Distributions of permutations generated by inhomogeneous Markov chains

Diplomarbeit

von

Thomas Theobald

eingereicht beim
Institut für Mathematik
der Goethe-Universität Frankfurt am Main

Betreuer: Prof. Dr. Ralph Neininger

December 1, 2009

Contents

List of Figures

Danksagung

Der erste Dank gilt Prof. Dr. Neininger und seinem Team für die stets vorbildliche Betreuung. Besonders danken möchten ich Dr. Grusea für ihren wichigen Rat bei der Erstellung des Algorithmus. Schließlich will ich mich bei meiner Familie bedanken, insbesondere bei meinen Geschwistern für das aufopferungsvolle Korrektur lesen.

Eidesstattliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig und nur unter Benutzung der angegebenen Literatur und Hilfsmittel angefertigt habe. Wörtlich übernommene Sätze und Satzteile sind als Zitate belegt, andere Anlehnungen hinsichtlich Aussage und Umfang unter den Quellenangaben kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen und ist nicht veröffentlicht.

Thomas Theobald

1 Introduction

Many articles have been published about Markov chain imbedding technique (MCIT) as an efficient instrument to obtain results especially for the distributions of runs and for the waiting times of patterns. Even more has been published concerning the cycle structures of permutations. This work connects both topics. As a final result program code is given that uses Markov chain imbedding technique to deliver proper numerical values for the discrete distribution of permutations according to their cycle structure. Simultaneously four other important distributions in the context of permutations are determined: the one of the number of cycles, the one of the rth longest and shortest cycle, among these the one of the longest and the shortest cycle and at last the one of the length of a random chosen cycle.
Apart from that all distributions are given for random permutations, describing the occurrence of each permutation as equiprobable, and for biased permutations where the occurrence is manipulated by a parameter θ such that special structures are more probable than others.
In section 2.1 the basic idea of Markov chain imbedding technique is presented where as in section 2.2 two articles containing applications of this technique are described in detail. Section 3 delivers the setting for the application on distributions in the context of permutations and presents numerical results. Section 4 finally compares these results with analytical formulas, if they exist. In this part it becomes clear that implementing these together in an algorithm does not lead to quicker results as they are generated by the algorithm using Markov chain imbedding technique. The code of the last is provided in the appendix. Irrespectively whether analytical formulas exist or not section 4 also deals with the limit distribution of the variables mentioned above. Finally the appendix contains more on technical details.

2 Markov chain imbedding technique (MCIT)

2.1 The basic idea

This section presents the finite Markov chain imbedding technique introduced by Fu and Koutras [4]. Consider a given integer n and a random variable X_n defined on the finite state space $\mathcal{O} = \{a_1, ..., a_l\}$, for which we either do not know its distribution or hope to find a quick way to obtain it.

Definition
X_n can be embedded into a finite Markov chain Y_t, $t \in I = \{0, ..., n\}$ if

1. there exists a finite Markov chain (Y_t) on a finite state space $\Omega = \{b_1, ..., b_m\}$ containing m different possible realizations.

2. there exists a finite partition C_x on the state space Ω where x corresponds to one of the possible realizations of X_n.

3. for every $x = a_1, ..., a_l$ the identity $P(X_n = x) = P(Y_n \in C_x)$ is valid.

If one is using Markov chain imbedding technique the distribution of the random variable X_n of interest is determined by the transition probabilities of the Markov chain. Note that the implicit use of this definition delivers the propositions of section 3.1, where the transition probabilities for cycle structures of permutations are given. In order to determine the probability of a certain realization of X_n in a common framework it is reasonable to cite a theorem also given by Fu and Koutras in [4].

Theorem
Transferring the setting of the definition above let PM_t be the $m \times m$ transition probability matrix of the finite Markov chain $(Y_t)_{t \in I}$. Furthermore a selection vector e_j is introduced with 1 at the j-th component and 0 elsewhere. According to the last item in the definition all b_j in C_x for $j = 1, \ldots, m$ must be selected, i.e. $e(C_x) := \sum_{j:b_j \in C_x} e_j$. If X_n can be embedded into a finite Markov chain, then

$$P(X_n = x) = \pi_0 \prod_{t=1}^{n} PM_t e(C_x),$$

where $\pi_0 = (P(Y_0 = b_1), \ldots, P(Y_0 = b_m))$ stands for the initial probability of the Markov chain.

Example
In order to get an idea how the proof is working in detail we add an example for $P(X_2 = x) = P(Y_2 = y_1)$ and where each of the variables in the Markov chain my only have two possible states to jump in. Based on Bay's theorem and Chapman-Kolmogorov equation we obtain

$$\begin{aligned}
&P(Y_2 = y_1) = \\
&P(Y_2 = y_1|Y_0 = y_1) P(Y_0 = y_1) + P(Y_2 = y_1|Y_0 = y_2) P(Y_0 = y_2) = \\
&P(Y_0 = y_1) (P(Y_1 = y_1|Y_0 = y_1) P(Y_2 = y_1|Y_1 = y_1) + \\
&P(Y_1 = y_2|Y_0 = y_1) P(Y_2 = y_1|Y_1 = y_2)) + \\
&P(Y_0 = y_2) (P(Y_1 = y_1|Y_0 = y_2) P(Y_2 = y_1|Y_1 = y_1) + \\
&P(Y_1 = y_2|Y_0 = y_2) P(Y_2 = y_1|Y_1 = y_2)),
\end{aligned}$$

which is exactly the same than

$$P\left(Y_2 = y_1\right) = \pi_0 PM_1 PM_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} P\left(Y_0 = y_1\right) & P\left(Y_0 = y_2\right) \end{pmatrix} \begin{pmatrix} p_{11}^{10} & p_{21}^{10} \\ p_{12}^{10} & p_{22}^{10} \end{pmatrix} \begin{pmatrix} p_{11}^{21} & p_{21}^{21} \\ p_{12}^{21} & p_{22}^{21} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} P\left(Y_0 = y_1\right) & P\left(Y_0 = y_2\right) \end{pmatrix} \begin{pmatrix} p_{11}^{10}p_{11}^{21} + p_{21}^{10}p_{12}^{21} & p_{11}^{10}p_{21}^{21} + p_{21}^{10}p_{22}^{21} \\ p_{12}^{10}p_{11}^{21} + p_{22}^{10}p_{12}^{21} & p_{12}^{10}p_{21}^{21} + p_{22}^{10}p_{22}^{21} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Thereby systematic notation $P\left(Y_1 = y_1 | Y_0 = y_1\right) = p_{11}^{10}$ is used to shorten the terms.

Proof
As it is shown in the example above it follows for each $b_j \in \Omega$ by Bay's theorem and Chapman-Kolmogorov equation that

$$P\left(Y_n = b_j\right) = \pi_0 \left(\prod_{t=1}^{n} PM_t\right) e_j.$$

Because X_n can be embedded into a finite Markov chain the rest of the proof is immediately delivered by the fact that the probability measure is σ-additive, i.e.

$$P\left(X_n\right) = P\left(Y_n \in C_x\right) = \sum_{b_j \in C_x} P\left(Y_n = b_j\right)$$

$$= \sum_{b_j \in C_x} \pi_0 \left(\prod_{t=1}^{n} PM_t\right) e_j = \pi_0 \left(\prod_{t=1}^{n} PM_t\right) e\left(C_x\right).$$

2.2 Outlook on different applications in literature

2.2.1 MCIT for distributions of Bernoulli trials

The first article that is presented here is the one 'Distribution Theory of Runs: A Markov Chain Approach' written by J.C. Fu and M.V. Koutras [4]. The term success run identifies a sequence of consecutive successes S interrupted by failures F, while the whole framework of alternating S and F build up a Bernoulli trial, which might be both identically or nonidentically independent distributed. Fu/Koutras have studied five frequently used statistics of success runs within such a trial of length n:

- $E_{n,k}$ the number of success runs of size exactly k,
- $G_{n,k}$ the number of success runs of size greater than or equal to k,
- $N_{n,k}$ the number of non overlapping consecutive k successes,
- $M_{n,k}$ the number of overlapping consecutive k successes,
- L_n the size of the longest success run.

