# The Use of Single Position Observations in Evaluating Mixing Times of Card Shuffles

Drew B. Sinha (407547)
Instructor: Professor Renato Feres
Submitted as a Requirement for Math 350 – Topics in Applied Math.: Monte Carlo Methods
19 December 2012

## Introduction

Given a stochastic process with a particular set of initial conditions (a probability distribution over some set of states), the speed at which the process approaches its steady-state distribution is a property of substantial practical importance. For instance, in the case of Monte Carlo Markov Chain simulation, the time necessary to converge to the desired (stationary) distribution that the user is attempting to sample from defines a lower bound for the run time necessary to obtain an sufficient number of realizations for the process {Diaconis, 1993}. More popularly, this problem has been addressed in the context of shuffling cards, in which a user starts from an unshuffled deck and considers the question of how many shuffles are necessary to thoroughly mix the deck {Aldous and Diaconis, 1986; Bayer and Diaconis, 1992; Diaconis, 1993; Mann}. Formally, this represents the amount of time necessary for the probability distribution over all possible orderings of cards in the deck to converge to the uniform distribution.

The notion of convergence implies the concept of distance in the space of probability distributions over card orderings (permutations). Commonly, this distance is defined in terms of the total variation norm, essentially an $L_1$ norm that defines the distance by considering the distributions as vectors {Aldous and Diaconis, 1987; Jonsson and Trefethen, 1997}. This norm is an intuitive way to understand the difference between distributions from a linear vector space perspective, particularly in the case of Markov chain processes. It is equally possible, however, to consider other distance metrics, such as other linear algebra inspired metrics (such as the $L_2$ matrix norm) or even information theoretic divergences/metrics.

In this project, the question of mixing time for card shuffling was assessed for a deck using both total variation and an information-based metric known as variation of information. These metrics were used to assess the effective mixing times observed when using the mixing of at a single card position as a proxy for mixing of the deck due to two different shuffles – the riffle shuffle and the overhand shuffle.

## Methods

Two types of shuffles were considered in this work. The first shuffle is a well-studied model of shuffling known as the riffle shuffle, credited to Gilbert, Shannon and Reeds and an algorithm is given in {Aldous and Diaconis, 1987}. This model of shuffling abstracts the traditional shuffle in which the deck is separated randomly into two piles and then interleaved together.

Algorithm 1 (Riffle Shuffle):

1. For a set of n cards, choose a value $c$ out of the binomial distribution Bin(n,1/2). Partition the deck into two subsets of size $c$ and $n$-$c$.
2. While there are still cards in both of these "subdecks"

   Choose the $i^{th}$ card of the new, shuffled deck as the bottom card of one of the subdecks with probability equivalent to the number of cards remaining in the deck (call them $c_1$ and $c_2$) divided by the total cards remaining in both decks:

$$P_{i,Deck\ 1} = \frac{c_1}{c_1 + c_2} \quad and \qquad P_{i,Deck\ 2} = \frac{c_2}{c_1 + c_2}$$

3. (Finishing Condition) When one of the subdecks has been exhausted, place the non-exhausted subdeck on top of the pile.

Algorithm 2 (Overhand Shuffle)

1. Choose a number of cut points in the deck of $n$ cards by choosing a number $n_{cuts}$ from a binomial distribution Bin(n,p) where p = 1/8. Then, sample a set $c_{pts}$ of size $n_{cuts}$ from a uniformly random permutation of the set {1,2, . . ., n-1} and sort it in ascending order.
2. For each partition of the deck separated by the cut points, reverse the order of cards within the partition. Assemble the new shuffled deck by combining the all of the partitions in the order in which they appeared in the original deck.

each process has dimension 52! x 52!. Since the dimension of this state space is extremely large, two methods were used assess the degree of mixing of the deck. First, the mixing of the deck was approximated by considering the mixing of a single card position (the 1st , 13th , or 26th card). Second, for the case of the riffle shuffle, it has been shown that the dimensionality of the state space can be reduced to a state space of order n by reformulating the shuffle as a Markov chain on the set of rising sequences within the deck (maximal subsets of the deck in which orderings of cards are increasing in value) {Jonsson and Trefethen, 1997}. This was used to compare the use a single card to approximate the mixing of the deck in the case of the riffle shuffle.

