The Kruskal Count

Jeffrey C. Lagarias

AT&T Labs - Research
Florham Park, NJ 07932

Eric Rains

AT&T Labs - Research
Florham Park, NJ 07932

Robert J. Vanderbei

Princeton University
Princeton, NJ 08544

(October 12, 2001)

Abstract

The Kruskal Count is a card trick invented by Martin Kruskal in which a magician "guesses" a card selected by a subject according to a certain counting procedure. With high probability the magician can correctly "guess" the card. The success of the trick is based on a mathematical principle related to coupling models for Markov chains. This paper analyzes in detail two simplified variants of the trick and estimates the probability of success. The results are compared with simulation data for several variants of the actual trick.

AMS Subject Classification (2000): 60J10 (Primary) 91A60 (Secondary)

Keywords: Markov chain, stopping time

1. Introduction

The *Kruskal Count* is a card trick invented by Martin D. Kruskal (who is most well known for his work on solitons) which is described in Fulves and Gardner [5] and Gardner [6],[7]. In this card trick a magician "guesses" one card in a deck of cards which is determined by a subject using a special counting procedure that we call *Kruskal's counting procedure.* The magician can with high probability identify the correct card.

The subject shuffles a deck of cards as many times as he likes. He mentally chooses a (secret) number between one and ten. Kruskal's counting procedure then goes as follows. The subject turns the cards of the deck face up one at a time, slowly, and places them in a pile. As he turns up each card he decreases his secret number by one and he continues to count this

way till he reaches zero. The card just turned up at the point when the count reaches zero is called the *first key card* and its value is called the *first key number.* Here the value of an Ace is one, face cards are assigned the value five, and all other cards take their numerical value. The subject now starts the count over, using the first key number to determine where to stop the count at the second key card. He continues in this fashion, obtaining successive key cards until the deck is exhausted. The last key card encountered, which we call the *tapped card*, is the card to be "guessed" by the magician.

The Kruskal counting procedure for selecting the tapped card depends on the subject's secret number and the ordering of cards in the deck. The ordering is known to the magician because the cards are turned face up, but the subject's secret number is unknown. It appears impossible for the magician to know the subject's secret number. The mathematical basis of the trick is that for most orderings of the deck most secret numbers produce the same tapped card. For any given deck two different secret numbers produce two different sequences of key cards, but if the two sequences ever have a key card in common, then they coincide from that point on, and arrive at the same tapped card. The magician therefore selects his own secret number and carries out the Kruskal counting procedure for it while the subject does his own count. The magician's "guess" is his own tapped card. The Kruskal Count trick succeeds with high probability, but if it fails the magician must fall back on his own wits to entertain the audience.

The problem of determining the probability of success of this trick leads to some interesting mathematical questions. We are concerned with the *ensemble success probability* averaged over all possible orderings of the deck (with the uniform distribution). Our objective in this paper is to estimate ensemble success probabilities for mathematical idealizations of such counting procedures. Then we numerically compare the ensemble success probabilities on a 52-card deck with that of the Kruskal Count trick itself. The success probability of the trick depends in part on the magician's strategy for choosing his own secret number. We show that the magician does best to always choose the first card in the deck as his first key card, i.e. to use secret number 1.

The general mathematical problem we consider applies the Kruskal counting procedure to a deck of N labelled cards with each card label a positive integer, in which each card has its label drawn independently from some fixed probability distribution on the positive integers $\mathbb{N}^+$. We call such distributions *i.i.d. deck distributions*; they are specified by the probabilities

$\{\pi_j : j \geq 1\}$ of a fixed card having value j. We assume that the subject chooses an initial secret number from an initial probability distribution on $\mathbb{N}^+ = \{1, \ 2, \ 3, \ \cdots\}$, and that the magician independently does the same from a possibly different initial probability distribution, and that thereafter each follows the Kruskal counting procedure. It is convenient to view the cards of the deck as turned over at unit times, so that the card in the M-th position is turned over at time M. If the M-th card is a key card for both magician and subject and no previous card is a key card for both, then we say that M is the *coupling time* for the sequences. Let t be a random variable denoting the coupling time on the resulting probability space with $t = +\infty$ if coupling does not occur. We wish to estimate the "failure probability" $\text{Prob}[t > N]$.

The set of permutations of a fixed deck (with uniform distribution) does not have the i.i.d. property, and is not Markovian, but it can be reasonably well approximated by such a distribution. The advantage of the simplifying assumption of an i.i.d. deck distribution is that the random variable t can be interpreted as a stopping time for a coupling method for a Markov chain, as is explained in §2.

The mathematical contents of the paper are determination of $\text{Prob}(t > n)$ for a geometric i.i.d. deck distribution, which is carried out in §3, and estimation of $\text{Prob}(t > n)$ for a uniform i.i.d. deck distribution, which is carried out in §4. The proofs of several results stated in §4 are given in an appendix.

In §5 we consider the actual Kruskal count trick, and compare its success probability with the approximations given by the models above. Because the Kruskal count trick using an actual deck of 52 cards involves a stochastic process that is not Markovian, we estimate the success probability by Monte Carlo simulation. We consider the effect on this success probability of varying the magician's strategy for choosing his key card, and of varying the value assigned to face cards. The magician should choose his key card value to be 1. Assuming this strategy for the magician, the success probability of the original Kruskal Count trick is just over 85%. Both the i.i.d. geometric distribution and i.i.d. uniform distribution models above give good approximations; the geometric distribution is off by less than 3%, and the uniform approximation is within 1%.

There has been some previous work on mathematical models of the Kruskal count. In 1975 Mallows [11] determined the expected value of the coupling time of i.i.d. sequences, and observed that especially simple formulae occur for the geometric distribution. Recently Haga and Robins [9] analyzed a simplified Markov chain model for the Kruskal count, which

is related to, but not the same as, the models considered here. We discuss their model further at the end of §4.

2. Coupling Methods for Markov Chains

The *coupling time random variable t* is a special case of a stopping time random variable t^* associated with a coupling method for studying a Markov chain. This motivates our terminology.

To explain this connection, consider a homogeneous Markov chain $(X_n : n \geq 0)$ on a countable discrete state space S. Given two initial probability distributions $\mathbf{p}$ and $\mathbf{p}'$ on S a *coupling method* constructs a bivariate process (X_n^1, X_n^2) consisting of two copies of process X_n with X_0^1 having distribution $\mathbf{p}$, X_0^2 having distribution $\mathbf{p}'$, and the two copies evolve independently until some (random) *stopping time* t^* at which $X_{t^*}^1 = X_{t^*}^2$ and then requires them to be equal thereafter, evolving as a single process X_n. The stopping time t^* is not necessarily required to be the *first* time t at which $X_t^1 = X_t^2$ occurs, and the particular rule for choosing t^* defines the coupling method. Let $\boldsymbol{\mu}_n$, $\boldsymbol{\mu}'_n$ denote the distribution at time n of the process X_n stating from the distribution $\mathbf{p}$, $\mathbf{p}'$ respectively, at time 0, and let the *variation distance* $||\mathbf{p} - \mathbf{p}'||$ between two distributions on S be

$$||\mathbf{p} - \mathbf{p}'|| := \sum_{s \in S} |p(s) - p'(s)| \ .$$

The *basic coupling inequality* is

$$\frac{1}{2}||\boldsymbol{\mu}_n - \boldsymbol{\mu}'_n|| \leq \mathrm{Prob}[t^* > n] \ . \tag{2.1}$$

Such inequalities can be used to prove ergodicity of a Markov chain and to bound the speed of convergence to the equilibrium distribution, by bounding the right side of the inequality.