To illustrate these definitions they give the example of a coin tossed ten times with the following result: SSFSSSSFFF. So the variables above receive the following values:
$E_{10,2} = 1, \quad G_{10,2} = 2, \quad N_{10,2} = 3, \quad M_{10,2} = 4, \quad L_{10} = 4.$

Between the variables above the following relations are valid:

$$\begin{aligned} &E_{n,k} \leq G_{n,k} \leq N_{n,k} \leq M_{n,k}, \\ &N_{n,k} = 0 \Rightarrow L_n < k. \end{aligned} \tag{1}$$

In (1) the first and the third inequality are clear since this connection between the variables is given by their definition. For the connection between $G_{n,k}$ and $N_{n,k}$ one has to think of a success run of size greater than k. First of all this one will automatically be captured in both variables. But if its sizes doubles k it will be captured twice in $N_{n,k}$ while $G_{n,k}$ still remains at one. In (1) the length of the longest success run is less than k if there does not occur at least one non overlapping consecutive k success run. But the second row in (1) might also be written with variable $G_{n,k}$.

In the following the distributions of $E_{n,k}, G_{n,k}, N_{n,k}$ and $M_{n,k}$ are identified by their transition matrices. In these one can find particular patterns such that these matrices can be built up quickly by the different elementary transition probabilities. If the transition matrices are constructed it will be easy to compute the distributions due to the theorem in section 2.1. For simplification it is assumed that the origin is almost sure equal to 0 and that the last state of the Markov chain is an absorbing state. This is equivalent to the last row of PM_t being equal to $(0, \ldots, 0, 1)$.

The distribution of $N_{n,k}$
It is useful to start with this variable since it is the easiest to embed. Consider the state space

$$\Omega\left(N_{n,k}\right) = \{(x,i): \quad x = 0, 1, \ldots, \lfloor \frac{n}{k} \rfloor, \quad i = 0, 1, \ldots, k-1\}$$

belonging to the Markov chain $(Y_t)_{t \in I}$. Thereby x stands for the number of non overlapping k successes and i for the modulo k check of the number of consecutive successes when counting backwards. Always when a k success is completed i jumps back to 0

and simultaneously x to $x+1$. Note that at most $\lfloor \frac{n}{k} \rfloor$ exist, which is the case when all entries of the success run are equal to S. Is for example n an odd integer where as k is even it will be necessary to round it off. That is where the floor function comes into play.

The partition $\bigcup_{x=0}^{\lfloor \frac{n}{k} \rfloor} C_x$ is given naturally where $C_x = \{(x,i) : \quad i = 0,1,\ldots,k-1\}$ for fixed x. With this setting $P(N_{n,k} = x) = P(Y_n \in C_x)$ is clear. Now one can think about possible transitions from $t-1$ to t and about the corresponding probabilities:

1. A transition from state (x,i) to the state $(x,i+1)$ can occur for $0 \leq x \leq \lfloor \frac{n}{k} \rfloor$ and for $0 \leq i \leq k-2$. This will happen with an additional S. So the transition probability is $P((x,i+1) \,|\, (x,i)) = p_t$.

2. A transition from state (x,i) to the state $(x,0)$ can occur for $0 \leq x \leq \lfloor \frac{n}{k} \rfloor$ and for $0 \leq i \leq k-1$. This will happen with an additional F. So the transition probability is $P((x,0) \,|\, (x,i)) = 1 - p_t = q_t$.

3. A transition from state $(x,k-1)$ to the state $(x+1,0)$ can occur for $0 \leq x \leq \lfloor \frac{n}{k} \rfloor - 1$. This will happen with an additional S. So the transition probability is $P((x+1,0) \,|\, (x,k-1)) = p_t$

In fact there is a implicit try-and-error mechanism in order to ensure that these are all possible transitions. Corresponding with this is that all other entries of the transition matrix are equal to 0. It is useful to consider an example first and to derive from this the general case. According to the transition probabilities above the transition matrix PM_5, i.e. the one from Y_4 to Y_5, for $N_{5,2}$ is given by

transition matrix		n	5	k	2	
from/to	(0,0)	(0,1)	(1,0)	(1,1)	(2,0)	(2,1)
(0,0)	qt	pt	0	0	0	0
(0,1)	qt	0	pt	0	0	0
(1,0)	0	0	qt	pt	0	0
(1,1)	0	0	qt	0	pt	0
(2,0)	0	0	0	0	qt	pt
(2,1)	0	0	0	0	0	1

Figure 1: Transition matrix for non overlapping k successes with $(n,k) = (5,2)$

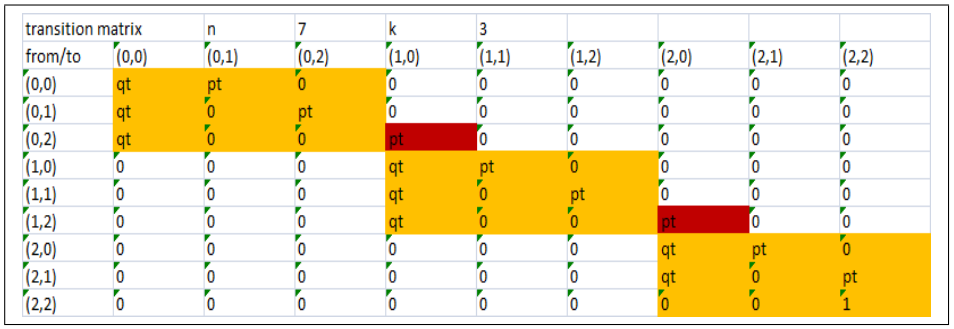
Concerning the state space it is clear that the total number of tuples (x,i) is equal to $\left(\lfloor \frac{n}{k} \rfloor + 1\right) k$. Considering that the target is a quadratic matrix the dimension of the transition matrix is equal to $\left(\lfloor \frac{n}{k} \rfloor + 1\right) k \;\times \left(\lfloor \frac{n}{k} \rfloor + 1\right) k$. It must already be mentioned here that this leads to different dimensions of the single matrices. Regarding the theorem above these must be transferred to the same dimension to ensure that a product of these matrices can be computed. For computation see the last item of this section and appendix D.

The example above is colored to make it easier to be aware of the patterns contained in all transition matrices. The yellow partial matrices have dimension $k \times k$ and deal on the one hand with (2.) and on the other with (1.) in the enumeration above. The red parts of dimension 1×1 arise from the case where x increases (3.). With this the construction of the transition probabilities can be easily done:

- Write down all possible elements of Ω to build up the framework of the transition matrix

- Put down the yellow partial matrices of dimension $k \times k$ along the main diagonal
- Put down the red parts at the southeast corner of the yellow matrices if this is possible

Applying this algorithm another example is given



transition matrix		n	7	k	3				
from/to	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)	(2,2)
(0,0)	qt	pt	0	0	0	0	0	0	0
(0,1)	qt	0	pt	0	0	0	0	0	0
(0,2)	qt	0	0	pt	0	0	0	0	0
(1,0)	0	0	0	qt	pt	0	0	0	0
(1,1)	0	0	0	qt	0	pt	0	0	0
(1,2)	0	0	0	qt	0	0	pt	0	0
(2,0)	0	0	0	0	0	0	qt	pt	0
(2,1)	0	0	0	0	0	0	qt	0	pt
(2,2)	0	0	0	0	0	0	0	0	1

Figure 2: Transition matrix for non overlapping k successes with $(n, k) = (7, 2)$

in order to show a pretended weakness. That is the fact that it is not possible to generate $(x, i) = (2, 2)$ in a binomial trial of length 7. Two success runs, each of length 3, demand at least six S, so that there is no space for two additional S. Note that this weakness has no computational consequence as long as the origin is almost sure $Y_0 = (0, 0)$. Otherwise the last row has to be transferred to $(0, \ldots, 0)$ and the next to last one to $(0, \ldots, 1)$. This correction also has to be done for all matrices M_i with $i < n$ when applying the theorem of section (2.1) since M_n contains states, in which Y_i with $i < n$ can not jump. Appendix D contains Maple computation according to the theorem in section 2.1. With this results for the explicit distribution of $N_{5,2}$ and $N_{7,3}$ are generated, when $p_t = \frac{1}{t+1}$:

$X_{n,k}$	0	1	2
$N_{5,2}$	0,7375	0,2486	0,0139
$N_{7,2}$	0,9386	0,0611	0,0003

Now we are ready to turn to the next variables. We limit our review here on the structure of the transition matrices. Similar computation as it was done for $N_{n,k}$ then delivers numerical values for the explicit distributions of the other variables. Such values are also given in [4].

