# The Bohnenblust–Spitzer algorithm and its applications

J. Michael Steele

*Department of Statistics, Wharton School, University of Pennsylvania, Steinberg Hall-Dietrich Hall 3000, Philadelphia, PA 19104, USA*

## Abstract

The familiar bijections between the representations of permutations as words and as products of cycles have a natural class of "data driven" extensions that permit us to use purely combinatorial means to obtain precise probabilistic information about the geometry of random walks. In particular, we show that the algorithmic bijection of Bohnenblust and Spitzer can be used to obtain means, variances, and concentration inequalities for several random variables associated with a random walk including the number of vertices and length of the convex minorant, concave majorant, and convex hull. © 2002 Elsevier Science B.V. All rights reserved.

*MSC:* primary: 60D05; secondary: 60C05; 68U05; 62G99

*Keywords:* Spitzer's combinatorial lemma; Random walk; Convex hull; Permutations; Cycle decomposition; Cycle lemma

## 1. Introduction

The most direct way to represent a permutation $\sigma$ of $\{1,2,\ldots,n\}$ is by the permutation word $(\sigma_1,\sigma_2,\ldots,\sigma_n)$ were $\sigma_k$ is simply the image of $k$ under the mapping $\sigma : [n] \rightarrow [n]$. Nevertheless, there are many other ways to represent a permutation, and there are also many different bijections between the structures that provide these representations. The purpose of this article is to show how the geometry of a random walk can be made tractable by exploiting a special bijection between the set of word representations and the set of representations as products of cycles. The most novel feature of this bijection is that it is "data driven" in a sense that will be soon be made explicit. The benefit of this class of bijections is that it permits us to establish identities that tell us about more than just the structure of the set of permutations. In particular, we will see how these data driven bijections can be used to obtain geometrical information about the convex minorant and concave majorant that make up the bottom and top of the convex hull of the graph of a random walk.

At the heart of our analysis is an algorithm that appears implicitly in the proof due to Bohnenblust of a combinatorial lemma due to Spitzer. Spitzer's combinatorial lemma has been widely used by probabilists, but probabilists seem to have ignored the fact that the algorithm actually gives us considerably more information than the particular lemma that it was designed to prove.

Before developing the Bohnenblust–Spitzer algorithm, we need to collect a few facts about the cycle structure of a random permutation. The required facts are well-known to experts, but the approach to these facts via record times is so simple and so powerful that it deserves to be known by everyone. We take up the Bohnenblust–Spitzer algorithm and suggest that one simple picture goes a long way toward making "everything obvious".

Next, we turn our attention to the applications of the Bohnenblust–Spitzer algorithm, and, in short order, we will collect a harvest of concrete results on the geometry of random walk. Finally, we point out some limitations of the Bohnenblust–Spitzer algorithm and mention several attractive open problems.

## 2. Standard cyclical representations and random permutations

Any permutation $\tau$ has a representation as a product of cyclic permutations

$$\tau = (a_1 a_2 \ldots a_{i_1})(a_{i_1+1} a_{i_1+2} \ldots a_{i_2}) \cdots (a_{i_{v-1}+1} a_{i_{v-1}+2} \ldots a_{i_v}),$$

but this representation is far from unique unless we impose some further restrictions. First, a cyclical permutation of any block does not change the represented permutation, so we can always take the first element of each block to be the largest element of the block. Also, the order in which the blocks are written does not change the permutation, so we can also take the blocks in that order that puts the leading elements into an increasing sequence. Cyclic representation that satisfy these two conditions are said to be *standard*, and the notion of a standard representation as a product of cyclical permutation leads us very quickly to a detailed understanding of the number $v = v(\tau)$ of cycles in a random permutation $\tau$.

The key combinatorial observation is that if we are given the word representation of a random permutation $\sigma$, then there is a simple way to associate $\sigma$ with the standard product cycle representation of another permutation $\tau$ (which typically differs from $\sigma$). Here is how the recipe goes. If $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, we say that $\sigma_k$ is a *record* provided that $\sigma_k > \max\{\sigma_j : j < k\}$, and, if we break the word $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ into blocks so that each new record starts a new block, then it is immediate that the resulting set of ordered blocks fulfills all of the requirements of a standard cyclical representation.

Next, there is a useful probabilistic observation. If we let $\xi_k$ be equal to one or zero accordingly as $\sigma_k$ is or is not a record of the random permutation word $\sigma$, then by direct calculation, one can show that the random variables $\{\xi_k : 1 \leqslant k \leqslant n\}$ are independent and $P(\xi_k = 1) = 1/k$. To see why the last equation holds, one just needs to note that of the $k!$ possible orderings of $\{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ in the permutation words that begin with such elements, there are exactly $(k-1)!$ for which the largest element occurs at the $k$th place. An analogous (but slightly longer) argument also establishes the independence of the set of random variables $\{\xi_k : 1 \leqslant k \leqslant n\}$.

We therefore find that the number of $v_n = v_n(\tau)$ of cycles in a random permutation $\tau$ of $[n]$ has the same distribution as the sum of $n$ independent random variables $\xi_1 + \xi_2 + \cdots + \xi_n$, and such

sums are very well understood. In particular, the mean and variance of $v_n$ have simple expressions in terms of the harmonic numbers $H_n$,

$$E[v_n] = \sum_{k=1}^{n} \frac{1}{k} \stackrel{\text{def}}{=} H_n \quad \text{and} \quad \text{Var}[v_n] = \sum_{k=1}^{n} \frac{1}{k}\left(1 - \frac{1}{k}\right) = H_n + O(1), \tag{1}$$

and the Feller–Lindeberg Central Limit Theorem can be applied to obtain the asymptotic distribution of $v_n$,

$$\lim_{n \to \infty} P\left((v_n - H_n)/\sqrt{H_n} \leqslant t\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} \, dx. \tag{2}$$

Finally, the characterization of $v_n$ as a sum of independent random variables can be used to obtain concentration results (or large deviation inequalities) for $v_n$. These will be given a bit later when they are needed.