The first coupling method was invented by Doeblin [3], and many other coupling methods have been proposed since, see Griffeath [8] for a survey. Applications to card shuffling and random walks on groups are described in Aldous and Diaconis [1] and Diaconis [2]. The basic coupling inequality (2.1) is also valid for non-ergodic Markov chains, e.g. null-recurrent or transient Markov chains on the state space $\mathbb{N}$, as was observed by Pitman [12]. Coupling methods are traditionally used as an auxiliary device to get information on the rate of convergence to equilibrium of an ergodic Markov chain. In this paper, we are interested in obtaining upper and lower bounds for the coupling probability itself, since it represents the failure probability

4

of the Kruskal Count trick. We do not use the basic coupling inequality, but instead in §4 use inequalities relating coupling probabilities for various different Markov chains.

For an i.i.d. deck the Kruskal counting procedure can be viewed as moving on a Markov chain $\mathcal{M}_{\boldsymbol{\pi}}$ on the state space $\mathbb{N}$ where a state j represents a current value of the Kruskal counting procedure, with state 0 representing being at a key card, and state j represents that the next key card be reached after exactly j more cards are turned over. Each transition of the Markov chain will correspond to turning over one card in the deck. Let the random variable X_n denote the state of the Markov chain at time n; it indicates the current Kruskal count value at location n of the deck, except that $X_n = 0$ indicates a key card at location n. The transition probability for this chain from state $j \geq 1$ is probability 1 to state $j - 1$ and 0 to all other states, and from state 0 to state j is probability π_{j+1}, where $\{\pi_j : j \geq 1\}$ is the distribution $\boldsymbol{\pi}$ of card labels. (That is, π_1 is the probability that the key card has value 1, and the chain transitions from state 0 to state 0.) The initial distribution of secret numbers are distributions $\mathbf{p}$, $\mathbf{p}'$ on the state space $\mathbb{N}$. We define the random variable $t = t(\mathbf{p}, \mathbf{p}')$ to be the stopping time associated to the coupling method that combines the chains X_n^1 and X_n^2 at the first time that $X_n^1 = X_n^2 = 0$. (This is not necessarily the first time that $X_n^1 = X_n^2$.) The basic coupling inequality (2.1) for $\mathcal{M}_\pi$ and t then gives

$$\frac{1}{2}||\boldsymbol{\mu}_n - \boldsymbol{\mu}'_n|| \leq \mathrm{Prob}[t > n] , \tag{2.2}$$

where $\boldsymbol{\mu}_n$ and $\boldsymbol{\mu}'_n$ are the n-step state probabilities for the chain $\mathcal{M}_\pi$ started with initial distributions $\mathbf{p}$ and $\mathbf{p}'$. We note that the Markov chain $\mathcal{M}_\pi$ is ergodic if $E[\boldsymbol{\pi}] = \sum_{j=1}^\infty j\pi_j$ is finite, and is null-recurrent otherwise. In the ergodic case the stationary distribution $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_0, \tilde{\pi}_1, \tilde{\pi}_2, \cdots)$ is given by

$$\tilde{\pi}_j = (1 - \pi_1 - \pi_2 - \cdots - \pi_j)(1 + E[\boldsymbol{\pi}])^{-1}$$

for $j \geq 0$. This chain is ergodic for the deck distributions that we consider, and our object is to estimate the "failure probability" $\mathrm{Prob}[t > n]$.

In the remainder of the paper, rather than considering Markov chains of the type M_π, we study simplified Markov chains that jump from one key card to the next, but which retain enough information for coupling methods to apply.

3. Geometric Distribution

We consider an idealized deck consisting of cards whose labels are independently and identically distributed random variables drawn from $\mathbb{N}^+ = \{1, \ 2, \ 3, \cdots\}$ with the geometric distribution $\mathcal{G}_p$ given by $\pi_k = (1-p)p^{k-1}$, $0 < p < 1$. The geometric distribution has mean

$$E[p] \quad = \quad \sum_{k=1}^{\infty} k\pi_k \quad = \quad \frac{1}{1-p}. \tag{3.1}$$

Let $\mathcal{G}_N(p)$ denote the deck distribution induced on a deck of N cards.

Assume that the magician and subject both pick a secret number drawn from the same geometric distribution $\mathcal{G}_p$. Let $\mathrm{Prob}[t > N]$ denote the probability (choosing a deck of cards at random as above) that the magician and subject have no common key card in positions 1 through N.

For the geometric deck distribution there is a simple exact formula for all coupling probabilities.

Theorem 3.1. *For the geometric deck distribution $\mathcal{G}_N(p)$ with initial geometric value distributions $\mathcal{G}_p$,*

$$\mathrm{Prob}[t > N] = p^N(2-p)^N . \tag{3.2}$$

Proof. We use the memorylessness property of the geometric distribution, which is that for a $\mathcal{G}_p$-distributed variable X the conditional probability $\mathrm{Prob}[u = k \mid u \geq \ell]$ satisfies

$$\mathrm{Prob}[u = k \mid u \geq \ell] = \mathrm{Prob}[u = k - \ell] . \tag{3.3}$$

By direct computation

$$\mathrm{Prob}[t > 1] = 1 - (1-p)^2 = p(2-p) .$$

Now for $N \geq 2$,

$$
\begin{aligned}
\mathrm{Prob}[t > N] \quad = \quad & \mathrm{Prob}[t > N | X_1^1 \geq 2 \text{ and } X_2^1 \geq 2] \ \ \mathrm{Prob}[X_1^1 \geq 2 \text{ and } X_2^1 \geq 2] \\
& + \mathrm{Prob}[t > N | X_1^1 = 1 \text{ and } X_1^2 \geq 2] \ \ \mathrm{Prob}[X_1^1 = 1 \text{ and } X_2^1 \geq 2] \quad (3.4) \\
& + \mathrm{Prob}[t > N | \ X_1^1 \geq 2 \text{ and } X_1^2 = 1] \ \ \mathrm{Prob}[X_1^1 \geq 2 \text{ and } X_2^1 = 1] \\
& + \mathrm{Prob}[t > N | \ X_1^1 = 1 \text{ and } X_1^2 = 1] \ \ \mathrm{Prob}[X_1^1 = 1 \text{ and } X_2^1 = 1],
\end{aligned}
$$

in which the last condition $X_1^1 = X_2^1 = 1$ has zero probability for $N \geq 2$. Now by (3.3)

$$\mathrm{Prob}[t > N | X_1^1 \geq 2 \text{ and } X_1^2 \geq 2] = \ \mathrm{Prob}[t > N - 1] .$$