The distribution of $G_{n,k}$
Consider the state space

$$\Omega\left(N_{n,k}\right) = \{(x, i): \quad x = 0, 1, \ldots, \lfloor \frac{n+1}{k+1} \rfloor, \quad i = -1, 0, 1, \ldots, k-1\} \backslash \{(0, -1)\}$$

belonging to the Markov chain $(Y_t)_{t \in I}$. Thereby x stands for the number of success runs greater or equal to k. This time it will not be enough for i to represent the number of consecutive successes when counting backwards in the same sense as before. The modulo k check has to be turned off. Let m count the consecutive successes backwards. If $m{=}k$, i will jump to -1 and x to $x + 1$. If then another S occurs $(x + 1, -1)$ will remain. Note that this state $(x + 1, -1)$ will hold until an F occurs. With this mechanics the state $(0, -1)$ can never be reached and the condition is fulfilled that a success run, how long it might ever be, at most only counts for one. If $m < k$, i can be as before. In order to

understand the maximal number of success runs greater or equal to k an example, here for $(n,k)=(7,2)$, is helpful

$$\bullet\bullet\,|\,\bullet\bullet|\bullet$$

The total number of signs is equal to 7 and the sequence contains 2 barriers. It becomes clear that each of the k success run except the last needs a F as a barrier. So the maximal number is given by $\lfloor\frac{n+1}{k+1}\rfloor$.

The partition turns into $C_0 \cup \bigcup_{x=1}^{\lfloor\frac{n+1}{k+1}\rfloor} C_x$, where $C_0 = \{(0,i): \quad i=0,1,\ldots,k-1\}$ and $C_x = \{(x,i): \quad i=-1,0,1,\ldots,k-1\}$. With this setting $P(G_{n,k}=x) = P(Y_n \in C_x)$ is clear. Now one can think about possible transitions from $t-1$ to t and about the corresponding probabilities:

1. A transition from state (x,i) to the state $(x,i+1)$ can occur for $0 \le x \le \lfloor\frac{n+1}{k+1}\rfloor$ and for $0 \le i \le k-2$. This will happen with an additional S. So $P((x,i+1)\,|\,(x,i)) = p_t$.

2. A transition from state (x,i) to the state $(x,0)$ can occur for $0 \le x \le \lfloor\frac{n+1}{k+1}\rfloor$ and for $-1 \le i \le k-1$. This will happen with an additional F. So $P((x,0)\,|\,(x,i)) = q_t$.

3. A transition from state $(x,k-1)$ to the state $(x+1,-1)$ can occur for $0 \le x \le \lfloor\frac{n+1}{k+1}\rfloor - 1$. This will happen with an additional S. So $P((x+1,-1)\,|\,(x,k-1)) = p_t$.

4. A transition from state $(x,-1)$ to the state $(x,-1)$ can occur for $1 \le x \le \lfloor\frac{n+1}{k+1}\rfloor$. This will happen with an additional S. So $P((x,-1)\,|\,(x,-1)) = p_t$.

Again we first consider an example and derive from this the general case. According to the transition probabilities above the transition matrix PM_5, i.e. the one from Y_4 to Y_5 for $G_{5,2}$ is given by

transition matrix	n	5	k	2			
from/to	(0,0)	(0,1)	(1,-1)	(1,0)	(1,1)	(2,-1)	(2,0)	(2,1)
(0,0)	qt	pt	0	0	0	0	0	0
(0,1)	qt	0	pt	0	0	0	0	0
(1,-1)	0	0	pt	qt	0	pt	0	0
(1,0)	0	0	0	qt	pt	0	0	0
(1,1)	0	0	0	qt	0	pt	0	0
(2,-1)	0	0	0	0	0	pt	qt	0
(2,0)	0	0	0	0	0	0	qt	pt
(2,1)	0	0	0	0	0	0	0	1

Figure 3: Transition matrix for the number of success runs of size at least k

Concerning the state space (with the impossible state $(0,-1)$) it is clear that the total number of tuples (x,i) is equal to $\left(\lfloor\frac{n+1}{k+1}\rfloor+1\right)(k+1)-1$. Considering that the target is a quadratic matrix the dimension of the transition matrix is equal to $\left(\lfloor\frac{n+1}{k+1}\rfloor+1\right)(k+1)-1 \times \left(\lfloor\frac{n+1}{k+1}\rfloor+1\right)(k+1)-1$. The problem of non-reachable states was sufficiently illustrated in the section about $N_{n,k}$. The same with computation of the explicit distribution.

Again the example is colored to be aware of the patterns in the matrix. The yellow parts along the main diagonal again cover (1.) and (2.) in the last enumeration. Its quadratic form is lost because of the additional transition from $(x,-1)$ to $(x,0)$. Again

the red parts are the one-dimensional ones, where x increases (3.). The one-dimensional violet ones arise from the fact that an additional S can not change x anymore, when it belongs to a success run, which has already increased x (4.). With this an algorithm to construct the matrix in analogy to the one in the section about $N_{n,k}$ can easily be given.

The distribution of $M_{n,k}$
Consider the state space

$$\Omega\left(M_{n,k}\right) = \{(x,i) : x = 0,1,\ldots,n-k, i = -1,0,1,\ldots,k-1\} \cup \{(n-k+1,-1)\} \backslash \{(0,-1)\}$$

belonging to the Markov chain $(Y_t)_{t\in I}$. Thereby x stands for the number of overlapping k successes. i is the number of consecutive successes when counting backwards as long as this number is lower or equal to $k-1$. Otherwise i again jumps to the value -1. Then with each further S i keeps the value -1 as long as a F occurs. But this time each further S also increases x because of the overlapping setting. Again the state $(0,-1)$ can not be reached. The maximum of x is given by $n-k+1$ since the first k of Ss deliver $x = 1$ and each further S increases x. Note that $x = n-k+1$ will only be true, if only Ss are given. Further $i = -1$ must hold. This is so since the considered variables are only non trivial for $k \leq n$.

