2}. \end{aligned}$$

An asymptotic analysis shows that if one chooses

$$(10.8) \quad q = q_n = e^{-\frac{\pi}{\sqrt{6n}}},$$

then  $\text{Var}_{Q_{q_n}}(N) \sim \frac{\sqrt{24}}{\pi} n^{\frac{3}{2}}$  and  $|n - E^{Q_{q_n}} N| = o(n^{\frac{3}{4}})$ . This shows that the random variable  $N$  under the measure  $Q_{q_n}$  concentrates around  $n$ . Indeed,

$$|N - n| \leq |n - E^{Q_{q_n}} N| + |N - E^{Q_{q_n}} N| = o(n^{\frac{3}{4}}) + |N - E^{Q_{q_n}} N|,$$

and by Chebyshev's inequality,

$$Q_{q_n}(|N - E^{Q_{q_n}} N| \geq n^{\frac{3}{4}} l_n) \leq \frac{O(n^{\frac{3}{2}})}{n^{\frac{3}{2}} l_n^2} \rightarrow 0, \text{ if } l_n \rightarrow \infty.$$

Thus, with probability approaching 1 as  $n \rightarrow \infty$ ,  $N$  will be within  $l_n n^{\frac{3}{4}}$  of  $n$  if  $l_n \rightarrow \infty$ . Using Fourier analysis, it was shown in [30] that  $Q_{q_n}(N = n) \sim (96n^3)^{-\frac{1}{4}}$ . From these results, the author was able to show that the Prohorov distance (a certain metric) between the distribution of  $X_{k_n}$  under the probability measure  $Q_{q_n}$  and its distribution under the measure  $Q_{q_n}(\cdot | N = n)$  converges to 0 as  $n \rightarrow \infty$ , if  $k_n = o(n^{\frac{1}{2}})$ . It then follows from (10.6) that the Prohorov distance between the distributions of  $X_{k_n}$  under  $Q_{q_n}$  and under  $P_n^{\text{part}}$  also converges to 0 as  $n \rightarrow \infty$ . From this, it follows that the  $P_n^{\text{part}}$ -probability appearing on the left hand side of Theorem 7 can be replaced by the  $Q_{q_n}$  probability. However, as noted, under  $Q_{q_n}$ , the random variables  $\{X_j\}_{j=1}^{\infty}$  are independent with geometric distributions; thus, everything can be calculated explicitly. Indeed, we have

$$Q_{q_n}\left(\frac{\pi}{\sqrt{6n}} k_n X_{k_n} \leq x\right) = Q_{q_n}(X_{k_n} \leq \frac{\sqrt{6nx}}{\pi k_n}) = \sum_{j=0}^{\lfloor \frac{\sqrt{6nx}}{\pi k_n} \rfloor} (1 - q_n^{k_n})(q_n^{k_n})^j =$$

$$1 - q_n^{k_n \lfloor \frac{\sqrt{6nx}}{\pi k_n} \rfloor} \xrightarrow{n \rightarrow \infty} 1 - e^{-x},$$

where in the final step we use (10.8).

## 11. THRESHOLD CALCULATIONS FOR ERDŐS-RÉNYI RANDOM GRAPHS USING THE FIRST AND SECOND MOMENT METHODS

We recall the definition of the Erdős-Rényi random graph  $G(n, p)$ , for  $n \in \mathbb{N}$  and  $p \in (0, 1)$ . This graph has  $n$  vertices and thus  $\binom{n}{2}$  possible edges. Independently, each of these edges is included in the graph with probability  $p$ , and not included with probability  $1 - p$ . We will analyze three so-called threshold properties of the random graphs  $G(n, p_n)$  as  $n \rightarrow \infty$ , in two cases with variable  $p_n$ , and in one case with fixed  $p_n = p$ . In what follows,  $P_n$  and  $E_n$  will be used for probabilities and expectations concerning  $G(n, p)$  or  $G(n, p_n)$ .

The method of proof we use in these threshold calculations is based on the first and second moment methods that were stated and proved in section 4—see (4.2) and (4.3).

**Threshold for disconnected vertices.** The first result we present concerns disconnected vertices in  $G(n, p_n)$ .

**Theorem 10.** *Let  $D_n$  be the random variable denoting the number of disconnected vertices in  $G(n, p_n)$ .*

- i. If  $p_n = \frac{\log n + c_n}{n}$ , and  $\lim_{n \rightarrow \infty} c_n = \infty$ , then  $D_n \xrightarrow{\text{dist}} 0$ ; equivalently,  $\lim_{n \rightarrow \infty} P_n(D_n = 0) = 1$ ;*
- ii. If  $p_n = \frac{\log n + c_n}{n}$ , and  $\lim_{n \rightarrow \infty} c_n = -\infty$ , then  $\frac{D_n}{E_n D_n} \xrightarrow{\text{dist}} 1$ . Also,  $\lim_{n \rightarrow \infty} P_n(D_n > M) = 1$ , for any  $M \in \mathbb{N}$ .*

*Proof.* For part (i), we use the first moment method. For  $1 \leq j \leq n$ , let  $D_{n,j}$  denote the indicator random variable for vertex  $j$  to be disconnected; that is,  $D_{n,j}$  is equal to 1 if vertex  $j$  is disconnected in  $G(n, p_n)$ , and equal to 0 otherwise. Then we can represent the random variable  $D_n$  as  $D_n = \sum_{j=1}^n D_{n,j}$ . For distinct vertices  $j, k \in [n]$ , let  $I_{n,j,k}$  denote the indicator random variable that is equal to 1 if  $G(n, p_n)$  contains an edge connecting  $j$  and  $k$ , and is equal to 0 otherwise. Then  $D_{n,j} = \prod_{k \in [n]; k \neq j} (1 - I_{n,j,k})$ . By the definition of the Erdős-Rényi graph, the random variables  $\{I_{n,j,k}\}_{k \in [n], k \neq j}$  are independent and  $E_n I_{n,j,k} = P_n(I_{n,j,k} = 1) = p_n$ . Thus,

$$(11.1) \quad E_n D_{n,j} = \prod_{k \in [n]; k \neq j} E(1 - I_{n,j,k}) = (1 - p_n)^{n-1}$$

and

$$(11.2) \quad E_n D_n = n(1 - p_n)^{n-1}.$$

Substituting  $p_n$  as in part (i) of the theorem into (11.2), it follows that  $\lim_{n \rightarrow \infty} E_n D_n = 0$ , and thus by the first moment method,  $D_n$  converges in distribution to 0; that is  $\lim_{n \rightarrow \infty} P_n(D_n \geq \epsilon) = 0$ , for all  $\epsilon > 0$ . Since  $D_n$  is integer-valued, this gives  $\lim_{n \rightarrow \infty} P(D_n = 0) = 1$ .

We use the second moment method for part (ii). We write

$$(11.3) \quad E_n D_n^2 = E \sum_{j=1}^n D_{n,j} \sum_{k=1}^n D_{n,k} = \sum_{j=1}^n E_n D_{n,j} + 2 \sum_{1 \leq j < k \leq n} E_n D_{n,j} D_{n,k}.$$

We have

$$(11.4) \quad \begin{aligned} E_n D_{n,j} D_{n,k} &= P_n(D_{n,j} = D_{n,k} = 1) = \\ P_n(I_{n,j,l} = I_{n,k,m} = 0, l \in [n] - \{j\}, m \in [n] - \{j, k\}) &= (1 - p_n)^{2n-3}, \quad j \neq k. \end{aligned}$$

From (11.1)-(11.4), we obtain

$$(11.5) \quad \sigma^2(D_n) = E_n D_n^2 - (E_n D_n)^2 = n(n-1)(1-p_n)^{2n-3} + n(1-p_n)^{n-1} - (n(1-p_n)^{n-1})^2.$$

Let  $p_n$  be as in part (ii) of the theorem. Then standard estimates give

$$n(n-1)(1-p_n)^{2n-3} \sim (n(1-p_n)^{n-1})^2 \sim e^{-2c_n} \xrightarrow{n \rightarrow \infty} \infty.$$

Using this with (11.5) and (11.2), we conclude that  $\sigma^2(D_n) = o((E_n D_n)^2)$  and  $\lim_{n \rightarrow \infty} E_n D_n = \infty$ . In particular then, by the second moment method,  $\frac{D_n}{E_n D_n} \xrightarrow{\text{dist}} 1$ . Since  $E_n D_n \xrightarrow{n \rightarrow \infty} \infty$ , it follows from the convergence in distribution that  $\lim_{n \rightarrow \infty} P_n(D_n > M) = 1$ , for any  $M \in \mathbb{N}$ .  $\square$

**Threshold for connectivity.** We now investigate the threshold for connectivity in  $G(n, p_n)$ . If  $G(n, p_n)$  has at least one disconnected vertex, then of course  $G(n, p_n)$  is disconnected. Thus, the threshold of  $p_n$  for connectivity is greater than or equal to that for no disconnected vertices. It turns out that the threshold for connectivity is asymptotically the same as for no disconnected vertices. The following result was first proved in [24].

