# Large Partial Random Permutations of Conditionally Convergent Series

Jiashun Jin, Joseph B. Kadane and Avranil Sarkar Department of Statistics Carnegie Mellon University

December 16, 2008

#### Abstract

This paper studies the behavior of sums of series under random permutations. The first result considers the case of uniformly distributed permutations on the set **N** of all integers, which requires finitely additive probability. The next result considers the limit of the first portion  $p, 0 \le p \le 1$ , of a uniform distribution on the first n integers, as n gets large. The focal point is when 0 , the permuted sum converge to a class of non-degeneratedistributions which to the best of our knowledge is unknown in the literature. Finally, weaddress the question of when the limiting distribution has a continuous density, includingthe well-known Erdös problem [3, 9] as a special case.

**Keywords**: Infinite Bernoulli convolution, limiting distribution, random permutation, Riemann's Rainbow problem.

AMS 2000 subject classifications: Primary-60F05, 60B-15; Secondary-60E05, 60E10.

Acknowledgments: The authors would like to thank Saul Jacka and Jon Warren for valuable pointers and discussion. Jiashun Jin was partially supported by NSF CAREER award DMS 09-08613.

## 1 Introduction

This investigation began by reconsidering Riemann's famous Rainbow Theorem, which addresses series that are convergent but not absolutely convergent (these series are also called conditionally convergent). A canonical example of such a series is the alternating harmonic series,

$$X_j = (-1)^{j+1} \frac{1}{j}, \qquad j = 1, 2, \dots,$$

Riemann's Theorem says that if one chooses a target sum T  $(-\infty < T < \infty)$ , there is a rearrangement of the terms of a conditionally convergent series that approaches T. Since a rearrangement of terms is a permutation of the set of integers, it occurred to us to ask what the distribution might be of a conditionally convergent series under a random permutation of the positive integers, i.e. what might be said about the distribution of

$$\sum_{j=1}^{\infty} X_{\pi(j)},\tag{1.1}$$

where  $\{X_j\}$  is a conditionally convergent series and  $\pi = \{\pi(1), \pi(2), \ldots\}$  is a random permutation of the positive integers. This formulation is intellectually connected to those in Jacka and Warren [5, 6] and has proved difficult for a number of reasons.

First, it is not clear exactly what is meant by a random permutation of the positive integers. The space of such permutations is large, indeed, larger than countable, as can be seen by the following argument: Imagine a mapping, from all the permutations of the integers, to the integers, and let  $\pi_{i,j}$  be the *j*-th element of the *i*-th permutation. Construct the permutation  $\pi = (\pi_1, \pi_2, ...)$  as follows: Let  $\pi_1$  be any integer different from  $\pi_{1,1}$ . In general, choose  $\pi_n$  to be different from  $\pi_{1,1}, \ldots, \pi_{n-1}$  and from  $\pi_{n,n}$ . Since only a finite number of integers are excluded at each stage, this can be done (with help from the axiom of choice). Now  $\pi$  is different from each of the permutations in the mapping, which is a contradiction. Therefore there is no such mapping, so the space of permutations of integers is not countable.

A second reason that (1.1) is not well formulated is that it isn't clear, whatever might be meant by a random permutation, whether they will typically converge. The standard construction in the poof of the Rainbow Theorem also proves the following generalization: Let  $-\infty \leq T_1 \leq T_2 \leq \infty$  be two numbers. Let  $X_1, X_2, \ldots$  be a conditionally convergent series. Then there is a rearrangement of the series such that the resulting partial sums have lim inf  $T_1$ and lim sup  $T_2$ . So convergence of (1.1) is likely to be "unusual", if we know what we mean by a random permutation of the integers. Theorem 2.1 addresses this case, under a very general finitely additive notion of a uniform distribution on the infinite permutations.

We can also reformulate our question in a (large) finite setting as follows: Let m be a finite integer (later we'll let  $m \to \infty$ ). Let  $n = n_m$  be an integer such that  $m \le n$  (also later  $m \to \infty$ ), let  $\pi$  be a random permutation of the first n integers. Then we study the limiting distribution of

$$\sum_{j=1}^{m} X_{\pi_n(j)},$$
(1.2)

as  $n_m \to \infty$  in such a way such that  $\frac{m}{n_m} \to p$ , for  $0 \le p \le 1$ . The results are given in Theorem 2.2 and 2.3.

Last, we discuss for which p and which sequences  $\{X_j\}_{j=1}^{\infty}$  the limiting distribution of that in (1.2) has a continuous density. We investigate three cases where correspondingly  $X_j$  decay polynomially fast, sub-exponentially fast, and exponentially fast. A rich literature exists on this topic, particularly in the exponentially decay case, where the problem is connected to Erdös problem on infinite Bernoulli convolutions [3, 9].

#### 2 Main results

Let  $X = \{X_j\}_{j=1}^{\infty}$  be a sequence of real numbers, **N** be the set of all integers  $\{1, 2, \ldots\}$ , and  $\pi$  is a random permutation from **N** to **N**. Define the partial sum as

$$S_m(\pi) = S_m(\pi; X) = \sum_{j=1}^m X_{\pi(j)}.$$

Our first claim says that under a mild summability condition on X, the partial sum converges to 0 in probability as  $m \to \infty$ .