The partition turns into $C_0 \cup \bigcup_{x=1}^{n-k} C_x \cup C_{n-k+1}$, where $C_0 = \{(0,i) : i = 0,1,\ldots,k-1\}$, $C_x = \{(x,i) : i = -1,0,1,\ldots,k-1\}$ and $C_{n-k+1} = \{(n-k+1,-1)\}$. With this setting $P\left(M_{n,k} = x\right) = P\left(Y_n \in C_x\right)$ is clear. Now one can think about possible transitions form $t-1$ to t and about the corresponding probabilities:

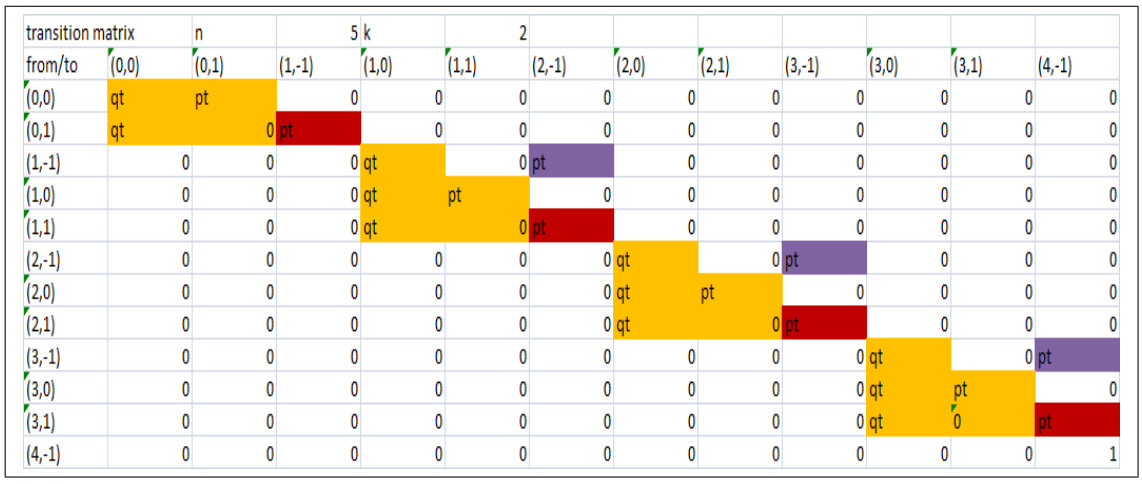
1. A transition from state (x,i) to state $(x,i+1)$ can occur for $0 \leq x \leq n-k$ and $0 \leq i \leq k-2$. This will happen with an additional S. So $P\left((x,i+1) \,|\, (x,i)\right) = p_t$.

2. A transition from state (x,i) to state $(x,0)$ can occur for $0 \leq x \leq n-k$ and $0 \leq i \leq k-1$. This will happen with an additional S. So $P\left((x,0) \,|\, (x,i)\right) = q_t$.

3. A transition from state $(x,k-1)$ to the state $(x+1,-1)$ can occur for $0 \leq x \leq n-k$. This will happen with an additional S. So $P\left((x+1,-1) \,|\, (x,k-1)\right) = p_t$

4. A transition from state $(x,-1)$ to the state $(x+1,-1)$ can occur for $0 \leq x \leq n-k$. This will happen with an additional S. So $P\left((x+1,-1) \,|\, (x,-1)\right) = p_t$.

Again we first consider an example and derive from this the general case. According to the transition probabilities above the transition matrix PM_5, i.e. the one from Y_4 to Y_5 for $M_{5,2}$ is given by



transition matrix	n		5	k		2						
from/to	(0,0)	(0,1)	(1,-1)	(1,0)	(1,1)	(2,-1)	(2,0)	(2,1)	(3,-1)	(3,0)	(3,1)	(4,-1)
(0,0)	qt	pt	0	0	0	0	0	0	0	0	0	0
(0,1)	qt	0	pt	0	0	0	0	0	0	0	0	0
(1,-1)	0	0	0	qt	0	pt	0	0	0	0	0	0
(1,0)	0	0	0	qt	pt	0	0	0	0	0	0	0
(1,1)	0	0	0	qt	0	pt	0	0	0	0	0	0
(2,-1)	0	0	0	0	0	0	qt	0	pt	0	0	0
(2,0)	0	0	0	0	0	0	qt	pt	0	0	0	0
(2,1)	0	0	0	0	0	0	qt	0	pt	0	0	0
(3,-1)	0	0	0	0	0	0	0	0	0	qt	0	pt
(3,0)	0	0	0	0	0	0	0	0	0	qt	pt	0
(3,1)	0	0	0	0	0	0	0	0	0	qt	0	pt
(4,-1)	0	0	0	0	0	0	0	0	0	0	0	1

Figure 4: Transition matrix for overlapping k successes

Concerning the state space it becomes clear that the total number of tuples (x, i) is equal to $(n-k+1)(k+1)$. Obviously it is enough to consider the elements of $\bigcup_{x=0}^{n-k} C_x$ since C_{n-k+1} and the impossible state $(0,-1)$ cancel each other out. The dimension of the corresponding quadratic matrix is equal to $(n-k+1)(k+1) \times (n-k+1)(k+1)$. The problem of non-reachable states was sufficiently illustrated in the section about $N_{n,k}$. The same with computation of the distribution.

Again the example is colored and again the yellow parts along the main diagonal cover (1.) and (2.) in the last enumeration. Its form is exactly the one from $G_{n,k}$. Now the red parts are the one-dimensional ones, where x increases and i changes from $k-1$ to -1 (3.). The violet parts are the one-dimensional ones, where x increases while i keeps its value (4.). This is the case when a S sequence is extended. With this an algorithm to construct the matrix in analogy to the one in the section about $N_{n,k}$ can easily be given.

The distribution of $E_{n,k}$
Consider the state space

$$\Omega(E_{n,k}) = \{(x,i) : x = 0, 1, \ldots, \lfloor \frac{n+1}{k+1} \rfloor, i = -2, -1, 0, \ldots, k-1\} \backslash \{(0,-2)\}$$

belonging to the Markov chain $(Y_t)_{t \in I}$. Thereby x stands for the number of success runs of size exactly k. i is the number of consecutive successes when counting backwards as long as this number is lower or equal to $k-1$. Introducing $i = -2$ generates the state where $x-1$ has jumped to x. This means the binomial trial exactly contains x success runs of size k. Introducing $i = -1$ generates the state where the number of consecutive successes is larger or equal to $k+1$. As soon as i jumps to -1, x decreases to $x-1$ since a success run, which does not contain exactly k successes is no good here. Fu and Koutras call this 'overflow state' [4]. With this it is clear that the impossible state turns to be $(0,-2)$. The maximal number of x is exactly the same as the one of $G_{n,k}$. This is so since again for each success run of size exactly k a F as a barrier is needed.

The partition turns into $C_0 \cup \bigcup_{x=1}^{\lfloor \frac{n+1}{k+1} \rfloor} C_x$, where $C_0 = \{(0,i) : \quad i = -1, 0, \ldots, k-1\}$ and $C_x = \{(x,i) : \quad i = -2, -1, 0, \ldots, k-1\}$. With this setting $P(E_{n,k} = x) = P(Y_n \in C_x)$ is clear. Now one can think about possible transitions from $t-1$ to t and about the corresponding probabilities:

1. A transition from (x,i) to the state $(x,i+1)$ can occur for $0 \leq x \leq \lfloor \frac{n+1}{k+1} \rfloor$ and $0 \leq i \leq k-2$. This will happen with an additional S. So $P\left((x,i+1)\,|\,(x,i)\right) = p_t$.

2. A transition from (x,i) to the state $(x,0)$ can occur for $0 \leq x \leq \lfloor \frac{n+1}{k+1} \rfloor$ and $-2 \leq i \leq k-1$. This will happen with an additional F. So $P\left((x,0)\,|\,(x,i)\right) = q_t$.

3. A transition from $(x,k-1)$ to the state $(x+1,-2)$ can occur for $1 \leq x \leq \lfloor \frac{n+1}{k+1} \rfloor$. This will happen with an additional S. So $P\left((x+1,-2)\,|\,(x,k-1)\right) = p_t$.

4. A transition from $(x,-2)$ to the state $(x-1,-1)$ can occur for $1 \leq x \leq \lfloor \frac{n+1}{k+1} \rfloor$. This will happen with an additional S. So $P\left((x-1,-1)\,|\,(x,-2)\right) = p_t$.

5. A transition from $(x,-1)$ to the state $(x,-1)$ can occur for $0 \leq x \leq \lfloor \frac{n+1}{k+1} \rfloor$. This will happen with an additional S. So $P\left((x,-1)\,|\,(x,-1)\right) = p_t$.

