AN EXPONENTIAL LIMIT SHAPE OF RANDOM q-PROPORTION BULGARIAN SOLITAIRE

Kimmo Eriksson
Mälardalen University, School of Education, Culture and Communication, Västerås, Sweden
kimmo.eriksson@mdh.se

Markus Jonsson
Mälardalen University, School of Education, Culture and Communication, Västerås, Sweden
markus.jonsson@mdh.se

Jonas Sjöstrand
Royal Institute of Technology, Department of Mathematics, Stockholm, Sweden
jonass@kth.se

Received: 2/15/17, Revised: 3/13/18, Accepted: 6/8/18, Published: 6/29/18

Abstract

We introduce *p_n-random q_n-proportion Bulgarian solitaire* ($0 < p_n, q_n \leq 1$), played on n cards distributed in piles. In each pile, a number of cards equal to the proportion q_n of the pile size rounded upward to the nearest integer are candidates to be picked. Each candidate card is picked with probability p_n, independently of other candidate cards. This generalizes Popov's random Bulgarian solitaire, in which there is a single candidate card in each pile. Popov showed that a triangular limit shape is obtained for a fixed p as n tends to infinity. Here we let both p_n and q_n vary with n. We show that under the conditions $q_n^2 p_n n/\log n \to \infty$ and $p_n q_n \to 0$ as $n \to \infty$, the p_n-random q_n-proportion Bulgarian solitaire has an exponential limit shape.

1. Introduction

The game of Bulgarian solitaire has received a great deal of attention, see reviews by Hopkins [10] and Drensky [2]. Bulgarian solitaire is played with a deck of n identical cards divided arbitrarily into a number of piles. A move consists of picking a card from each pile and letting these cards form a new pile. If piles are sorted in order of decreasing size, every position in the solitaire is equivalent to a Young diagram of an integer partition of n.

Popov [15] considered a random version of Bulgarian solitaire defined by a probability $p \in (0, 1]$, such that one card from each pile is picked with probability p, independently of the other piles. We will refer to this stochastic process on configurations as *p-random* Bulgarian solitaire. The probabilities of configurations converge to a stationary distribution. Popov showed that as n grows to infinity and configuration diagrams are downscaled by $\sqrt{n}$ in both dimensions, the stationary probability of the set of configurations that deviate from a triangle with slope p by more than $\varepsilon > 0$ tends to zero. In this sense, random configurations have a limit shape.

The objective of the present paper is to study such limit shapes in a generalization of random Bulgarian solitaire.

1.1. q_n-proportion Bulgarian Solitaire

Olson [13] introduced a generalization of Bulgarian solitaire in which the number of cards that are picked from a pile of size h is given by some nonnegative valued function $\sigma(h)$. Eriksson, Jonsson and Sjöstrand [3] recently studied the special case when σ is well-behaved in the sense that $\sigma(1) = 1$ and both $\sigma(h)$ and $h - \sigma(h)$ are non-decreasing functions of h. In particular, they studied a special case that they called *q_n-proportion Bulgarian solitaire*, defined by the rule $\sigma(h) = \lceil q_n h \rceil$. This means that from each pile we pick a number of cards given by the proportion q_n of the pile size rounded upward to the nearest integer. To illustrate the effect of the parameter q_n, set it to 0.3 and consider the configuration $(6, 2, 2, 1)$. From the first pile we pick $\lceil 0.3 \times 6 \rceil = 2$ cards; similar calculations give that 1 card is picked from each of the other three piles. Note that for $q_n \leq 1/n$ exactly one card is always picked from each pile, and we retrieve the ordinary Bulgarian solitaire.

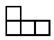
As n tends to infinity, Eriksson, Jonsson and Sjöstrand [3] determined limit shapes of stable configurations of q_n-proportion Bulgarian solitaire: in case $q_n^2 n \to 0$, the limit shape is triangular, which generalizes the limit shape result for the ordinary Bulgarian solitaire. For other asymptotic behavior of q_n, other limit shapes were obtained. Specifically, in case $q_n^2 n \to \infty$, the limit shape is exponential. The intermediate case $q_n^2 n \to C > 0$ produces a family of limit shapes interpolating between the triangular and the exponential shape.

1.1.1. p_n-random q_n-proportion Bulgarian Solitaire

We shall examine a p_n-random version of q_n-proportion Bulgarian solitaire, in which the proportion q_n (rounded upward) of cards in a pile are only *candidates* to be picked, each of which is picked only with probability p_n, independently of all other candidate cards. This process will be denoted by $\mathcal{B}(n, p_n, q_n)$. Note that in the special case of a fixed p and for $q_n \leq 1/n$, this process is equivalent to Popov's p-random Bulgarian solitaire.

Our focus will be on establishing a regime in which p_n-random q_n-proportion Bulgarian solitaire has an exponential limit shape.

2. The Concept of Limit Shapes

In this section we give the precise definitions of the limit shapes we consider. Let $\mathcal{P}(n)$ be the set of integer partitions of n. For any partition $\lambda \in \mathcal{P}(n)$ with $N = N(\lambda)$ positive parts $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N > 0$, define $\lambda_i = 0$ for $i > N(\lambda)$, and the *diagram* of λ as the Young diagram oriented such that the parts of λ are represented by left and bottom aligned columns, weakly decreasing in height from left to right. For example,  is the diagram of the partition $(2, 1, 1)$. We define the *diagram-boundary function* of λ as the nonnegative, weakly decreasing and piecewise constant function $\partial\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ describing the boundary of λ, given by

$$\partial\lambda(x) = \lambda_{\lfloor x \rfloor + 1}.$$

Following [4] and [16], the diagram is downscaled using some *scaling factor* $a_n > 0$ such that all row lengths are multiplied by $1/a_n$ and all column heights are multiplied by a_n/n, yielding a constant area of 1.

Thus, given a partition λ, define the *a_n-scaled diagram-boundary function* of λ as the nonnegative, real-valued, weakly decreasing and piecewise constant function $\partial^{a_n}\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ given by

$$\partial^{a_n}\lambda(x) = \frac{a_n}{n}\partial\lambda(a_n x) = \frac{a_n}{n}\lambda_{\lfloor a_n x \rfloor + 1}. \tag{1}$$

The p_n-random q_n-proportion Bulgarian solitaire $\mathcal{B}(n, p_n, q_n)$ (with $p_n, q_n \in (0, 1]$) can be regarded as a Markov chain on the finite state-space $\mathcal{P}(n)$. Let us denote the sequence of visited states by $(\lambda^{(0)}, \lambda^{(1)}, \ldots)$. In the truly random case of $p_n < 1$, it is straightforward to verify that this Markov chain is aperiodic and irreducible. It is well known that an aperiodic and irreducible Markov chain on a finite state space has a unique stationary distribution π and that, starting from any initial state, the distribution of the ith state $\lambda^{(i)}$ converges to π as i tends to infinity. We denote by π_{n,p_n,q_n} the stationary measure of the Markov chain $(\lambda^{(0)}, \lambda^{(1)}, \ldots)$ on $\mathcal{P}(n)$ given by $\mathcal{B}(n, p_n, q_n)$ for $p_n < 1$.[1] When we refer to a limit shape of the process $\mathcal{B}(n, p_n, q_n)$ for $p_n < 1$ as n grows to infinity, we shall mean the limit shape of the stationary measure π_{n,p_n,q_n}. The intuitive sense of this concept is that *when the solitaire is played on a sufficiently large number of cards for sufficiently long,*

[1] Readers acquainted with the limit shape literature may wonder whether the stationary measure has the property of being *multiplicative*, in the sense of interpretable as the product measure on the space of integer sequences restricted to a certain affine subspace [7]. The multiplicative property is useful in limit shape problems and related problems [1, 5, 8, 14, 16]. However, such techniques will not be used here as π_{n,p_n,q_n} is unlikely to be multiplicative in general.

the configuration will with a high probability be very close to the limit shape after suitable downscaling. Following Vershik [16], a sequence $\{\pi_n\}$ of probability distributions on $\mathcal{P}(n)$ is said to have a limit shape ϕ under the scaling a_n if the a_n-scaled diagrams approach ϕ in probability as n grows to infinity. The exact condition for convergence can vary. We shall use the definition that

$$\lim_{n\to\infty} \pi_n\{\lambda \in \mathcal{P}(n) : \sup_{x>0} |\partial^{a_n}\lambda(x) - \phi(x)| < \varepsilon\} = 1 \tag{2}$$

for any $\varepsilon > 0$.[2]

3. The Approach of Ordering Piles by Time of Creation

It will sometimes be useful to explicitly order piles by time of creation rather than by size. Here we develop this approach.