## 3. Bohnenblust–Spitzer algorithm

Now we are ready to develop the Bohnenblust–Spitzer algorithm and to see how it compares with the basic record-time bijection. The data that drives the algorithm is simply a sequence of $n$ real numbers $x_1, x_2, \ldots, x_n$ that we assume to be linearly independent over $\mathbb{Z}$; that is, we assume that whenever we have $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$ for some $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{Z}^n$ then we must have $\mathbf{a} = 0$.

Here is the algorithm. We take any permutation $\sigma$ with word representation $(\sigma_1, \sigma_2, \ldots, \sigma_n)$, and we first form the partial sums $s_k(\sigma)$ given by

$$s_k(\sigma) = x_{\sigma_1} + x_{\sigma_2} + \cdots + x_{\sigma_k} \quad \text{for } 1 \leqslant k \leqslant n.$$

Next, we consider the set of $n$ points $p_k = (k, s_k(\sigma)) \in \mathbb{R}^2$, and we compute the least concave majorant $M(\sigma)$ of these points. Since these points are automatically given to us in order of increasing $x$-coordinate, the convex hull algorithm of Graham and Yao [10] tells us that one can compute $M(\sigma)$ in linear time. At this point one should consider Fig. 1 which gives a simple example of such a majorant.

The last step of the algorithm is to break the word representation for $\sigma$ into blocks that correspond to the faces of $M(\sigma)$. Specifically, we let $i_1, i_2, \ldots, i_{v-1}$ denote the set of indices at which $M(\sigma)$ makes a turn, and, to close out the list, we set $i_v = n$. Finally, we define the permutation $v$ to be the product of cycles given by

$$(\sigma_1, \sigma_2, \ldots, \sigma_{i_1})(\sigma_{i_1+1}, \sigma_{i_1+2}, \ldots, \sigma_{i_2}) \cdots (\sigma_{i_{v-1}+1}, \sigma_{i_{v-1}+2}, \ldots, \sigma_{i_v}).$$

This representation for $\tau$ completes the algorithm for the mapping $\sigma \mapsto \tau$, but there are two important geometric features of this representation that one should note.

First, by the concavity of the majorant, the slopes of the successive faces form a decreasing sequence, and in terms of the sequence $\{x_k\}$ this tells us that

$$\frac{1}{i_1} \sum_{j=1}^{i_1} x_{\sigma_j} > \frac{1}{i_2 - i_1} \sum_{j=i_1+1}^{i_2} x_{\sigma_j} > \cdots > \frac{1}{i_v - i_{v-1}} \sum_{j=i_{v-1}+1}^{i_v} x_{\sigma_j}.$$
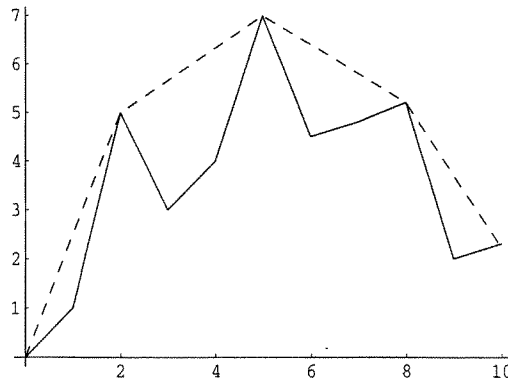
Fig. 1. The concave majorant guides us to the cycle cuts.

Second, from the definition of the concave majorant, we see that if we set $i_0 = 1$, then for all $0 \leqslant k < v$ and all $i_k < j < i_{k+1}$ the points

$$(j, s_j(\sigma)) = (j, x_{\sigma_1} + x_{\sigma_2} + \cdots + x_{\sigma_j})$$

all lie below the line from the point $(i_k, s_{i_k}(\sigma))$ to the point $(i_{k+1}, s_{i_{k+1}}(\sigma))$. We now just need to check that these two properties are enough to guarantee that the mapping $\sigma \mapsto \tau$ is an honest bijection.

## 4. Why does it work?

We need to show that given any permutation $\tau$ and any representation of $\tau$ as a product of cycles, there is a unique $\sigma$ whose word representation is mapped to $\tau$ by the Bohnenblust–Spitzer algorithm. An important element of this inversion process is an elementary "cycle lemma". The surveys of Dershowitz and Zaks [7] and Snevily and West [17] show that such lemmas come in many flavors and have many different combinatorial consequences. Here we only need a simple geometric version.

**Lemma 1** (Cycle lemma). *If $\{x_1, x_2, \ldots, x_n\}$ is a sequence of $n$ real numbers that are linearly independent over $\mathbb{Z}$, then there is a unique $k$ such that the sequence $y_1, y_2, \ldots, y_n$ given by*

$$y_1 = x_k, \quad y_2 = x_{k+1}, \quad \ldots, \quad y_{n-k} = x_n, \quad y_{n-k+1} = x_1, \quad \ldots, \quad y_n = x_{k-1}$$

*has the property that for each $1 \leqslant j \leqslant n$ the point $(j, y_1 + y_2 + \cdots + y_j) \in \mathbb{R}^2$ lies above the line from $(0, 0)$ to $(n, y_1 + y_2 + \cdots + y_n)$.*

The proof of this lemma is almost immediate if we make the right normalization and if we draw the right picture. For the normalization, we simply let $s$ denote the sum of the $\{x_k\}$ and set $z_k = x_k - s/n$ for $1 \leqslant k \leqslant n$. The sum of the $z_k$ is then equal to zero, and a typical graph of the partial sums $z_1 + z_2 + \cdots + z_k$ for $0 \leqslant k \leqslant n$ is given in Fig. 2

Now, by the linear independence of the $\{x_j\}$, there is a unique minimum in the set of partial sums of the $\{z_j\}$. Also, from the figure it is evident that if we take $k$ to be the $x$-coordinate of that minimum, then the incremental partial sums $z_{k+1}, z_{k+1} + z_{k+2}, \ldots$ on up to $z_{k+1} + z_{k+2} + \cdots +$
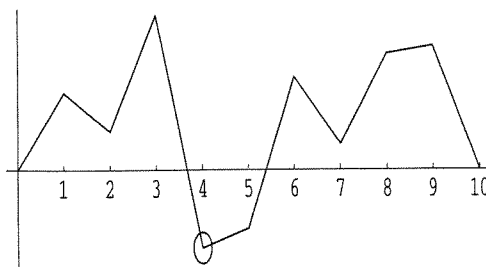
Fig. 2. The minimum partial sum of the $\{z_j\}$ tells us where to start our cycle to get a path that stays positive.