Again we first consider an example and derive form this the general case. According to the transition probabilities above the transition matrix PM_5, i.e. the one from Y_4 to Y_5 for $E_{5,2}$ is given by

transition matrix		n	5	k	2						
from/to	(0,-1)	(0,0)	(0,1)	(1,-2)	(1,-1)	(1,0)	(1,1)	(2,-2)	(2,-1)	(2,0)	(2,1)
(0,-1)	pt	qt	0	0	0	0	0	0	0	0	0
(0,0)	0	qt	pt	0	0	0	0	0	0	0	0
(0,1)	0	qt	0	pt	0	0	0	0	0	0	0
(1,-2)	pt	0	0	0	0	qt	0	0	0	0	0
(1,-1)	0	0	0	0	pt	qt	0	0	0	0	0
(1,0)	0	0	0	0	0	qt	pt	0	0	0	0
(1,1)	0	0	0	0	0	qt	0	pt	0	0	0
(2,-2)	0	0	0	0	pt	0	0	0	0	qt	0
(2,-1)	0	0	0	0		0	0	0	pt	qt	0
(2,0)	0	0	0	0		0	0	0	0	qt	pt
(2,1)	0	0	0	0		0	0	0	0	0	1

Figure 5: Transition matrix for the number of success runs of size exactly k

Concerning the state space (with the impossible state $(0,-2)$) it is clear that the total number of tuples (x,i) is equal to $\left(\lfloor \frac{n+1}{k+1} \rfloor + 1\right)(k+2) - 1$. The quadratic matrix then has dimension $\left(\lfloor \frac{n+1}{k+1} \rfloor + 1\right)(k+2) - 1 \times \left(\lfloor \frac{n+1}{k+1} \rfloor + 1\right)(k+2) - 1$. The problem of non-reachable states was sufficiently illustrated in the section about $N_{n,k}$. The same with computation of the distribution.

Again the example is colored and again the yellow parts along the main diagonal cover (1.) and (2.) in the last enumeration. These are not quadratic any longer because of the additional transitions from $(x,-1)$ to $(x,0)$ and from $(x-2)$ to $(x,0)$. Again the red parts are one-dimensional ones, where x increases (3.). The one-dimensional violet ones correspond to the first time, when too many S occur (4.). The one-dimensional green ones stand for a continuation of a S sequence that is already too long to increase x (5.). With this an algorithm to construct the matrix in analogy to the one in the section about $E_{n,k}$ can easily be given.

The distribution of L_n
Note that it is possible to embed L_n in a similar way as it was done for the other variables above, but applying (1) delivers a quicker way to obtain numerical values for the distribution of the size of the longest success run for $k = 0, 1 \ldots, n$, i.e.

$$\begin{aligned} P(L_n = k) &= P(L_n < k+1) - P(L_n < k) \\ &= P(N_{n,k+1} = 0) - P(N_{n,k} = 0). \end{aligned}$$

That is why an illustration of the imbedding mechanics for L_n is set aside here. If $k = 0$, this will correspond to a binomial trial that only contains failures.

Waiting time distribution of the success run
Note that it is reasonable to express the waiting time for the mth success run of length k in a binomial trial as the number of time steps to get there. Fu and Koutras [4] define $T_{m,k}$ in this way and use the non overlapping setting. A first statement can be easily given with using that the Bernoulli variables are stochastically independent,i.e.

$$P(T_{m,k} = mk) = \prod_{t=1}^{mk} p_t. \tag{2}$$

This is so since such a binomial trial only consists of Ss. A second statement is given by the following: The waiting time will only be equal to $n > mk$ if the last realizations of the binomial trial are all successes and if exactly $m-1$ (non overlapping) k successes have been realized before, i.e. according to the theorem in section 2.1

$$\begin{aligned} P(T_{m,k} = n) &= P(Y(N_{n-k,k}) = (m-1, 0), X_{n-k+1} = S, \ldots, X_n = S) \\ &= \pi_0 \prod_{t=1}^{n-k} PM_t(N_{n-k}, k)\, e'_{m-1} \prod_{n-k+1}^{n} p_t. \end{aligned} \tag{3}$$

Thereby e'_{m-1} is not the classical unit vector with 1 in the $m-1$th component, but instead of this a vector which has 1 in all components that correspond to states where $x = m-1$. Elsewhere its components are equal to 0. So after all the demanded probabilities are given by (2) and (3).

2.2.2 MCIT for quality control schemes

The second article that is presented is the one 'On the average run lengths of quality control schemes using a Markov chain approach' by J.C. Fu, F.A. Spiring and H. Xie [5]. In manufacturing processes quality schemes are used to guarantee the procedure. Such procedures can be filling or cutting of goods. Classical quality control schemes are

1. Cumulative sum (CUSUM), which is defined by

$$\begin{aligned} S_0 &= 0 \\ S_n^1 &= \left(S_{n-1} + \frac{X_n - \mu_0}{\sigma_0} \right)^+ \quad \text{for} \quad n = 1, 2, \ldots \\ S_n^2 &= \left(S_{n-1} + \frac{X_n - \mu_0}{\sigma_0} \right)^- \quad \text{for} \quad n = 1, 2, \ldots, \end{aligned}$$

where $(\cdot)^+ = \max\{0, \cdot\}$ and $(\cdot)^- = \min\{0, \cdot\}$. X_i for $i = 1, \ldots, n$ are the independent identically distributed procedure variables with mean μ, variance σ^2, target mean μ_0 and target variance σ_0^2. S_n^1 is called upper side CUSUM, where as S_n^2 is the lower side CUSUM.

2. Exponentially weighted moving average (EWMA), which is defined by

$$S_0 = 0$$
$$S_n = (1-\lambda)\, S_{n-1} + \lambda \frac{X_n - \mu_0}{\sigma_0} \quad \text{for} \quad n = 1, 2, \ldots \quad \text{and} \quad 0 < \lambda \leq 1.$$

The X_i are the same as above.

Then the manufacturing is out of control after n steps, when for $h > 0$

$$\begin{aligned} -h < S_i < h \quad &\text{for} \quad i = 0, 1, \ldots, n-1 \\ S_n \leq -h \quad &\text{or} \quad S_n \geq h. \end{aligned}$$

h is called control limit. Note that it is also possible to define h depending on X instead of fixed. As the case may be it could then be monotonically increasing or decreasing, but that is not considered here. For a given h define

$$N = \inf\{n \geq 1 : S_n \leq h \quad \text{or} \quad S_n \geq h\}$$

as the run length of the procedure variable. It also represents a waiting time measured in procedure steps.

As we can already see the probability to jump into a out of control state only depends on where the process has been in the last step before. So S_n can be seen as a homogeneous Markov chain since these probabilities do not change. First of all S_n may have infinite states. In order to use a finite state Markov chain it is necessary that the number of states then tends to ∞, i.e.

$$\lim_{n \to \infty} S_n(m) = S_n. \tag{4}$$

So the article deals with constructing a finite Markov chain fulfilling condition (4) and with determining distribution and moments of N by this simplification. For that it has to be shown that $N(m)$ tends almost sure to N.

In the following for simplification upper side CUSUM is taken. After all it is easy to adjust the terms for the other quality control schemes. Consider the state space

$$\Omega = \{a_0, \ldots, a_m, a_{m+1}\}$$

of $S_n(m)$ with size-ordered $0 = a_0 < a_1 < \ldots < a_m < a_{m+1}$. Define $a_i = (i - 0.5)\,\frac{h}{m+1}$ for $i = 1, 2, \ldots, m$, i.e. the distance up to h is well partitioned. The corresponding transition matrix can be written as

$$\begin{aligned} T &= \left(\begin{array}{cccccc|c} p_{00} & p_{01} & \cdot & \cdot & \cdot & p_{0m} & p_{0m+1} \\ p_{00} & p_{01} & \cdot & \cdot & \cdot & p_{0m} & p_{0m+1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{m0} & p_{m1} & \cdot & \cdot & \cdot & p_{mm} & p_{mm+1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \\ &= \left(\begin{array}{c|c} A(m) & B(m) \\ - \quad - & - \\ 0 & 1 \end{array}\right). \end{aligned}$$

Obviously $A(m)$ is a $(m+1) \times (m+1)$ matrix and $B(m)$ is $(m+1) \times 1$ vector. The last row in the transition matrix detects a_{m+1} as an absorbing state corresponding to the fact that the manufacturing process hast to stopped when it is out of control.

$$N(m) = \inf\{n \geq 1 : \quad S_n(m) = a_{m+1} | S_0(m) = a_0\}$$

is the corresponding run length and the initial state is almost sure equal to 0. Furthermore let I be the identity matrix.