To do so, we introduce the  $\mathcal{L}_r$ -space. Fix  $r \in (0, \infty]$ , the  $\mathcal{L}_r$ -space contains all sequences that have a finite  $\ell^r$ -norm,

$$\mathcal{L}_{r} = \begin{cases} \{X = \{X_{j}\}_{j=1}^{\infty}, \ \sum_{j=1}^{\infty} |X_{j}|^{r} < \infty\}, & r < \infty, \\ \{X = \{X_{j}\}_{j=1}^{\infty}, \ \max_{\{1 \le j \le \infty\}} \{|X_{j}|\} < \infty\}, & r = \infty. \end{cases}$$

We have to specify what we mean by a random permutation on the (countably infinite) set  $\mathbf{N}$ . As in the case of uniform distributions on  $\mathbf{N}$  (see Kadane and O'Hagan [4] and Schirokauer and Kadane [8]), there are many possible ways to do this in the setting of finite additivity. Here we impose only the condition that the probability of every finite set has probability 0, and of course the set of all permutations on  $\mathbf{N}$  has probability 1. While this specifies that every co-finite set has probability 1, it leaves unspecified the probability of every infinite set of permutations whose complement is also infinite.

The following theorem summarizes the main findings and is proved in Section 4.

**Theorem 2.1** Fix  $r \in (0, \infty)$  and a sequence  $X = \{X_j\}_{j=1}^{\infty}$ . If  $X \in \mathcal{L}_r$ , then  $S_m(\pi)$  converges to 0 in probability.

Underlying the proof of the theorem is the intuition that, with overwhelming probability, the indices  $\pi(1), \pi(2), \ldots, \pi(m)$  are uniformly large. Given the summability condition, all elements in the summation of  $S_m(\pi)$  are uniformly and arbitrarily small. Hence  $S_n(\pi)$  converges to 0 in probability.

Note that when  $X \in \mathcal{L}_{\infty}$ ,  $S_m(\pi)$  does not have to converge to 0 as m tends to  $\infty$ . Consider for example the sequence of

$$X_j = (-1)^{j+1}, \qquad j = 1, 2, \dots$$

It is clear that  $|S_{m+1}(\pi) - S_m(\pi)| = 1$  for all m, so  $S_m(\pi)$  diverges with probability 1.

The key in the the proof of Theorem 2.1 is that, compared to  $\mathbf{N}$ , any finite set has a negligible cardinality, so with overwhelming probability, the indices  $\pi(1), \pi(2), \ldots, \pi(m)$  are uniformly large. This motivates us to look into a a more delicate situation, where instead of considering the permutation from  $\mathbf{N}$  to  $\mathbf{N}$ , we consider the case where  $\pi = \pi_n$  is a permutation from the set of  $\{1, 2, \ldots, n\}$  to itself. The question is then, what happens to the partial sum  $S_m$  when both m and n tend to  $\infty$  (but of course  $1 \le m \le n$ ). We discuss this in next subsection.

## **2.1** When $\pi = \pi_n$ is a random permutation from $\{1, 2, ..., n\}$ to itself

When  $\pi = \pi_n$  is a random permutation from  $\{1, 2, ..., n\}$  to itself, the partial sum is modified into

$$S_m(\pi_n) = S_m(\pi_n; X) = \sum_{j=1}^m X_{\pi_n(j)} \cdot \mathbf{1}_{\{\pi_n(j) \le m\}}$$

Clearly, the ratio of m/n plays a key role. When m ranges between 1 and  $\infty$ , we let  $n = n_m$  such that  $1 \le m \le n_m$ . We discuss the degenerate case where the ratio  $m/n_m$  tends to either 0 or 1 and the non-degenerate case where the ratio is bounded away from 0 and 1 separately.

Consider the first case. Write  $n_m$  as n for short whenever there is no confusion. The following theorem says that when m/n tends to either 0 or 1,  $S_m(\pi_n)$  converges to a point mass in probability, under the condition that the sequence  $X = \{X_j\}_{j=1}^{\infty}$  is both summable and square summable. We note the harmonic sequence satisfies both these summability conditions.

**Theorem 2.2** Given a sequence  $\{X_j\}_{j=1}^{\infty}$  which is both summable and square summable, but not necessary absolutely summable. Let  $n = n_m$  be a sequence of indices such that  $n_m \ge m$  for all

m. When  $m \to \infty$ ,

$$S_m(\pi_{n_m}) \longrightarrow \begin{cases} 0, & m/n_m \to 0, \\ \sum_{j=1}^{\infty} X_j, & m/n_m \to 1, \end{cases} \quad in \text{ probability.}$$

Note that Theorem 2.1 can be thought of as a special case of Theorem 2.2, where  $n = \infty$ . Theorem 2.2 is proved in Section 4, where the key is the direct calculation of the variance of  $S_m(\pi_n)$ , which tends to 0 when  $m/n \to 0$  or 1.

We now move the second case, where the ratio m/n tends to some parameter  $p \in (0, 1)$ , and the results are more interesting. For simplicity, we calibrate  $n = n_m$  as

$$n = m/p. \tag{2.1}$$

Let  $b_j$  be iid Bernoulli random variables with success probability of p, i.e. Bernoulli(p). Interestingly, it turns out that  $S_m(\pi_n)$  behaves asymptotically as the sum of

$$\sum_{j=1}^{\infty} (b_j \cdot X_j), \tag{2.2}$$

partially attributed to the fact that  $1_{\{\pi_n(j) \leq m\}}$  is distributed as Bernoulli(p).