Finally, two metrics were used to assess deviation of the probability distribution over deck permutations from the uniform distribution. First, the total variation norm is defined for a matrix as:

$$\|P_1 - P_2\|_1 = \frac{1}{2}\max_i \sum_{j=1}^{m} \left|P_{1,ij} - P_{2,ij}\right|$$

The total variation matrix norm was used to calculate the analytical solution found by using the method suggested by {Trefthen, 1997}. In the case, where a single card was queried for mixing, $A_1$ and $A_2$ reduce to vectors and the relation above reduces to the $L_1$ norm of the difference. The second metrics used was the variation of information:

$$I = |H(P_1) - H(P_2)|$$

where H denotes the entropy of the distribution. This definition is different that that given by other authors due to absolute value added to make $I$ a legitimate metric (versus just a divergence if the absolute value was not present). When using each of these metrics for the case where a single card is queried, the probability distribution associated with the identity of the card was produced for each shuffle by using a Monte Carlo Monte Carlo sampling method to simulate $10^4$ decks shuffled with each type of shuffle (code given in Appendix).

## Results

### *Using One Card as a Proxy for Mixing in the Riffle Shuffle*

Shuffling a deck of cards can be described as a Markov chain over the set of permutations of the deck whose distribution asymptotically approaches the uniform distribution. In general, for an n-card deck, the dimension of the transition matrix for this process is n! x n!. This makes analysis of the process substantially difficult for the realistic case of n = 52 cards. To bypass the issue of high-dimensionality, the use of mixing for a single card was considered as a proxy for the mixing of the deck.

To assess the consequences of using this approximation, single card mixing during riffle shuffles (simulated by Markov Chain Monte Carlo simulation – MCMC) was compared to mixing of the deck from an analytical solution given by {Jonsson and Trefethen, 1997} using the total variation metric, shown in Figure 1. It is evident that the mixing of any single card in the deck is a lower bound for the mixing occurring in the entire deck. Further, three positions in the deck were queried for comparison to assess whether the observed mixing of a single card was dependent on which position is queried. As shown, mixing times are shorter for cards closer to the center of the deck when only a few shuffles are done, though on the first shuffle, all cards are mixed equally well.

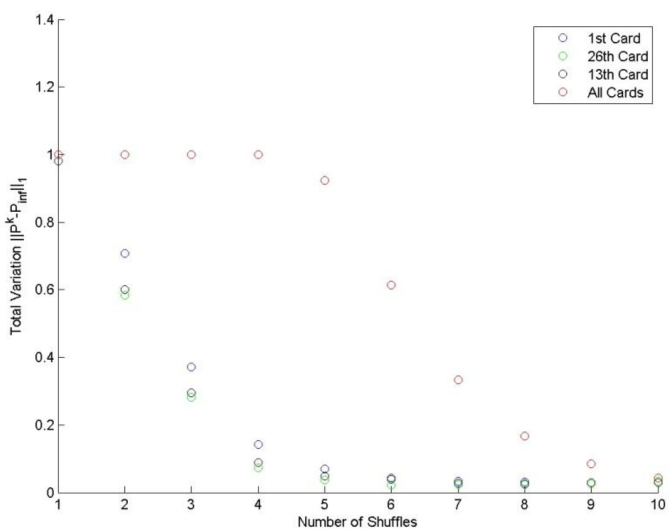


Figure 1: Single Card Mixing is Faster Closer to the Center of the Deck and Lower Bounds Mixing of the Deck. Three card positions in a 52 card deck were queried for degrees of mixing as a function of the number of shuffles as measured by total variation metric, comparing the distribution of singles cards to the uniform distribution (blue, green, and black dots). Single card distributions were obtained by MCMC simulation of $10^4$ decks. An analytic solution is given for the entire deck (red dots) as given by {Jonsson and Trefethen, 1997}.

*Use of Variation of Information as a Metric for Mixing*

The distribution of single cards was then used to determine the degree of mixing using an information theoretic metric known as the variation of information, in which distance is quantified by the absolute value of the difference between the entropy of the distribution of the single card $X_i$ and the entropy of the uniform distribution. The results of MCMC simulation are shown in Figure 2. In general, the same qualitative trends found in Figure 1 between the degree of mixing and card position still apply. By the 6$^{th}$ shuffle, nearly all information related to the identity of the queried card has been lost.