**Theorem 11.** *i. If  $p_n = \frac{\log n + c_n}{n}$ , and  $\lim_{n \rightarrow \infty} c_n = \infty$ , then  $\lim_{n \rightarrow \infty} P_n(G(n, p_n) \text{ is connected}) = 1$ ;*  
*ii. If  $p_n = \frac{\log n + c_n}{n}$ , and  $\lim_{n \rightarrow \infty} c_n = -\infty$ , then  $\lim_{n \rightarrow \infty} P_n(G(n, p_n) \text{ is connected}) = 0$ .*

The proof of part (ii) of course follows from part (ii) of Theorem 10, so we only need to prove part (i). The proof is not via the simple first moment method presented above, but rather via a slightly more involved first moment technique [12].

*Proof of part (i).* We may and will assume that  $n \geq 3$  and  $1 \leq c_n \leq \log n$ . For each  $A \subset [n]$ , let  $I_{A,n}$  be the indicator random variable that is equal to one if in  $G(n, p_n)$  every vertex in  $A$  is disconnected from the set of vertices  $[n] - A$ , and that is equal to zero otherwise. Then

$$(11.6) \quad P_n(I_{A,n} = 1) = (1 - p_n)^{k(n-k)}, \text{ if } |A| = k.$$

Clearly,

$$(11.7) \quad P_n(G(n, p_n) \text{ is disconnected}) = P_n(\cup_{A \subset [n], |A| \leq \lfloor \frac{n}{2} \rfloor} \{I_{A,n} = 1\}).$$

Since the probability of the union of events is less than or equal to the sums of the probabilities of the individual events composing the union, we have from (11.6) and (11.7) that

$$(11.8) \quad P_n(G(n, p_n) \text{ is disconnected}) \leq \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{k} (1 - p_n)^{k(n-k)}.$$

We break the sum on the right hand side of (11.8) into two parts. Substituting for  $p_n$  from part (i) of the theorem, using the estimate  $\binom{n}{k} \leq (\frac{en}{k})^k$  from Stirling's formula, using the inequality  $1 - x \leq e^{-x}$ , for  $x \geq 0$ , and recalling our additional assumption  $1 \leq c_n \leq \log n$  from the first line of the proof, we have

$$\begin{aligned} \sum_{k=1}^{\lfloor n^{\frac{3}{4}} \rfloor} \binom{n}{k} (1 - p_n)^{k(n-k)} &\leq \sum_{k=1}^{\lfloor n^{\frac{3}{4}} \rfloor} \left(\frac{en}{k}\right)^k e^{-k(\log n + c_n)} e^{2k^2 \frac{\log n}{n}} = \\ &e^{(1-c_n)} e^{\frac{2 \log n}{n}} + \sum_{k=2}^{\lfloor n^{\frac{3}{4}} \rfloor} \left(\frac{e^{1-c_n} e^{2k \frac{\log n}{n}}}{k}\right)^k \leq \\ &e^{(1-c_n)} e^{\frac{2 \log n}{n}} + e^{2(1-c_n)} \sum_{k=2}^{\lfloor n^{\frac{3}{4}} \rfloor} \left(\frac{e^{2n^{-\frac{1}{4}} \log n}}{2}\right)^k. \end{aligned}$$

Consequently, for sufficiently large  $n$ ,

$$(11.9) \quad \sum_{k=1}^{\lfloor n^{\frac{3}{4}} \rfloor} \binom{n}{k} (1 - p_n)^{k(n-k)} \leq e^2 e^{-c_n} + e^{2(1-c_n)}.$$

For the second part of the sum, using two of the inequalities used for the first part, using the fact that  $k(n-k) \geq \frac{kn}{2}$  (which of course holds over the entire original range of  $k$ ), and substituting for  $p_n$ , we have

$$(11.10) \quad \begin{aligned} \sum_{k=\lfloor n^{\frac{3}{4}} \rfloor + 1}^n \binom{n}{k} (1 - p_n)^{k(n-k)} &\leq \sum_{k=\lfloor n^{\frac{3}{4}} \rfloor + 1}^n \left(\frac{en}{k}\right)^k e^{-\frac{k \log n}{2}} \leq \\ &\sum_{k=\lfloor n^{\frac{3}{4}} \rfloor + 1}^n (en^{\frac{1}{4}})^k n^{-\frac{k}{2}} = \sum_{k=\lfloor n^{\frac{3}{4}} \rfloor + 1}^n (en^{-\frac{1}{4}})^k. \end{aligned}$$

(Note that for this part of the sum we have ignored the term  $c_n$  in  $p_n$  as it is not needed.) From (11.8)-(11.10), we conclude that

$$\lim_{n \rightarrow \infty} P_n(G(n, p_n) \text{ is disconnected}) = 0,$$

if  $p_n$  is as in part (i) of the theorem.  $\square$

**Threshold for clique size.** We now consider a problem regarding  $G(n, p)$  with  $p \in (0, 1)$  fixed. Recall that a *clique* of size  $k$  in  $G(n, p)$  is a complete subgraph on  $k$  vertices, and that the *clique number* of the graph is the largest  $k$  for which a clique of size  $k$  exists. Let  $\log_{\frac{1}{p}}$  denote the logarithm in base  $\frac{1}{p}$  and let  $\log_{\frac{1}{p}}^{(2)} = \log_{\frac{1}{p}} \log_{\frac{1}{p}}$ . The following result reveals the close to deterministic nature of the clique number in  $G(n, p)$ . See [12, 55].

**Theorem 12.** *Let  $L_n$  be the random variable denoting the clique number of  $G(n, p)$ . Then*

$$\lim_{n \rightarrow \infty} P_n(L_n \geq 2 \log_{\frac{1}{p}} n - c \log_{\frac{1}{p}}^{(2)} n) = \begin{cases} 0, & \text{if } c < 2; \\ 1, & \text{if } c > 2. \end{cases}$$

Theorem 12 is an immediate corollary of the following theorem [55]. For each  $k \in [n]$ , define the random variable  $N_{n,k}$  to be the *number* of cliques in  $G(n, p)$  of size  $k$ .

**Theorem 13.** *i. If  $k_n \geq 2 \log_{\frac{1}{p}} n - c \log_{\frac{1}{p}}^{(2)} n$ , for some  $c < 2$ , then*

$$(11.11) \quad \lim_{n \rightarrow \infty} E_n N_{n, k_n} = 0.$$

*Thus,  $N_{n, k_n} \xrightarrow{\text{dist}} 0$ ; equivalently,  $\lim_{n \rightarrow \infty} P_n(N_{n, k_n} = 0) = 1$ ;*

*ii. If  $k_n \leq 2 \log_{\frac{1}{p}} n - c \log_{\frac{1}{p}}^{(2)} n$ , for some  $c > 2$ , then*

$$(11.12) \quad \lim_{n \rightarrow \infty} E_n N_{n, k_n} = \infty.$$

*Also,*

$$(11.13) \quad \frac{N_{n, k_n}}{E N_{n, k_n}} \xrightarrow{\text{dist}} 1,$$

*and*

$$(11.14) \quad \lim_{n \rightarrow \infty} P_n(N_{n, k_n} > M) = 1, \text{ for any } M \in \mathbb{N}.$$

Note that the claims in the sentence following (11.11) are a consequence of the first moment method along with the fact that  $N_{n,k_n}$  is integer-valued. Note also that (11.14) follows immediately from (11.12) and (11.13). We will give the proofs of (11.11) and (11.12); they are of course first moment calculations. The proof of (11.13) can be given by the second moment method, but the calculations are much more involved than they were in the case of Theorem 10. We will begin the second moment calculation so as to reveal the considerations that arise, and then refer the interested reader elsewhere for a complete proof.