When parts are not sorted by size, a configuration is not represented by an integer partition but by a *weak integer composition*: an infinite sequence $\alpha = (\alpha_1, \alpha_2, \dots)$, not necessarily decreasing, of nonnegative integers adding up to n. Let $\mathcal{W}(n)$ denote the set of weak compositions of n. We define the *diagram*, the *diagram-boundary function* $\partial\alpha$, and the *a_n-scaled diagram-boundary function* $\partial^{a_n}\alpha$ of a weak composition α in exact analogy to the way we defined them for partitions in Section 2. For example, the diagram of $\alpha = (3, 0, 2, 4, 1, 0, 0, \dots)$ and the corresponding function graph $y = \partial\alpha(x)$ are shown in Figure 1. Also, for a weak composition $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N, 0, 0, \dots)$ we define the number of parts $N = N(\alpha)$ disregarding the trailing zeros.

3.1. Connecting the Limit Shapes of Compositions and Partitions

We shall now connect compositions with partitions. For any $\alpha \in \mathcal{W}(n)$, define the operator ord as the ordering operator that arranges the parts of α in descending order, thus yielding a partition. We shall now prove that such sorting of the piles by size conserves the convergence to a limit shape. The proof uses some basic theory of symmetric-decreasing rearrangements, see for example [9, Ch. 10] or [11, Ch. 3].

---

[2] Vershik [16] and Erlihson and Granovsky [6] used a weaker condition for convergence toward a limit shape, namely that

$$\lim_{n\to\infty} \pi_n\{\lambda \in \mathcal{P}(n) : \sup_{x\in[a,b]} |\partial^{a_n}\lambda(x) - \phi(x)| < \varepsilon\} = 1$$

should hold for any compact interval $[a, b] \subset (0, \infty)$ and any $\varepsilon > 0$. Yakubovich [17] and Eriksson and Sjöstrand [4] used the even weaker condition that

$$\lim_{n\to\infty} \pi_n\{\lambda \in \mathcal{P}(n) : |\partial^{a_n}\lambda(x) - \phi(x)| < \varepsilon\} = 1$$

should hold for any $x > 0$ and any $\varepsilon > 0$.

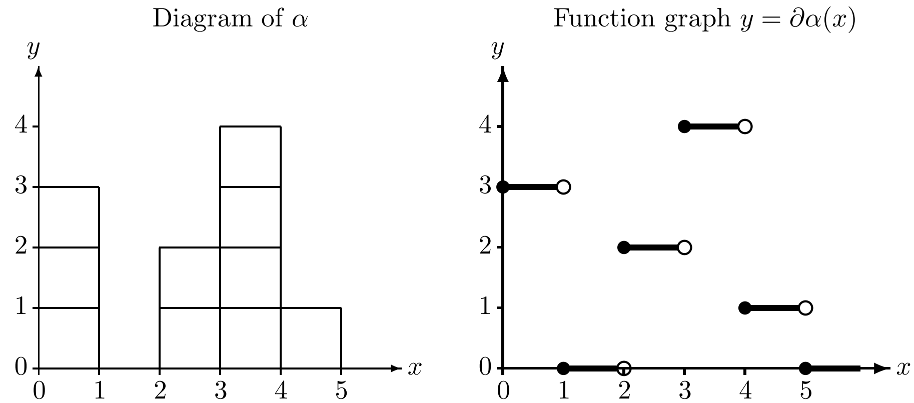


Figure 1: The composition $\alpha = (3, 0, 2, 4, 1, 0, 0, \dots) \in \mathcal{W}(10)$.

For any measurable function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that $\lim_{x\to\pm\infty} f(x) = 0$, there is a unique function $f^* : \mathbb{R} \to \mathbb{R}_{\geq 0}$, called the *symmetric-decreasing rearrangement* of f, with the following properties:

- f^* is symmetric, that is, $f^*(-x) = f^*(x)$ for all x,
- f^* is weakly decreasing on the interval $[0, \infty)$,
- f^* and f are equimeasurable, that is,

$$\mathcal{L}(\{x : \ f(x) > t\}) = \mathcal{L}(\{x : \ f^*(x) > t\})$$

  for all $t > 0$, where $\mathcal{L}$ denotes the Lebesgue measure,
- f^* is lower semi-continuous.

In particular, if f is a symmetric function that is weakly decreasing and right-continuous on $[0, \infty)$ and tends to 0 at infinity, then $f^* = f$.

Lemma 1. *Let $\alpha \in \mathcal{W}(n)$ be a weak composition of n and let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a right-continuous and weakly decreasing function such that $f(x) \to 0$ as $x \to \infty$. The diagram-boundary functions before and after sorting of the weak composition satisfy the inequality*

$$\|\partial \mathsf{ord}\, \alpha - f\|_\infty \leq \|\partial\alpha - f\|_\infty,$$

where $\|\cdot\|_\infty$ denotes the max norm $\|f\|_\infty = \sup\big\{|f(x)| : x \geq 0\big\}$.

Proof. The intuition of the lemma should be obvious from Figure 2. To be able to use the standard machinery of symmetric rearrangements, we consider the functions f, $\partial\alpha$, and $\partial\mathsf{ord}\,\alpha$ as being defined on the entire real axis by letting $f(x) = f(|x|)$ and analogously for $\partial\alpha$, and $\partial\mathsf{ord}\,\alpha$.

Since $f(x) \to 0$ as $x \to \infty$, its symmetric-decreasing rearrangement f^* is defined and, since f is weakly decreasing and lower semi-continuous, we have $f^* = f$. Similarly, $\partial\mathsf{ord}\,\alpha(x) \to 0$ as $x \to \infty$ and is weakly decreasing, so $(\partial\mathsf{ord}\,\alpha)^* = \partial\mathsf{ord}\,\alpha$. Moreover, $(\partial\alpha)^* = \partial\mathsf{ord}\,\alpha$ must hold because the operator ord arranges the composition parts in descending order.

Now, since symmetric rearrangements decrease L^p-distances for any $1 \le p \le \infty$ (see for example [11], Section 3.4), we obtain

$$\|\partial\mathsf{ord}\,\alpha - f\|_\infty = \|(\partial\alpha)^* - f^*\|_\infty \le \|\partial\alpha - f\|_\infty.$$

□

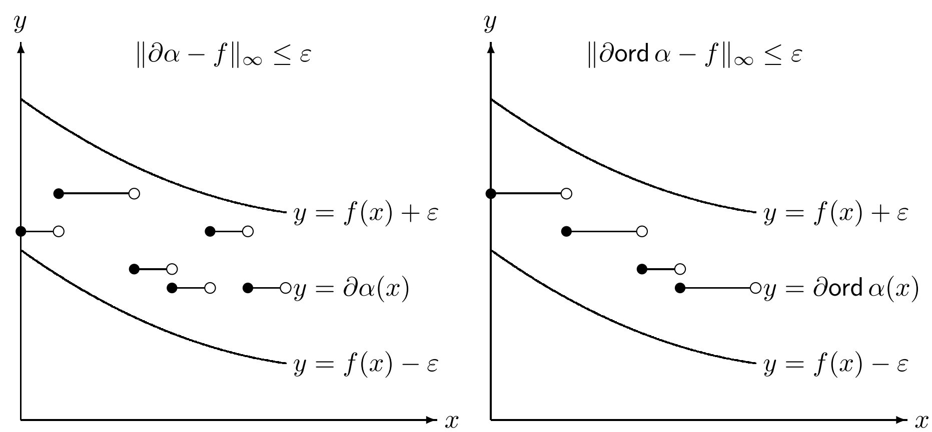


Figure 2: An example of a composition α and a decreasing function f showing that if $\partial\alpha(x)$ is enclosed between $f(x) - \varepsilon$ and $f(x) + \varepsilon$, then so is $\partial\mathsf{ord}\,\alpha(x)$, an immediate consequence of Lemma 1.

Lemma 2. *For any distribution π_n on $\mathcal{W}(n)$, define a corresponding distribution $\tilde{\rho}^{(n)}$ on $\mathcal{P}(n)$ by*

$$\tilde{\rho}^{(n)}(\lambda) = \sum_{\substack{\alpha \in \mathcal{W}(n) \\ \mathsf{ord}\,\alpha = \lambda}} \pi_n(\alpha). \tag{3}$$

If ϕ is a limit shape of π_n on $\mathcal{W}(n)$ under some scaling a_n, then ϕ is also a limit shape of $\tilde{\rho}^{(n)}$ on $\mathcal{P}(n)$ under the same scaling.