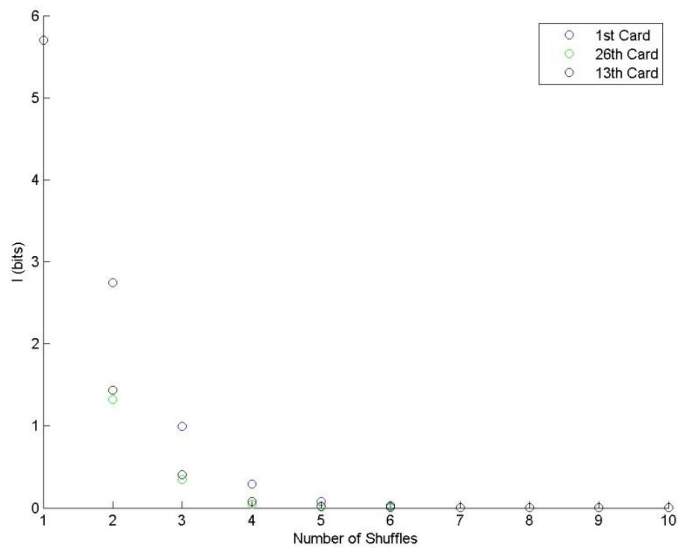


Figure 2: Single Card Mixing Destroys Information on Card Identity by the 6$^{th}$ Shuffle. Cards closer to the center of the deck mix faster than those on the end. Mixing was assessed by the variation of information method (measured relative to the asymptotic uniform distribution).

Finally, using these two metrics, the mixing of a deck was assessed during repeated overhand shuffles by querying single card positions. For the total variation metric (Figure 3A), the deck still has a total difference in probability of 0.2 after 10 shuffles, whereas, for the variation of information metric (Figure 3B), nearly all of the information encoded by the identity of a single card has been lost by the tenth shuffle. The same qualitative relationships between the position of the card and the relative deviation from the uniform distribution are the same as in the case of the riffle shuffle.

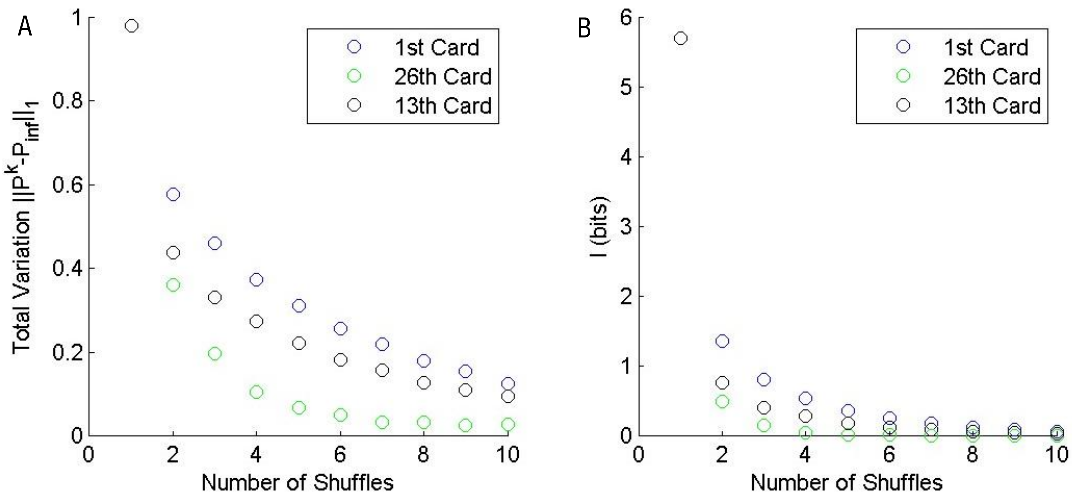


Figure 3: Single Card Mixing for an Overhand Shuffle. Deviation of the deck ordering from the uniform distribution as measured by (A) the total variation and (B) the variation of information metric.

## Discussion

In this project, the mixing of cards in a deck was explored for two different shuffles and was assessed with two different metrics of probabilistic distance. This was achieved primarily by using the approximation afforded by observations of mixing for single card positions within the deck.

It was found that the use of a single card approximation established a lower bound on the time needed for mixing. Moreover, the mixing of a single card showed qualitative differences compared to the analytic solution given by {Jonsson and Trefethen, 1997}, namely the absence the so-called "cut-off phenomenon" in single card mixing. This phenomenon is characterized by relatively little change in deviation from uniformity for increasing shuffle number, followed by a sharp decline. This phenomenon is observed ubiquitously in many Markov chains {Diaconis 1993} and has been attributed to a combination of the use of total variation norm and the spectrum of the particular transition matrix, causing transient behavior (i.e. for few shuffles) uncharacteristic of the underlying system{Jonsson and Trefethen, 1997}. In light of this fact, the observation of single card mixing shows an immediate decline, possibly pointing to the relatively few degrees of freedom/influence associated with investigating only a single card versus an entire deck as the factor for not seeing this rich dynamic. Alternatively, an investigation of the how decline in total variation changes as a function of the number of cards may be an interesting avenue to pursue in a future experiment.