*Proofs of (11.11) and (11.12).* We will prove the two results simultaneously. For (11.11), fix  $c < 2$  and for each  $n$  sufficiently large, let  $c_n \leq c$  be such that  $k_n := 2 \log_{\frac{1}{p}} n - c_n \log_{\frac{1}{p}}^{(2)} n$  is an integer. For (11.12), fix  $c > 2$  and for each  $n$  sufficiently large, let  $c_n \geq c$  be such that  $k_n := 2 \log_{\frac{1}{p}} n - c_n \log_{\frac{1}{p}}^{(2)} n$  is an integer. We need to show that (11.11) holds for the first choice of  $k_n$  and that (11.12) holds for the second choice of  $k_n$ .

There are  $\binom{n}{k_n}$  different subsets of  $[n]$  of size  $k_n$ . For each such subset  $K$ , let  $I_K$  be the indicator random variable that is equal to 1 if the vertices in  $K$  form a clique in  $G(n, p)$ , and equal to 0 otherwise. Then

$$N_{n,k_n} = \sum_{K \subset [n], |K|=k_n} 1_K.$$

From the definition of the Erdős-Rényi graph, we have

$$(11.15) \quad E_n 1_K = p^{\binom{k_n}{2}}, \quad |K| = k_n.$$

Thus,

$$E_n N_{n,k_n} = \binom{n}{k_n} p^{\binom{k_n}{2}}.$$

It is not hard to show that as long as  $k_n = o(n^{\frac{1}{2}})$ , then

$$(11.16) \quad \binom{n}{k_n} \sim \frac{n^{k_n}}{k_n!}.$$

From (11.15) and (11.16) along with Stirling's formula, we obtain

$$(11.17) \quad E_n N_{n,k_n} \sim \frac{n^{k_n} p^{\frac{k_n(k_n-1)}{2}}}{k_n^{k_n} e^{-k_n} \sqrt{2\pi k_n}}.$$

Taking logarithms in base  $\frac{1}{p}$  on both sides of (11.17), we have

$$(11.18) \quad \log_{\frac{1}{p}} E_n N_{n,k_n} \sim k_n \log_{\frac{1}{p}} n - \frac{1}{2} k_n^2 + \frac{1}{2} k_n - k_n \log_{\frac{1}{p}} k_n + k_n \log_{\frac{1}{p}} e - \frac{1}{2} \log_{\frac{1}{p}} 2\pi k_n.$$

We have

$$(11.19) \quad \log_{\frac{1}{p}} k_n = \log_{\frac{1}{p}} \left( 2 \log_{\frac{1}{p}} n - c_n \log_{\frac{1}{p}}^{(2)} n \right) = \log_{\frac{1}{p}} \left( (\log_{\frac{1}{p}} n) \left( 2 - \frac{c_n \log_{\frac{1}{p}}^{(2)} n}{\log_{\frac{1}{p}} n} \right) \right) =$$

$$\log_{\frac{1}{p}}^{(2)} n + \log_{\frac{1}{p}} \left( 2 - \frac{c_n \log_{\frac{1}{p}}^{(2)} n}{\log_{\frac{1}{p}} n} \right) = \log_{\frac{1}{p}}^{(2)} n + O(1).$$

The sum of the three terms of highest orders on the right hand side of (11.18) is  $k_n \log_{\frac{1}{p}} n - \frac{1}{2} k_n^2 - k_n \log_{\frac{1}{p}} k_n$ . Using (11.19) to substitute for  $\log_{\frac{1}{p}} k_n$ , and using the definition of  $k_n$  to substitute for  $k_n$  where it appears elsewhere in this three term sum, we have

$$(11.20) \quad k_n \log_{\frac{1}{p}} n - \frac{1}{2} k_n^2 - k_n \log_{\frac{1}{p}} k_n = (2 \log_{\frac{1}{p}} n - c_n \log_{\frac{1}{p}}^{(2)} n) \log_{\frac{1}{p}} n -$$

$$\frac{1}{2} (2 \log_{\frac{1}{p}} n - c_n \log_{\frac{1}{p}}^{(2)} n)^2 - (2 \log_{\frac{1}{p}} n - c_n \log_{\frac{1}{p}}^{(2)} n) (\log_{\frac{1}{p}}^{(2)} n + O(1)) =$$

$$(c_n - 2) (\log_{\frac{1}{p}} n) \log_{\frac{1}{p}}^{(2)} n + O(\log_{\frac{1}{p}} n).$$

The sum of the rest of the terms on the right hand side of (11.18) satisfies

$$(11.21) \quad \frac{1}{2} k_n + k_n \log_{\frac{1}{p}} e - \frac{1}{2} \log_{\frac{1}{p}} 2\pi k_n = O(\log_{\frac{1}{p}} n).$$

From (11.18), (11.20) and (11.21), in the case of the first choice of  $k_n$ , for which  $c_n \leq c < 2$ , we have

$$\lim_{n \rightarrow \infty} \log_{\frac{1}{p}} E_n N_{n,k_n} = -\infty,$$

and thus, (11.11) holds, while in the case of the second choice of  $k_n$ , for which  $c_n \geq c > 2$ , we have

$$\lim_{n \rightarrow \infty} \log_{\frac{1}{p}} E_n N_{n,k_n} = \infty,$$

and thus (11.12) holds.  $\square$

We now begin the second moment calculation to give an idea of what is involved. Since  $\sigma^2(N_{n,k_n}) = E_n N_{n,k_n}^2 - (E_n N_{n,k_n})^2$ , the requirement

$\sigma^2(N_{n,k_n}) = o((E_n N_{n,k_n})^2)$  that is needed for the second moment method can be written as

$$E_n(N_{n,k_n})^2 = (E_n N_{n,k_n})^2 + o((E_n N_{n,k_n})^2).$$

We label the subsets  $K \subset [n]$  satisfying  $|K| = k_n$  according to the vertices they contain. Thus  $K_{i_1, \dots, i_{k_n}}$  denotes the subset  $\{i_1, \dots, i_{k_n}\}$ . Let  $I_{K_{i_1, \dots, i_{k_n}}}$  denote the indicator random variable that is equal to one if the vertices in  $K_{i_1, \dots, i_{k_n}}$  form a clique in  $G(n, p)$ , and equal to zero elsewhere. Then we can write  $N_{n,k_n}$  in the form

$$N_{n,k_n} = \sum_{1 \leq i_1 < i_2 < \dots < i_{k_n} \leq n} I_{K_{i_1, \dots, i_{k_n}}}.$$

Consequently,

$$E_n N_{n,k_n}^2 = \sum_{\substack{1 \leq i_1 < i_2 < \dots < i_{k_n} \leq n \\ 1 \leq l_1 < l_2 < \dots < l_{k_n} \leq n}} E_n I_{K_{i_1, \dots, i_{k_n}}} I_{K_{l_1, \dots, l_{k_n}}}.$$