Proposition
If $I - A$ is invertible, then we have

$$\begin{aligned} T^n &= \begin{pmatrix} A^n(m) & | & W_n B(m) \\ - & - & - \\ 0 & | & 1 \end{pmatrix} \\ W_n &= I + A(m) + A^2(m) + \ldots + A^{n-1}(m) \\ \lim_{n\to\infty} W_n &= (I - A(m))^{-1}. \end{aligned} \tag{5}$$

Proof
The first statement is linear algebra. Nevertheless an example with $n = 2$ illustrates how the multiplication of matrices works in this case. On the one hand we have

$$\begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} p_{00}^2 + p_{01}p_{10} & p_{00}p_{01} + p_{01}p_{11} & p_{00}p_{02} + p_{01}p_{12} + p_{02} \\ p_{10}p_{00} + p_{11}p_{10} & p_{10}p_{01} + p_{11}^2 & p_{10}p_{02} + p_{11}p_{12} + p_{12} \\ 0 & 0 & 1 \end{pmatrix}.$$

On the other we have with self-explaining last row

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} p_{00}^2 + p_{01}p_{10} & p_{00}p_{01} + p_{01}p_{11} \\ p_{10}p_{00} + p_{11}p_{10} & p_{10}p_{01} + p_{11}^2 \end{pmatrix}$$

$$\begin{pmatrix} p_{00} + 1 & p_{01} \\ p_{10} & p_{11} + 1 \end{pmatrix} \begin{pmatrix} p_{02} \\ p_{12} \end{pmatrix} = \begin{pmatrix} p_{00}p_{02} + p_{02} + p_{01}p_{12} \\ p_{10}p_{02} + p_{11}p_{12} + p_{12} \end{pmatrix}.$$

Putting $W = I + A(m) + A^2(m) + \ldots$ and using a shift to the right we have

$$\begin{aligned} \lim_{n\to\infty} W_n = I + A(m) + A^2(m) + \ldots = I + A(m)W \Leftrightarrow \\ W - A(m)W = I \Leftrightarrow \\ (I - A)\lim_{n\to\infty} W_n = I, \end{aligned}$$

which completes the proof.

Proposition
With the initial distribution given above and T as transition matrix it follows that

$$P(S_n(m) < a_{m+1} | S_0(m) = a_0) = \pi_0 A^n(m)\, \hat{e}(m)_2' \quad \text{for} \quad n = 1, 2, \ldots. \tag{6}$$

Proof
Let $\pi_0 = (1, 0, \ldots, 0)$,$\hat{e}(m) = (1, \ldots, 1, 0)$ and $\hat{e}(m)_2 = (1, \ldots, 1)$. With the theorem from section 2.1 we have

$$\begin{aligned} P(S_n(m) < a_{m+1} | S_0(m) = a_0) = \\ \pi_0 T^n \hat{e}(m)' = \\ \pi_0 A^n(m) \hat{e}(m)_2'. \end{aligned}$$

Theorem
Let $S_n(m)$ be as described above. Then it follows

1. $P(N(m) = n | S_0(m) = 0) = \pi_0 A^{n-1}(m) (I - A(m)) \hat{e}(m)_2'$
2. $\phi_{N(m)}(t) = 1 + \left(e^t - 1\right) \pi_0 \left(I - e^t A(m)\right)^{-1} \hat{e}(m)_2'$
3. $E(N(m)) = \pi_0 (I - A(m))^{-1} \hat{e}(m)_2'$
4. $E\left(N^2(m)\right) = \pi_0 (I + A(m)) (I - A(m))^{-2} \hat{e}(m)_2'$.

Proof
Because of a_{m+1} being an absorbing state it follows

$$S_n(m) < a_{m+1} \Leftrightarrow N(m) > n.$$

1. This leads to

$$\begin{aligned} P(N(m) = n | S_0(m) = a_0) = \\ P(N(m) > n-1 | S_0(m) = a_0) - P(N(m) > n | S_0(m) = a_0) = \\ P(S_{n-1}(m) < a_{m+1} | S_0(m) = a_0) - P(S_n(m) < a_{m+1} | S_0(m) = a_0) = \\ \pi_0 A^{n-1}(m) \hat{e}(m)_2' - \pi_0 A^n(m) \hat{e}(m)_2' = \\ \pi_0 A^{n-1}(m) (I - A(m)) \hat{e}(m)_2'. \end{aligned}$$

2. Applying definition of moment generating function leads to

$$\begin{aligned} \phi_{N(m)}(t) &= E\left(e^{tN(m)}\right) \\ &= \sum_{n=1}^{\infty} e^{nt} P(N(m) = n | S_0(m) = 0) \\ &= \sum_{n=1}^{\infty} e^{nt} \pi_0 A^{n-1}(m) \hat{e}(m)_2' - \sum_{n=1}^{\infty} e^{nt} \pi_0 A^n(m) \hat{e}(m)_2' \\ &= (*). \end{aligned}$$

   Changing the index in the first sum from n to $n-1$, subtracting the case $n = 0$ in the second and applying the proposition above leads to

$$\begin{aligned} (*) &= \left(e^t - 1\right) \pi_0 \left(\sum_{n=0}^{\infty} e^{nt} A^n(m)\right) \hat{e}(m)_2' + 1 \\ &= 1 + \left(e^t - 1\right) \pi_0 \left(I - e^t A(m)\right)^{-1} \hat{e}(m)_2'. \end{aligned}$$

3. Differentiating and putting $t = 0$ leads to

$$\begin{aligned} E(N(m)) &= \phi'_{N(m)}(0) \\ &= \pi_0 (I - A(m))^{-1} \hat{e}(m)_2'. \end{aligned}$$

4. Differentiating a second time delivers then

$$\begin{aligned} E\left(N^2(m)\right) &= \phi''_{N(m)}(0) \\ &= \pi_0 (I + A(m)) (I - A(m))^{-2} \hat{e}(m)'_2 . \end{aligned}$$

With this variance can be computed by $E\left(N^2(m)\right) - E(N(m))^2$.

Let $f(x)$ be the probability density function of all normalized variables $\frac{X_i - \mu_0}{\sigma_0}$. Because of the definition of the a_i we have $a_{i+1} = a_i + \frac{h}{m+1}$. That is why it is comprehensible, when Fu et al. define the transition probabilities of the upper side CUSUM by

$$\begin{aligned} p_{ij} &= \int_{a_j - 0.5\frac{h}{m+1}}^{a_j + 0.5\frac{h}{m+1}} f(x - a_i)\, dx \quad \text{for} \quad i = 0, 1, \ldots, m, \quad j = 1, \ldots, m \\ p_{i0} &= \int_{-\infty}^{0.5\frac{h}{m+1}} f(x - a_i)\, dx \quad \text{for} \quad i = 0, 1, \ldots, m \\ p_{im+1} &= \int_{a_m + 0.5\frac{h}{m+1}}^{\infty} f(x - a_i)\, dx \quad \text{for} \quad i = 0, 1, \ldots, m \\ p_{m+1j} &= 0 \quad \text{for} \quad j = 0, \ldots, m \\ p_{m+1m+1} &= 1. \end{aligned} \tag{7}$$

Since p_{ij} stands for the transition form a_i to a_j, the last two equations indicate the absorbing state.