To shed light on this, we write the partial sum as

$$S_m(\pi_n) = \sum_{j=1}^n X_j \cdot 1_{\{\pi_n(j) \le m\}}.$$

First, under mild summability conditions, the sum over the indices starting, say, at N + 1 (here  $N = N_m$ ), to n is negligible,

$$\left|\sum_{j=N+1}^{n} X_{j} \cdot 1_{\{\pi_{n}(j) \le m\}}\right| \le \sum_{j=N+1}^{n} |X_{j}| \approx 0,$$

so the partial sum is approximately equals

$$S_m(\pi_n) \approx \sum_{j=1}^N X_j \cdot 1_{\{\pi_n(j) \le m\}}.$$
 (2.3)

Second, somewhat surprisingly, when  $N \ll \sqrt{n}$ , the sequence of random variables  $\{1_{\{\pi_n(j) \leq m\}}, 1 \leq j \leq N\}$  are asymptotically distributed as iid Bernoulli(p). In detail, fix  $1 \leq N \leq n$  and  $1 \leq k \leq N$ , let  $\Theta_N(k)$  be the set of N-dimensional vectors of 0's and 1's whose  $\ell^1$ -norm is k,

$$\Theta_N(k) = \{ \theta = (\theta_1, \theta_2, \dots, \theta_N), \ \theta_j \in \{0, 1\}, \ \sum_{j=1}^N \theta_j = k \}.$$

The following lemma is proved in Section 4.

**Lemma 2.1** Fix  $p \in (0,1)$ , let m = pn,  $a_p = \max\{\frac{1}{p}, \frac{1}{1-p}\}$ . For any  $1 \le N \le \frac{n}{2a_p}$ ,  $1 \le k \le N$ , and vector  $\theta \in \Theta_N(k)$ ,

$$e^{-a_p N^2/n} \le \frac{P\{\mathbf{1}_{\{\pi_n(j) \le m\}} = \theta_j, \ 1 \le j \le N\}}{\prod_{j=1}^N P\{\mathbf{1}_{\{\pi_n(j) \le m\}} = \theta_j\}} \le e^{N^2/n}.$$

Note that the denominator  $\prod_{j=1}^{N} P\{1_{\{\pi_n(j) \le m\}} = \theta_j\} = p^k (1-p)^{N-k}$ . Lemma 2.1 implies that, if we choose  $N = N_m$  and  $n = n_m$  in a way such that

$$\frac{N^2}{n} \to 0, \qquad n \to \infty,$$

then the sequence of random variables  $\{1_{\{\pi_n(j) \leq m\}}, 1 \leq j \leq N\}$  are asymptotically independent. Putting this together with (2.3), we expect to have

$$S_m(\pi_n) \approx \sum_{j=1}^N X_j \cdot \mathbb{1}_{\{\pi_n(j) \le m\}} \approx_{in \ distribution} \sum_{j=1}^N b_j X_j,$$

where we recall  $b_j$  are iid Bernoulli(p). Finally, by the summability condition on  $X_j$ ,

$$\sum_{j=1}^{N} b_j X_j \approx \sum_{j=1}^{\infty} b_j X_j,$$

so we expect to have

$$S_m(\pi_n) \approx_{in \ distribution} \sum_{j=1}^{\infty} b_j X_j.$$

This is formalized in the following theorem.

**Theorem 2.3** Let  $\{X_j\}_{j=1}^{\infty}$  be a sequence that is both summable and square summable, but not necessary absolutely summable. Fix  $p \in (0,1)$  and let  $n = n_m = m/p$ . Then as  $n \to \infty$ ,  $S_{m,n}$  converges in law to the distribution of

$$\sum_{j=1}^{\infty} (b_j \cdot X_j), \tag{2.4}$$

where  $b_j$  are iid Bernoulli(p).

Denote the distribution in (2.4) by  $\nu_0(p, X)$ . A special case is where for a number  $\lambda \in (0, 1)$ ,

$$X_j = (-1)^{j+1} \lambda^j, \qquad j \ge 1.$$

In this case,  $\nu_0(p; X)$  is closely related to the infinite Bernoulli convolution which is well-known in the literature. See [3, 9] and Section 3 for more discussion.



Figure 1: Comparison of density (top) and characteristic function (bottom; reported is modulus) for different polynomially decaying X (p = 1/2,  $\alpha \in \{0.6, 1, 2\}$ ). A fast decaying sequence Xtends to give a more wiggly density function and a heavier tail in the characteristic function than a slowly decaying one does.

## **3** When does $\nu_0(p, X)$ have a continuous density?

An interesting question is for which range of p and for which sequences  $X = \{X_j\}_{j=1}^{\infty}$  the distribution has a continuous density. The theory along this direction is surprisingly rich, and we consider the following three cases, where the rate at which  $X_j$  tends to 0 ranges from polynomial to exponential, and therefore covers the whole range of interest.

- (Polynomial decay).  $X_j = (-1)^{j+1}/j^{\alpha}, \, \alpha > 1/2.$
- (Sub-exponential decay).  $X_j = (-1)^{j+1} \exp(-j^{1/\alpha}), \alpha > 1.$
- (Exponential decay).  $X_j = (-1)^{j+1} \lambda^j, \ 0 < \lambda < 1.$

For the polynomial decay or sub-exponential decay, the main result in this section generally remains valid if we change the sequences without changing the rate of decay. For the exponential decay sequence, it is however unclear if the main result remains true if we change the sequence without changing the rate. In this case, the existence of an absolutely continuous density is a well-known challenging problem that goes back to as early as 1940's (see for example Erdös [2, 3]).