In the second case $X_2^2 - 1$ is geometrically distributed, hence by (3.3) again

$$\text{Prob}[t > N | X_1^1 = 1 \text{ and } X_1^2 \geq 2] = \text{Prob}[t > N - 1] \ .$$

The same holds for the third case, so (3.4) becomes

$$\begin{aligned}
\text{Prob}[t > N] &= \text{Prob}[t > N - 1]\text{Prob}[\max(X_1^1, X_1^2) \geq 2] \\
&= p(2 - p)\text{Prob}[t > N - 1] \ .
\end{aligned}$$

The theorem follows. ∎

For the geometric distribution the magician can improve his chances by always selecting the first card. Let t' denote the coupling time for this process where the subject draws his secret number from $\mathcal{G}_p$. Then one finds by a similar calculation that

$$\text{Prob}[t' > N] = p(p(2 - p))^{N-1} = p^N(2 - p)^{N-1} \ , \tag{3.5}$$

which is smaller than (3.2) by a factor $1/(2 - p)$.

4. Uniform Distribution

Consider a deck of N cards having a uniform i.i.d. distribution of card values drawn from $[1, B]$. We estimate $\text{Prob}[t > N]$ where t is the coupling time assuming that both the magician and the subject draw a secret value *uniformly* from $[1, B]$.

For our analysis we introduce two auxiliary finite state Markov chains. The first of these is a chain $\mathcal{L}_B$ that we call the *leapfrog chain.* View the subject and magician as performing the Kruskal counting procedure on two independently drawn decks. The subject will use a white pebble to mark the location of key cards and the magician will use a black pebble, according to their decks, and simultaneously each moves to their respective first key card. After this is done, the person having his pebble furthest behind in the deck moves it to his next key card. In case of a tie, where both pebbles are in the same relative position in the deck, a move consists of both persons simultaneously moving their pebbles to their next key cards, respectively. (Since the players have separate decks, the next key card values of the two players need not be the same.) The states of the chain $\mathcal{L}_B$ represent the distance the white pebble is currently ahead of or behind the black pebble in the card numbering, so there are $2B - 1$ states i with $-(B-1) \leq i \leq B - 1$. A transition occurs whenever a pebble is moved; a transition from state 0 corresponds to both pebbles moving (independently), while a transition from any other

		exit state j						
		-3	-2	-1	0	1	2	3
	3	0	0	1/4	1/4	1/4	1/4	0
	2	0	1/4	1/4	1/4	1/4	0	0
entering	1	1/4	1/4	1/4	1/4	0	0	0
state	0	1/16	2/16	3/16	1/4	3/16	2/16	1/16
i	-1	0	0	0	1/4	1/4	1/4	1/4
	-2	0	0	1/4	1/4	1/4	1/4	0
	-3	0	1/4	1/4	1/4	1/4	0	0

Table 4.1:Leapfrog Chain $\mathcal{L}_4$

state corresponds to exactly one pebble being moved. A transition often involves one pebble leapfrogging over the other, hence the choice of name for $\mathcal{L}_B$. The transition probabilities p_{ij} are determined by the uniform distribution on card values. For $i \neq 0$ the transition from i to j is determined by the value v of the key card by

$$v = \text{sign}(i)(i - j) \ , \tag{4.1}$$

so that

$$p_{ij} = \begin{cases} \dfrac{1}{B} & \text{if} \quad 1 \leq \ \text{sign}(i)(i - j) \leq B \ , \\[2ex] 0 & \text{otherwise} \ , \end{cases} \tag{4.2a}$$

while for $i = 0$ the transition probabilities are

$$\pi_j := p_{0j} = \frac{B - |j|}{B^2} \ . \tag{4.2b}$$

This chain is ergodic, and it is easy to check that π_j in (4.2b) gives the stationary distribution for $\mathcal{L}_B$. Table 4.1 gives the state transition matrix $[p_{ij}]$ for $\mathcal{L}_4$.

Now consider the case that the subject and magician perform the Kruskal counting procedure on the *same* deck. As long as their sequences of key cards remain disjoint, these key card values are independent random variables, and their relative positions of current key cards are described by transitions of the leapfrog chain. This persists until they have a key card in common, i.e. until the state 0 is reached on the leapfrog chain. Thus $\text{Prob}[t > N]$ corresponds to the probability of those sequences of transitions in the leapfrog chain starting from 0 that avoid the 0 state until one pebble has moved to a position beyond N. We can keep track of sequences that never visit 0 by forming the *reduced leapfrog chain* $\bar{\mathcal{L}}_B$ obtained by deleting the 0 state and assigning new transition probabilities

$$\bar{P}_{ij} := (1 - p_{i0})^{-1} p_{ij} \ . \tag{4.3}$$

8

		exit state j					
		-3	-2	-1	1	2	3
	3	0	0	1/3	1/3	1/3	0
	2	0	1/3	1/3	1/3	0	0
entering	1	1/3	1/3	1/3	0	0	0
state	-1	0	0	0	1/3	1/3	1/3
i	-2	0	0	1/3	1/3	1/3	0
	-3	0	1/3	1/3	1/3	0	0

Table 4.2: Reduced Leapfrog Chain $\bar{\mathcal{L}}_4$

For $\mathcal{L}_B$ the probability of going to 0 is a constant, hence

$$\bar{P}_{ij} = \left(1 - \frac{1}{B}\right)^{-1} p_{ij} \,, \tag{4.4}$$

so that all values $\bar{P}_{ij}$ are either $\frac{1}{B-1}$ or 0. Table 4.2 gives the state transition probabilities $[P_{ij}]$ for $\bar{\mathcal{L}}_4$.

The *initial state distribution* on the reduced leapfrog chain $\bar{\mathcal{L}}_B$ corresponds to that after one transition of the leapfrog chain from the 0 state, conditioned on not staying at 0, which is

$$\bar{\pi}_j := \left(1 - \frac{1}{B}\right)^{-1} \frac{B - |j|}{B^2} \,, \quad \text{for } 1 \le |j| \le B - 1 \,. \tag{4.5}$$

This chain is ergodic and has $\bar{\pi}_j$ as its stationary distribution.

We next define a random variable $t^{**}_{N,B}$ which counts the total number of key cards produced during the Kruskal count by the subject and magician, up to and including the first key card that occupies a position exceeding N. We call $t^{**}_{N,B}$ the *travel time beyond position N*. To determine the travel time, we require as additional data the position i of the *top key card,* which we define to be that key card which is closest to the top of the deck. Given that the initial state of the chain is in state j the conditional probability r_{ij} that the top key card is in position i is

$$r_{ij} = \frac{1}{B - |j|} \quad 1 \le i \le B - |j| \,, \tag{4.6}$$

and is 0 otherwise. The position of the top key card together with the sequences of successive states of $\bar{\mathcal{L}}_B$ allow the reconstruction of all moves during the Kruskal count, and the determination of the travel time $t^{**}_{N,B}$.