Proof. The assumption that ϕ is a limit shape of the distribution π_n on $\mathcal{W}(n)$ under the scaling a_n means that

$$\lim_{n\to\infty} \pi_n\{\alpha \in \mathcal{W}(n) : \sup_{x>0} |\partial^{a_n}\alpha(x) - \phi(x)| < \varepsilon\} = 1$$

for any $\varepsilon > 0$. By virtue of Lemma 1 we can replace α with $\mathsf{ord}\,\alpha$ in this formula:

$$\lim_{n\to\infty} \pi_n\{\alpha \in \mathcal{W}(n) : \sup_{x>0} |(\partial^{a_n} \mathsf{ord}\,\alpha)(x) - \phi(x)| < \varepsilon\} = 1. \tag{4}$$

The set $A := \{\alpha \in \mathcal{W}(n) : \sup_{x>0} |(\partial^{a_n} \mathsf{ord}\,\alpha)(x) - \phi(x)| < \varepsilon\}$ can be written as a disjoint union of equivalence classes with respect to sorting:

$$A = \bigcup_{\lambda\in L} \{\alpha \in \mathcal{W}(n) :\ \mathsf{ord}\,\alpha = \lambda\}$$

where $L = \{\lambda \in \mathcal{P}(n) :\ \sup_{x>0} |\partial^{a_n}\lambda(x) - \phi(x)| < \varepsilon\}$. The π_n-probability measure of A is

$$\begin{aligned}
\pi_n(A) &= \pi_n\left(\bigcup_{\lambda\in L} \{\alpha \in \mathcal{W}(n) :\ \mathsf{ord}\,\alpha = \lambda\}\right)\\
&= \sum_{\lambda\in L} \pi_n\{\alpha \in \mathcal{W}(n) :\ \mathsf{ord}\,\alpha = \lambda\}\\
&= \sum_{\lambda\in L} \sum_{\substack{\alpha\in\mathcal{W}(n)\\ \mathsf{ord}\,\alpha=\lambda}} \pi_n(\alpha)\\
&= \sum_{\lambda\in L} \tilde{\rho}^{(n)}(\lambda) && \text{(by (3))}\\
&= \tilde{\rho}^{(n)}(L).
\end{aligned}$$

From (4) we have that $\lim_{n\to\infty} \pi_n(A) = 1$. Because $\pi_n(A) = \tilde{\rho}^{(n)}(L)$, we can conclude that also $\lim_{n\to\infty} \tilde{\rho}^{(n)}(L) = 1$, that is,

$$\lim_{n\to\infty} \tilde{\rho}^{(n)}\{\lambda \in \mathcal{P}(n) : \sup_{x>0} |\partial^{a_n}\lambda(x) - \phi(x)| < \varepsilon\} = 1$$

for any $\varepsilon > 0$. This means that ϕ is a limit shape of the distribution π_n on $\mathcal{P}(n)$. $\square$

4. Three Regimes

Recall from Section 1.1 the q_n-proportion Bulgarian solitaire developed in [3], where the limit shape is triangular when $q_n^2 n \to 0$, exponential when $q_n^2 n \to \infty$ and an interpolation between the two when $q_n^2 n \to C > 0$.

The p_n-random q_n-proportion Bulgarian solitaire seems to share this property of three regimes of limit shapes, based on the asymptotic behavior of $p_n q_n^2 n$. The focus in this paper is the exponential regime of the p_n-random q_n-candidate Bulgarian solitaire, i.e. the case $p_n q_n^2 n \to \infty$ as $n \to \infty$. However, with the proof technique we employ we will prove the stronger statement that the limit shape holds even when the configurations are considered elements of $\mathcal{W}(n)$, i.e. even without sorting the piles of a configuration according to size to create a partition in $\mathcal{P}(n)$. We will instead require the stronger condition $p_n q_n^2 n/\log n \to \infty$ as $n \to \infty$. By virtue of Lemma 2, the limit shape will also hold for partitions.

In Section 7 we conjecture the limit shape to be triangular when $p_n q_n^2 n \to 0$, exponential when $p_n q_n^2 n \to \infty$ and an interpolation between the two (a piecewise linear function graph that depends on C) when $p_n q_n^2 n \to C > 0$.

5. The Exponential Limit Shape

Here we investigate the limit shape of configurations in the p_n-random q_n-candidate Bulgarian solitaire $\mathcal{B}(n, p_n, q_n)$ in the regime

$$\frac{p_n q_n^2 n}{\log n} \to \infty \text{ as } n \to \infty. \tag{5}$$

Our main result, Theorem 1, says that, under the additional asymptotic property $p_n q_n \to 0$ as $n \to \infty$, the (properly downscaled) boundary function of the diagram obtained after playing sufficiently many moves (for each fixed n) will converge to the exponential shape e^{-x} in probability as $n \to \infty$. See Figure 3. Throughout this section, the asymptotic notations o and O will always be with respect to $n \to \infty$.

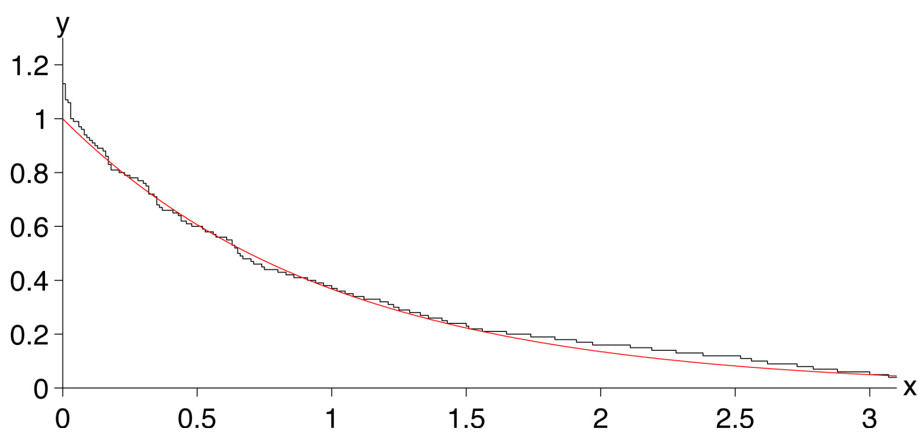


Figure 3: The result of a computer simulation after 200 moves of p_n-random q_n-proportion Bulgarian solitaire in the case $q_n = 1$, with $n = 10^5$ cards and $p_n = 0.01$, starting from a triangular configuration. The jagged curve is the 100-scaled diagram-boundary function of the resulting configuration and the smooth curve is the limit shape $y = e^{-x}$.

We shall see that the condition $p_n q_n^2 n/\log n \to \infty$ implies that the rounding effect in computing the number of candidate cards is negligible in the sense that the number of candidate cards will be $q_n n(1 + o(1))$. This in turn means that the expected number of picked cards will be $p_n q_n n(1 + o(1))$, thus the new pile created in a move of the solitaire will have size $\lambda_1 = p_n q_n n(1 + o(1))$. It will turn out that we need the scaling factor $a_n = (p_n q_n)^{-1}$ to obtain our exponential limit shape. If $p_n q_n$ were bounded away from zero, this scaling would not be able to transform the jumpy boundary diagrams into a smooth limit shape. Therefore, we also require

$$p_n q_n \to 0 \text{ as } n \to \infty. \tag{6}$$

On the other hand, if $p_n q_n$ tends to zero too fast, the pile sizes will be small and their random fluctuations will be large. For instance, the new pile after each move has a size drawn from the binomial distribution $\mathrm{Bin}(K, p_n)$, where $K \approx q_n n$ is the number of candidate cards, with relative standard deviation $\sim 1/\sqrt{p_n q_n n}$. The requirement (5) guarantees that $p_n q_n$ does not tend to zero too fast.

Theorem 1. *For each positive integer n, pick q_n and p_n with $0 < p_n, q_n \leq 1$ and a (possibly random) initial configuration $\lambda^{(0)} \in \mathcal{P}(n)$. Let $(\lambda^{(0)}, \lambda^{(1)}, \dots)$ be the Markov chain on $\mathcal{P}(n)$ defined by $\mathcal{B}(n, p_n, q_n)$, and denote its stationary measure by π_{n,p_n,q_n}. Suppose*

$$p_n q_n \to 0 \quad \textit{and} \quad \frac{p_n q_n^2 n}{\log n} \to \infty \quad \textit{as } n \to \infty.$$

Then π_{n,p_n,q_n} has the limit shape e^{-x} under the scaling $a_n = (p_n q_n)^{-1}$.