Note that  $E_n I_{K_{i_1, \dots, i_{k_n}}} I_{K_{l_1, \dots, l_{k_n}}}$  equals 1 or 0 depending on whether or not the graph  $G(n, p)$  possesses an edge between every pair of vertices in  $K_{i_1, \dots, i_{k_n}}$  and between every pair of vertices in  $K_{l_1, \dots, l_{k_n}}$ . It is easy to see that the expression  $\sum_{1 \leq l_1 < l_2 < \dots < l_{k_n} \leq n} E_n I_{K_{i_1, \dots, i_{k_n}}} I_{K_{l_1, \dots, l_{k_n}}}$  is independent of the particular set  $K_{i_1, \dots, i_{k_n}}$ . Thus, choosing the set  $K_{1, \dots, k_n}$ , we have

$$E_n N_{n,k_n}^2 = \binom{n}{k_n} \sum_{1 \leq l_1 < l_2 < \dots < l_{k_n} \leq n} E_n I_{K_{1, \dots, k_n}} I_{K_{l_1, \dots, l_{k_n}}}.$$

Let  $J = J(l_1, \dots, l_{k_n}) = |K_{1, \dots, k_n} \cap K_{l_1, \dots, l_{k_n}}|$  denote the number of vertices shared by  $K_{1, \dots, k_n}$  and  $K_{l_1, \dots, l_{k_n}}$ . It is not hard to show that

$$E_n I_{K_{1, \dots, k_n}} I_{K_{l_1, \dots, l_{k_n}}} = \begin{cases} p^{2\binom{k_n}{2} - \binom{J}{2}}, & \text{if } J = J(l_1, \dots, l_{k_n}) \geq 2; \\ p^{\binom{k_n}{2}}, & \text{if } J = J(l_1, \dots, l_{k_n}) \leq 1. \end{cases}$$

Thus, we can write

$$E_n N_{n,k_n}^2 = \binom{n}{k_n} \sum_{\substack{1 \leq l_1 < l_2 < \dots < l_{k_n} \leq n \\ J(l_1, \dots, l_{k_n}) \leq 1}} p^{\binom{k_n}{2}} + \binom{n}{k_n} \sum_{\substack{1 \leq l_1 < l_2 < \dots < l_{k_n} \leq n \\ J(l_1, \dots, l_{k_n}) \geq 2}} p^{2\binom{k_n}{2} - \binom{J}{2}}.$$

It turns out that the first term on the right hand side above is equal to  $(E_n N_{n,k_n})^2 + o((E_n N_{n,k_n})^2)$ , while the second term is equal to  $o((E_n N_{n,k_n})^2)$ .

However the proofs of these statements require considerable additional calculation. We refer the reader to [55], and note that the notation there is a little different from the notation here.

A deeper and more difficult result than the above ones is the identification and analysis of a striking phase transition in the connectivity properties of  $G(n, p_n)$  between the cases  $p = \frac{c}{n}$  with  $c \in (0, 1)$  and  $p_n = \frac{c}{n}$  with  $c > 1$ . Note that by Theorem 11, for such values of  $p_n$  the probability of  $G(n, p_n)$  being connected approaches zero as  $n \rightarrow \infty$ . The phase transition that occurs for  $p_n$  close to  $\frac{1}{n}$  concerns the size of the largest connected component, which makes a transition from logarithmic to linear as a function of  $n$ . We only state the result, which was proved by Erdős and Rényi in 1960 [24]. For more recent proofs, see [4, 55].

**Theorem 14.** *i. Let  $p_n = \frac{c}{n}$ , with  $c \in (0, 1)$ . Then there exists a  $\gamma = \gamma(c)$  such that the size  $C_n^{lg}$  of the largest connected component of  $G(n, p_n)$  satisfies*

$$\lim_{n \rightarrow \infty} P_n(C_n^{lg} \leq \gamma \log n) = 1;$$

*ii. Let  $p = \frac{c}{n}$  with  $c > 1$ . Then there exists a unique solution  $\beta = \beta(c) \in (0, 1)$  to the equation  $1 - e^{-cx} - x = 0$ . For any  $\epsilon > 0$ , the size  $C_n^{lg}$  of the largest connected component of  $G(n, p_n)$  satisfies*

$$\lim_{n \rightarrow \infty} P((1 - \epsilon)\beta n \leq C_n^{lg} \leq (1 + \epsilon)\beta n) = 1.$$

A detailed analysis of the largest connected component in the case that  $p_n = \frac{1}{n} + o(\frac{1}{n})$  can be found in [13].

## 12. ARCSINE LAW FOR RANDOM WALKS AND A COMBINATORIAL LEMMA OF SPARRE-ANDERSEN

Consider the simple symmetric random walk (SSRW)  $\{S_n^\pm\}_{n=0}^\infty$  on  $\mathbb{Z}$ . It is constructed from a sequence  $\{X_n^\pm\}_{n=1}^\infty$  of IID random variables with the following Bernoulli distribution with parameter  $\frac{1}{2}$ :  $P(X_n^\pm = 1) = P(X_n^\pm = -1) = \frac{1}{2}$ . One defines  $S_0^\pm = 0$  and  $S_n^\pm = \sum_{j=1}^n X_j^\pm$ ,  $n \geq 1$ . It is well-known that the SSRW is *recurrent* [22], [55]:

$$P(\liminf_{n \rightarrow \infty} S_n^\pm = -\infty \text{ and } \limsup_{n \rightarrow \infty} S_n^\pm = +\infty) = 1.$$

Let

$$A_{+,n} = A_{+,n}(\{S_j^\pm\}_{j=1}^n) = |\{j \in [n] : S_j^\pm > 0\}|$$

denote the number of steps up until step  $n$  that the SSRW spends on the positive half axis. If one wants to consider things completely symmetrically, then one can define instead

$$A_{+,n}^{\text{sym}} = A_{+,n}^{\text{sym}}(\{S_j^\pm\}_{j=1}^n) = |\{j \in [n] : S_j^\pm > 0 \text{ or } S_j^\pm = 0 \text{ and } S_{j-1}^\pm > 0\}|.$$

There are several ways, some more probabilistic and some more combinatoric, to prove the following result.

**Theorem 15.** *For SSRW,*

$$(12.1) \quad P(A_{+,2n}^{\text{sym}} = 2k) = \frac{\binom{2k}{k} \binom{2n-2k}{n-k}}{2^{2n}}, \quad k = 0, 1, \dots, n.$$

See, for example, [26] for a somewhat probabilistic approach and see [55] for a completely combinatorial approach using Dyck paths. Stirling's formula gives

$$\frac{\binom{2k}{k} \binom{2n-2k}{n-k}}{2^{2n}} \sim \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}},$$

and for any  $\epsilon \in (0, 1)$ , the above estimate is uniform for  $k \in [\epsilon n, (1 - \epsilon)n]$  as  $n \rightarrow \infty$ . Using this estimate with (12.1), a straightforward calculation [55] reveals that

$$(12.2) \quad \lim_{n \rightarrow \infty} P\left(\frac{1}{2n} A_{+,2n}^{\text{sym}} \leq x\right) = \frac{2}{\pi} \arcsin \sqrt{x}, \quad x \in [0, 1].$$

Of course it then follows that (12.2) also holds with  $n$  in place of  $2n$ . It is not hard to show that the proportion of time up to  $n$  that the SSRW is equal to zero converges in distribution to 0. See, for example, [55, Exercise 4.3]. Thus, we also have

$$(12.3) \quad \lim_{n \rightarrow \infty} P\left(\frac{1}{n} A_{+,n} \leq x\right) = \frac{2}{\pi} \arcsin \sqrt{x}, \quad x \in [0, 1].$$

The distribution whose distribution function is  $\frac{2}{\pi} \arcsin \sqrt{x}$ , for  $x \in [0, 1]$ , is called the *Arcsine distribution*. The density of this distribution is given by  $\frac{d}{dx} \left( \frac{2}{\pi} \arcsin \sqrt{x} \right) = \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$ . We note that in (12.1), the most likely values of the distribution are  $k = 0$  and  $k = n$ , and the least likely values are  $k = \frac{n}{2}$ , if  $n$  is even, and  $k = \lfloor \frac{n}{2} \rfloor \pm 1$ , if  $n$  is odd. In the limit, the resulting density is unbounded at  $0^+$  and  $1^-$ , and attains its minimum at  $x = \frac{1}{2}$ . Thus, counter-intuitively, it turns out that the most likely proportions of time that the SSRW is on the positive half axis are 0 and 1, and the least likely proportion is  $\frac{1}{2}$ . For a discussion of this, see [26].