$z_n + z_1 + z_2 + \cdots + z_{k-1}$ are all strictly positive. Finally, if we undo our normalization, we see that the positivity of these partial sums and the uniqueness of $k$ are precisely the assertions of the lemma.

Now suppose we have a permutation $\tau$, and a representation of $\tau$ as a product of cyclic permutations,

$$\tau = (a_1 a_2 \ldots a_{i_1})(a_{i_1+1} a_{i_1+2} \ldots a_{i_2}) \cdots (a_{i_{v-1}+1} a_{i_{v-1}+2} \ldots a_{i_v}).$$

Without changing the permutation $\tau$ we can rearrange the blocks of the representation so that the average value of the $\{x_i\}$ over the blocks is in decreasing order; in other words, we can assume without loss of generality that we have

$$\frac{1}{i_1} \sum_{j=1}^{i_1} x_{a_j} > \frac{1}{i_2 - i_1} \sum_{j=i_1+1}^{i_2} x_{a_j} > \cdots > \frac{1}{i_v - i_{v-1}} \sum_{j=i_{v-1}+1}^{i_v} x_{a_j}. \tag{3}$$

Next, by the cycle lemma (or rather its "below the line" version), we note that by a rotation within each of the blocks of $\tau$, we can assume that for all $0 \leqslant k < v$ and all $i_k + 1 \leqslant s < i_{k+1}$ then we have that

$$\left( s - i_k, \sum_{j=i_k+1}^{s} x_{a_j} \right) \text{ lies below the line from } (0,0) \text{ to } \left( i_{k+1} - i_k, \sum_{j=i_k+1}^{i_{k+1}} x_{a_j} \right). \tag{4}$$

The linear independence of the $\{x_i\}$ and the uniqueness assertion of the cycle lemma are all that one needs in order to check that the product cycle representation for $\tau$ is uniquely specified when the block sums decrease according to the inequalities (3) and when the majorization property (4) holds for each of the blocks. Finally, the uniqueness of this representation for $\tau$ makes it immediate that the permutation $\sigma$ with word representation $(a_1 a_2 \ldots a_n)$ is the unique permutation that is mapped to $\tau$ by the Bohnenblust–Spitzer algorithm.

These observations tell us that Bohnenblust–Spitzer algorithm does indeed provide a bijection of the set of permutations, but this just begins the story. The real task is to see how this bijection fits together with data set $\{x_i\}$ that drives it.

## 5. Interpreting the picture

The power of the Bohnenblust–Spitzer bijection comes from the fact that many of the geometrical features of Fig. 1 have two interpretations. Such features can be expressed equally well in terms of the word representation for $\sigma$, or in terms of the product cycle representation of $\tau$. For example, if we write $C \in \tau$ to indicate that $C$ is a cycle of $\tau$, then the length of the concave majorant of the point set $B_\sigma = \{(0,0),(1,s_1(\sigma)),\ldots,(n,s_n(\sigma))\}$ can be written as

$$\sum_{C \in \tau} \left( |C|^2 + \left( \sum_{i \in C} x_i \right)^2 \right)^{1/2} \tag{5}$$

and, in fact, almost all of the geometric quantities in Fig. 1 can be written in the general form

$$Z_f(\tau) = \sum_{C \in \tau} f\left( |C|, \sum_{i \in C} x_i \right). \tag{6}$$

Specifically, if we take $f(k,y) \equiv 1$, then $Z_f(\tau)$ is equal to the number of faces in the concave majorant of $B_\sigma$, and if we take $f(k,y) = y/k$ then $Z_f(\tau)$ is equal to the sum of their slopes.

One of the most interesting choices of $f$—and the only one considered by Spitzer [19]—is simply $f(k,y) = \max(0,y)$. For this choice, the concavity of the majorant implies that $Z_f(\tau)$ simply equals the sum of the rises in the majorant until it reaches its maximum. In other words, we have

$$\max(0,s_1(\sigma),\ldots,s_n(\sigma)) = \sum_{C \in \tau} \left( \sum_{i \in C} x_i \right)^+, \tag{7}$$

where $y^+$ is used as shorthand for $\max(0,y)$. This marvelous identity of Spitzer [19] has many important consequences, and it has inspired much subsequent work. Nevertheless, the preceding discussion should make clear that this identity contains just part of the geometric information that one can draw from the Bohnenblust–Spitzer bijection.

## 6. Computing expectations

To see how one can exploit formulas like (7) and its geometric cousin (5), it is a good exercise to show that for any sequence of independent identically distributed random variables $\{X_i : 1 \leq i \leq n\}$, the partial sums $S_k = X_1 + X_2 + \cdots + X_k$ satisfy the identity,

$$E[\max(0,S_1,S_2,\ldots,S_n)] = \sum_{k=1}^{n} \frac{1}{k} E[S_k^+]. \tag{8}$$

The critical trick that brings the Bohnenblust–Spitzer bijection into play is simply prerandomization of the order of the summands, an idea that is familiar in the theory of algorithms as a way to guarantee the expected running time of QuickSort.

There is also one small technical issue needed in the proof of the expectation identity (8). Since the Bohnenblust–Spitzer bijection requires that the $\{x_i\}$ be independent over $\mathbb{Z}$, we will first prove

the identity under the assumption that the random variables $\{X_i\}$ have a density. This assumption is more than enough to guarantee that with probability one, the set of real numbers $\{X_i(\omega)\}$ will form an independent set over $\mathbb{Z}$.