The proof of Theorem 1 heavily relies on the following version of Chernoff bounds. For a proof, see for example [12].

Chernoff Bound. *For $n \geq 1$ and $0 < p \leq 1$, let $X \sim \mathrm{Bin}(n, p)$ and set $\mu = E(X) = np$. Then, for any $0 < \gamma < \mu$,*

$$P(|X - \mu| \geq \gamma) \leq 2 \exp\left(-\frac{\gamma^2}{3\mu}\right). \tag{7}$$

The idea of the proof of Theorem 1 is the following.

We will use the approach developed in Section 3, i.e. card configurations in the solitaire will be represented by weak integer compositions and the piles are ordered with respect to creation time, i.e. if $\alpha \in \mathcal{W}(n)$ is the current configuration in the solitaire, then α_1 was the last formed pile, α_2 the pile that was formed two moves ago, etc. With this representation, some piles may be empty, so one may imagine each pile being placed in a bowl and the bowls are lined up in a row on the table. In each move of the solitaire, the new (possibly empty) pile is put in a new bowl to the left of all old bowls. As mentioned in Section 4, we shall prove Theorem 1 as a limit shape result for diagram-boundary functions of *compositions*. Thus, throughout this section, each configuration of n cards will be represented by an element of $\mathcal{W}(n)$. Also, in the following we may abbreviate $p = p_n$ and $q = q_n$ unless the dependency on n is crucial.

Assume a configuration $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N, 0, 0, \ldots)$ of n cards with $N = N(\alpha)$ piles (so that $\sum_{i=1}^N \alpha_i = n$) in the solitaire $\mathcal{B}(n, p, q)$. The number of candidate cards in the next move is $\kappa := \sum_{i=1}^N \lceil q\alpha_i \rceil$. We denote the rounding effect in pile $1 \le i \le N$ by $R_i := \lceil q\alpha_i \rceil - q\alpha_i$ and the total rounding effect by $R := \kappa - qn = \sum_{i=1}^N R_i$.

Clearly, $R < N$ (since $R_i < 1$ for any i), i.e. the total rounding effect is bounded above by the number of piles. The first thing we will do is to show that after a sufficient number D of moves from the initial configuration $\alpha^{(0)}$, the number of piles $N(\alpha^{(D)})$ in the resulting configuration $\alpha^{(D)}$ is typically much smaller than qn (so that the number of candidate cards κ is approximately qn and thus the total rounding effect R is negligible). In Lemma 4 we show that it is possible to choose such a D, namely $D = c\frac{\log n}{pq}$ for any $c \ge 14$.

We also need to make sure that the number of piles stays $o(qn)$ for sufficiently many additional moves M, long enough to establish the convergence of the overall shape. Lemma 4 will also guarantee that $M = \lceil n^2/p \rceil$ suffices for this purpose.

Thus, in the following we shall use

$$D = \left\lceil 14\frac{\log n}{pq} \right\rceil \quad \text{and} \quad M = \left\lceil \frac{n^2}{p} \right\rceil. \tag{8}$$

If the number of piles stays $o(qn)$ during M moves so that the number of candidate cards stays $qn(1+o(1))$, the newly formed pile in each of these moves will have expected size $pqn(1 + o(1))$. Our proof technique involves studying the evolution of such a pile (which will follow an exponential decay in size). Therefore we need to additionally make sure that no old piles (which could potentially be much larger than pqn) remain after these M moves. Lemma 3 shows that, in fact, after M moves all piles in the starting configuration have disappeared with a probability tending to one.

Lemma 3. *Let M be given by (8). From any initial configuration $\alpha \in \mathcal{W}(n)$, after M moves in the solitaire $\mathcal{B}(n, p_n, q_n)$, the probability that all piles in α have been consumed tends to 1 as $n \to \infty$.*

Proof. Consider a pile of size n. The size of this pile after M moves is statistically dominated by $\max(n-X, 0)$ where $X \sim \text{Bin}(M, p_n)$ whose expected value is $E(X) = Mp_n = \lceil n^2/p_n \rceil p_n > n$. Therefore, the probability that the pile remains after M moves is $P(X \le n)$ with the bound

$$\begin{aligned} P(X \le n) &\le P(|X - Mp_n| \ge |Mp_n - n|) \le 2\exp\left(-\frac{(Mp_n - n)^2}{3Mp_n}\right) \\ &\le 2\exp\left(-\frac{n^2}{3}(1 + o(1))\right), \end{aligned}$$

where we used the Chernoff bound (7). Thus, since any given pile in α has size $\leq n$, and the number of non-empty piles (of any size in any configuration) is $\leq n$, the probability that *all* piles in α have been consumed after M moves is at least

$$1 - 2n \exp\left(-\frac{n^2}{3}(1+o(1))\right) \to 1$$

which concludes the proof. $\square$

Lemma 4. *Let n, p_n, q_n and an initial configuration $\alpha^{(0)}$ be given in the solitaire $\mathcal{B}(n, p_n, q_n)$. Then*

$$\frac{1}{q_n n} \max\left\{N(\alpha^{(D+1)}), \ldots, N(\alpha^{(D+M)})\right\} \to 0 \textit{ in probability,}$$

where D and M are given by (8).

Proof. We will first prove that all piles of size at most $q^{-1}\log n$ disappear with probability tending to one after D moves, making sure that there are not many small piles in $\alpha^{(D)}$. Consider a pile of size at most $q^{-1}\log n$ in $\alpha^{(0)}$. Note that every nonempty pile decreases by at least 1 with probability at least p in each move. Therefore, after D moves the number of picked cards from this pile statistically dominates $X \sim \mathrm{Bin}(D, p)$ with expected value $Dp \geq 14q^{-1}\log n$. Using the Chernoff bound (7), the probability that this pile remains after D moves is at most

$$\begin{aligned} P_1 &:= P\left(X < \frac{\log n}{q}\right) \leq P\left(|X - Dp| > \left|Dp - \frac{\log n}{q}\right|\right) \\ &\leq 2\exp\left(-\frac{\left(Dp - \frac{\log n}{q}\right)^2}{3Dp}\right) \leq 2\exp\left(-\frac{(\frac{14\log n}{q} - \frac{\log n}{q})^2}{3\cdot 14\frac{\log n}{q}}\right) = 2n^{-\frac{13^2}{42}\frac{1}{q}} < 2n^{-4}. \end{aligned}$$

Since there can be at most n piles of size at most $q^{-1}\log n$, the probability that not *all* piles of size at most $q^{-1}\log n$ have disappeared after D moves is bounded by

$$nP_1 < 2n^{-3}.$$

Let us now turn our attention to *the number of* piles after these D moves. By the above, all piles smaller than $q^{-1}\log n$ have disappeared with probability tending to one. Clearly, the number of piles *larger than* $q^{-1}\log n$ can never be more than $\frac{n}{q^{-1}\log n} = \frac{qn}{\log n}$. Also, during the process of these D moves, at most D new piles have been formed. (Exactly D piles have been formed but some may have disappeared in the process.) Thus, for the total number of piles $N(\alpha^{(D)})$ in the configuration $\alpha^{(D)}$ after D moves, with probability at least $1 - 2n^{-3}$, we have

$$N(\alpha^{(D)}) \leq \frac{qn}{\log n} + D = \frac{qn}{\log n} + 14\frac{\log n}{pq} = qn\left(\frac{1}{\log n} + 14\frac{\log n}{pq^2 n}\right) = o(qn),$$

where we used assumption (5) in the last step. It follows that, for any $\varepsilon > 0$,

$$\frac{1}{qn} \max\left\{N(\alpha^{(D+1)}), \ldots, N(\alpha^{(D+M)})\right\} < \varepsilon$$

with probability at least $1 - 2n^{-3}M \geq 1 - 2n^{-3}(n^2/p + 1) \to 1$ as $n \to \infty$, since $pn \to \infty$. (That $pn \to \infty$ is also a consequence of assumption (5).) $\square$

Lemma 4 asserts that the number of piles is $o(qn)$ in probability during the M moves from $\alpha^{(D)}$ to $\alpha^{(D+M)}$, hence the number of candidate cards remains to be $qn(1 + o(1))$ in probability during the same moves. Therefore the number of picked cards (which equals the size of the newly formed pile) remains of *expected* size $pqn(1 + o(1))$. In Lemma 5 we prove that the *actual* number of picked cards in each of these M moves does not deviate (relatively) from pqn.