Now consider a general symmetric random walk on  $\mathbb{R}$ . It is constructed from an IID sequence of random variables  $\{X_n\}_{n=1}^\infty$  with an arbitrary symmetric distribution:  $P(X_n \geq x) = P(X_n \leq -x)$ , for  $x \geq 0$ . We define  $\mathcal{S}_0 = 0$  and  $\mathcal{S}_n = \sum_{j=1}^n X_j$ ,  $n \geq 1$ . We will assume that the distribution is continuous; that is  $P(X_n = x) = 0$ , for all  $x \in \mathbb{R}$ . It then follows that  $P(\mathcal{S}_n = x) = 0$ , for all  $x \in \mathbb{R}$  and  $n \in \mathbb{N}$ . In the sequel we will use the fact that  $P(\mathcal{S}_n > 0) = 1 - P(\mathcal{S}_n < 0)$ . Here is a truly remarkable result concerning  $A_{+,n} = A_{+,n}(\{\mathcal{S}_j\}_{j=1}^n)$ :

**Theorem 16.** *For any symmetric random walk  $\{\mathcal{S}_n\}_{n=0}^\infty$  generated from a continuous distribution,*

$$(12.4) \quad P(A_{+,n} = k) = \frac{\binom{2k}{k} \binom{2n-2k}{n-k}}{2^{2n}}, \quad k = 0, 1, \dots, n.$$

Thus, (12.3) not only holds for the SSRW on  $\mathbb{Z}$ , but also for general symmetric random walks on  $\mathbb{R}$ . Theorem 16 was first proved by Sparre-Andersen in 1949 [6]. The very complicated original proof has since been very much simplified by combinatorial arguments. This simplification is an example par excellence of the use of combinatorial tools to prove probabilistic results. Our exposition is a streamlined version of the proof in [27]. For a somewhat different approach, see [41].

A combinatorial lemma will show that Theorem 16 is equivalent to the following result, of independent interest.

**Theorem 17.** *Let  $\{\mathcal{S}_n\}_{n=0}^\infty$  be a symmetric random walk on  $\mathbb{R}$  generated by a continuous distribution. Let  $L_n$  denote the index (that is, location) of the first maximum of  $\{\mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_n\}$ . (Actually, by the continuity assumption, the maximum almost surely occurs at only one location.) Then*

$$(12.5) \quad P(L_n = k) = \frac{\binom{2k}{k} \binom{2n-2k}{n-k}}{2^{2n}}, \quad k = 0, 1, \dots, n.$$

**Remark.** Thus,

$$\lim_{n \rightarrow \infty} P\left(\frac{1}{n}L_n \leq x\right) = \frac{2}{\pi} \arcsin \sqrt{x}, \quad x \in [0, 1].$$

*Proof of Theorem 17.* We write the event  $\{L_n = k\}$  as

$$\{L_n = k\} = A_k \cap B_{n-k}, \quad k = 0, \dots, n,$$

where

$$A_k = \{\mathcal{S}_k > \mathcal{S}_0, \dots, \mathcal{S}_k > \mathcal{S}_{k-1}\}, \quad k = 1, \dots, n$$

and

$$B_{n-k} = \{\mathcal{S}_k \geq \mathcal{S}_{k+1}, \dots, \mathcal{S}_k \geq \mathcal{S}_n\}, \quad k = 0, \dots, n-1,$$

and  $A_0$  and  $B_0$  are equal to the entire probability space, so  $P(A_0) = P(B_0) = 1$ . The events  $A_k$  and  $B_{n-k}$  are independent since  $A_k$  depends only on  $\{X_j\}_{j=1}^k$  and  $B_{n-k}$  depends only on  $\{X_j\}_{j=k+1}^n$ . Thus,

$$(12.6) \quad P(L_n = k) = P(A_k)P(B_{n-k}), \quad k = 0, \dots, n.$$

From the equidistribution of the  $\{X_n\}_{n=1}^\infty$  and the continuity and symmetry of the distribution, it follows that

$$(12.7) \quad \begin{aligned} P(A_k) &= q_k, \quad k = 1, \dots, n; \\ P(B_{n-k}) &= q_{n-k}, \quad k = 0, \dots, n-1, \end{aligned}$$

where

$$(12.8) \quad q_n = P(\mathcal{S}_1 > 0, \dots, \mathcal{S}_n > 0), \quad n = 1, 2, \dots.$$

Define  $q_0 = 1$  and let  $H(s)$  be the generating function for  $\{q_n\}_{n=0}^\infty$ :

$$H(s) = \sum_{n=0}^{\infty} q_n s^n.$$

From (12.6) and (12.7) we have

$$1 = \sum_{k=0}^n P(L_n = k) = \sum_{k=0}^n q_k q_{n-k},$$

from which we obtain  $H^2(s) = (1-s)^{-1}$ . Thus,

$$H(s) = (1-s)^{-\frac{1}{2}}.$$

One has

$$\frac{d^n H}{ds^n}(0) = \frac{(2n)!}{n!2^{2n}}, \quad n \geq 0.$$

Thus,

$$(12.9) \quad q_n = \frac{1}{2^{2n}} \binom{2n}{n}, \quad n \geq 0.$$

Theorem 17 follows from (12.6), (12.7) and (12.9).  $\square$

From (12.8) and (12.9) we obtain the following corollary of interest, which is not needed for the proof of Theorem 16.

**Corollary 1.**

$$(12.10) \quad P(\mathcal{S}_1 > 0, \dots, \mathcal{S}_n > 0) = \frac{1}{2^{2n}} \binom{2n}{n}, \quad n \geq 1.$$

We now turn to the combinatorial lemma. Let  $n \in \mathbb{N}$ , let  $\{x_j\}_{j=1}^n$  be real numbers, denote their partial sums by  $\{s_j\}_{j=1}^n$ , and let  $s_0 = 0$ . We say that the first maximum among the partial sums occurs at  $j \in \{0, \dots, n\}$  if  $s_j > s_i$ , for  $i = 0, \dots, j - 1$ , and  $s_j \geq s_i$ , for  $i = j + 1, \dots, n$ . For each permutation  $\sigma \in S_n$ , let  $s_0^\sigma = 0$  and  $s_j^\sigma = \sum_{i=1}^j x_{\sigma_i}$ ,  $j = 1, \dots, n$ , denote the partial sums of the permuted sequence  $\{x_{\sigma_j}\}_{j=1}^n$ .

**Combinatorial Lemma.** *Let  $r$  be an integer satisfying  $0 \leq r \leq n$ . The number  $A_r$  of permutations  $\sigma \in S_n$  for which exactly  $r$  of the partial sums are strictly positive is equal to the number  $B_r$  of permutations  $\sigma \in S_n$  for which the first maximum among these partial sums occurs at position  $r$ .*

*Proof.* We prove the lemma by induction. The result is true for  $n = 1$ . Indeed, if  $x_1 > 0$ , then  $A_1 = B_1 = 1$  and  $A_0 = B_0 = 0$ , while if  $x_1 \leq 0$ , then  $A_1 = B_1 = 0$  and  $A_0 = B_0 = 1$ . Now let  $n \geq 2$  and assume that the result is true for  $n - 1$ . Denote by  $A_r^{(k)}$  and  $B_r^{(k)}$  the values corresponding to  $A_r$  and  $B_r$  when the  $n$ -tuple  $(x_1, \dots, x_n)$  is replaced by the  $(n - 1)$ -tuple obtained by deleting  $x_k$ . By the inductive hypothesis,  $A_r^{(k)} = B_r^{(k)}$ , for  $k = 1, \dots, n$  and  $r = 0, \dots, n - 1$ . This is also true for  $r = n$  since trivially  $A_n^{(k)} = B_n^{(k)} = 0$ .