Now, if we let $S_n$ denote the set of all permutations of $[n]$, then the fact that the random vectors $(X_{\sigma_1}, X_{\sigma_2}, \ldots, X_{\sigma_n})$ have the same distribution for all $\sigma \in S_n$ tells us that we can write $E[\max(0, S_1, S_2, \ldots, S_n)]$ as

$$\frac{1}{n!} E\left[ \sum_{\sigma \in S_n} \max(0, X_{\sigma_1}, X_{\sigma_1} + X_{\sigma_2}, \ldots, X_{\sigma_1} + X_{\sigma_2} + \cdots + X_{\sigma_n}) \right],$$

so the Bohnenblust–Spitzer bijection and the combinatorial identity (7) tell us that the last expectation is also equal to

$$\frac{1}{n!} E\left[ \sum_{\tau \in S_n} \sum_{k=1}^{n} \sum_{C \in \tau, |C|=k} \left( \sum_{i \in C} X_i \right)^+ \right].$$

Since there are exactly $(n-k)!(k-1)!\binom{n}{k}$ permutations that contain the cycle $C$ when $|C| = k$, the sum over the permutations may be further simplified to

$$\frac{1}{n!} \sum_{k=1}^{n} (n-k)!(k-1)! \binom{n}{k} E\left[ \left( \sum_{i=1}^{k} X_i \right)^+ \right] = \sum_{k=1}^{n} \frac{1}{k} E[S_k^+].$$

This identity completes the proof of the target formula (8) in the case when the $\{X_i\}$ have a density. Its validity for independent random variables $\{X_i\}$ with an arbitrary distribution then follows by standard approximation arguments.

Variations on this theme can be created almost at will, yet we content ourselves with just two further examples, the length $L_n$ of the concave majorant and the sum $A_n$ of the slopes of its faces. Although the function $f(k, y) = \sqrt{k^2 + y^2}$ that represents the length functional in the form (6) depends on both $y$ and $k$, this function is no harder to handle than the univariate function $f(y) = y^+$ that corresponds to $\max(0, S_1, S_2, \ldots, S_n)$. The additional parameter does not interfere with our earlier computation, and one finds without difficulty that

$$E[L_n] = \sum_{k=1}^{n} \frac{1}{k} E\left[ \sqrt{k^2 + S_k^2} \right]. \tag{9}$$

Now, without some additional distributional assumptions, this expectation cannot be significantly simplified, but, if we assume that the $\{X_i\}$ satisfy $E[|X_i|] < \infty$ and $E[X_i] = \mu$, then we can get some useful asymptotic information. If we take $k$ inside the radical and use the strong law of large numbers and the uniform integrability of $\{S_k/k\}$, we easily find

$$E[L_n] \sim n\sqrt{1 + \mu^2} \quad \text{as } n \to \infty.$$

This interesting geometric formula tells us that the expected length of the concave majorant grows exactly like the length of the line from $(0, 0)$ to the point $(0, E[S_n]) = (n, n\mu)$.

The slope functional also has an informative interpretation in the limit. Here we need to assume that the $\{X_i\}$ have a density as well as a first moment $\mu$, but, under these conditions, the general representation (6) and the now familiar expectation calculation tell us that we also have

$$E[A_n] = \sum_{k=1}^{n} \frac{1}{k} E[S_k/k] = \mu H_n \sim \mu \log n. \tag{10}$$

In the next section we will find that the expected number of faces of the concave majorant exactly equals the harmonic number $H_n$, so, in an average sense, the slope formula (10) tells us that the average slope of the concave majorant is $\mu$. One can view this result as a very weak form of the law of large numbers for the slope sizes.

## 7. Looking harder at the faces

The methods of the last section can be used to calculate the expectation of the number $F_n$ of faces of the concave majorant, and the simple choice of $f(k,y) \equiv 1$ in the representation (6) quickly brings us to the nice *distribution free* formula

$$E[F_n] = \sum_{k=1}^{n} \frac{1}{k} \equiv H_n \sim \log n. \tag{11}$$

Nevertheless, in this case there is an alternative derivation of the formula (11) that is much more informative. In particular, this alternative approach will give us the exact variance and asymptotic distribution of $F_n$; results that are far out of reach for $L_n$ and $A_n$.

The key observation is that one can show that for independent random variables $\{X_k\}$ with a density, the number of faces $F_n$ of the concave majorant of the random walk will have exactly the same distribution as $R_n$, the number of record times of the $\{X_k\}$. Since we saw in Section 2 that $R_n$ is equal in distribution to a sum of independent Bernoulli random variables, the link between $R_n$ and $F_n$ gives us a powerful tool for the analysis of $F_n$, though we will find that there are some instructive limitations to the inferences one can draw from this link.

If we let $F(x_{\sigma_1}, x_{\sigma_2}, \ldots, x_{\sigma_n})$ denote the number of faces of the concave majorant of the set of points $B_\sigma = \{(0,0), (1, s_1(\sigma)), \ldots, (n, s_n(\sigma))\}$, where $s_k(\sigma)$ denotes the partial sum $x_{\sigma_1} + x_{\sigma_2} + \cdots + x_{\sigma_k}$ for $1 \leqslant k \leqslant n$ and where the real numbers $\{x_k\}$ are independent over $\mathbb{Z}$, then the Bohnenblust–Spitzer bijection tells us that for all $\sigma \in S_n$ we have

$$F(x_{\sigma_1}, x_{\sigma_2}, \ldots, x_{\sigma_n}) = \sum_{C \in \tau} 1. \tag{12}$$

This formula and the one-to-one correspondence between the $\sigma$'s and the $\tau$'s combine to tell us that if $\sigma$ is chosen at random then $F(x_{\sigma_1}, x_{\sigma_2}, \ldots, x_{\sigma_n})$ is equal to the number of cycles in the random permutation $\tau$, and we already know from Section 2 that the number of cycles in a random permutation is equal in distribution to $T_n$ where

$$T_n = \sum_{k=1}^{n} \xi_k, \quad P(\xi_k = 1) = \frac{1}{k}, \quad P(\xi_k = 0) = 1 - \frac{1}{k} \tag{13}$$

and the $\{\xi_k\colon 1 \leqslant k \leqslant n\}$ are independent. We would like to deduce from these facts that $F_n = F(X_1, X_2, \ldots, X_n)$ is equal in distribution to $T_n$, and an honest derivation of this intuitive fact seems easiest with the use of characteristic functions.