Lemma 5. *Let n, p_n, q_n and an initial configuration $\alpha^{(0)}$ be given in the solitaire $\mathcal{B}(n, p_n, q_n)$. Let D and M be given by* (8)*. Then*

$$\max_{k\in[D+1,D+M]} \frac{|\alpha_1^{(k)} - p_n q_n n|}{p_n q_n n} \to 0 \textit{ in probability}$$

as $n \to \infty$.

Proof. Let $\varepsilon > 0$ and let κ be the number of candidate cards in $\alpha^{(k-1)}$ for some $k = D + 1, \ldots, D + M$. Recall that the total rounding effect in computing the number of candidate cards is bounded above by the number of piles. It therefore follows from Lemma 4 that $\kappa = nq(1 + o(1))$. The new pile size is $\alpha_1^{(k)} \sim \mathrm{Bin}(\kappa, p)$. Then, using the triangle inequality and the Chernoff bound (7) we have

$$\begin{aligned}
P_2 := P(|\alpha_1^{(k)} - pqn| > \varepsilon pqn) &\leq P(|\alpha_1^{(k)} - \kappa p| > \varepsilon pqn - |\kappa p - pqn|) \\
&< 2\exp\left(-\frac{(\varepsilon pqn - |\kappa p - pqn|)^2}{3\kappa p}\right) \\
&= 2\exp\left(-\frac{\varepsilon^2}{3} pqn(1 + o(1))\right) \\
&= o(1/M)
\end{aligned}$$

where the last equality is derived as follows. By (5), $\log n = o(pqn)$ and hence $\log n^a = o(pqn)$ for any $a \geq 1$. Since $pqn \to \infty$, this means that $\exp(-pqn)$ tends to zero faster than $\exp(-\log n^a)$, i.e., $\exp(-pqn) = o(1/n^a)$. Since $np \to \infty$, we therefore also have $\exp(-pqn) = o(p/n^a) = o(1/M)$. The next to the last equality follows from the fact that εpqn dominates over $|pqn - \kappa p|$ (since $|pqn - \kappa p| = |pqn - nq(1 + o(1))p| = pqn \cdot o(1)$).

Therefore, the probability that $|\alpha_1^{(k)} - pqn| > \varepsilon pqn$ for *any* k during the entire process of M moves is bounded by $MP_2 = M \cdot o(1/M) = o(1)$. $\square$

While playing the solitaire, there is a risk that at some point there will be too many piles, and thereby the number of candidate cards will be significantly larger than qn (and thus the size of the newly formed pile will be bigger than pqn). Lemmas 3 and 4 ensure that the probability tends to zero that this ever happens during the entire process of M moves from $\alpha^{(D)}$ to $\alpha^{(D+M)}$.

There is also a risk that, even if there are suitably many (qn) candidate cards, the number of *picked* cards among them will deviate from pqn due to random fluctuations (and thereby the size of the newly formed pile will deviate from pqn). Lemma 5 ensures that the probability of this ever happening during the same period of M moves tends to zero.

Therefore, after $m := D + M$ moves, with probability tending to one

- all current piles have been formed during the last M moves, and
- all current piles had size approximately pqn when they were formed.

At this point, i.e. in the configuration $\Gamma := \alpha^{(m)}$, the leftmost pile (of size Γ_1) was formed one move ago, the second pile from the left (of size Γ_2) was formed two moves ago, and so on. We shall prove that, with probability tending to one, the size Γ_k of the pile that was formed k moves ago for any $k \in \{1, 2, \ldots, m\}+$ is $\Gamma_k \sim \Gamma_1(1 - pq)^k \sim pqn(1 - pq)^k$, i.e. the size decreases exponentially with k with decay factor $1 - pq$.

We will now consider the evolution of a given pile of size A_1 during $r \geq 1$ steps in the p-random q-proportion Bulgarian solitaire in the following way. We will need to keep track of each individual card in this pile. To this end, we label the cards remaining in a given pile $1, 2, \ldots, A_1$ starting from the top, and each card will keep their label throughout the process. Let $X_{i,k} \in \{0, 1\}$ where $i = 1, \ldots, A_1$ and $k = 1, \ldots, r$ be independent Bernoulli random variables with $P(X_{i,k} = 1) = p$.

Consider the following process. Let A_{k+1} be the number of cards remaining in a given pile after k moves. In each move $k \in \{1, 2, \ldots, r\}$, we remove the card with label i if $X_{i,k} = 1$ and this card belongs to the candidate cards, i.e., the $\lceil qA_k \rceil$ topmost remaining cards. We will call this process a *q-process*. This process describes the evolution of a pile of size A_1 in the p-random q-proportion Bulgarian solitaire.

Using the same Bernoulli variables, for any real number $0 \leq s \leq 1$, we define an *s-threshold process* in the following way. In each move $k = 1, 2 \ldots, r$, we remove the card with label i if $X_{i,k} = 1$ and $i \leq \lceil sA_1 \rceil$. In this process, we let $A_{k+1}^{[s]}$ denote the number of remaining cards after k moves. When it is relevant to indicate the initial pile size, an s-threshold process is called an (s, A_1)-threshold process and the number of remaining cards after k moves is denoted by $A_{k+1}^{[s,A_1]}$.

In the proof of Theorem 1, we will use two different s-threshold processes (for two different values of s) to over- and underestimate the sizes of $r + 1$ consecutive piles in Γ (corresponding to the r steps in an s-threshold process). Both these processes will have the same desired limit shape and thus the limit shape of our solitaire will follow by the squeeze theorem. We first need a combinatorial lemma giving sufficient conditions for overestimation and for underestimation.

Lemma 6. *(i) If* $\lceil sA_1 \rceil \le \lceil qA_1 \rceil$*, then* $A_k^{[s]} \ge A_k$ *for* $k = 1, \ldots, r+1$.
(ii) If $(1-q)A_{r+1}^{[s]} \ge A_1 - \lceil sA_1 \rceil$*, then* $A_k^{[s]} \le A_k$ *for* $k = 1, \ldots, r+1$.

Proof. (i) A card that is removed at some step ℓ during the s-threshold process must have label $i \le \lceil sA_1 \rceil$, so in the q-process it belongs to the $\lceil qA_1 \rceil$ candidate cards in the initial pile and hence it belongs to the candidate cards also at step ℓ and will be removed. Thus, every card removed in the s-threshold process is removed in the q-process too, and it follows that $A_k^{[s]} \ge A_k$ for $k = 1, \ldots, r+1$.

(ii) We show by induction over r that, after r steps, the remaining cards in the s-threshold process is a subset of the remaining cards in the q-process. Suppose $(1-q)A_{r+1}^{[s]} \ge A_1 - \lceil sA_1 \rceil$. Since $A_r^{[s]} \ge A_{r+1}^{[s]}$ we have $(1-q)A_r^{[s]} \ge A_1 - \lceil sA_1 \rceil$ which by the induction hypothesis implies that $A_k^{[s]} \le A_k$ for $1 \le k \le r$. It follows that $(1-q)A_r \ge A_1 - \lceil sA_1 \rceil$ which in turn implies that $A_r - (A_1 - \lceil sA_1 \rceil) \ge \lceil qA_r \rceil$. This latter inequality means that the $\lceil qA_r \rceil$ topmost cards before step r in the q-process all have labels no larger than $\lceil sA_1 \rceil$. Thus, if a card is removed in step r in the q-process it is also removed in step r or earlier in the s-threshold process. This concludes the induction step. The base step $r = 0$ is trivial. □

Recall that we are considering the configuration $\Gamma = \alpha^{(m)}$ after $m = M + D$ moves in the solitaire from the initial configuration $\alpha^{(0)}$. We will compare the sizes of $r+1$ consecutive piles in Γ to the $r+1$ pile sizes in an s-threshold process. In order to make the comparison for *all* piles in Γ, this will be done for $r+1$ consecutive piles (which we will call an r-*chunk*) at a time. In each r-chunk the initial pile size is the corresponding pile size in the solitaire. In other words, $\Gamma_1, \Gamma_2, \ldots, \Gamma_{r+1}$ will be compared to the pile sizes in an (s, Γ_1)-threshold process (with initial pile size Γ_1); and $\Gamma_{r+2}, \Gamma_{r+3}, \ldots, \Gamma_{2(r+1)}$ will be compared to the pile sizes in an (s, Γ_{r+2})-threshold process (with initial pile size Γ_{r+2}), and so on. Let us call the resulting union of s-threshold processes an (r, s)-*union* process. Thus, if we denote the pile sizes in this (r, s)-union process by $U_1, U_2, \ldots$, we have

$$\begin{aligned} &U_1 = \Gamma_1 = A_1^{[s,\Gamma_1]}, && U_2 = A_2^{[s,\Gamma_1]}, && \ldots, && U_{r+1} = A_{r+1}^{[s,\Gamma_1]}, \\ &U_{r+2} = \Gamma_{r+2} = A_1^{[s,\Gamma_{r+2}]}, && U_{r+3} = A_2^{[s,\Gamma_{r+2}]}, && \ldots, && U_{2(r+1)} = A_{r+1}^{[s,\Gamma_{r+2}]}, \ldots. \end{aligned}$$