When using the variation of information metric, it was found that convergence to the uniform distribution occurred after about six shuffles. Interestingly, the analytic solution described by {Trefethen and Trefethen, 2000} for the variation of information of the entire deck as a function of shuffle number also converged within approximately six shuffles. While this may be a coincidence, it may also point to the use of a distance metric that is invariant to the number of objects being observed during a mixing process. Another investigation into how single card and entire deck observations agree across a variety of $n$ would be suitable for substantiating this claim.

Finally, single card mixing observations were used to assess how well a deck is mixed based on an overhand shuffle. In general, it was found that the overhand shuffle performed worse compared to the riffle shuffle (in terms of mixing time). In hindsight, given that mixing time for the overhand shuffle is $O(n^2 \log(n))$ {Jonasson, 2006} and the mixing time for the riffle shuffle is $O(n \log(n))$ {Aldous and Diaconis, 1987; Bayer and Diaconis, 1992} this is not unexpected.

## Acknowledgements

I would like to thank Prof. Renato Feres for his guidance throughout this project – it has been a long journey, but I think I found something in this project that I really enjoyed. I would also like to thank my classmates for their support, particularly Zeyang Yu and Jinsoo Park.

## Works Cited


Aldous, D. and P. Diaconis 1986. Shuffling cards and stopping times. The American Mathematical Monthly 93(5): 333-348.

Bayer, D. and P. Diaconis 1992. Trailing the dovetail shuffle to its lair. The Annals of Applied Probability 2(2): 294-313.

Diaconis, P. 1996. The cutoff phenomenon in finite markov chains. Proceedings of the National Academy of Sciences 93(4): 1659-1664.

Jonasson, J. 2006. The overhand shuffle mixes in $O(n^2 \log n)$ steps. The Annals of Applied Probability: 231-243.

Jonsson, G. and L. N. Trefethen 1997. A numerical analyst looks at the "cutoff phenomenon" in card shuffling and other markov chains. Higham et al.[HWG98]: 150-178.