Theorem
Let $S_n(m)$ be as described above. Then it follows

$$\lim_{n \to \infty} P(N(m) > n | S_0(m) = a_0) = P(N > n | S_0 = a_0) .$$

Proof
The connection between $S_n(m)$ and $N(m)$ can also be repeated for S_n and N, i.e.

$$\begin{aligned} P(N(m) > n | S_0(m) = a_0) &= P(S_n(m) < h | S_0(m) = a_0) \\ P(N > n | S_0 = a_0) &= P(S_n < h | S_0 = a_0) . \end{aligned} \tag{8}$$

Out of the construction of the transition probabilities in (7) we have that $S_n(m)$ converges to S_n at least weakly for $m \to \infty$, since $\frac{h}{m+1} \to 0$. What follows is

$$\lim_{n \to \infty} P(S_n(m) < h | S_0(m) = a_0) = P(S_n < h | S_0 = a_0) .$$

By using the equations in (8) the proof is completed.

In contrast to the article presented in section 2.2.1 we leave out numerical results. They are available in [5]. However the topic is quite similar to all what follows since the Markov chain imbedding technique represents a rather simple and efficient instrument to produce such results. Next we turn to the permutations to use that. A general overlook about Markov chain imbedding technique with several of its applications is given in the books of Fu [7] and Koutras [2].

3 Imbedding distributions of permutations

3.1 The distribution of permutations according to their cycle structure

Consider permutations in their canonical cycle notation. As it is mentioned in the following propositions one can generate all permutations of length $n+1$ out of the ones of length n by opening a new cycle or including the new number to the right of each number in the existing cycles. To illustrate this a simple example is given for the step from two to three:

original permutation		generated permutation	
linear notation	cycle notation	cycle notation	linear notation
(12)	(1) (2)	(13) (2)	(321)
		(1) (23)	(132)
		(1) (2) (3)	(123)
(21)	(12)	(132)	(312)
		(123)	(231)
		(12) (3)	(213)

Concerning the extension of existing cycles and the insertion of a new one in the setting presented above we differ two stochastic dynamics:

- Each constellation is equally likely, i.e. each of the possible scenarios of transforming a given permutation of the length n into one of the length $n+1$ occurs with probability $\frac{1}{n+1}$ apart from the fact whether there was opened a new cycle or not.
- A parameter θ is included such that a new cycle is opened with probability $\frac{\theta}{n+\theta}$. Furthermore a constellation arising from the extension of an existing cycle occurs with probability $\frac{1}{n+\theta}$ because there are just n integers in the existing cycles, which can be extended.

The first case is identified in this framework with the term 'uniformly distributed' or 'uniform case' while the second case often is called 'biased'. One can imagine more complicated methods for constructing permutations but that shall not be the next step. See section 5. Instead of this we take the previous cycle notation in a new form, which will make it possible to construct the Markov chain of interest that delivers the base for applying the Markov chain imbedding technique.
Suppose we are interested in permutations up to length n. So we display a special permutation by a row vector, which contains in its i-th component k_i the number of cycles of length i that occur if we write down the permutation in its cycle notation. To illustrate this the example above is given in the new form:

original permutation		generated permutation	
linear notation	cycle notation	cycle notation	linear notation
(12)	$[2,0]$	$[1,1,0]$	(321)
		$[1,1,0]$	(132)
		$[3,0,0]$	(123)
(21)	$[0,1]$	$[0,0,1]$	(312)
		$[0,0,1]$	(231)
		$[1,1,0]$	(213)

Obviously parts of the information are reduced when using this structure, e.g. (13)(2) and (1)(23) lead to the same expression [1,1,0]. This fact must be taken into account regarding

the transition probabilities for permutations from length n to $n+1$:

- uniform distribution: Now one has to differ between the scenario for opening a new cycle and the other ones. Since there is still only one possibility to open the cycle this transition probability keeps its value $\frac{1}{n+1}$. The effect for the existing cycles can be caught by $\frac{k_i i}{n+1}$ because the product $k_i i$ summarizes all extension possibilities within two categories. On the one hand writing the new integer at the same place in different cycles, e.g. (13)(2) and (1)(23) lead both to [1,1,0], on the other writing the new integer at different places in the same cycle, e.g. (132) and (123) lead both to [0,0,1].
- biased distribution: Summarizing all possibilities as known from the uniformly distributed case leads to $\frac{\theta}{n+\theta}$ for the probability of opening a new cycle and to $\frac{k_i i}{n+\theta}$ for the probability of extending existing cycles.

Now we can fix our results in the following propositions, where e_i identifies the row vector with 1 at the i-th component and 0 elsewhere:

Proposition for the uniform distribution
Define $Y_t := \left(y_1^t, \ldots, y_n^t\right)$ on the n-dimensional state space $\Omega = \mathbb{N}_0^n$ of non negative integers, where y_i^t stands for the number of cycles of length i of a permutation of length t and $y_i^t = 0$ for $i > t$ by definition. K,K' shall denote realizations of $(Y_t)_{1\leq t\leq n}$ and k_i those of y_i^t for $1 \leq i \leq t$. Then $(Y_t)_{1\leq t\leq n}$ is a finite, inhomogeneous Markov chain with

1. trivial initial distribution, i.e. $P(Y_1 = (1, 0, \ldots, 0)) = 1$,
2. two possible transition probabilities listed below
   - $P(Y_t = K'|Y_{t-1} = K) = \frac{1}{t}$ for the transition to $K' = K + e_1$,
   - $P(Y_t = K'|Y_{t-1} = K) = \frac{k_i i}{t}$ for the transition to $K' = K - e_i + e_{i+1}$,

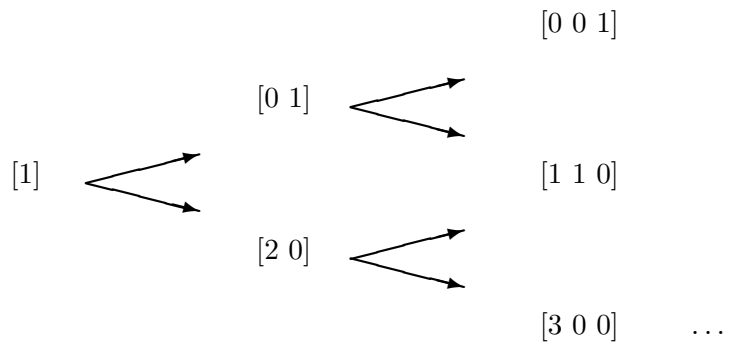
   $\forall i \in \{1, \ldots, t-1\}$ with $k_i > 0$.

Proposition for the biased distribution
Define $Y_t := \left(y_1^t, \ldots, y_n^t\right)$ on the n-dimensional state space $\Omega = \mathbb{N}_0^n$ of non negative integers, where y_i^t stands for the number of cycles of length i of a permutation of length t and $y_i^t = 0$ for $i > t$ by definition. K,K' shall denote realizations of $(Y_t)_{1\leq t\leq n}$ and k_i those of y_i^t for $1 \leq i \leq t$. θ is a parameter greater than 0. Then $(Y_t)_{1\leq t\leq n}$ is a finite, inhomogeneous Markov chain with

1. trivial initial distribution, i.e. $P(Y_1 = (1, 0, \ldots, 0)) = 1$,
2. two possible transition probabilities listed below
   - $P(Y_t = K'|Y_{t-1} = K) = \frac{\theta}{t-1+\theta}$ for the transition to $K' = K + e_1$,
   - $P(Y_t = K'|Y_{t-1} = K) = \frac{k_i i}{t-1+\theta}$ for the transition to $K' = K - e_i + e_{i+1}$,

   $\forall i \in \{1, ..., t-1\}$ with $k_i > 0$.