It is well-known that the smoothness of a density depends on the tail-behavior of its characteristic function. The characteristic function corresponding to  $\nu_0(p, X)$  is

$$\varphi_X(t;p) = \prod_{j=1}^{\infty} (1-p+pe^{itX_j}).$$

The key insight in the tail behavior of  $\varphi(t)$  is as follows. Introduce the grid  $\{k\pi : k \ge 1\}$ . Observe that when  $tX_j$  is bounded away from the grid by a constant  $\delta \in (0, \pi/2)$ , then

$$|1 - p + pe^{itX_j}| = \sqrt{(1 - p)^2 + 2p(1 - p)\cos(\delta) + p^2} < 1.$$
(3.1)

Suppose we fix t and p, and let the sequence range. By (3.1), the more indices at which  $tX_j$  are bounded away from the grid, the smaller the correspondingly value of  $|\varphi(t)|$  is. Now, consider a sequence that decays sufficiently slowly. Once  $tX_j$  manages to get away from the grid for some j, it remains away from the grid for many other subsequent j. Consider a sequence that decays fast enough. The value of  $tX_j$  changes rapidly, and this property does not hold any longer. In comparison, we have more j such that  $tX_j$  are bounded away from the grid with a slower decaying sequence. As a result, we expect to have a smoother density for a slowly decaying sequence.

The main result in this section can be summarized as follows. For the case of polynomial decay and the sub-exponential decay, a smooth density (i.e. the derivative of any order exists and is continuous) exists. For the exponential decay sequence, there is however a phase change. Define

$$\lambda_0 = \lambda_0(p) = p^p (1-p)^{1-p}.$$

Then for a.e.  $\lambda \ge \lambda_0$ , an absolutely continuous density exists; for all  $\lambda < \lambda_0$ , the distribution is singular. We now elaborate case by case.

Consider the case of polynomial decay first. In this case,  $tX_j$  and  $tX_{2j}$  differ only by a constant. Therefore, once  $t|X_j|$  is bounded away from the grid for  $j = j_0$ , it remains bounded away for at least  $O(j_0)$  subsequent indices. Since  $j_0$  can be as large as  $O(t^{1/\alpha})$ , we expect to



Figure 2: Comparison of density (top) and characteristic function (bottom; reported is modulus) for different sub-exponentially decaying X (p = 1/2,  $\alpha \in \{4, 2, 4/3\}$ . A fast decaying sequence X tends to give a more wiggly density function and a heavier tail in the characteristic function than a slowly decaying one does.

have at least  $O(|t|^{1/\alpha})$  many  $tX_j$ , each of which is bounded away from the grid by a constant. Therefore,  $|\varphi(t)|$  decays in a rate of  $\exp(-O(|t|^{1/\alpha}))$ . This is the following theorem which is proved in Section 4.

**Theorem 3.1** Fix  $\alpha > 1$  and  $p \in (0,1)$ , let X be the sequence  $X_j = (-1)^{j+1}/j^{\alpha}$ . There is a constant  $c_0 = c_0(p, \alpha) > 0$  such that for sufficiently large |t|,

$$|\varphi(t)| \le \exp(-c_0|t|^{1/\alpha}).$$

As a result, the distribution  $\nu_0(p, X)$  has a smooth density.

Consider the sub-exponential case next. This case turns out to be much more delicate, and a much more careful proof is necessary. The following theorem is proved in Section 4. **Theorem 3.2** Fix  $\alpha > 1$  and  $p \in (0,1)$ , let X be the sequence  $X_j = (-1)^{j+1} \exp(-j^{1/\alpha})$ . There is a constant  $c_1(p, \alpha) > 0$  such that for sufficiently large |t|,

$$|\varphi(t)| \le \exp(-(\log |t|)^{(1+c_1)}).$$

As a result, the distribution  $\nu_0(p, X)$  has a smooth density.

We now consider the case of exponential decay. In this case,  $\nu_0(p, X)$  is called the *infinite* Bernoulli convolution and there is a rich literature studying when  $\nu_0(p, X)$  has an absolutely continuous density.

In particular, the case of p = 1/2 was extensively studied, as early as 1940's. In fact, in 1940 Erdös [2, 3] shows that that  $\nu_0(p, X)$  has an absolutely continuous density for a.e.  $\lambda$  in a neighborhood of 1. Recently, the results were extended by Solomyak [9], showing that the distribution has an absolutely continuous density for a.e.  $\lambda$  in (1/2, 1), and the distribution is singular if  $\lambda > 1/2$ . The case of  $\lambda = 1/2$  is especially interesting. In this case,  $X_j = (-1)^{j+1}2^{-j}$ and  $\nu_0(1/2, X)$  is the uniform distribution on (0, 1).

The extension to the case of  $p \neq 1/2$  were considered in Peres and Solomyak [7], where a similar phenomenon is found for  $p \in [1/3, 2/3]$ , but the problem remains open otherwise. This is the following theorem, whose proof can be derived using that in [7] so we omit it.