Trefethen, L. N. and L. M. Trefethen 2000. How many shuffles to randomize a deck of cards? Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 456(2002): 2561-2568.

```
%{
    D. Sinha
    Project_Code.m

    Driver code for single card observation shuffling project
%}

%%
% Generate the transition matrix over rising sequences
num_cards = 52;
[P A Pinf] = Riffle(num_cards);

% Get the total variation distance from the uniform distr. & entropy for the riffle.
max_k = 10;
A_K_norm = zeros(max_k,1);
I_dist = zeros(max_k,1);

P_K = P;
for ii = 1:max_k
    A_K_norm(ii) = 0.5*max(sum(abs(P_K-Pinf),2));
    P_K = P_K*P;
end

% Compute logs of Eulerian numbers
n=num_cards;
a = zeros(1,n);
anew = zeros(1,n);
for j = 2:n
    anew(2:j-1) = log((2:j-1).*exp(a(2:j-1)-a(1:j-2))+(j-1:-1:2))+a(1:j-2);
    a = anew;
end



% Calculate entropy & I of distribution relative to uniform distribution
% NOT OPERATIONAL!!!!
P_i = zeros(size(P,1),1)'; P_i(1) = 1;
for ii = 1:max_k
    P_i = P_i*P;
    %{
    [V L] = eig(P_K');
    idx = find(abs(diag(L) - 1) < 0.01);
    P_i = V(:,idx)/sum(V(:,idx));
    %}
    P_adj = (P_i.*exp(a))/sum(P_i.*exp(a));
    H_p = - P_adj .* log2(P_adj); H_p(P_adj == 0) = 0; H_p = sum(H_p);
    I_dist(ii) = sum(log2(1:num_cards)) - H_p;
end
%%
% MC for Riffle One-Card
pos_tocheck = [1 26 13];
num_cards = 52;
num_decks = 10^4; num_shuffles = max_k;
cards = repmat((1:num_cards)',1,num_decks);
```

```
cards_inpos_riff = zeros(num_shuffles,num_decks,length(pos_tocheck));
for s_num = 1:num_shuffles
    % Record card in each desired position
    for pos_num = 1:length(pos_tocheck)
        cards_inpos_riff(s_num,:,pos_num) = cards(pos_tocheck(pos_num),:);
    end

    % Do Riffle Shuffle
    for d_num = 1:num_decks
        div_val = binornd(num_cards,1/2);
        temp = zeros(num_cards,1); pos_first = 1; pos_second = 1; pos_temp = 1;
        while pos_first <= div_val && pos_second <= (num_cards-div_val)
            choose_deckone = rand < (div_val-pos_first + 1)/((div_val-pos_first+1)+(num_cards-
div_val-pos_second+1));
            if(choose_deckone)
                temp(pos_temp) = cards(pos_first,d_num);
                pos_first = pos_first + 1;
            else
                temp(pos_temp) = cards(div_val+pos_second,d_num);
                pos_second = pos_second+1;
            end
            pos_temp = pos_temp + 1;
        end
        if(pos_first > div_val)
            temp(pos_temp:end) = cards(div_val+pos_second:end,d_num);
        else
            temp(pos_temp:end) = cards(pos_first:div_val,d_num);
        end
        cards(:,d_num) = temp;
    end
end

distr_riff = zeros(num_shuffles,num_cards,length(pos_tocheck));
A_K_MC_riff = zeros(num_shuffles,length(pos_tocheck));
I_MC_riff = zeros(num_shuffles,length(pos_tocheck));
for pos_num = 1:length(pos_tocheck)
    for s_num = 1:num_shuffles
        distr_riff(s_num,:,pos_num) = hist(squeeze(cards_inpos_riff(s_num,:,pos_num))',[1:
num_cards]);  distr_riff(s_num,:,pos_num) = distr_riff(s_num,:,pos_num)/sum(distr_riff(s_num,:,
pos_num));
        A_K_MC_riff(s_num,pos_num) = 0.5*sum(abs(squeeze(distr_riff(s_num,:,pos_num))-1/
num_cards*ones(1,num_cards)));
        H_p = distr_riff(s_num,:,pos_num).*log2(distr_riff(s_num,:,pos_num)); H_p(distr_riff(
s_num,:,pos_num) == 0) = 0; H_p = squeeze(H_p);
        I_MC_riff(s_num,pos_num) = log2(num_cards) + sum(H_p);
    end
end
%%
% MC for Overhand Shuffle
p_os = 1/8;

pos_tocheck = [1 26 13];
num_cards = 52;
```

```
num_decks = 10^4; num_shuffles = max_k;
cards = repmat((1:num_cards)',1,num_decks);
cards_inpos_over = zeros(num_shuffles,num_decks,length(pos_tocheck));
for s_num = 1:num_shuffles
    % Record card in each desired position
    for pos_num = 1:length(pos_tocheck)
        cards_inpos_over(s_num,:,pos_num) = cards(pos_tocheck(pos_num),:);
    end

    % Do Overhand Shuffle
    for d_num = 1:num_decks
        n_pts = binornd(num_cards-1,p_os);
        if(n_pts > 0)