From formula (12) and the known distribution of the number of cycles in a random permutation, we have for all $\theta \in \mathbb{R}$ that

$$\frac{1}{n!} \sum_{\sigma \in S_n} \exp(i\theta F(x_{\sigma_1}, x_{\sigma_2}, \ldots, x_{\sigma_n})) = \frac{1}{n!} \sum_{\tau \in S_n} \exp\left(i\theta \sum_{C \in \tau} 1\right) = E[\exp(i\theta T_n)]$$

and, in terms of the random variables $\{X_k\colon 1 \leqslant k \leqslant n\}$, this tells us that with probability one we have

$$\frac{1}{n!} \sum_{\sigma \in S_n} \exp(i\theta F(X_{\sigma_1}(\omega), X_{\sigma_2}(\omega), \ldots, X_{\sigma_n}(\omega))) = E[\exp(i\theta T_n)]. \tag{14}$$

When we take the expectation on both sides of the pointwise identity (14) we find that $E[\exp(i\theta F_n)] = E[\exp(i\theta T_n)]$, so $F_n$ and $T_n$ are indeed equal in distribution.

For each $n$, $1 \leqslant n < \infty$, the number of cycles $v_n$, the number of records $R_n$, the number of faces $F_n$, and the Bernoulli sum $T_n$ all have the same distribution, so the mean–variance formulas (1) and central limit theorem (2) for $v_n$ may seem to provide quick answers for all of the natural questions one might ask about the sequences $\{F_n\}$ and $\{R_n\}$. Nevertheless, a curious gap emerges when we consider the strong law of large numbers.

## 8. Of processes and strong laws

The relationship between the number of records $R_n$ and the Bernoulli sums $T_n$ turns out to be much stronger than the relationship we have found between $F_n$ and $T_n$. In fact, when one looks hard at the argument of Section 2, or simply consults the original article of Rényi [15], then one finds that $\{R_n\colon 1 \leqslant n < \infty\}$ and $\{T_n\colon 1 \leqslant n < \infty\}$ are *equivalent as processes*; that is, all of the joint distributions of these random variables are equal. Among other things, this equivalence as processes tell us that the strong law of large numbers for $R_n$,

$$P\left(\lim_{n \to \infty} R_n / \log n = 1\right) = 1,$$

will follow from the corresponding law for the Bernoulli sum $T_n$. As Rényi [15] observes, the strong law for the process $\{T_n\}$ is an immediate consequence of Kolmogorov's Three Series Theorem, so we do indeed have a strong law for the process $\{R_n\}$.

In the case of the number of faces $F_n$, we know that for each $n$ the random variable $F_n$ has the same distribution as $T_n$, but the processes $\{F_n\colon 1 \leqslant n < \infty\}$ and $\{T_n\colon 1 \leqslant n < \infty\}$ certainly are *not* equivalent. To see this, we just need to note that $T_1 \leqslant T_2 \leqslant \cdots \leqslant T_n$ with probability one, but no such monotonicity relationship holds for the sequence $\{F_n\}$.

One consequence of the nonequivalence of $\{T_n\}$ and $\{F_n\}$ is that the strong law for $\{T_n\}$ does not imply a corresponding strong law for $\{F_n\}$. Nevertheless, one might hope to prove such a strong law by first obtaining good bounds on the tail probabilities $P(T_n \geqslant t)$ and then exploiting the marginal equivalence $P(T_n \geqslant t) = P(F_n \geqslant t)$. We pursue this plan in the next section.

## 9. Tail probabilities for the number of faces

The classic inequality of Bennett [4, Eq. (8b)] tells us that if the independent random variables $\{Y_k: 1 \leqslant k \leqslant n\}$ satisfy

$$E[Y_k] = 0, \qquad \sup_k \|Y_k\|_\infty \leqslant a \quad \text{and} \quad \sum_{k=1}^{n} E[Y_k^2] \leqslant b^2,$$

then we have the tail bound

$$P\left(\sum_{k=1}^{n} Y_k \geqslant t\right) \leqslant \exp\left\{\frac{t}{a} - \left[\frac{t}{a} + \frac{b^2}{a^2}\right] \log\left(1 + \frac{at}{b^2}\right)\right\} \quad \text{for all } t \geqslant 0. \tag{15}$$

If we let $Y_k = \xi_k - 1/k$ where the $\xi_k$ are the summands of $T_n$, then both $Y_k$ and $-Y_k$ satisfy the conditions of Bennett's inequality with

$$a = 1 \quad \text{and} \quad b^2 = \sum_{k=1}^{n} \frac{1}{k} \equiv H_n,$$

so the elementary inequality $\log(1 + x) \geqslant x - x^2/2$ for $0 \leqslant x \leqslant 1$ tells us we can apply Bennett's inequality to $T_n$ and $-T_n$ to deduce from $F_n \overset{d}{=} T_n$ that $F_n$ satisfies

$$P(|F_n - H_n| \geqslant \varepsilon H_n) \leqslant 2 \exp(-\varepsilon^2 H_n/4) \quad \text{for all } 0 \leqslant \varepsilon \leqslant \tfrac{1}{2}. \tag{16}$$

This inequality offers a useful complement to the central limit theorem for $\{F_n\}$, and it provides us with a quantitative version of the weak law of large numbers for $\{F_n\}$. Nevertheless, the estimate (16) still falls short of the type of tail bound that might be combined with the Borel–Cantelli lemma to get a strong law for the ratios $\{F_n/\log n\}$.