Lemma 4.1. *If $N \ge B$ then*

$$\text{Prob}[t > N] = \sum_{j=1}^{N} \left(1 - \frac{1}{B}\right)^{j-1} \text{Prob}[t^{**}_{N,B} = j] \,. \tag{4.7}$$

9

Proof. The event $[t > N]$ corresponds to all sequences of state transitions in $\bar{\mathcal{L}}_B$ starting at state 0 that never return to 0 before some pebble moves to a position $\geq N + 1$. Such a sequence of transitions is matched (after the first move) by corresponding state transitions in $\bar{\mathcal{L}}_B$. The probabilities between $\mathcal{L}_B$ and $\bar{\mathcal{L}}_B$ differ by a multiplicative factor $\left(1 - \frac{1}{B}\right)$. There is one less factor of $\left(1 - \frac{1}{B}\right)$ than $t_{N,B}^{**}$ counts because the initial state of $\bar{\mathcal{L}}_B$ counts as two key cards, but corresponds to only one transition in $\mathcal{L}_B$. ∎

Lemma 4.1 is useful because the distribution of the travel time $t_{N,B}^{**}$ is strongly peaked and relatively tractable to estimate. Since no move of a pebble is larger than B, and since both pebbles are within B cards of the N-th card at the stopping time $t_{N,B}^{**}$, one has

$$t_{N,B}^{**} \geq \frac{2N}{B} - 1 . \tag{4.8}$$

Lemma 4.1 then yields

$$\text{Prob}[t > N] \leq \left(1 - \frac{1}{B}\right)^{\frac{2N}{B} - 2} . \tag{4.9}$$

This shows the (well-known) fact that $\text{Prob}[t > N]$ decreases exponentially as a function of N.

Using large-deviation theory we can obtain the asymptotic behavior of $\text{Prob}[t > N]$ as $N \to \infty$.

Theorem 4.1. *For fixed B there is a positive constant α_B such that*

$$\text{Prob}[t > N] = \exp(-\alpha_B N(1 + o(1))) \tag{4.10}$$

as $N \to \infty$.

We relegate the proof of this result to the appendix, where we also give a variational formula for α_B. We easily obtain from (4.9) the inequality

$$\alpha_B \geq (\frac{2}{B})|\log(1 - \frac{1}{B})| = \frac{2}{B^2} + O(\frac{1}{B^3}). \tag{4.11}$$

It is intuitively clear that the expected value of a key card is $\geq \frac{B}{2}$ in all states, hence one expects that $\text{Prob}\left[t_{N,B}^{**} \leq \frac{N}{B}\right] \geq \frac{1}{2}$, which with Lemma 4.1 would imply that $\alpha_B \leq \frac{4}{B^2} + O(\frac{1}{B^3})$. Theorem 4.2 below shows that $\frac{B^2}{\alpha_B} \to 4$ as $B \to \infty$, see (4.19).

We next obtain upper and lower bounds for $\text{Prob}[t_{N,B}^{**} > k]$ by approximating the reduced leapfrog chain $\bar{\mathcal{L}}_B$ with two simpler Markov chains $\mathcal{L}_B^+$ and $\mathcal{L}_B^-$, as follows. These chains both describe the leapfrog motion of two colored pebbles at most B units apart, with the states representing the current distance the white pebble is ahead.

(i) In $\mathcal{L}_B^-$ the pebble further behind jumps v units with v drawn uniformly from the range $[1,\ B-1]$.

(ii) In $\mathcal{L}_B^+$ the pebble further behind jumps v units with v drawn uniformly from the range $[2,\ B]$.

The chain $\mathcal{L}_B^-$ is exactly like the leapfrog chain $\mathcal{L}_{B-1}$ except that in state 0 only the white pebble jumps. The chain $\mathcal{L}_B^-$ has $2B-1$ states labelled by $|i| \leq B-1$, while the chain $\mathcal{L}_B^+$ has $2B+1$ states labelled by $|i| \leq B$. Both chains have the property that the card values drawn are independent of the current state. For the chain $\mathcal{L}_B^-$ we define a *travel time* $t_{N,B}^-$ beyond position N, which is obtained by starting the chain in state 0, with both pebbles in position 0, associating a movement of pebbles on a line with each state transition, and counting the total number of state transitions up to and including the first time that a pebble is moved beyond position N. For the chain $\mathcal{L}_B^+$ we define a *travel time* $t_{N,B}^+$ beyond position N similarly.

Lemma 4.2. *For all N, B and k, one has*

$$\mathrm{Prob}[t_{N,B}^+ > k] \ \geq \ \mathrm{Prob}[t_{N,B}^{**} > k] \ \geq \ \mathrm{Prob}[t_{N,B}^- > k] \ . \tag{4.12}$$

We give the proof of Lemma 4.2 in the appendix. Lemmas 4.1 and 4.2 when combined yield the bounds

$$P_{N,B}^+ \geq \mathrm{Prob}[t > N] \geq P_{N,B}^- \tag{4.13}$$

where

$$P_{N,B}^\pm := \sum_{j=1}^N \left(1 - \frac{1}{B}\right)^{j-1} \mathrm{Prob}[t_{N,B}^\pm = j] \ .$$

The simple form of the chains $\mathcal{L}_B^+$ and $\mathcal{L}_B^-$ allows the asymptotic behavior of $P_{N,B}^+$ and $P_{N,B}^-$ to be explicitly determined, as follows.

Theorem 4.2. *For fixed B as $N \to \infty$ one has*

$$P_{N,B}^\pm = \exp(-\alpha_B^\pm N(1 + o(1))) \tag{4.14}$$

where $\frac{1}{2}\alpha_B^-$ is the unique root α of

$$\sum_{i=1}^{B-1} \exp(i\alpha) = B \ , \tag{4.15}$$

and $\frac{1}{2}\alpha_B^+$ is the unique root of

$$\sum_{i=1}^{B-1} \exp((i+1)\alpha) = B \ . \tag{4.16}$$

11

As $B \to \infty$ these quantities satisfy

$$\alpha_B^+ = \frac{4}{B^2} - \frac{20/3}{B^3} + O(B^{-4}) \tag{4.17}$$

$$\alpha_B^- = \frac{4}{B^2} + \frac{4/3}{B^3} + O(B^{-4}). \tag{4.18}$$

The proof of this result is given in the appendix. Theorem 4.1 together with the inequalities (4.13) shows that for large B one has

$$\mathrm{Prob}[t > N] = \exp\left(-\left(\frac{4}{B^2} + O\left(\frac{1}{B^3}\right)\right)(1 + o(1))N\right) \tag{4.19}$$

as $N \to \infty$.