            cpts = randperm(num_cards-1); cpts = [1 sort(cpts(1:n_pts)) num_cards];
            temp = zeros(num_cards,1); temp_pos = 1;
            for sect_num = 1:length(cpts)-1
                temp(temp_pos:temp_pos+(cpts(sect_num+1)-cpts(sect_num))) = flipud(cards(cpts(
sect_num):cpts(sect_num+1),d_num));
                temp_pos = cpts(sect_num+1)+1;
            end
            cards(:,d_num) = temp(1:52);
        end
    end
end

distr_over = zeros(num_shuffles,num_cards,length(pos_tocheck));
A_K_MC_over = zeros(num_shuffles,length(pos_tocheck));
I_MC_over = zeros(num_shuffles,length(pos_tocheck));
for pos_num = 1:length(pos_tocheck)
    for s_num = 1:num_shuffles
        distr_over(s_num,:,pos_num) = hist(squeeze(cards_inpos_over(s_num,:,pos_num))',[1:
num_cards]);  distr_over(s_num,:,pos_num) = distr_over(s_num,:,pos_num)/sum(distr_over(s_num,:,
pos_num));
        A_K_MC_over(s_num,pos_num) = 0.5*sum(abs(squeeze(distr_over(s_num,:,pos_num))-1/
num_cards*ones(1,num_cards)));
        H_p = distr_over(s_num,:,pos_num).*log2(distr_over(s_num,:,pos_num)); H_p(distr_over(
s_num,:,pos_num) == 0) = 0; H_p = squeeze(H_p);
        I_MC_over(s_num,pos_num) = log2(num_cards) + sum(H_p);
    end
end

%%
figure; hold on;
scatter(1:max_k,[A_K_MC_riff(:,1)],'MarkerEdgeColor',[0 0 1]);
scatter(1:max_k,[A_K_MC_riff(:,2)],'MarkerEdgeColor',[0 1 0]);
scatter(1:max_k,[A_K_MC_riff(:,3)],'MarkerEdgeColor',[0 0 0]);
scatter(1:max_k,[A_K_norm],'MarkerEdgeColor',[1 0 0]);
xlabel('Number of Shuffles'); ylabel('Total Variation ||P^k-P_{inf}||_1');
legend([{'1st Card'} {'26th Card'} {'13th Card'} {'All Cards'}]);
print(gcf,'-djpeg90','FigA_Riffle_TV.jpg');
%%
figure; hold on;
scatter(1:max_k,[I_MC_riff(:,1)],'MarkerEdgeColor',[0 0 1]);
```

```
scatter(1:max_k,[I_MC_riff(:,2)],'MarkerEdgeColor',[0 1 0]);
scatter(1:max_k,[I_MC_riff(:,3)],'MarkerEdgeColor',[0 0 0]);
xlabel('Number of Shuffles'); ylabel('I (bits)');
legend([{'1st Card'} {'26th Card'} {'13th Card'}]);
print(gcf,'-djpeg90','FigB_Riffle_I.jpg');
%%
figure;
subplot(1,2,1); hold on; axis square;
scatter(1:max_k,[A_K_MC_over(:,1)],'MarkerEdgeColor',[0 0 1]);
scatter(1:max_k,[A_K_MC_over(:,2)],'MarkerEdgeColor',[0 1 0]);
scatter(1:max_k,[A_K_MC_over(:,3)],'MarkerEdgeColor',[0 0 0]);
xlabel('Number of Shuffles'); ylabel('Total Variation ||P^k-P_{inf}||_1');
legend([{'1st Card'} {'26th Card'} {'13th Card'}]);

subplot(1,2,2);  hold on; axis square;
scatter(1:max_k,[I_MC_over(:,1)],'MarkerEdgeColor',[0 0 1]);
scatter(1:max_k,[I_MC_over(:,2)],'MarkerEdgeColor',[0 1 0]);
scatter(1:max_k,[I_MC_over(:,3)],'MarkerEdgeColor',[0 0 0]);
xlabel('Number of Shuffles'); ylabel('I (bits)');
legend([{'1st Card'} {'26th Card'} {'13th Card'}]);
print(gcf,'-djpeg90','FigC_Over_Both.jpg');
```

```
%{
    D. Sinha
    RIFFLE.m

    Original Authors: Jonsson & Trefthen from "A Numerical Analyist Looks at the
    'Cutoff Phenomenon' in Card Shuffling and Other Markov Chains."
    Numerical Analysis 1997

    Input: n - number of objects being riffle shuffled
    Output:
        P - Transition matrix corresponding to riffle shuffle
        A - Decay matrix corresponding to A = P-P^inf
%}

function [P A Pinf] = Riffle(n)
% Compute logs of Eulerian numbers
a = zeros(1,n);
anew = zeros(1,n);
for j = 2:n
    anew(2:j-1) = log((2:j-1).*exp(a(2:j-1)-a(1:j-2))+(j-1:-1:2))+a(1:j-2);
    a = anew;
end

% Compute logs of binomial coefficients
b = zeros(1,n+2);
bnew = zeros(1,n+2);
for j = 2:n+2
    bnew(2:j-1) = log(exp(b(2:j-1)-b(1:j-2))+1)+b(1:j-2);
    b = bnew;
end

% Construct transition matrix P
b = b-n*log(2);
r = [b(1) -Inf*ones(1,n-1)];
c = [b -Inf*ones(1,n-2)]';
T = toeplitz(c,r);
P = T(2:2:2*n,:);
P = exp(P-a'*ones(1,n)+ones(n,1)*a);

% Compute stationary distribution and use it to construct P^inf
v = eye(1,n);
vnew = eye(1,n);
for j = 2:n
    vnew(1) = v(1);
    vnew(2:j) = (2:j).*v(2:j)+(j-1:-1:1).*v(1:j-1);
    v = vnew/j;
end
Pinf = ones(n,1)*v;
A = P -Pinf;
end % riffle.m
```