To be sure, if we set $N(s) = \lfloor s^{8/\varepsilon^2} \rfloor$ for $s = 1, 2, \ldots$, then the inequality (16) and the Borel–Cantelli lemma imply that

$$\lim_{s \to \infty} \frac{1}{H_{N(s)}} F_{N(s)} = 1 \quad \text{a.s.} \tag{17}$$

and, if the sequence $\{F_n\}$ were monotone like $\{T_n\}$, then the limit (17) would imply

$$\lim_{n \to \infty} \frac{1}{H_n} F_n = 1 \quad \text{a.s.} \tag{18}$$

Unfortunately, as we noted earlier, the sequence $\{F_n\}$ is not monotone increasing, and the issue of a strong law for the ratios $\{F_n/\log n\}$ remains open.

In fact, there is even room to doubt if such a law is true, despite the validity of the law for the close cousins $\{T_n\}$ and $\{R_n\}$. When the increments have zero mean, so $E[X_k] = 0$ for $= 1, 2, \ldots$, one source of doubt comes the almost sure recurrence of the random walk $\{S_n\}$, the infinite expected time between the recurrence times, and the curious behavior explained by the arcsin laws. Also, we will find in the next section that even when $E[X_k] \neq 0$ the process $\{F_n\}$ is likely to have recurring instances of large percentagewise decrease. Therefore, for the moment at least, the heuristic arguments for and against the strong law for $\{F_n/\log n\}$ are well balanced.

## 10. Face strides and the cycle structure

The last several sections provide answers to the most direct geometrical questions concerning the geometry of the concave majorant, but there is a further geometric quantity introduced by Suidan [21] where the Bohnenblust–Spitzer bijection can prove very useful. For each face of the concave majorant, we consider the length of the projection of that face on the $x$-axis, and we then consider these lengths in decreasing order

$$\lambda_1(n) \geqslant \lambda_2(n) \geqslant \cdots \geqslant \lambda_k(n) \geqslant \cdots .$$

To avoid conflicts with our standing terms, we will call the $\{\lambda_k(n)\}$ the *strides* of the concave majorant, and, in particular, $\lambda_1(n)$ is the length of the longest stride. For a random walk for with increments $\{X_k: 1 \leqslant k \leqslant n\}$ that have a density, the Bohnenblust–Spitzer algorithm then tells us that set of strides $\{\lambda_k(n)\}$ is exactly equal in distribution of the set of cycle lengths of a random permutation.

From the work of Lloyd and Shepp [16], Arratia and Tavare [1], and others, the probability theory of such cycle lengths is extremely well understood. To give one inference from among many that can be read off from known results, we just note that Eq. (14) of Lloyd and Shepp [16] gives us

$$\lim_{n\to\infty} \frac{1}{n} E[\lambda_1(n)] = \int_0^\infty \exp\left[-x - \int_x^\infty \frac{e^{-y}}{y}\,dy\right]\,dx = 0.6243\ldots,$$

so, in the typical case, one single stride covers more than half of the observations. One consequence of the existence of such dominating strides is that the periodic emergence of a new one seems likely to cause a major decrease in $F_n$. In turn, this fact offers room for doubting the validity of a strong law for $\{F_n/\log n\}$.

## 11. The convex hull

We are now ready to consider what is surely one of the most natural geometric features of the graph of a random walk—the convex hull. If $\{X_k: 1 \leqslant k < \infty\}$ is any sequence of independent absolutely continuous random variables, we now let $F_n^{\mathrm{hull}}$ denote the number of faces of the convex hull of the random point set $B = \{(0,0),(1,S_1)),\ldots,(n,S_n)\}$ determined by the random walk $S_k = X_1 + X_2 + \cdots + X_k$. To complement this notation we can also let $F_n^{\mathrm{top}}$ and $F_n^{\mathrm{bottom}}$, respectively, denote the number of faces of the concave majorant and convex minorant of $B$. Since $F_n^{\mathrm{top}}$ is just the variable that we denoted earlier simply by $F_n$ and since the distribution of $F_n$ did not depend on the distribution of the $\{X_k\}$, we now see that $F_n^{\mathrm{top}}$ and $F_n^{\mathrm{bottom}}$ are equal in distribution. Oddly enough, it seems hard to judge if this fact should be regarded as surprising or as obvious; there is support for each perspective.

In any case, we have the trivial identity

$$F_n^{\mathrm{hull}} = F_n^{\mathrm{top}} + F_n^{\mathrm{bottom}} \tag{19}$$

and, as a consequence, we find the nice expectation formula

$$E[F_n^{\mathrm{hull}}] = 2\sum_{k=1}^n \frac{1}{k}. \tag{20}$$

As it happens, this formula is a special case of a more general result of Baxter [3], and Baxter's proof even turns out to be simpler than the derivation we have given here for the distribution of $F_n$. Nevertheless, we will shortly see that the present methods yield many results that seem to escape Baxter's method.

Baxter's key observation was that Lemma 1 (the cycle lemma) may be generalized to genuinely two-dimensional random walks without any real change in the proof. Again, the trick is to start one's cycle at the lowest point of the path. By building on this observation, Baxter was able to prove that if $\{Z_k: 1 \leqslant k \leqslant n\}$ is an independent sequence of random vectors with values in $\mathbb{R}^2$, then the expected value of the number of faces of the convex hull of the partial sums $Z_1 + Z_2 + \cdots + Z_k$, $1 \leqslant k \leqslant n$ is exactly equal to $2 \sum_{k=1}^{n} 1/k$ provided that with probability one the set $\{Z_k(\omega): 1 \leqslant k \leqslant n\}$ is linearly independent over $\mathbb{Z}$.

The reason for our somewhat labored statement of this result is that we want to stress that Baxter's theorem applies to the vectors $Z_k = (1, X_k)$ where the $\{X_k\}$ are independent random variables with a density. It is for this choice that Baxter's expectation gives us the formula (20) for the convex hull of the graph of a one-dimensional random walk. In this important case, the $\{Z_k\}$ are not absolutely continuous, but they are still almost surely linearly independent over $\mathbb{Z}$.