We relate these results to the model of Haga and Robins [9]. The Markov chain studied by Haga and Robins is obtained from the leapfrog chain by identifying states k and $-k$ for all $k \geq 1$; thus it has exactly B states. The resulting chain factors out the action of the involution sending k to $-k$ under which the chain probabilities are invariant, and this loses the "leapfrog" information which is necessary for computing exact coupling probabilities. Haga and Robins estimate instead the probability of avoiding absorption in the absorbing state 0 in the first M transitions of the resulting factor chain. This probability asymptotically decays like $O((\lambda_B)^M)$ as $M \to \infty$, where λ_B is the modulus of the second largest eigenvalue of the characteristic polynomial of their Markov chain. The characteristic polynomial of the transition matrix of the Haga-Robins Markov chain is $p_B(x) := (x + 1/B)^B - (1 + 1/B)^B x^{B-1}$, and it can be shown that the modulus of its second largest eigenvalue satisfies

$$\lambda_B = 1 - \frac{2}{B} + O(\frac{1}{B^2}) \tag{4.20}$$

as $B \to \infty$. To relate λ_B to the asymptotic coupling probability decay rate $exp(-\alpha_B)$ in Theorem 4.1, we note that the expected size of a step in the Haga-Robins chain is about $B/2$, so that after M steps the location of the chain should be around the position $N \approx MB/2$. One should therefore compare $(\lambda_B)^{2/B}$ and $exp(-\alpha_B)$, and one finds that both of these quantities are asymptotic to $1 - \frac{4}{B^2} + O(\frac{1}{B^3})$ as $B \to \infty$, using (4.20) and Theorem 4.2.

5. Numerical Results: The Kruskal Count

We compare predictions obtained from the two models studied in this paper with the performance of the actual Kruskal Count trick.

For the actual Kruskal count we consider a standard deck of 52 cards, and we assume that the subject draws a key card using a uniform distribution from the set of available key card values. We study the effects of varying the magician's strategy on the success probability of the Kruskal Count trick. The magician has the freedom to choose his key card, and he has also has the extra freedom to specify a rule for assigning values to the "face cards" J, Q, K. We study three possible rules variations:

(a) Assign the values 11, 12, 13 to J, Q, K, respectively.

(b) Assign the value 10 to each of J, Q, K,

(c) Assign the value 5 to each of J, Q, K.

The first two of these rules variations are presented as "straw men" useful for comparison with the models of this paper. To obtain numerical values for the Kruskal count trick we used a Monte Carlo simulation with 10^6 trials for each data point. For simulations of the i.i.d. uniform deck distribution, an "exact" calculation was done using an enlarged Markov chain which kept a running total of the value of the position N of the leading pebble, and with an absorbing state whenever a pebble jumps past the end of the deck. Since the smallest step size is 1, this chain reaches an absorbing state after a number of steps equal to the size of the deck; consequently, we it suffices to compute the state of the chain after that number of steps. Simulations of the i.i.d. "semiuniform" distributions for rules variations (b) and (c) were done similarly to the i.i.d. uniform case.

The rules variation (a) corresponds to the uniform distribution on $\{1, 2, ..., 13\}$. The average key card size is 7. We therefore consider as an approximation the i.i.d. geometric deck distribution with $p = \frac{6}{7}$, which has mean key card size 7. According to Theorem 3.1 the failure probability for $N = 52$ the magician drawing his first key card according to the geometric distribution is

$$FP(\mathcal{G}_a) = (\frac{6}{7})^{52}(\frac{8}{7})^{52} = 0.342254. \tag{5.1}$$

If the magician chooses the first card to be his first key card, by (3.5) his failure probability

	Kruskal	uniform
1	0.315180	0.319486
2	0.318564	0.322994
3	0.321975	0.326492
4	0.325298	0.329981
5	0.328794	0.333461
6	0.332235	0.336929
7	0.336055	0.340385
8	0.339264	0.343827
9	0.342758	0.347251
10	0.346780	0.350655
11	0.349464	0.354034
12	0.353367	0.357385
13	0.357044	0.360703
avg	0.335906	0.340276

Table 5.1: Failure probabilities for rules variation (a)

for $N = 52$ is

$$FP(\mathcal{G'}_a) = \frac{7}{8}FP(\mathcal{G}_a) = 0.299472 \tag{5.2}$$

Table 5.1 presents data for rules variation (a) for the Kruskal Count and the i.i.d. uniform deck distribution on $\{1, 2, ..., 13\}$. The table gives failure probabilities in which the magician's strategy is to choose as first key card the j-th card, for $1 \leq j \leq 13$. plus a final row that gives the failure probability when the magician draws a card uniformly in $\{1, 2, ...13\}$. The data in Table 5.1 show that the magician does best to choose $j = 1$ as his key card. The non-Markovian nature of the actual deck causes the failure probabilities to differ from the i.i.d. uniform deck distribution; the effect is a decrease of about 0.3%. We also see that the failure probability for the i.i.d. geometric distribution is an overestimate of the failure probability for the Kruskal Count when the magician picks a random card as first key card, and underestimates the failure probability when the magician picks the first card as key card.

We next consider the rules variations (b) and (c). For rules variation (b) the expected key card size is $\frac{85}{13}$, so for comparison we consider the i.i.d. geometric deck distribution with $p = \frac{72}{85}$. If the magician chooses his first key card according to the same geometric distribution, then the failure probability is

$$FP(\mathcal{G}_b) = (\frac{72}{85})^{52}(\frac{98}{85})^{52} = 0.292064, \tag{5.3}$$

	Kruskal (b)	semiuniform (b)	Kruskal (c)	semiuniform (c)	uniform
1	0.277869	0.284060	0.146238	0.152658	0.150944
2	0.280756	0.287235	0.148801	0.155266	0.153684
3	0.284330	0.290447	0.151204	0.157847	0.156407
4	0.287163	0.293623	0.153736	0.160399	0.159109
5	0.290317	0.296782	0.156075	0.162918	0.161789
6	0.293557	0.299920	0.159744	0.166357	0.164444
7	0.296910	0.303034	0.162474	0.168973	0.167070
8	0.300023	0.306118	0.164977	0.171553	0.169665
9	0.303194	0.309171	0.167735	0.174094	0.172225
10	0.306383	0.312185	0.170064	0.176591	0.174747
avg	0.292050	0.298258	0.158105	0.164666	0.163008