We intend to use the (r, s)-union process to estimate the pile sizes in Γ. In an s-threshold process, starting with a pile of size A_1, among the $\lceil A_1 s \rceil$ topmost cards, the number of remaining cards B after r moves is binomially distributed: $B \sim \mathrm{Bin}(\lceil A_1 s \rceil, (1-p)^r)$. See Figure 4. Therefore we need to choose $r = r_n$ and $s = s_n$ in such a way that the following hold with high probability in each s-threshold process:

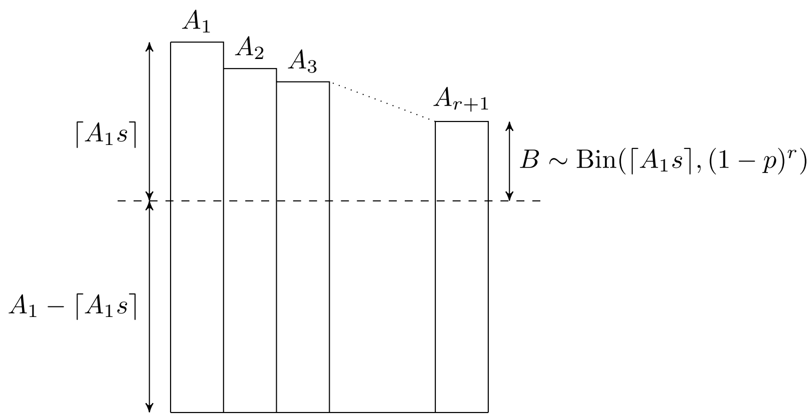


Figure 4: The r steps of an (s, A_1)-threshold process.

(I) The pile size A_{k+1} is close to $A_1(1-pq)^k$ for all $k = 1, \ldots, r$.

(II) At the same time s must be close enough to q to make the over- and under-estimations tight enough.

To accomplish (II) we shall see that $s = q$ will suffice for the overestimation and $s = q(1+2pr) = q(1+o(1))$ for the underestimation.

Let us choose

$$r_n = \left\lceil \rho_n^{-1/3} p_n^{-1} \right\rceil \text{ where } \rho_n = \frac{p_n q_n^2 n}{1+\log n}. \tag{9}$$

Lemma 7. *The following equations hold for our choice of r_n under assumption (5).*

$$(p_n r_n)^2 \frac{p_n q_n^2 n}{\log n} = \frac{p_n^3 q_n^2 n r_n^2}{\log n} \to \infty \textit{ as } n \to \infty. \tag{10}$$

$$p_n(r_n - 1) \to 0 \textit{ as } n \to \infty. \tag{11}$$

Proof. The limit (11) holds since $p_n(r_n-1) \le \rho_n^{-1/3} \to 0$, and (10) holds since

$$(p_n r_n)^2 \frac{p_n q_n^2 n}{\log n} > \rho_n^{-2/3} \frac{p_n q_n^2 n}{\log n} = \frac{(p_n q_n^2 n)^{1/3}}{(1+\log n)^{-2/3} \log n} > \left(\frac{p_n q_n^2 n}{\log n}\right)^{1/3} \to \infty$$

as $n \to \infty$ by (5). $\square$

Our next lemma, Lemma 8, will bound the probability P' that an initial pile of size $I_n := O(pqn)$ will, after r_n moves in an s_n-threshold process, deviate from its expected size when $s_n = q_n(1+o(1))$.

Since the number of piles is $\approx (p_n q_n)^{-1}$, the number of r-chunks is $\approx (p_n q_n r_n)^{-1}$. When using Lemma 8 we need the bound P' to hold for *each* chunk during *all* M

moves (where M is given by (8)), specifically $P'M/(p_nq_nr_n) \to 0$ as $n \to \infty$. The probability in Lemma 8 is therefore bounded by $o(p_nq_nr_n/M) = o(p_n^2q_nr_n/n^2)$. This is also why the pile size deviation εnp_nq_n is scaled with the number of chunks, resulting in the deviation $(\varepsilon p_nq_nn)(p_nq_nr_n) = \varepsilon p_n^2q_n^2nr_n$.

Lemma 8. *Let $(p_n)_n$ and $(q_n)_n$ be real sequences such that $0 < p_n, q_n \leq 1$ and $p_nq_n \to 0$ as $n \to \infty$. For each n, also let $B_n \sim \mathrm{Bin}(F_ns_n, (1-p_n)^{r_n})$ where $(F_n)_n$ and $(s_n)_n$ are real sequences such that*

$$F_n = O(p_nq_nn) \quad \textit{and} \quad s_n = q_n(1+o(1)), \tag{12}$$

and F_ns_n is an integer for every n. Let $(r_n)_n$ be the sequence of positive integers in (9). Then, for all $\varepsilon > 0$ we have

$$P\left(\left|B_n + F_n(1-s_n) - F_n(1-p_nq_n)^{r_n}\right| > \varepsilon p_n^2q_n^2nr_n\right) = o(p_n^2q_nr_n/n^2).$$

Proof. Let us abbreviate $F = F_n$, $B = B_n$, $p = p_n$, $q = q_n$, $r = r_n$ and $s = s_n$. Thus, we want to prove that

$$P := P\left(\left|B + F(1-s) - F(1-pq)^r\right| > \varepsilon p^2q^2nr\right) = o(p^2qr/n^2).$$

We first note that the expected value $E(B) = Fs(1-p)^r$. Using the triangle inequality $|B+F(1-s)-F(1-pq)^r| \leq |B-E(B)|+|E(B)-F(1-pq)^r+F(1-s)|$ we obtain

$$P \leq P(|B-E(B)| > \varepsilon p^2q^2nr - |E(B) - F(1-pq)^r + F(1-s)|).$$

By the Chernoff bound (7) we get

$$P \leq 2\exp\left(-\frac{(\varepsilon p^2q^2nr - |E(B) - F(1-pq)^r + F(1-s)|)^2}{3E(B)}\right). \tag{13}$$

For the indices n for which $r > 1$ we have $rp \leq 2(r-1)p \to 0$ and hence

$$(1-p)^r = 1 - pr + o(pr) \quad \text{and} \quad (1-pq)^r = 1 - pqr + o(pqr). \tag{14}$$

For the indices n for which $r = 1$, the relations in (14) are trivially true.

This means

$$\begin{aligned}
|E(B) - F(1-pq)^r + F(1-s)| &= |Fs(1-p)^r - F(1-pq)^r + F(1-s)| \\
&= |Fs(1-pr+o(pr)) - F(1-pqr+o(pqr)) + F(1-s)| \\
&= F\Big(pr(q-s) + s\cdot o(pr) + o(pqr)\Big) \\
&= o(Fpqr) = o(p^2q^2nr). && \text{(by (12))}
\end{aligned}$$

Thus the numerator in (13) can be written $[(\varepsilon+o(1))p^2q^2nr]^2$. By the assumptions in (12), the denominator in (13) can be written

$$3E(B) = 3Fs(1-p)^r = 3\cdot O(pqn)\cdot q(1+o(1))\cdot O(1) = O(pq^2n).$$

Putting these together, the bound (13) on P can be written

$$-\frac{1}{\log P} = O\left(\frac{O(pq^2n)}{[(\varepsilon+o(1))p^2q^2nr]^2}\right) = O\left(\frac{1}{p^3q^2nr^2}\right) = o\left(\frac{1}{\log n}\right)$$

where (10) was used in the last step. Since $pqnr \to \infty$ (also by (10)) and $pqr \to 0$ (by (6) and (11)), we have $\frac{1}{pqr} = o(n)$ and hence $\log\frac{1}{pqr} = o(\log n)$. Therefore

$$-\frac{1}{\log P} = o\left(\frac{1}{\log n + \log\frac{1}{pqr}}\right) = o\left(\frac{1}{\log\frac{n}{pqr}}\right).$$