Baxter's method is well-focussed for the computation of expectations, and it can also be used to get information on the expected length of the convex hull, a problem first studied by Spitzer and Widom [20]. These authors based their proof on a geometric formula going back to Cauchy and on a formula given by Kac [12] whose proof Kac attributes to Dyson. Strangely enough, neither the method of Baxter nor that of Spitzer and Widom seems to be able to get information about the number of faces or length of the concave majorant. The ability to look at the convex hull problem "one half at a time" seems to one of the ways that the Bohnenblust–Spitzer bijection gains power over the more direct applications of the cycle lemma.

A second drawback of Baxter's method is that it does not offer any information about the variances or any other measures of concentration, although such information sometimes can be obtained by more direct probabilistic means. For example, Snyder and Steele [18] used the representation of Spitzer and Widom [20] together with martingale methods to obtain variance estimates and concentration inequalities for the length of the hull of the two-dimensional random walk.

One of the interesting features of the face count functional $F_n^{\text{hull}}$ is that it resists any such attack by the martingale bounded difference method. The problem is that if one changes one of the $\{X_i\}$ there is no control over the change that is made in $F_n^{\text{hull}} = F_n^{\text{hull}}(X_1, X_2, \ldots, X_n)$. Nevertheless, our understanding of $F_n^{\text{top}}$ and $F_n^{\text{bottom}}$ can be used to get variance bounds and concentration inequalities.

Since we have no information at present about the correlation between $F_n^{\text{top}}$ and $F_n^{\text{bottom}}$, the best that one can deduce from the representation (19) is simply

$$\text{Var}[F_n^{\text{hull}}] \leqslant 4 \sum_{k=1}^{n} \frac{1}{k}\left(1 - \frac{1}{k}\right) \leqslant 4H_n.$$

We might have hoped to obtain an exact formula for $\text{Var}[F_n^{\text{hull}}]$, but there are still reasons to be grateful. First, this bound is likely to be of the right order, and, in any case, it provides the first nontrivial bound for the variance of the number of faces of the convex hull of the graph of a one-dimensional random walk. It remains a challenge to see if one can obtain an analogous estimate for the number of faces of the convex hull of the two-dimensional random walk $\{Z_1 + Z_2 + \cdots + Z_k: 1 \leqslant k \leqslant n\}$ in the case where Baxter's method provides the expected value.

One should also note that the concentration inequality (16) also provides information about the concentration of $F_n^{\text{hull}}$ about its mean, and just for the record we note that it implies

$$P(|F_n^{\text{hull}} - E[F_n^{\text{hull}}]| \geqslant \varepsilon E[F_n^{\text{hull}}]) \leqslant 4 \exp(-\varepsilon^2/4) \quad \text{for all } 0 \leqslant \varepsilon \leqslant \tfrac{1}{2}. \tag{21}$$

Finally, since we have a central limit theorem (2) for $F_n^{\text{top}}$ and $F_n^{\text{bottom}}$, we might expect to have a corresponding theorem for their sum. Such a result would be immediate if we had an asymptotic understanding of the correlation of $F_n^{\text{top}}$ and $F_n^{\text{bottom}}$, but, at present, no such understanding of the joint distribution is available.

Finally, we should note that the article by Barndorff-Nielsen and Baxter [2] provides interesting information about one of the most natural functionals of the convex hull of the two-dimensional random walk, the area. Regrettably, the bijection provided by the Bohenblust–Spitzer algorithm does not seem to be able to contribute to the understanding of the area functional, even when we restrict our attention to the convex hull of the graph of a one-dimensional random walk.

## 12. Concluding remarks

The only way that Spitzer used the bijection given by the Bohnenblust–Spitzer algorithm was in his study of the maximal process $M_n = \max(0, S_1, S_2, \ldots, S_n)$ for the random walk $\{S_k\}$, but the aim of this article has been to show that the bijection can be used to obtain useful information about many other geometrical features. Specifically, we have seen that the Bohnenblust–Spitzer bijection can be used to obtain information on the length of the concave majorant and the number of faces of this majorant. For the latter, we also found the sequence of distributional identities

$$F_n \stackrel{\mathrm{d}}{=} R_n \stackrel{\mathrm{d}}{=} v_n \stackrel{\mathrm{d}}{=} T_n$$

that relates $F_n$ to the number of records in a random permutation, number of cycles in a random permutation, and the random variable $T_n$ given by the simple Bernoulli sum (13). Almost all of the information one might want about $F_n$ is easily read from these identities, except for the crucial information about the joint distributions of the process $\{F_n\}$ that might help us resolve the interesting question of a strong law $\{F_n/\log n\}$.

Traditionally, the result from [19] that most people have regarded as key has not been the Bohnenblust–Spitzer bijection but rather the analytical formula

$$1 + \sum_{n=1}^{\infty} \phi_n(\theta) t^n = \exp\left[ \sum_{n=1}^{\infty} \frac{\psi_n(\theta)}{n} t^n \right] \tag{22}$$

that relates the pair of characteristic functions

$$\psi_n(\theta) = E[\exp(i\theta S_n^+)] \quad \text{and} \quad \phi_n(\theta) = E[\exp(i\theta M_n)].$$

This formula has many important consequences, and, as the essay of Kesten [13] makes clear, almost all of the investigators who follow Spitzer [19] have taken this formula as their starting point. One example in this tradition that prompted much further work is the instructive article of Wendel [22] which first gives a direct proof of Spitzer's Formula by analytic methods and subsequently shows how

one may reverse Spitzer's argument to prove Spitzer's combinatorial lemma from the exponential formula (22).

Here we have focused entirely on the combinatorial, algorithmic, and geometric aspects of the Bohnenblust–Spitzer bijection without any attention to the analytic approach, but we should also note that there are also interesting geometric results that can be cast in an analytic form analogous to Spitzer's identity. In particular, one can show that the exponential relation (22) also holds for the more geometric pair

$$\tilde{\psi}_n(\theta) = E\left[\exp(i\theta\sqrt{k^2 + S_k^2})\right] \quad \text{and} \quad \tilde{\phi}_n(\theta) = E[\exp(i\theta L_n)], \tag{23}$$

where, as before, $L_n$ denotes the Euclidean length of the concave majorant.