We break up the rest of the proof into two cases. The first case is when  $\sum_{k=1}^n x_k \leq 0$ . We construct the  $n!$  permutations of  $(x_1, \dots, x_n)$  by first selecting  $k \in [n]$  and placing  $x_k$  in the last position, and then permuting the remaining  $n - 1$  numbers in the first  $n - 1$  positions. Since the  $n$ th partial sum for every permutation is equal to  $\sum_{k=1}^n x_k$ , which is non-positive, it is clear that the number of positive partial sums and the index of the first maximal partial sum depend only on the numbers in the first  $n - 1$  positions. Thus,  $A_r = \sum_{k=1}^n A_r^{(k)}$  and  $B_r = \sum_{k=1}^n B_r^{(k)}$ , and consequently,  $A_r = B_r$  by the inductive hypothesis.

The second case is when  $\sum_{k=1}^n x_k > 0$ . In this case, the  $n$ th partial sum is positive, and thus the previous argument shows that  $A_r = \sum_{k=1}^n A_{r-1}^{(k)}$ , for  $r = 1, \dots, n$ , and  $A_0 = 0$ . To obtain an analogous formula for  $B_r$ , construct the  $n!$  permutations of  $(x_1, \dots, x_n)$  by first selecting  $k \in [n]$  and placing  $x_k$  in the first position, and then permuting the remaining  $n - 1$  numbers in the last  $n - 1$  positions. Such a permutation is of the form  $(x_k, x_{j_1}, \dots, x_{j_{n-1}})$ . Since the  $n$ th partial sum is positive for every permutation, it follows that  $s_0^\sigma = 0$  is not a maximal partial sum for any permutation  $\sigma$ , so  $B_0 = 0$ .

Thus,  $A_0 = B_0$ . Clearly, the first maximal partial sum for the permuted order  $(x_k, x_{j_1}, \dots, x_{j_{n-1}})$  occurs at index  $r \in [n]$  if and only if the first maximum of the partial sums of  $(x_{j_1}, \dots, x_{j_{n-1}})$  occurs at  $r - 1$ . Thus,  $B_r = \sum_{k=1}^n B_{r-1}^{(k)}$ , for  $r = 1, \dots, n$ . By the inductive hypothesis, it then follows that  $A_r = B_r$ .  $\square$

We now prove Theorem 16 using the Combinatorial Lemma and Theorem 17.

*Proof of Theorem 16.* For  $n \in \mathbb{N}$ , label the permutations in  $S_n$  from 1 to  $n!$ , with the first one being the identity permutation. Consider the random variables  $\{X_k\}_{k=1}^n$  and their partial sums  $\{\mathcal{S}_k\}_{k=1}^n$ . Denote the partial sums for the  $j$ th permutation of  $\{X_k\}_{k=1}^n$  by  $\{\mathcal{S}_k^{(j)}\}_{k=1}^n$ ; thus,  $\mathcal{S}_k^{(1)} = \mathcal{S}_k$ . For a fixed integer  $r$  satisfying  $0 \leq r \leq n$ , let  $Z_{n,r}^{(j)}$  be the indicator random variable equal to 1 if permutation number  $j$  has exactly  $r$  positive sums, and equal to 0 otherwise. By symmetry, the random variables  $\{Z_{n,r}^{(j)}\}_{j=1}^{n!}$  have the same distribution. Thus, we have

$$(12.11) \quad P(A_{+,n} = r) = P(Z_{n,r}^{(1)} = 1) = EZ_{n,r}^{(1)} = \frac{1}{n!} E \sum_{k=1}^n Z_{n,r}^{(k)}.$$

Similarly, let  $W_{n,r}^{(j)}$  be the indicator random variable equal to 1 if for permutation number  $j$ , the first maximal partial sum has index  $r$ , and equal to 0 otherwise. By symmetry, the random variables  $\{W_{n,r}^{(j)}\}_{j=1}^{n!}$  have the same distribution. Thus, we have

$$(12.12) \quad P(L_n = r) = P(W_{n,r}^{(1)} = 1) = EW_{n,r}^{(1)} = \frac{1}{n!} E \sum_{k=1}^n W_{n,r}^{(k)},$$

where  $L_n$  is as in Theorem 17. By the Combinatorial Lemma,  $\sum_{k=1}^n Z_{n,r}^{(k)} = \sum_{k=1}^n W_{n,r}^{(k)}$ , for every realization of  $\{X_j\}_{j=1}^n$ . Thus, the right hand sides of (12.11) and (12.12) are equal. Thus  $P(A_{+,n} = r) = P(L_n = r)$ ,  $r = 0, \dots, n$ , which in conjunction with Theorem 17 proves Theorem 16.  $\square$

The material in this section is one of the cornerstones of a larger subject which includes path decomposition and the Wiener-Hopf factorization for random walks and for Lévy processes. See [26] for some basic material on Wiener-Hopf factorization. For more modern and more comprehensive work, see for example [20, 44, 62] and references therein. The exploitation of symmetries, of which the proofs of Theorems 16 and 17 are examples, is an important recurring theme in probability theory; see for example [42].

## 13. THE PROBABILISTIC METHOD

I cannot imagine a survey of the bridge between combinatorics and probability that does not include the probabilistic method, popularized by Erdős beginning in the 1940s. The method gives only a one-sided bound, and that bound is usually not tight, yet frequently it is the only known method to yield a non-trivial result. Almost every combinatorist is familiar with Erdős' use of the method to get a lower bound on the Ramsey number  $R(s, s)$ . A small tweaking of the most basic use of the method ([4], [55]) gives the lower bound  $R(k, k) \geq (1 + o(1)) \frac{k}{\sqrt{2e}} 2^{\frac{k}{2}}$ , leaving a wide gap with the best known upper bound which is around  $4^k$ . We illustrate the probabilistic method with two other examples, one entirely straightforward and simplistic, the other a bit more creative. For much more on the probabilistic method, see [4].

**Two-colorings with no monochromatic sets.** Consider  $N$  sets of objects, not necessarily disjoint, each of size  $k$ , with  $N, k \geq 2$ . Then the total number of distinct objects in the union of the sets is at least  $k$  and at most  $Nk$ . Given  $k$ , how large can  $N$  be so that no matter what the configuration of the objects, it is always possible to assign one of two colors to each object in such a way that none of the  $N$  sets is monochromatic?

*Example.* Let  $N = 3$  and  $k = 2$ . Consider three objects, labelled 1, 2 and 3, and define three sets of size two by  $\{1, 2\}, \{1, 3\}, \{2, 3\}$ . Then every two-coloring will produce a monochromatic set.

**Proposition 3.** *Consider  $N$  sets of not necessarily disjoint objects, where each set contains  $k$  objects. If*

$$N\left(\frac{1}{2}\right)^{k-1} < 1,$$

*then it is always possible to choose a two-coloring of the objects in such a way that none of the  $N$  sets is monochromatic.*

*Proof.* Call the two colors black and white. To use the probabilistic method, one colors all the objects independently and at random so that each object is colored black or white with equal probabilities  $\frac{1}{2}$ . Let  $A_j$  be the event that the  $j$ th set is monochromatic. Then  $P(A_j) = \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^{k-1}$ . Let  $A = \cup_{j=1}^N A_j$  be the event that at least one of the  $N$  sets is monochromatic. Then  $P(A) \leq \sum_{j=1}^N P(A_j) = N\left(\frac{1}{2}\right)^{k-1}$ . If  $N\left(\frac{1}{2}\right)^{k-1} < 1$ , then  $P(A) < 1$ , which guarantees that  $A^c \neq \emptyset$ .  $\square$

**Maximal antichains and Sperner's Theorem.** Let  $n \in \mathbb{N}$ . Recall that a family  $\mathcal{F}_n$  of subsets of  $[n]$  is called an *antichain* in  $[n]$  if no set belonging to  $\mathcal{F}_n$  is contained in another set belonging to  $\mathcal{F}_n$ .