**Theorem 3.3** Fix  $p \in [1/3, 2/3]$ , consider the sequence  $X_j = (-1)^{j+1} \lambda^j$ .

- For a.e.  $\lambda$  in the interval  $[p^p(1-p)^{1-p}, 1)$ , the distribution  $\nu_0(p, X)$  has an absolutely continuous density.
- For all  $\lambda < p^p(1-p)^{1-p}$ , the distribution  $\nu_0(p, X)$  is singular.

We remark that the "a.e." part can not be removed for that is was shown [2, 7] that when  $\lambda$  is the reciprocal of a Pisot-Vijayaraghavan (PV)-number [1],  $\nu_0(p, X)$  is singular. The question of whether the set of exceptional  $\lambda$ 's is countable remains open, to the best or our knowledge.

#### 3.1 Figures on the density function

In this section, we display some figures on the characteristic function and density function associated with  $\nu_0(p, X)$ , in hopes of learning the role of p and decaying rate of X more vividly.

First, recall that the characteristic function associated with  $\nu_0(p, X)$  is

$$\varphi_X(t;p) = \prod_{j=1}^{\infty} (1-p+pe^{itX_j}),$$



Figure 3: Comparison of density (top) and characteristic function (bottom; reported is modulus) for different exponentially decaying X (p = 1/2,  $\alpha \in \{0.95, 0.75, 0.55\}$ ). A fast decaying sequence X tends to give a more wiggly density function and a heavier tail in the characteristic function than a slowly decaying one does.

Second, once the characteristic function is given, the density function  $f_X(x;p)$  can be calculated conveniently using the inverse Fourier transform,

$$f_X(x;p) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_X(t) dt.$$

We have posted Matlab code on www.stat.cmu.edu/jiashun/Research/software/RandomPermuta tion. Input the type of the sequence X (polynomial, sub-exponential, exponential), the parameter  $\alpha$  (or  $\lambda$  in the exponential case), and the parameter p, the code outputs the density function and the characteristic function.

Below, we report the results on four example as follows, where we investigate the role of decaying rate of X in the first three, and investigate the role of p in the last one.

**Example A.** For p = 1/2 and each of the  $\alpha$  in  $\{0.6, 1, 2\}$ , let  $X_j = (-1)^{j+1}/j^{\alpha}$ , and plot the

density function as well as the modulus of the characteristic function (i.e.  $|\varphi_X(t;p)|$ ). The results are displayed in Figure 1, which suggests that a fast decaying sequence tends to gives a heavier tail in the characteristic function and a more wiggly density function than a slowly decaying one does.

- **Example B.** For p = 1/2 and every  $\alpha \in \{4, 2, 4/3\}$ , let X be the exponentially decaying sequence  $X_j = (-1)^{j+1} \exp(-j^{1/\alpha})$ . Plot the density function and the modulus of the characteristic function in Figure 2, which suggests a similar phenomenon as in Example A.
- **Example C.** For p = 1/2 and every  $\lambda \in \{0.95, 0.75, 0.55\}$ , let X be the exponentially decaying sequence  $X_j = (-1)^{j+1}\lambda^j$ . Plot the density function and the modulus of the characteristic function in Figure 3, which suggests a similar phenomenon as in Example A.
- **Example D.** Let X be the harmonic sequence  $X_j = (-1)^j/j$ ,  $p \in \{0.5, 0.7, 0.9\}$ . Plot the density function and the characteristic function. The results are displayed in Figure 4, which suggests that those p close to 1 tend to give a heavier tail in the characteristic function and a more wiggly density function than those close to 1/2.

## 4 Proofs

In this section, we prove all the theorems in the order that they appear. The proof of Lemma 2.1 is given in the end of this section.

#### 4.1 Proof of Theorem 2.1

Since  $\mathcal{L}_{r_1} \subset \mathcal{L}_{r_2}$  for any  $r_1 < r_2$ , without loss of generality, we assume r > 1. Clearly,

$$\left|\sum_{j=1}^{m} X_{\pi_{(j)}}\right| \le \left(m^{r-1} \sum_{j=1}^{m} |X_{\pi(j)}|^r\right)^{1/r} \equiv (T_m(\pi))^{1/r},$$

so it is sufficient to show that

$$T_m(\pi) \to 0,$$
 in probability. (4.1)

To do so, we introduce

$$A_k = A_k(X; r) = \sum_{j=k}^{\infty} |X_j|^r, \qquad k = 1, 2, \dots,$$



Figure 4: Comparison of density function (top) and characteristic function (bottom; real part in red, imaginary part in green) for different p. Left to right: p = 0.5, 0.75, 0.95. X is the harmonic sequence. The figure suggests that the closer that p is to 0.5, the less wiggly the density function. and

$$K_m = K_m(\pi) = \inf_{\{1 \le j \le m\}} \pi(j)$$

Let  $N = N_m$  be the smallest integer such that

$$A_N \le m^{-r},$$

we have

$$T_m(\pi) \le m^{r-1} \left( A_1 P\{K_m \le N\} + A_N P\{K_m \ge N\} \right) \le A_1 m^{r-1} P\{K_m \le N\} + \frac{1}{m}.$$
 (4.2)

Note that

$$P\{K_m \le N\} \le \sum_{j=1}^m P\{\pi(j) \le N\} = 0.$$
(4.3)

Inserting (4.3) into (4.2) gives

$$T_m(\pi) \le \frac{1}{m},$$

and (4.1) follows directly.