Table 5.2: Failure probabilities for rules variations (b) and (c)

while if the magician draws the first card as his key card, then

$$FP(\mathcal{G}'_b) = \frac{85}{98}FP(\mathcal{G}_b) = 0.253320. \tag{5.4}$$

For rules variation (c) the expected key card size is $\frac{70}{13}$, so for comparison we consider the i.i.d. geometric deck distribution with $p = \frac{57}{70}$. If the magician chooses his first key card with the same geometric distribution, then the failure probability is

$$FP(\mathcal{G}_c) = (\frac{57}{70})^{52}(\frac{83}{70})^{52} = 0.161197, \tag{5.5}$$

while if the magician chooses the first card as his key card, the failure probability is

$$FP(\mathcal{G}'_c) = \frac{70}{83}FP(\mathcal{G}_c) = 0.135949. \tag{5.6}$$

Table 5.2 presents failure probability data for rules variations (b) and (c) for the Kruskal Count and for the i.i.d. semiuniform deck distributions which have the card values $\{1, 2, ..., 10\}$ chosen with the same probabilities as rules variations (b) and (c) impose on the actual deck. The non-Markovian nature of the actual deck results in the Kruskal count failure probabilities differing from the corresponding i.i.d. deck distributions; they are smaller by about 0.6%. The failure probability for the i.i.d. geometric distribution when the magician chooses the first card as first key card gives an underestimate for the failure probabilities of the Kruskal Count in rules variations (b) and (c). The numerical results show that the magician should choose the first card as his key card. The effect of the choice of the magician's key card on the failure probability is small, at most 2.5%. In comparing rules variations (b) and (c) we see that the

choice to have face cards take the value 5 rather than 10 has a much larger effect on the failure probability than the magician's choice of first key card position. The final column of Table 5.2 presents the failure probabilities for the i.i.d. uniform deck distribution on $\{1, 2, ..., 10\}$. One expects this i.i.d. uniform distribution to be comparable with rules variation (c) rather than (b), because the expected key card size is similar to case (c). (The Kruskal Count (c) mean value is slightly lower.)

To conclude: The rules variation to count face cards as having value 5 rather than 10 is important to the success of the Kruskal Count trick in practice; the choice of the first card as key card offers a further small improvement in success probability.

6.   Appendix: Proofs of Theorem 4.1, Lemma 4.2 and Theorem 4.2

Proof of Theorem 4.1. In view of Lemma 4.1, one has

$$M_N \leq \mathrm{Prob}[t < N] \leq NM_N \tag{6.1}$$

where

$$M_N := \max_{1 \leq k \leq N} \left\{ \left(1 - \frac{1}{B} \right)^k \mathrm{Prob}\left[t^{**}_{N,B} = k \right] \right\} . \tag{6.2}$$

The maximum will occur with $k \approx \gamma N$ for some constant $\gamma = \gamma(N, B)$. To estimate γ, we note that the travel time $t^{**}_{N,B}$ beyond position N depends on the successive transitions of the chain $\bar{\mathcal{L}}_B$. We convert this to a problem about successive states of the *jump chain* $\bar{\mathcal{L}}^J_B$ having $2B(B-1)$ states which correspond to all possible transitions of the chain $\bar{\mathcal{L}}_B$. A jump chain state (i, v) will mean state i of $\bar{\mathcal{L}}_B$ together with an allowable key card value v which determines the next state of $\bar{\mathcal{L}}_B$. The allowable values are $1 \leq v \leq B$ with $v \neq |i|$. The transition probability $P_{s,s'}$ from $s = (i, v)$ to $s' = (j, v')$ is $\frac{1}{B-1}$ when j is uniquely determined by (4.1) and $1 \leq v' \leq B$ with $v' \neq j$, and is 0 otherwise.

We let $\{(i_k, v_k) : k+1, 2, \cdots\}$ denote a sequence of states of $\bar{\mathcal{L}}^J_B$, and introduce the modified travel time

$$\tilde{t}_{N,B} := \min\{k : v_1 + \ldots + v_k \geq 2N\} . \tag{6.3}$$

Now $\mathrm{Prob}[\tilde{t}_{N,B} \leq \gamma N]$ can be estimated using large deviation theory, using the following special case of Theorem 1 of Donsker and Varadhan [4].

Theorem A.1. *For fixed B and γ one has as $N \to \infty$,*

$$\mathrm{Prob}[\tilde{t}_{N,B} \leq \gamma N] = \exp(-f_B(\gamma)N + o(N)) \tag{6.4}$$

16

where the function $f_B(\gamma) = \gamma I^*(\gamma)$ where

$$I^*(\gamma) := \inf\{I(\mu) : \text{weight}(\mu) \geq \frac{2}{\gamma}\} \ . \tag{6.5}$$

Here μ runs over the set of probability measures on the state space S of the chain $\bar{\mathcal{L}}_B^J$, and

$$\text{weight}(\mu) := \sum_{s=(i, \ v)\in S} v\mu((i, \ v)) \ ,$$

is the expected card value for the measure μ, and

$$I(\mu) := -\inf\left\{\sum_{s\in S} \log\left(\frac{\pi_u(s)}{u(s)}\right) \mu(s) : u : S \to \mathbb{R}^+\right\} \ , \tag{6.6}$$

where $\pi_u(s) := \sum_{s'\in S} p_{s,s'}u(s')$.

It is easy to show that

$$t_{N,B}^{**} \leq \tilde{t}_{N,B} \leq t_{N,B}^{**} + B \ ,$$

and this yields

$$\text{Prob}[\tilde{t}_{N,B} \leq k] \leq \text{Prob}[t_{N,B}^{**} \leq k] \leq \text{Prob}[\tilde{t}_{N,B} \leq k - B] \ . \tag{6.7}$$

Using Theorem A.1 and (6.7) we see that the quantities $t_{N,B}^{**}$ and $\tilde{t}_{N,B}$ have the same asymptotic behavior, with

$$\text{Prob}[t_{N,B}^{**} \leq \gamma N] = \exp(-f_B(\gamma)N + o(N)) \tag{6.8}$$

as $N \to \infty$. Now (6.2) leads us to define

$$-\alpha_B := \max_{0<\gamma<\frac{B+1}{2}} \left[\gamma \log\left(1 - \frac{1}{B}\right) - f_B(\gamma)\right] \ . \tag{6.9}$$

Using the fact that $f_B(\gamma)$ is a strictly convex function, it can be shown there is a unique value of γ attaining the maximum on the right-hand side. With some further work this fact and (6.1), (6.2) and (6.8) imply that

$$M_N = \exp(-\alpha_B N + o(N))$$

as $N \to \infty$. The bound of Theorem 4.1 follows. ∎

Proof of Lemma 4.2. We exhibit a $1 - 1$ correspondence between a sequence of states and admissible transitions for the three chains, corresponding to moving two colored pebbles on the line $\{n : n \geq 0\}$ starting with both at zero.

We compare $\mathcal{L}_B^-$ and $\bar{\mathcal{L}}_B$. The position after two pebble moves of $\mathcal{L}_B^-$ corresponds to an initial position for $\bar{\mathcal{L}}_B$ together with the top key card and its color. Let q_{ij}^-, q_{ij} denote the associated probability distributions of the locations (i, j) of the white and black pebbles for the chains $\bar{\mathcal{L}}_B$, $\mathcal{L}_B^-$ respectively, which are: $q_{ij}^- + \frac{1}{(B-1)^2}$ for $1 \le i,\ j \le B-1$ and 0 otherwise, $q_{ij} + \frac{1}{B(B-1)}$ for $1 \le i,\ j \le B$ with $i \ne j$ and 0 otherwise.