**Theorem 18.** *Let  $\mathcal{F}_n$  be an antichain in  $[n]$ . Then*

$$(13.1) \quad \sum_{A \in \mathcal{F}_n} \frac{1}{\binom{n}{|A|}} \leq 1.$$

As a corollary of Theorem 18, we obtain *Sperner's Theorem*.

**Corollary 2.** *Let  $\mathcal{F}_n$  be an antichain in  $[n]$ . Then*

$$(13.2) \quad |\mathcal{F}_n| \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

*Proof of Corollary.* The function  $a \rightarrow \binom{n}{a}$  is maximized at  $a = \lfloor \frac{n}{2} \rfloor$ . Thus, from (13.1), we have  $|\mathcal{F}_n| \frac{1}{\binom{n}{\lfloor \frac{n}{2} \rfloor}} \leq 1$ .  $\square$

Sperner's theorem originally appeared in [69]. See [3] for a short non-probabilistic proof.

*Proof of Theorem 18.* We follow the argument in [4]. Fix an antichain  $\mathcal{F}_n$  in  $[n]$ . For each permutation  $\sigma \in S_n$ , define the family  $\mathcal{C}_\sigma$  of subsets of  $[n]$  by

$$\mathcal{C}_\sigma = \{ \{ \sigma_i : 1 \leq i \leq j \} : 1 \leq j \leq n \}.$$

Let  $X_n(\sigma) = |\mathcal{F}_n \cap \mathcal{C}_\sigma|$  denote the number of subsets of  $[n]$  common to both  $\mathcal{F}_n$  and  $\mathcal{C}_\sigma$ . We now consider  $S_n$  with the uniform probability measure  $P_n$ ; this turns  $X_n = X_n(\sigma)$  into a real-valued random variable and  $\mathcal{C} = \mathcal{C}_\sigma$  into a family-of-sets-valued random variable on the probability space  $(S_n, P_n)$ . We represent  $X_n$  as a sum of indicator random variables. For  $A \in \mathcal{F}_n$ , let  $I_A$  equal 1 if  $A \in \mathcal{C}$  and 0 otherwise. Then

$$X_n = \sum_{A \in \mathcal{F}_n} I_A$$

and

$$(13.3) \quad E_n X_n = \sum_{A \in \mathcal{F}_n} E_n I_A = \sum_{A \in \mathcal{F}_n} P_n(A \in \mathcal{C}).$$

The random family  $\mathcal{C}$  of subsets of  $[n]$  contains exactly one subset of size  $|A|$ . Since  $\sigma$  is a uniformly random permutation, this one subset of size  $|A|$  is distributed uniformly over all subsets of size  $|A|$ . Thus,

$$(13.4) \quad P_n(A \in \mathcal{C}) = \frac{1}{\binom{n}{|A|}}.$$

By construction, for any  $\sigma \in S_n$ , the family of subsets  $\mathcal{C}_\sigma$  forms a *chain*; that is for  $A, B \in \mathcal{C}_\sigma$ , either  $A \subset B$  or  $B \subset A$ . Thus, since  $\mathcal{F}_n$  is an antichain, it follows that  $X_n(\sigma) = |\mathcal{F}_n \cap \mathcal{C}_\sigma| \leq 1$ , for all  $\sigma \in S_n$ . In particular then,

$$(13.5) \quad E_n X_n \leq 1.$$

The theorem follows from (13.3)-(13.5).  $\square$

**Acknowledgement.** I thank Jim Pitman, as this article was much enhanced through the generous advice and constructive criticism he proffered upon viewing a first draft and then again upon viewing a second draft.

#### REFERENCES

- [1] Aldous, D., *The continuum random tree. I*, Ann. Probab. **19** (1991), 1-28.
- [2] Aldous, D., *The continuum random tree. II. An overview*, Stochastic analysis (Durham, 1990), 23-70, London Math. Soc. Lecture Note Ser., 167, Cambridge Univ. Press, Cambridge, (1991).
- [3] Aigner, M. and Ziegler, G., *Proofs From The Book*, Fourth edition, Springer-Verlag, Berlin, (2010).
- [4] Alon, N. and Spencer, J., *The Probabilistic Method*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., New York, (1992).
- [5] Arratia, R., Barbour, A. and Tavaré, S., *Logarithmic Combinatorial Structures: A Probabilistic Approach*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zurich, (2003).
- [6] Andersen, E. S., *On the number of positive sums of random variables*, Skand. Aktuarietidskr. **32** (1949), 27-36.
- [7] Baik, J., Deift, P. and Johansson, K., *On the distribution of the length of the longest increasing subsequence of random permutations*, J. Amer. Math. Soc. **12** (1999), 1119-1178.
- [8] Barbour, A., Holst, L. and Janson, S., *Poisson Approximation*, Oxford Studies in Probability, **2**, Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, (1992).
- [9] Bhatnagar, N. and Peled, R., *Lengths of monotone subsequences in a Mallows permutation*, Probab. Theory Related Fields **161** (2015), 719-780.
- [10] Bertoin, J. and Miermont, G., *The cut-tree of large Galton-Watson trees and the Brownian CRT*, Ann. Appl. Probab. **23** (2013), 1469-1493.
- [11] Billingsley, P., *The probability theory of additive arithmetic functions*, Ann. Probability **2** (1974), 749-791.
- [12] Bollobás, B., *Modern Graph Theory*, Graduate Texts in Mathematics **184**, Springer-Verlag, New York, (1998).
- [13] Bollobás, B., *The evolution of random graphs*, Trans. Amer. Math. Soc. **286** (1984), 257-274.

- [14] Bóna, M., *Combinatorics of Permutations*. With a foreword by Richard Stanley. Second edition, Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, (2012).
- [15] Buchstab, A., *An asymptotic estimation of a general number-theoretic function*, Mat. Sbornik, **44** (1937), 1239-1246.
- [16] Comtet, L. *Advanced Combinatorics. The Art of Finite and Infinite Expansions*. Revised and enlarged edition, D. Reidel Publishing Co., Dordrecht, (1974).
- [17] David, F. and Barton, D. *Combinatorial Chance*, Hafner Publishing Co., New York (1962).
- [18] Dickman, K., *On the frequency of numbers containing prime factors of a certain relative magnitude*, Ark. Math. Astr. Fys., **22** (1930), 1-14.
- [19] Dobiński, G., *Summierung der Reihe  $\sum \frac{n^m}{m!}$  für  $m = 1, 2, 3, 4, 5, \dots$* , Grunert's Archiv, **61** (1877), 333-336.
- [20] Doney, R., *Fluctuation Theory for Lévy Processes*, Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6-23, 2005. Edited and with a foreword by Jean Picard. Lecture Notes in Mathematics, **1897**, Springer, Berlin, (2007).
- [21] Drmota, M., *Random Trees. An Interplay Between Combinatorics and Probability*, Springer Wien New York, Vienna, (2009).
- [22] Durrett, R., *Probability Theory and Examples*, Third edition, Brooks/Cole, Belmont, CA (2005).
- [23] Erdős, P. and Lehner, J., *The distribution of the number of summands in the partitions of a positive integer*, Duke Math. J. **8** (1941), 335-345.
- [24] Erdős, P. and Rényi, A., *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17-61.
- [25] Feller, W., *The fundamental limit theorems in probability*, Bull. Amer. Math. Soc. **51** (1945), 800-832.
- [26] Feller, W., *An Introduction to Probability Theory and its Applications, Vol. I*, Third edition, John Wiley & Sons, Inc., New York-London-Sydney (1968).
- [27] Feller, W., *An Introduction to Probability Theory and its Applications, Vol. II*, Second edition, John Wiley & Sons, Inc., New York-London-Sydney (1971).
- [28] Féray, V., *Asymptotic behavior of some statistics in Ewens random permutations*, Electron. J. Probab. **18** (2013), 32 pp.
- [29] Flajolet, P. and Sedgewick, R., *Analytic Combinatorics*, Cambridge University Press, Cambridge, (2009).
- [30] Fristedt, B., *The structure of random partitions of large integers*, Trans. Amer. Math. Soc. **337** (1993), 703-735.
- [31] Gladkikh, A. and Peled, R., *On the cycle structure of Mallows permutations*, Ann. Probab. **46** (2018), 1114-1169.
- [32] Gnedin, A. and Olshanski, G., *The two-sided infinite extension of the Mallows model for random permutations*, Adv. in Appl. Math. **48** (2012), 615-639.