From this follows

$$\log P = o\left(\log\frac{pqr}{n}\right) = o\left(\log\frac{p^2qr}{n^2}\right),$$

and thus $P = o(p^2qr/n^2)$. □

Note that Lemma 8 concerns an s-threshold process, i.e. only r steps. In other words, it asserts that

$$P\left(\left|A_{r+1} - A_1(1-p_nq_n)^{r_n}\right| > \varepsilon p_n^2q_n^2nr_n\right) = o(p_n^2q_nr_n/n^2), \tag{15}$$

where $A_1 = O(p_nq_nn)$ is the first pile size in an r-chunk and $A_{r+1} = (1-s_n)A_1 + B_n$ the last (see Figure 4). However, the deviation and the probability were chosen in such a way that they can be added over all r-chunks. This is done in Lemma 9 which bounds the probability of deviation for the *entire* union process. Specifically, we will show that, for any $C > 0$, the piles in Γ formed at most $\frac{C}{pq}$ moves ago, i.e. Γ_k for $k \le \frac{C}{pq}$, will follow an exponential decay with probability tending to one. The sizes of the piles formed more than $\frac{C}{pq}$ moves ago ($k > \frac{C}{pq}$) will be shown to be sufficiently small to be close enough to the tail in the exponential limit shape.

Lemma 9. *Let $U_1, U_2, \ldots$ be the pile sizes in an (r_n, s_n)-union process corresponding to $\mathcal{B}(n, p_n, q_n)$, where the initial pile size is $U_1 = O(p_nq_nn)$, and r_n is given by (9) and $s_n = q_n(1+o(1))$. Let $M = \lceil n^2/p_n\rceil$. Then, for all $C, \varepsilon > 0$ and $k < \frac{C}{p_nq_n}$,*

$$P(|U_{k+1} - U_1(1-p_nq_n)^k| > \varepsilon p_nq_nn) = o(1/M) = o(p_n/n^2).$$

Proof. As in the proof of Lemma 8, for the simplicity of notation we do not indicate in p, q, r and s the dependence on n. Let $C, \varepsilon > 0$ and $\varepsilon' = \varepsilon/C$. By the triangle inequality,

$$\begin{aligned}
&|U_{k+r+1} - U_1(1-pq)^{k+r}| \\
&\qquad \le |U_{k+1} - U_1(1-pq)^k|(1-pq)^r + |U_{k+r+1} - U_{k+1}(1-pq)^r| \\
&\qquad \le |U_{k+1} - U_1(1-pq)^k| + |U_{k+r+1} - U_{k+1}(1-pq)^r|.
\end{aligned}$$

Lemma 8 is now applicable for the first pile in each r-chunk (since $U_1 \ge U_2 \ge \cdots$ and $U_1 = O(p_n q_n n)$), so by (15), $|U_{k+r+1} - U_{k+1}(1-pq)^r| < \varepsilon' p^2 q^2 rn$ with probability $1 - o(p^2 qr/n^2)$. Thus,

$$|U_{k+r+1} - U_1(1-pq)^{k+r}| < |U_{k+1} - U_1(1-pq)^k| + \varepsilon' p^2 q^2 rn \tag{16}$$

with probability $1 - o(p^2 qr/n^2)$. We now note that the first term in the right hand side has the same form as the left hand side, only shifted by r piles. Thus, by induction it follows that, for any positive integer d, we have

$$|U_{dr+1} - U_1(1-pq)^{dr}| < d\varepsilon' p^2 q^2 rn$$

with probability $1 - o(dp^2 qr/n^2)$. Thus, adding the probabilities for deviation for $k = r, 2r, \ldots, \eta r$, where $\eta = \lfloor \frac{C}{pqr} \rfloor$ we get

$$\begin{aligned}
P\big(\exists k \in \{r, 2r, \ldots, \eta r\} : |U_{k+1} - U_1(1-pq)^k| > \eta\varepsilon' p^2 q^2 rn \ge \varepsilon pqn\big) \\
= \eta d \cdot o(p^2 qr/n^2) = o(p/n^2).
\end{aligned} \tag{17}$$

We have thereby proved the claim in the lemma for $k = r, 2r, \ldots, \eta r$. If k is not a multiple of r, suppose $dr < k < (d+1)r$ for some positive integer d. Then, since $pqr \to 0$ as $n \to \infty$ (which follows from (5) and (9)), we have $(1-pq)^r = 1 - pqr + o(pqr)$ and hence

$$|U_1(1-pq)^{(d+1)r} - U_1(1-pq)^{dr}| = O(pqn)(1-pq)^{dr}|pqr + o(pqr)| < \varepsilon pqn$$

for sufficiently large n. The lemma then follows by (17) and the fact that $U_{dr} \le U_k \le U_{(d+1)r}$. $\square$

6. Proof of Theorem 1

Below follows the proof of Theorem 1, stated in Section 5.

Proof. First, as in the previous section, let us consider $\mathcal{B}(n, p_n, q_n)$ as a process on $\mathcal{W}(n)$ rather than on $\mathcal{P}(n)$, and let $\alpha^{(0)} \in \mathcal{W}(n)$ be the weak composition representing the initial configuration of cards in the solitaire. Also, let M and D be given by (8).

Let $(r_n)_n$ be the sequence of positive integers given by (9) and let $(s_n)_n$ be the sequence $s_n = q_n(1 + 2p_n r_n)$. By Lemma 3 applied to $\alpha^{(D)}$, all piles present in $\alpha^{(D)}$ have disappeared in $\Gamma := \alpha^{(D+M)}$ with probability tending to one. Let $\Gamma_k = \alpha_k^{(D+M)}$ for $1 \leq k \leq M$ be the number of cards in the pile that was formed k moves ago. By Lemma 5, each of these piles had size $O(np_n q_n)$ in probability when they were formed. Let $F_n := O(np_n q_n)$ be a sequence such that $F_n s_n$ is an integer for each n. Let $0 < \varepsilon < 1$ and choose C_n such that $C_n > \frac{p_n q_n \log \varepsilon}{\log(1-p_n q_n)}$.

Let $\check{U}_1, \check{U}_2, \ldots$ be the pile sizes in the (r_n, s_n)-union process with initial pile size Γ_1. Using the fact that $p_n q_n r_n \to 0$, it is a straightforward calculation to show that $s_n = q_n(1 + 2p_n r_n)$ implies $(1 - q_n)((1 - p_n r_n)^{r_n} - \varepsilon p_n q_n r_n) > 1 - s_n$ and therefore

$$(1 - q_n)\big(A(1 - p_n r_n)^{r_n} - \varepsilon A p_n q_n r_n\big) > (1 - s_n)A \quad \text{for any } A > 0.$$

By Lemma 8, the probability that $\check{U}_1(1 - p_n r_n)^{r_n} - \varepsilon \check{U}_1 p_n q_n r_n < \check{U}_{r_n+1}$ is $P_1 := 1 - o(p_n^2 q_n r_n / n^2)$. Thus, with probability P_1 we have $(1 - q_n)\check{U}_{r_n+1} > (1 - s_n)\check{U}_1 \geq \check{U}_1 - \lceil s\check{U}_1 \rceil$ so by Lemma 6(ii), the pile sizes $\check{U}_1, \check{U}_2, \ldots, \check{U}_{r_n+1}$ in the first r-chunk of the (r_n, s_n)-process underestimate the pile sizes $\Gamma_1, \Gamma_2, \ldots, \Gamma_{r_n+1}$ with probability P_1. In the next chunk, we have a new absolute threshold $s\Gamma_{r_n+2} = s\check{U}_{r_n+2}$. Since $\Gamma_{r_n+2} \leq \Gamma_1$, we have $(1 - q_n)\check{U}_{2r_n+2} > (1 - s_n)\check{U}_{r_n+2}$ with probability at least P_1, making Lemma 6(ii) applicable also for the second chunk to conclude that $\check{U}_{r_n+2}, \ldots, \check{U}_{2r_n+2}$ underestimate $\Gamma_{r_n+2}, \ldots, \Gamma_{2r_n+2}$ with probability at least P_1. Continuing in the same manner for the first $C_n(p_n q_n r_n)^{-1}$ chunks, we conclude that the (r_n, s_n)-union process underestimates the solitaire with high probability:

$$P(\check{U}_k > \Gamma_k \text{ for all } k < \frac{C_n}{p_n q_n}) < (1 - P_1)C_n(p_n q_n r_n)^{-1} = o(p_n/n^2).$$

Let $\widehat{U}_1, \widehat{U}_2, \ldots$ be the pile sizes in the (r_n, q_n)-union process with initial pile size Γ_1. By Lemma 6(i) (with $s_n = q_n$), the (r_n, q_n)-union process surely overestimates the solitaire in each chunk.