#### 4.2 Proof of Theorem 2.2

Write for short  $n = n_m$ , we have

$$S_m(\pi_n) = \sum_{j=1}^m X_{\{\pi_n(j)\}} = \sum_{j=1}^n X_j \mathbb{1}_{\{\pi_n(j) \le m\}}.$$

By direct calculations,

$$E[1_{\{\pi_n(j) \le m\}}] = m/n,$$

and

$$E[1_{\{\pi_n(j) \le m\}} 1_{\{\pi_n(k) \le m\}}] = \begin{cases} \frac{m}{n}, & j = k, \\ \frac{m(m-1)}{n(n-1)}, & j \ne k. \end{cases}$$

Hence,

$$E[S_m(\pi_n)] = \frac{m}{n} \sum_{j=1}^n X_j,$$
(4.4)

and

$$E[(S_m(\pi_n))^2] = \sum_{j=1}^n \frac{m}{n} X_j^2 + \sum_{\{1 \le j, k \le n, j \ne k\}} X_j X_k \frac{m(m-1)}{n(n-1)}$$

Write  $\frac{m}{n} = \frac{m(m-1)}{n(n-1)} + \frac{m(n-m)}{n(n-1)}$ . It follows from basic algebra that

$$E[(S_m(\pi_n))^2] = \frac{m(n-m)}{n(n-1)} \sum_{j=1}^n X_j^2 + \frac{m(m-1)}{n(n-1)} (\sum_{j=1}^n X_j)^2.$$
(4.5)

Now, combining (4.5) with (4.4) gives

$$\operatorname{var}(S_m(\pi_n)) = \frac{m(n-m)}{n(n-1)} \sum_{j=1}^n X_j^2 - \frac{m(n-m)}{n^2(n-1)} (\sum_{j=1}^n X_j)^2,$$
(4.6)

and the claim follows easily from (4.4) and (4.6).

Pick  $N_n = n^{1/3}$ , denote for short  $N = N_n$  and  $m = m_n$ . Write

$$S_m(\pi_n, X) = \sum_{j=1}^N X_j \cdot \mathbf{1}_{\{\pi_n(j) \le m\}} + \sum_{j=N+1}^n X_j \cdot \mathbf{1}_{\{\pi_n(j) \le m\}} \equiv I + II.$$
(4.7)

Let  $\varphi_{1,N}(t)$  and  $\varphi_0(t)$  denote the characteristic functions of I and  $\nu_0(p, X)$ , respectively. It is sufficient to show that as  $n \to \infty$ ,

$$|\varphi_{1,N}(t) - \varphi_0(t)| \to 0,$$
 for every fixed  $t,$  (4.8)

and

$$E[II] \to 0, \qquad E[(II)^2] \to 0.$$
 (4.9)

Consider (4.8) first. Introducing a bridging quantity  $\varphi_{0,N}(t)$ , which is the characteristic function of  $\sum_{j=1}^{N} b_j X_j$ . Write  $\sum_{j=1}^{N} b_j X_j + \sum_{j=N+1}^{\infty} b_j X_j$ . Direct calculations show that both mean and the variance of the second part tend to 0, so we have

$$|\varphi_{0,N}(t) - \varphi_0(t)| \to 0,$$
 for every fixed t.

Using triangle inequality, to show (4.8), it is sufficient to show

$$|\varphi_{1,N}(t) - \varphi_{0,N}(t)| \to 0,$$
 for every fixed t. (4.10)

To do so, we note that

$$\varphi_{1,N}(t) = \sum_{\{\theta \in \Theta_N(k)\}} e^{it \sum_{j=1}^N \theta_j X_j} P\{1_{\{\pi_n(j) \le m\}} = \theta_j, \ 1 \le j \le N\},\$$

and

$$\varphi_{0,N}(t) = \sum_{\{\theta \in \Theta_N(k)\}} e^{it \sum_{j=1}^N \theta_j X_j} P\{b_j = \theta_j, \ 1 \le j \le N\},\$$

where all except the probability terms are the same. Note that for any  $\theta \in \Theta_N(k)$ ,

$$P\{b_j = \theta_j, \ 1 \le j \le N\} = p^k (1-p)^{N-k},$$

and that by Lemma 2.1,

$$P\{I_N = \lambda\} = \left(1 + O(\frac{N^2}{n})\right) \cdot p^k (1-p)^{N-k}.$$

Combining these gives

$$|\varphi_{1,N}(t) - \varphi_{0,N}(t)| = O(\frac{N^2}{n}) \cdot \sum_{\lambda \in S_N(k)\}} p^k (1-p)^{N-k} = O(n^{-1/3}),$$

where in the last equality we used that  $\sum_{\lambda \in S_N(k)} p^k (1-p)^{N-k} = 1$  and that  $N = n^{1/3}$ . This shows (4.8).