Claim. *There is a mapping ϕ^- of the probability mass q_{ij} on $(i,\ j)$ for $\bar{\mathcal{L}}_B$ to various (i', j') having $i' \le i$ and $j' \le j$ whose image is the distribution q_{ij}^-.*

Let $(i,\ j) \succ (i',\ j')$ mean $\min(i,j) \ge \min(i',j')$ and $\max(i,j) \ge \max(i',j')$, i.e. the pebbles $(i,\ j)$ are both moved further along the line then $(i',\ j')$, ignoring their colors. Assuming the claim to be true for the moment, we have a (stochastic) pairing of pebble positions $(i,\ j) \xrightarrow{\phi} (i',\ j')$ such that $(i,\ j) \succ (i',\ j')$ between $\bar{\mathcal{L}}_B$ and $\mathcal{L}_B^-$. For each subsequent move, both claims have $B-1$ possible transitions with probabilities each $\frac{1}{B-1}$. For $\bar{\mathcal{L}}_B$ in state $k+i-j$ the admissible value of the next move is $\{1,\ 2, \ldots, B\} - \{|k|\}$. We map these transitions to transitions of $\mathcal{L}_B^-$ in linear order, with a mapping $\psi_{|k|}$ having $\psi_{|k|}(i) = i$ for $i \le |k|$ and $\psi_j(i) + i - 1$ for $i \ge |k| + 1$. One easily sees that if pebbles in $\bar{\mathcal{L}}_B$ are at $(i_1,\ j_1)$ and the corresponding ones are at $(i_2,\ j_2)$ with $(i_1,\ j_1) \succ (i_2,\ j_2)$, and if the pebble closer to the origin is moved i resp. $\phi_{|k|}(i)$ for the two claims resulting in positions $(i_1^*,\ j_1^*)$, $(i_2^*,\ j_2^*)$ then $(i_1^*,\ j_1^*) \succ (i_2^*,\ j_2^*)$. This gives a stochastic pairing of pebble positions at all subsequent moves, with both pebbles of $\mathcal{L}_B^-$ always being behind those of $\bar{\mathcal{L}}_B$ in the ordering $\succ$. Consequently

$$\mathrm{Prob}[t_{N,B}^{**} > t] \ge \mathrm{Prob}[t_{N,B}^- > t] \tag{6.10}$$

for all t, which is the right side of (4.12).

It remains to prove the claim. Here we remark if $q_{ij}(n)$ and $q_{ij}^-(n)$ denote the probability distribution of pebble locations of $\bar{\mathcal{L}}_B$ and $\mathcal{L}_B^-$ after n moves then

$$\mathrm{Prob}[t_{N,B}^{**} > t] = 1 - \sum_{\substack{i \le N \\ j \le N}} q_{ij}(t)$$

with a similar formula for $\mathrm{Prob}[t_{N,B}^- > t]$. The proof above actually establishes the majorization inequalities

$$\sum_{\substack{i \le i_0 \\ j \le j_0}} q_{ij}^-(t) \ge \sum_{\substack{i \le i_0 \\ j \le j_0}} q_{ij}(t) \ , \quad \text{all} \quad i_0 \ge 1,\ j_0 \ge 1\ , \tag{6.11}$$

for all $t \ge 2$, and the special case $i_0 = j_0 + N$ yields (6.10). The claim to be established is equivalent to proving that (6.11) holds for the case $t = 2$. Since the probabilities $q_{ij} = q_{ij}(2)$

18

and $q'_{ij} = q'_{ij}(2)$ are explicitly known, verifying (6.11) is an easy calculation. The equivalence of the inequalities (6.11) to the existence of a coordinate-monotone probability rearrangement ϕ is a two-dimensional majorization inequality, see Marshall and Olkin [10]. (One can also prove the claim by explicitly constructing a suitable mapping ϕ rather easily.)

The inequality (4.12) relating $\bar{\mathcal{L}}_B$ and $\mathcal{L}_B^+$ is proved in similar fashion. If $q_{ij}^+(t)$ is the probability that the pebbles are at (i, j) after t steps, then

$$\sum_{\substack{i \leq i_0 \\ j \leq j_0}} q_{ij}(t) \geq \sum_{\substack{i \leq i_0 \\ j \leq j_0}} q_{ij}^+(t) , \quad \text{all } i_0 \geq 1 , \quad j_0 \geq 1 , \tag{6.12}$$

for all $t \geq 2$. ∎

Proof of Theorem 4.2. We let $A(N) \approx B(N)$ mean $A(N) = B(N)^{1+o(1)}$ as $N \to \infty$.

Consider first $P_{N,B}^-$. Let $\tilde{t}_B^-(M)$ denote the travel time for the chain $\mathcal{L}_B^-$, which counts the number of transitions up to and including the transition at which the sum of the jumps of the chain exceeds M. Then for any fixed sequence of transitions

$$t_B^-(2N) \geq t_{N,B}^- \geq t_B^-(2N - N) . \tag{6.13}$$

Hence

$$P_N^- \leq N \max_{1 \leq j \leq N} \left\{ \left(1 - \frac{1}{B} \right)^{j-1} \text{Prob}[\tilde{t}_B^-(2N) + j] \right\} \tag{6.14}$$

and

$$P_N^- \geq \max_{1 \leq j \leq N} \left\{ \left(1 - \frac{1}{B} \right)^{j-1} \text{Prob}[\tilde{t}_B^-(2N - B) = j] \right\} . \tag{6.15}$$

It's easy to check that

$$Q_N^- := \max_{1 \leq j \leq N} \left\{ \left(1 - \frac{1}{B} \right)^{j-1} \text{Prob}[\tilde{t}_B^-(2N) = j] \right\}$$
$$\approx \max_{1 \leq j \leq N} \left\{ \left(1 - \frac{1}{B} \right)^{j-1} \text{Prob}[\tilde{t}_B^-(2N - B) + j] \right\} \tag{6.16}$$

using $\text{Prob}[t_B^-(2N - B) \geq j] \geq \text{Prob}[t_B^-(2N) \geq j + B]$. It suffices to asymptotically estimate Q_N^-. One has

$$Q_N^- \approx \sup \left\{ B^{-\gamma N} \left(\frac{\gamma N}{\gamma_1 N \gamma_2 N \cdots \gamma_{B-1} N} \right) : \gamma_i \geq 0 , \quad \sum_{i=1}^{B-1} \gamma_i = \gamma , \quad \sum_{i=1}^{B-1} i \gamma_i = 2 \right\} . \tag{6.17}$$