To be sure, some of the results that we have found by other means can also be obtained via the exponential relations for $(\psi_n, \phi_n)$ and $(\tilde{\psi}_n, \tilde{\phi}_n)$. For example, Spitzer obtained the expectation

$$E[\max(0, S_1, S_2, \ldots, S_n)] = \sum_{k=1}^{n} \frac{1}{k} E[S_k^+]. \tag{24}$$

by differentiating the identity (22) and setting $\theta = 0$, while we obtained this expectation (24) in Section 6 by summing over cycles. Also, in parallel to Spitzer's derivation of (24), one can differentiate the exponential formula for the pair $(\tilde{\psi}_n, \tilde{\phi}_n)$ to obtain an alternative derivation of our formula

$$E[L_n] = \sum_{k=1}^{n} \frac{1}{k} E\left[\sqrt{k^2 + S_k^2}\right]. \tag{25}$$

Nevertheless, the exponential formulas do not give up their secrets willingly, and progress was slow even in the case considered by Spitzer where special features of $S_n^+$ bring powerful, well-studied, Wiener–Hopf techniques smoothly into play. Thus, despite its attractive prospects, the analytical investigation of the exponential formula for the geometric pair $(\tilde{\psi}_n, \tilde{\phi}_n)$ is best left for another time and place.

This exposition of the Spitzer–Bohnenblust algorithm and its applications has touched on the theory of record values at several points, but there is much more that could be said. In particular, the early survey of Glick [8] and the up-to-date survey of Bunge and Goldie [6] give many beautiful properties of record times that suggests cognate questions for the geometry of the concave majorant. On the other hand, there have been no combinatorial extensions of Spitzer's lemma except for that in [5], which provides circumstances where generalized means may replace the simple averages used by Spitzer. Moreover, Goldie [9, p. 172], has observed that Brunk's theorem can be used to give an alternative approach to equality of the distributions of $F_n$ and $R_n$, and this naturally suggests that Brunk's results may help with the investigation of other features of the convex majorant. Nevertheless, Brunk's theorem is rather complicated, and such investigations do not look easy. Finally, it may be useful to note that Groeneboom [11] and Pitman [14] have discovered a number of delicate connections between the convex minorant (or concave majorant) and the geometry of Brownian motion paths, although, for the moment, one cannot say what these interesting continuous-time results may tell us about the Bohnenblust–Spitzer algorithm, or vice-versa.

# References

[1] R. Arratia, S. Tavaré, The cycle structure of random permutations, Ann. Probab. 20 (1992) 1567–1591.

[2] O. Barndorff-Nielsen, G. Baxter, Combinatorial lemmas in higher dimensions, Ann. Math. Statist. 34 (1963) 313–325.

[3] G. Baxter, A combinatorial lemma for complex numbers, Ann. Math. Statist. 32 (1961) 901–904.

[4] G. Bennett, Probability inequalities for sums of independent random variables, J. Amer. Statist. Soc. 57 (1962) 33–45.

[5] H.D. Brunk, A generalization of Spitzer's combinatorial lemma, Z. Wahrscheinlichkeitstheorie 2 (1964) 395–405.

[6] J. Bunge, C.M. Goldie, Record sequences and their applications, in: D.N. Shanbhag, C.R. Rao (Eds.), Handbook of Statistics, Vol. 19, Elsevier, Amsterdam, 2000.

[7] B. Dershowitz, S. Zaks, The cycle lemma and some applications, European J. Combin. 11 (1990) 35–40.

[8] N. Glick, Breaking records and breaking boards, Amer. Math. Monthly 85 (1978) 2–26.

[9] C.M. Goldie, Records, permutations and greatest convex minorants, Math Proc. Cambridge Philos. Soc. 106 (1989) 169–177.

[10] R.L. Graham, F.F. Yao, Finding the convex hull of a simple polygon, J. Algorithms 4 (1983) 324–331.

[11] P. Groeneboom, The concave minorant of Brownian motion, Ann. Probab. 11 (1983) 1016–1027.

[12] M. Kac, Toeplitz matrices, translation kernels, and a related problem in probabililty theory, Duke Math. J. 21 (1954) 501–509.

[13] H. Kesten, Frank Spitzer's work on random walk and Brownian motion, Ann. Probab. 21 (1993) 593–607.

[14] J.W. Pitman, Remarks on the convex minorant of Brownian motion, in: E. Çinlar, K.L. Chung, R.K. Getoor (Eds.), Seminar on Stochastic Processes, 1982, Progress in Probability, Vol. 5, Birkhäuser, Basel, 1983, pp. 219–227.

[15] A. Rényi, On the extreme elements of observations, MTA III Oszt. Közl 12 (1962) 105–121 (also available in P. Turán (Ed.), Selected Papers of Alfréd Rényi, Vol. 3, Akademiai Kaidó, Budapest, 1976, pp. 50–66).

[16] L.A. Shepp, S.P. Lloyd, Ordered cycle lengths in a random permutation, Trans. Amer. Math. Soc. 121 (1966) 340–357.

[17] H.S. Snevily, D.B. West, The bricklayer problem and the strong cycle lemma, Amer. Math. Monthly 105 (1998) 131–143.

[18] T.L. Snyder, J.M. Steele, Convex hulls of random walks, Proc. Amer. Math. Soc. 117 (1993) 1165–1173.

[19] F. Spitzer, A combinatorial lemma and its application to probability theory, Trans. Amer. Math. Soc. 82 (1956) 323–339.

[20] F. Spitzer, H. Widom, The circumference of a convex polygon, Trans. Amer. Math. Soc. 12 (1961) 506–509.

[21] T. Suidan, Statistical properties of convex minorants of random walks and Brownian motions, Preprint, Department of Applied and Computational Mathematics, Princeton University (www.math.washington.edu/prob/Letters/letter_56.html).

[22] J.G. Wendel, Spitzer's formula: a short proof, Proc. Amer. Math. Soc. 82 (1958) 905–908.