We now consider (4.9). The first claim follows directly from the summability of  $\{X_j\}_{j=1}^{\infty}$ , so we only need to show the second claim. Direct calculations show that

$$E[1_{\{\pi_n(j) \le m\}} 1_{\{\pi_n(k) \le m\}}] = \begin{cases} \frac{m}{n}, & j \ne k, \\ \frac{m(m-1)}{n(n-1)}, & j \ne k, \end{cases}$$

we hence have

$$E[(II)^{2}] = \sum_{j=N+1}^{n} \frac{m}{n} X_{j}^{2} + \sum_{\{N+1 \le j,k \le n, j \ne k\}} X_{j} X_{k} \frac{m(m-1)}{n(n-1)}.$$

Write  $\frac{m}{n} = \frac{m(m-1)}{n(n-1)} + \frac{m(n-m)}{n(n-1)}$ , it follows from basic algebra that

$$E[(II)^{2}] = \frac{m(n-m)}{n(n-1)} \sum_{j=N+1}^{n} X_{j}^{2} + \frac{m(m-1)}{n(n-1)} (\sum_{j=N+1}^{n} X_{j})^{2},$$

which tends to 0 by the summability and the square summability of the sequence  $\{X_j\}_{j=1}^{\infty}$ . This concludes the proof of (4.9) as well as the theorem.

#### 4.4 Proof of Theorem 3.1

Once the first claim is proved, the second claim follows by elementary Fourier analysis. We now show the first claim. Let  $j_0 = j_0(t; \alpha)$  be the smallest integer such that  $t \cdot j^{-\alpha} < 2$ . It is seen that for large t,

$$j_0 \sim (t/2)^{1/\alpha}$$
. (4.11)

and for any  $j_0 \leq j \leq 2^{1/\alpha} j_0$ ,

$$t \cdot j^{-\alpha} \ge \frac{1}{2}t \cdot j_0^{-\alpha} \ge 1.$$

Note that when  $1 \le x \le 2$ ,  $|\cos(x)| \le 0.6$ , it follows that for sufficiently large t,

$$|\varphi(t)| \le \prod_{j=j_0}^{2^{1/\alpha} j_0} |1 - p + p\cos(tj^{-\alpha})| \le (1 - p + 0.6p)^{(2^{1/\alpha} j_0)},$$

and the claim follows from (4.11).

#### 4.5 Proof of Theorem 3.2

All we need to show is the first claim. With X as in the theorem,

$$\varphi_X(t;p) = \prod_{j=1}^{\infty} \left( 1 - p + \cos\left(\frac{t}{\exp(j^{1/\alpha})}\right) \right).$$

The idea is to count for how many n the corresponding terms  $t \cdot \exp(-n^{1/\alpha})$  are bounded away from the grid  $\{k\pi : k \ge 1\}$  by a distance of  $\pi/4$  (say). In fact, suppose there are N(t) such j. Note that for each such n,

$$|1 - p + p\cos\left(\frac{t}{\exp(j^{1/\alpha})}\right)| \le (1 - p + \frac{1}{\sqrt{2}}p),$$

 $\mathbf{SO}$ 

$$|\varphi(t)| \le \left(1 - p + \frac{1}{\sqrt{2}}p\right)^{N(t)},$$

г	-	-	
L			
L			

and it suffices to show that there are constant  $C_1(\alpha) > 1$  and  $C_2(\alpha) > 0$  such that for sufficiently large t,

$$N(t) \ge C_2(\alpha) (\log(|t|))^{C_1(\alpha)}.$$
(4.12)

Towards this end, let

$$s = s(\alpha) = \frac{\max\{3 - \alpha, 1\}}{2},$$
 (note that  $1/2 \le s < 1$ ).

Introduce the set of indices

$$\Omega(t;s) = \left\{ j : \frac{3\pi}{4} \le \frac{t}{\exp(j^{1/\alpha})} \le \pi \cdot \exp(\log^s(t)) \right\},\$$

and its subsets

$$\Omega_k(t;s) = \left\{ j \in \Omega(t;s) : |\frac{t}{\exp(j^{1/\alpha})} - k\pi| \le \frac{\pi}{4} \right\}, \qquad k = 1, 2, \dots$$

Note that by the definitions,  $\Omega_k(t;s)$  is only non-empty when

$$k \le \pi \cdot \exp(\log^s(t)) + \frac{\pi}{4}.$$
(4.13)

So if we let  $j_0 = j_0(t; s)$  be the largest integer that satisfies (4.13), then for any index j in the set

$$\Omega(t;s) \setminus \bigcup_{k=1}^{k_0} \Omega_k(t;s),$$

the term  $t \cdot \exp(-j^{1/\alpha})$  is bounded away from the grid  $\{k\pi : k \ge 1\}$  by a distance of  $\pi/4$ . Denote the cardinality of a set  $\Omega$  by  $|\Omega|$ , this implies that

$$\left|N_{0}(t)\right| \geq \left|\Omega(t;s) \setminus \bigcup_{k=1}^{k_{0}} \Omega_{k}(t;s)\right|.$$

$$(4.14)$$

We now calculate the cardinality of  $\Omega(t; s)$  and  $\bigcup_{k=1}^{k_0} \Omega_k(t; s)$ . Consider the first one first. By definition,  $j \in \Omega(t)$  if and only if

$$\left(\log\left(\frac{t}{\pi \exp(\log^s(t))}\right)\right)^{\alpha} < j < \left(\log\left(\frac{t}{3\pi/4}\right)\right)^{\alpha}$$

By elementary calculus, for large t,

$$|\Omega(t;s)| \ge \left[ \left( \log\left(\frac{t}{3\pi/4}\right) \right)^{\alpha} - \left( \log\left(\frac{t}{\pi \exp(\log^s(t))}\right) \right)^{\alpha} - 2 \right] \sim \frac{1}{\alpha} \log^{(\alpha - 1 + s)}(t).$$
(4.15)