19

(Here $j \approx \gamma N$ and $B^{-\gamma N}$ arises as $\left(1 - \frac{1}{B}\right)^{\gamma N} (B-1)^{-\gamma N}$.) Using Stirling's formula, one obtains $Q_N^- \approx \exp(\frac{1}{2}\alpha_B^- N)$ where $\frac{1}{2}\alpha_B^-$ is the optimal value of the constrained maximization problem (M^-) given by:

$$\text{maximize } Z = -\gamma \log B + \gamma \log \gamma - \sum_{i=1}^{B-1} \gamma_i \log \gamma_i$$

subject to

$$\sum_{i=1}^{B-1} i\gamma_i = 2 \ , \tag{6.18}$$

$$\sum_{i=1}^{B-1} \gamma_i = \gamma \ , \tag{6.19}$$

$$\gamma_i \geq 0 \quad \text{for} \quad 1 \leq i \leq B-1 \ . \tag{6.20}$$

Introducing Lagrange multipliers λ_1, λ_2 for the two equality constraints, and setting

$$G = \gamma \log B + \gamma \log \gamma - \sum_{i=1}^{B-1} \gamma_i \log \gamma_i - \lambda_1 \left(\sum_{i=1}^{B-1} i\gamma_i - 2\right) - \lambda_2 \left(\sum_{i=1}^{B-1} \gamma_i - \gamma\right) \ .$$

Necessary conditions for an interior extremal are:

$$\frac{\partial G}{\partial \gamma} = 1 + \log \gamma - \log B + \lambda_2 = 0 \ , \tag{6.21}$$

$$\frac{\partial G}{\partial \gamma_i} = -1 - \log \gamma_i - i\lambda_1 - \lambda_2 = 0 \ . \tag{6.22}$$

These yield

$$\gamma = B \exp(\lambda_2 - 1) \ , \tag{6.23}$$

$$\gamma_i = \exp(\lambda_2 - 1)\exp(i\lambda_1) \ , \quad 1 \leq i \leq B-1. \tag{6.24}$$

Substituting these expressions into (6.19) and cancelling $\exp(\lambda_2 - 1)$ from both sides yields

$$\sum_{i=1}^{B-1} \exp(i\lambda_1) = B \ . \tag{6.26}$$

Substituting the values above into (6.21) and using this formula yields

$$\exp(\lambda_2 - 1) = \frac{2}{\sum_{i=1}^{B-1} i\exp(i\lambda_1)} \tag{6.27}$$

Now (6.23) and (6.24) become

$$\gamma = \frac{2B}{\sum_{i=1}^{B-1} i\exp(i\lambda_1)}$$

$$\gamma_i = \frac{2i\exp(i\lambda_1)}{\sum_{i=1}^{B-1} i\exp(i\lambda_1)} \ .$$

Using these formulas the objective function value Z is evaluated (with $F = \sum_{i=1}^{B-1} i\exp(i\lambda_1)$)

as

$$
\begin{aligned}
Z &= \frac{2B}{F}\left(\log\frac{2B}{F} - \log B\right) - \sum_{i=1}^{B-1} \frac{2\exp(i\lambda_1)}{F}\log\left(\frac{2\exp(i\lambda_1)}{F}\right) \\
&= \frac{2B}{F}\log\frac{2}{F} - \frac{2}{F}\sum_{i=1}^{B-1}\exp(i\lambda_1)\left[\log\frac{2}{F} + i\lambda_1\right] \\
&= \frac{2\lambda_1}{F}\sum_{i=1}^{B-1} i\exp(i\lambda_1) = -2\lambda_1 \ . \tag{6.28}
\end{aligned}
$$

Combining this with (6.26) gives (4.14) for $P_{N,B}^-$ with $\frac{1}{2}\alpha_B^-$ given by (4.15), provided the maximum of (M^-) occurs at an interior point where all $\gamma_i > 0$. We omit the details of checking that boundary extremals having some $\gamma_i = 0$ do not give the absolute maximum in (M^-).

The case of $P_{N,B}^+$ is handled by analogous arguments. One reduces it to solving the constrained maximization problem (M^+) given by:

$$\text{maximize}\ \ Z = -\gamma\log B + \gamma\log\gamma - \sum_{i=1}^{B-1}\gamma_i\log\gamma_i$$

subject to

$$\sum_{i=1}^{B-1}(i+1)\gamma_i = 2 \ , \tag{6.29}$$

$$\sum_{i=1}^{B-1}\gamma_i = \gamma \ , \tag{6.30}$$

$$\gamma_i \geq 0 \ \ \text{for}\ \ 1 \leq i \leq B \ . \tag{6.31}$$

Again $Z = -2\lambda_1$ at the extremal point, and λ_1 is determined by

$$\sum_{i=1}^{B-1}\exp(i+1)\lambda_i = B \ ,$$

which is (4.16).

The asymptotic formulae (4.17) and (4.18) are obtained from the formulas (4.15) and (4.16), by setting $\exp(\frac{1}{2}\alpha) = 1 + \frac{2}{B^2} + \frac{\delta}{B^3} + O(B^{-4})$, in which δ is an unknown to be determined. We find it by noting that

$$\exp(\tfrac{1}{2}i\alpha) = 1 + i\left(\frac{2}{B^2} + \frac{\delta}{B^3}\right) + i^2\frac{2}{B^4} + O\left(B^{-3}\left(1 + \frac{i}{B}\right)^{-1}\right) ,$$

and solving for the appropriate value of δ. $\blacksquare$

References

[1] D. Aldons and P. Diaconis, Shuffling Cards and Stopping Times, Amer. Math. Monthly **93** (1986), 333–348.

[2] P. Diaconis, *Group Representations in Probability and Statistics,* IMS Lecture Notes – Monograph Series No. **11**, Institute of Math. Statistics, Hayward, Calif. 1988.

[3] W. Doeblin, Exposé de la theorie des chaines simple constantes de Markov á un nombre fini d'etats, Rev. Math Union Interbalkanique **2** (1938), 77–105.

[4] M. D. Donsker and S. R. S. Varadhan, Asymptotic Evaluation of Certain Markov Process Expectations for Large Time I, Comm. Pure. Appl. Math. **28** (1975), 1–47.

[5] C. Fulves and M. Gardner, The Kruskal Principle, The Pallbearers Review, June 1975.

[6] M. Gardner, Mathematical Games, Sci. Amer. **238** (1978) No. 2 (February), 19–32.

[7] M. Gardner, *From Penrose Tiles to Trapdoor Ciphers,* W. H. Freeman Co., New York, 1988. (Chapter 19)

[8] D. Griffeath, Coupling Methods for Markov Processes, in: *Studies in Probability and Ergodic Theory* (G. C. Rota, Ed.), Academic Press, New York, 1978, pp. 1–43.

[9] W. Haga and S. Robins, On Kruskal's Principle, in *Organic Mathematics*, (J. Borwein, P. Borwein, L Jorgenson, and R. Corless, Eds.), Canadian Math. Society Conference Proceedings, vol. 20, AMS: Providence, RI, 1997, pp. 407-412.

[10] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications,* Academic Press, New York, 1979.

[11] C. L. Mallows, On a probability problem suggested by M. D. Kruskal's card trick, Bell Laboratories memorandum, April 18, 1975, unpublished.

[12] J. Pitman, On coupling of Markov Chains, Z. Wahrscheinlichkeitheorie **35** (1976), 315–322.

```
email:  jcl@research.att.com
        rains@research.att.com
        rvdb@princeton.edu
```