Taking the results for the (r_n, s_n)-union process and the (r_n, q_n)-union process together we have

$$P(\check{U}_k \leq \Gamma_k \leq \widehat{U}_k \text{ for all } k < \frac{C_n}{p_n q_n}) > 1 - o(p/n^2).$$

Now, applying Lemma 9 to both the pile sizes $\check{U}_k$ and to the pile sizes $\widehat{U}_k$ and using the squeeze theorem, we obtain

$$P(|\Gamma_k - \Gamma_1(1-p_nq_n)^k| > \varepsilon np_nq_n) < o(p_n/n^2)$$

for all $\varepsilon > 0$ and $k < \frac{C_n}{p_nq_n}$. Thus, the probability that $|\Gamma_k - \Gamma_1(1-p_nq_n)^k| < \varepsilon np_nq_n$ for the first $\frac{C_n}{p_nq_n}$ piles throughout *all* M moves from $\alpha^{(D)}$ to $\alpha^{(D+M)}$ is $1 - M \cdot o(1/M) = 1 - o(1)$.

For piles $k > \frac{C_n}{p_nq_n}$, the exponential decrease (with decay factor $1-p_nq_n$) in pile size will yield piles smaller than $np_nq_n(1-p_nq_n)^{\frac{C_n}{p_nq_n}} < \varepsilon np_nq_n$ (by our choice of C_n). Thus, the pile sizes themselves are below εnp_nq_n.

In summary, after sufficiently many moves of $\mathcal{B}(n, p_n, q_n)$, the resulting composition diagram will converge uniformly in probability to the boundary diagram of the composition α where $\alpha_k = np_nq_n(1-p_nq_n)^{k-1}$ for all $k = 1, 2, \ldots$. The corresponding boundary function is $\partial\alpha(x) = np_nq_n(1-p_nq_n)^{\lfloor x \rfloor}$. The corresponding downscaled boundary function, with the given scaling factor $a_n = (p_nq_n)^{-1}$, is

$$\partial^{a_n}\alpha(x) = (1-p_nq_n)^{\lfloor \frac{x}{p_nq_n} \rfloor} \to e^{-x}$$

uniformly on $[0, \infty)$ since $p_nq_n \to 0$ as $n \to \infty$.

Setting $m := D + M$, and letting π_n^m denote the probability distribution on $\mathcal{W}(n)$ for $\alpha^{(m)}$, we have

$$\lim_{n\to\infty} \pi_n^m\{\alpha \in \mathcal{W}(n) : \sup_{x>0} |\partial^{a_n}\alpha(x) - e^{-x}| < \varepsilon\} = 1,$$

for all $\varepsilon > 0$, in accordance with (2). By virtue of Lemma 2, the same limit shape holds when configurations in the solitaire $\mathcal{B}(n, p_n, q_n)$ are represented by partitions $\mathcal{P}(n)$.

Since π_{n,p_n,q_n} is the stationary distribution of the Markov chain $(\lambda^{(0)}, \lambda^{(1)}, \ldots)$, if we start with a partition $\lambda^{(0)}$ sampled from π_{n,p_n,q_n} and play m moves, the resulting partition $\lambda^{(m)}$ will also be sampled from π_{n,p_n,q_n}. Thus, the theorem follows by choosing $\lambda^{(0)}$ as a stochastic partition sampled from the stationary distribution. $\square$

7. Conjectures

Recall that Theorem 1 was proved with $\mathcal{B}(n, p_n, q_n)$ being considered a process on $\mathcal{W}(n)$, and by virtue of Lemma 2 it also holds in $\mathcal{P}(n)$. We imposed the condition $\frac{p_nq_n^2n}{\log n} \to \infty$. Here we conjecture that the weaker condition $p_nq_n^2n \to \infty$ suffices in order for Theorem 1 to hold in $\mathcal{P}(n)$.

Conjecture 1. Theorem 1 also holds when the condition $np_nq_n^2/\log n \to \infty$ is replaced by the weaker condition $np_nq_n^2 \to \infty$.

The reason for this conjecture can be understood by considering the example $q_n = 1$ and $p_n n = \log\log n$. For this example it is easy to prove that there is no limit shape when sorting is not performed. Since $q_n = 1$, the number of picked cards in each move, and thus the expected size of a new pile, is $\mathrm{Bin}(n, p_n)$, with expected value np_n and standard deviation $\sigma \approx \sqrt{np_n}$. A pile of size np_n will, after $1/p_n$ moves, have the expected size $E = np_n(1-p_n)^{1/p_n}$, and clearly we have $E/(e^{-1}np_n) \to 1$ as $n \to \infty$.

Thus, the probability of a “visible” deviation (i.e. greater than $d = \sqrt{np_n}$ standard deviations) is $P(\text{deviation} \geq d\sigma) = e^{-np_n}$, so for $1/p_n$ piles, the expected number of such large deviations is approximately $\frac{e^{-np_n}}{p_n} = \frac{n}{\log n \log(\log n)} \to \infty$ as $n \to \infty$. This makes it impossible to achieve a convergence in probability towards a limit shape. However, from simulations we have reason to believe that the process converges towards a limit shape when sorting is performed.

Further, recall from Section 4 the other regimes $np_nq_n^2 \to 0$ and $np_nq_n^2 \to C$ for some constant $C > 0$. We conjecture that the limit shapes in the p_n-random q_n-proportion Bulgarian solitaire in these regimes are the same as in the deterministic q-proportion Bulgarian solitaire developed in [3].

Conjecture 2. If $p_nq_n^2n \to 0$ as $n \to \infty$, the limit shape of the p_n-random q_n-proportion Bulgarian solitaire is triangular.

Conjecture 3. If $p_nq_n^2n \to C$ as $n \to \infty$ for some constant $C > 0$, the limit shape of the p_n-random q_n-proportion Bulgarian solitaire is a piecewise linear shape that depends on the value of C.

References

[1] S. Corteel, B. Pittel, C. Savage, and H. Wilf, On the multiplicity of parts in a random partition, *Random Structures Algorithms* **14** (1999), 185–197.

[2] V. Drensky, The Bulgarian solitaire and the mathematics around it, *arXiv preprint 1503.00885* (2015).

[3] K. Eriksson, M. Jonsson, and J. Sjöstrand, Limit shapes of stable configurations of a generalized Bulgarian solitaire, *arXiv preprint 1703.07099* (2017).

[4] K. Eriksson and J. Sjöstrand, Limiting shapes of birth-and-death processes on Young diagrams, *Adv. in Appl. Math.* **48** (2012), 575 – 602.

[5] M. Erlihson and B. Granovsky, Reversible coagulation–fragmentation processes and random combinatorial structures: Asymptotics for the number of groups, *Random Structures Algorithms* **25** (2004), 227–245.

[6] M. Erlihson and B. Granovsky, Limit shapes of Gibbs distributions on the set of integer partitions: The expansive case, *Ann. Inst. H. Poincaré Probab. Stat.* **44** (2008), 915–945.

[7] B. Fristedt, The structure of random partitions of large integers, *Trans. Amer. Math. Soc.* **337** (1993), 703–735.

[8] W. Goh and P. Hitczenko, Random partitions with restricted part sizes, *Random Structures Algorithms* **32** (2008), 440–462.

[9] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1934.

[10] B. Hopkins, 30 years of Bulgarian solitaire, *College Math. J.* **43** (2012), 136–141.

[11] E. Lieb and M. Loss, *Analysis, Graduate Studies in Mathematics, Volume 14*, American Mathematical Society, Providence, Rhode Island, 2001.

[12] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, New York, 2005.

[13] J. Olson, Variants of Bulgarian solitaire, *Integers* **16** (2016), A8.

[14] B. Pittel, On a likely shape of the random Ferrers diagram, *Adv. in Appl. Math.* **18** (1997), 432–488.

[15] S. Popov, Random Bulgarian solitaire, *Random Structures Algorithms* **27** (2005), 310–330.

[16] A. Vershik, Statistical mechanics of combinatorial partitions, and their limit shapes, *Funct. Anal. Appl.* **30** (1996), 90–105.

[17] Y. Yakubovich, Ergodicity of multiplicative statistics, *J. Combin. Theory, Series A* **119** (2012), 1250–1279.