Consider the second one. For each  $1 \le k \le k_0$ , it follows from similar argument and elementary calculus that

$$|\Omega_k(t;s)| \lesssim \log^{(\alpha-1)}(t) \left(\frac{1}{2\alpha k} + \frac{C}{k^3}\right),$$

where C > 0 is a constant that only depends on  $\alpha$ . Therefore,

$$\left|\cup_{k=1}^{k_0}\Omega_k(t;s)\right| \lesssim \log^{(\alpha-1)}(t) \sum_{k=1}^{k_0} \left(\frac{1}{2\alpha k} + \frac{C}{k^3}\right) \lesssim \frac{1}{2\alpha} \log^{\alpha-1}(t) \cdot \log(k_0).$$

Recalling  $\log(k_0) \sim \log^s(t)$ , it follows that

$$\left|\bigcup_{k=1}^{k_0}\Omega_k(t;s)\right| \lesssim \frac{1}{2\alpha}\log^{\alpha-1+s}(t).$$
(4.16)

Combining (4.15) and (4.16) gives

$$\left|\Omega(t;s) \setminus \bigcup_{k=1}^{k_0} \Omega_k(t;s)\right| \gtrsim \frac{1}{2\alpha} \log^{\alpha - 1 + s}(t).$$
(4.17)

Finally, note that

$$\alpha - 1 + s = \frac{\max\{\alpha + 1, 2\alpha - 1\}}{2} > 1,$$

inserting (4.17) into (4.14) and letting  $C_1(\alpha) = \alpha - 1 + s$  and  $C_2(\alpha) = \frac{1}{4\alpha}$  give (4.12) and conclude the claim.

### 4.6 Proof of Lemma 2.1

Write for short  $a = a_p$ ,  $\pi = \pi_n$ . By elementary combinatorics, for any  $\theta \in \Theta_N(k)$ ,

$$P\{1_{\{\pi_n(j) \le m\}} = \theta_j, \ 1 \le j \le N\} = \frac{\binom{pn}{k} \binom{(1-p)n}{N-k} (n-N)! (N-k)! k!}{n!} \equiv \frac{B_{p,k,N,n}}{A_{p,k,N,n}},$$

with

$$A_{p,k,N,n} = n(n-1)\dots(n-N+1),$$

and

$$B_{p,k,N,n} = (pn)(pn-1)\dots(pn-k+1)((1-p)n)((1-p)n-1)\dots((1-p)n-(N-k)+1).$$

First,

$$(n-N)^N \le A_{p,k,N,n} \le n^N.$$

Note that  $\log(1-x) \ge -x$  for all  $x \in [0, 1/2]$ . It follows that  $(n-N)^N \ge n^N e^{N \log(1-N/n)} \ge e^{-N^2/n} n^N$ , and so

$$e^{-N^2/n}n^N \le A_{p,k,N,n} \le n^N.$$
 (4.18)

Second,

$$(pn-N)^{k}((1-p)n-N)^{N-k} \le B_{p,k,N,n} \le (pn)^{k}((1-p)n)^{N-k}.$$
(4.19)

Note that  $pa \ge 1$  and  $(1-p)a \ge 1$ , so

$$(pn-N)^{k}((1-p)n-N)^{N-k} \ge (pn-paN)^{k}((1-p)n-a(1-p)N)^{N-k} = p^{k}(1-p)^{N-k}(n-aN)^{N}.$$
(4.20)

By a similar argument,  $(n - aN)^N \ge e^{-aN^2/n}n^N$ , combining this with (4.19)-(4.20) gives

$$e^{-aN^2/n}p^k(1-p)^{N-k}n^N \le B_{p,k,N,n} \le n^N.$$
(4.21)

Combining (4.18) and (4.21) gives the lemma.

## References

- BERTIN, M.J., Decomps-Guilloux, A., Grandet-Hugot, M., Pathiaux-Delefosse, M. and J.P. Schreiber, J. P. (1992). Pisot and Salem Numbers. Birkhäuser.
- [2] ERDÖS, Paul. (1939). On a family of symmetric Bernoulli convolutions. Amer. J. Math, 61 974–975.
- [3] ERDÖS, Paul. (1940). On the smoothness properties of a family of Bernoulli convolutions. *Amer. J. Math* 62 180–186.
- [4] KADANE, J. B. and O'HAGAN, A. (1995). Using finitely additive probability: uniform distribution on the natural numbers, J. Amer. Statist. Assoc. 90 626–631.
- [5] JACKA, S. and WARREN, J. (2007). Random orderings of the integers and card shuffling Stoch. Proc. and Appl. 117 708–719.
- [6] JACKA, S. and WARREN, J. (2007). On shuffling an infinite pack of cards. Preprint, University of Warwick.
- [7] PERES, Y. and SOLOMYAK, B. (1998). Self-similar measures and intersection of cantor sets. Tran. Amer. Math. Soc. 350 (10) 4065–4087.
- [8] SCHIROKAUER, O. and KADANE, J. B. (2007). Uniform distributions on the natural numbers J. Theor. Probab., 20, 429–441.
- [9] SOLOMYAK, Boris. (1995). On the random series  $\Sigma \pm \lambda^n$ , Ann. Math. 142 (3), 611-625.