- [33] Gontcharoff, V., *Du domaine de l'analyse combinatoire*, (Russian) Bull. Acad. Sci. URSS Sér. Math. [Izvestia Akad. Nauk SSSR] **8** (1944), 3-48.
- [34] Hammersley, J. M., *A few seedlings of research*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics, pp. 345-394. Univ. California Press, Berkeley, Calif., (1972).
- [35] Harper, L., *Stirling behavior is asymptotically normal*, Ann. Math. Statist. **38** (1967), 410-414.
- [36] Hoffman, C., Rizzolo, D. and Slivken, E., *Pattern-avoiding permutations and Brownian excursion Part I: shapes and fluctuations*, Random Structures Algorithms **50** (2017), 394-419.
- [37] Hoffman, C., Rizzolo, D. and Slivken, E., *Pattern-avoiding permutations and Brownian excursion, part II: fixed points*, Probab. Theory Related Fields **169** (2017), 377-424.
- [38] Janson, S., *Patterns in random permutations avoiding the pattern 132*, Combin. Probab. Comput. **26** (2017), 24-51.
- [39] Janson, S., *Patterns in random permutations avoiding the pattern 321*, Random Structures Algorithms **55** (2019), 249-270.
- [40] Janson, S., *Patterns in random permutations avoiding some sets of multiple patterns*, Algorithmica **82** (2020), 616-641.
- [41] Kallenberg, O., *Foundations of Modern Probability*, Second edition. Probability and its Applications (New York). Springer-Verlag, New York, (2002).
- [42] Kallenberg, O., *Probabilistic Symmetries and Invariance Principles*, Probability and its Applications (New York). Springer, New York, (2005).
- [43] Kolčín, V., *A certain problem of the disribution of particles in cells, and cycles of random permutation* (Russian), Teor. Verojatnost. i Primenen. **16** (1971), 67-82.
- [44] Kyprianou, A., *Fluctuations of Lévy Processes With Applications*, Introductory lectures. Second edition. Universitext. Springer, Heidelberg, (2014).
- [45] Le Gall, J.-F., *Random trees and applications* Probab. Surv. **2** (2005), 245-311.
- [46] Le Gall, J.-F. and Miermont, G., *Scaling limits of random trees and planar maps. Probability and statistical physics in two and more dimensions*, 155-211, Clay Math. Proc., **15**, Amer. Math. Soc., Providence, RI, (2012).
- [47] Logan, B. and Shepp, L., *A variational problem for random Young tableaux*, Advances in Math. **26** (1977), 206-222.
- [48] Mansour, T. *Combinatorics of Set Partitions*, Chapman & Hall/CRC an imprint of Taylor & Francis LLC, Discrete Mathematics and Its Applications Series, (2012).
- [49] Mansour, T. and Schork, M., *Commutation Relations, Normal Ordering and Stirling Numbers*, Chapman & Hall/CRC an imprint of Taylor & Francis LLC, Discrete Mathematics and Its Applications Series, (2015).
- [50] Marckert, J.-F. and Miermont, G., *The CRT is the scaling limit of unordered binary trees*, Random Structures Algorithms **38** (2011), 467-501.

- [51] Montgomery, H. and Vaughan, R., *Multiplicative Number Theory. I. Classical Theory*, Cambridge Studies in Advanced Mathematics **97**, Cambridge University Press, Cambridge, (2007).
- [52] Mueller, C. and Starr, S., *The length of the longest increasing subsequence of a random Mallows permutation*, J. Theoret. Probab. **26** (2013), 514-540.
- [53] Najnudel, J. and Pitman, J., *Feller coupling of cycles of permutations and Poisson spacings in inhomogeneous Bernoulli trials*, Electron. Commun. Probab. **25** (2020), Paper No. 73.
- [54] Pemantle, R. and Wilson, M., *Analytic Combinatorics in Several Variables*, Cambridge Studies in Advanced Mathematics, **140**, Cambridge University Press, Cambridge, (2013).
- [55] Pinsky, R., *Problems From the Discrete to the Continuous. Probability, Number Theory, Graph Theory, and Combinatorics*, Universitext. Springer, Cham, (2014).
- [56] Pinsky, R., *Some connections between permutation cycles and Touchard polynomials and between permutations that fix a set and covers of multisets*, Electron. Commun. Probab. **22**, (2017), Paper No. 17, 9 pp.
- [57] Pinsky, R., *On the strange domain of attraction to generalized Dickman distribution for sums of independent random variables*, Electron. J. Probab. **23** (2018), Paper No. 3, 17 pp.
- [58] Pinsky, R., *Permutations avoiding a pattern of length three under Mallows distributions*, Random Structures Algorithms **58** (2021), 676-690.
- [59] Pitman, J., *Some probabilistic aspects of set partitions*, Amer. Math. Monthly **104** (1997), 201-209.
- [60] Pitman, J., *Random mappings, forests, and subsets associated with Abel-Cayley-Hurwitz multinomial expansions*, Sém. Lothar. Combin. **46** (2001/02), Art. B46h.
- [61] Pitman, J., *Combinatorial Stochastic Processes*, Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7-24, 2002. With a foreword by Jean Picard. Lecture Notes in Mathematics, **1875** Springer-Verlag, Berlin, (2006).
- [62] Pitman, J. and Tang, W., *Extreme order statistics of random walks*, preprint arXiv:2007.13991 (2020).
- [63] Riordan, J., *An Introduction to Combinatorial Analysis*. Reprint of the 1958 original, Dover Publications, Inc., Mineola, NY, (2002).
- [64] Romik, D., *The Surprising Mathematics of Longest Increasing Subsequences*, Institute of Mathematical Statistics Textbooks, **4**, Cambridge University Press, New York, (2015).
- [65] Rota, G., *The number of partitions of a set*, Amer. Math. Monthly **71** (1964), 498-504.
- [66] Sachkov, V., *Random partitions of sets*, Theory Probab. Appl., **19** (1973), 184-190.
- [67] Sachkov, V., *Probabilistic Methods in Combinatorial Analysis*. Translated from the Russian. Revised by the author. Encyclopedia of Mathematics and its Applications, **56**, Cambridge University Press, Cambridge, (1997).
- [68] Shepp, L. and Lloyd, S., *Ordered cycle lengths in a random permutation*, Trans. Amer. Math. Soc. **121** (1966), 340-357.

- [69] Sperner, E., *Ein Satz über Untermengen einer endlichen Menge*, (German) Math. Z. **27** (1928), 544-548.
- [70] Stam, A., *Generation of a random partition of a finite set by an urn model*, J. Combin. Theory Ser. A **35** (1983), 231-240.
- [71] Stanley, R., *Enumerative Combinatorics. Volume 1*, Second edition, Cambridge Studies in Advanced Mathematics **49**, Cambridge University Press, Cambridge, (2012).
- [72] Tenenbaum, G., *Introduction to Analytic and Probabilistic Number Theory*, Cambridge Studies in Advanced Mathematics **46**, Cambridge University Press, Cambridge, (1995).
- [73] Veršik, A. and Kerov, S., *Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux*, (Russian) Dokl. Akad. Nauk SSSR **233** (1977), 1024-1027.
- [74] Wilf, H., *generatingfunctionology*, Third edition. A K Peters, Ltd., Wellesley, MA, (2006).

DEPARTMENT OF MATHEMATICS, TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY,  
HAIFA, 32000, ISRAEL

*Email address:* `pinsky@math.technion.ac.il`

*URL:* `http://www.math.technion.ac.il/~pinsky/`