Note that the start transitions of the cycle structure can be written as a binary tree:



With this one has

$$P\left(Y_1 = 1, Y_2 = K_2, \ldots, Y_n = K_n\right) = P\left(Y_1 = 1\right) \prod_{i=2}^{n} P\left(Y_i = K_i | Y_{i-1} = K_{i-1}\right),$$

which is equivalent to Markov property by

$$\begin{aligned} P\left(Y_n = K_n | Y_{n-1} = K_{n-1}, \ldots, Y_2 = K_2, Y_1 = 1\right) = \\ \frac{P\left(Y_1 = 1, Y_2 = K_2, \ldots, Y_n = K_n\right)}{P\left(Y_1 = 1, Y_2 = K_2, \ldots, Y_{n-1} = K_{n-1}\right)} = \\ P\left(Y_n = K_n | Y_{n-1} = K_{n-1}\right). \end{aligned}$$

Notice that the tree above does not continue so easily. For example [1 1 1 0 0 0] has three forerunners. These are [1 2 0 0 0], [2 0 1 0 0] and [0 1 1 0 0]. Furthermore [1 1 1 0 0 0] creates more than two successors. These are [2 1 1 0 0 0 0], [0 2 1 0 0 0 0], [1 0 2 0 0 0 0] and [1 1 0 2 0 0 0]. Otherwise there would be just n states after n steps. This does not change that Markov property holds.

With the propositions above we could just generate for all transitions from $t-1$ to t the cycle structures and write down the corresponding transition probabilities in transition probability matrices. According to the theorem of section 2.1 multiplying these would deliver the result we are looking for. Note that this matrix multiplication up to $n = 50$ with alone $a\left(50\right) = 204.226$ possible states for cycle structures of permutations with length 50 takes too long, although a lot of the entries in the matrices are equal to 0. That is why we use the structure of the tree above.

For each step it is enough to store the actual cycle structures as well as their probabilities. The new states and their probabilities can be computed by the propositions given above. Challenging is that the algorithm must realize which states lead to the same successor, i.e. the fact that the tree above is irregularly recombining. Thereby it should be used as little storage as possible. After that storage is overwritten with the new information. Somehow it could be said that the algorithm even emphasizes Markov property. For implementation in Matlab see appendix C.

3.2 Implementing further distributions of permutations

The distributions in this section are an automatic result of the propositions given in section 3.1. If the probability of a certain cycle structure (and of course the corresponding cycle structure) is known there are only elementary operations needed in order to generate out of this the following distributions.

The structure of the following sections is as follows: First the mechanism of the elementary operations is described. Then distributions for permutations of length 50 regarding the uniform and the biased case are presented. Last but not least some comments about the results are given.

Concerning the biased case on the one hand θ is chosen equal to 0.5 on the other equal to 1.5. It is reasonable to consider results for $\theta < 1$ as well as for $\theta > 1$ because of the fact that the uniform case is given by $\theta = 1$. According to section 3.1

$$\frac{1}{n+1} \qquad \text{must be compared with} \qquad \frac{\theta}{n+\theta}, \tag{9}$$

where (9) represents the probability for opening a new cycle. Note that $f(\theta) = \frac{\theta}{n+\theta}$ is monotonically increasing for $n \in \mathbb{N}, \theta > 0$ since for these assumptions $f'(\theta) = \frac{n}{(n+\theta)^2} > 0$. That is why

$$\frac{\theta_{<1}}{n+\theta_{<1}} < \frac{1}{n+1} < \frac{\theta_{>1}}{n+\theta_{>1}}.$$

Or in other words for the choice of $\theta < 1$ one expects fewer cycles. Because of

$$\sum_{i=1}^{n} k_i i = n \tag{10}$$

for fixed n and all cycle structures $[k_1, \ldots, k_n]$ higher probability on fewer cycles also implies higher probability on longer cycles. Contrary the choice of $\theta > 1$ corresponds to more mass on more and shorter cycles compared with the uniform case as a benchmark. Note that (10) plays an important role in the so called Conditioning Relation mentioned in section 4.

3.2.1 The distribution of the number of cycles

For the number of cycles it is enough to accumulate the entries of a cycle structure and to add the corresponding probabilities when the sum is identical for different cycle structures.

The following figure presents the distributions of the number of cycles for permutations of length $n = 50$ with different parameter choice $\theta_1 = 0.5, \theta_2 = 1$ and $\theta_3 = 1.5$.

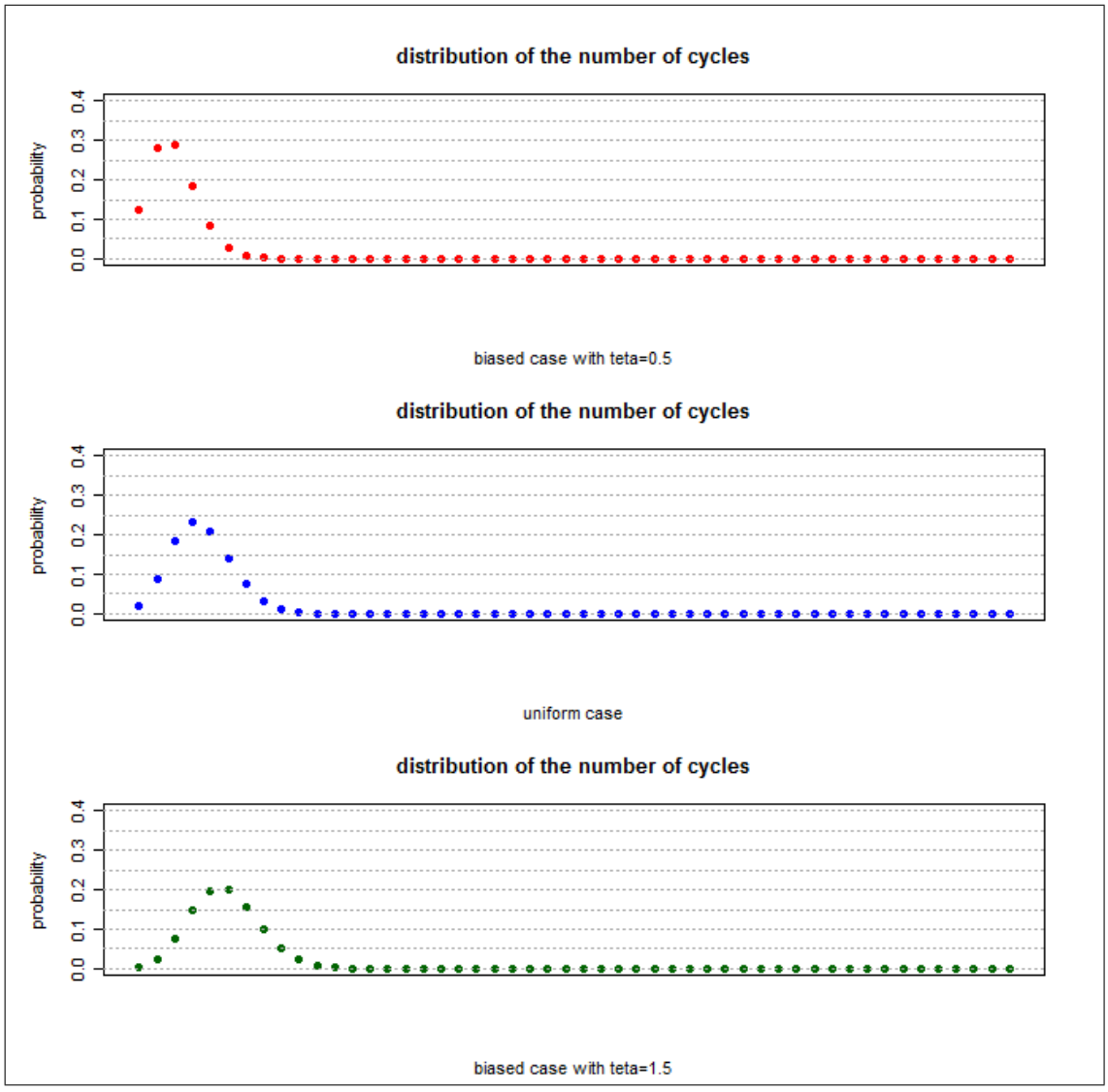


Figure 6: Explicit distribution for the number of cycles with $n = 50$

As expected the mass on more cycles increases with increasing θ. With this the normal limit distribution of section 4.2.3 can already be recognized.

3.2.2 The distribution of the rth longest and shortest cycle

We differ here between longest, rth longest, shortest, rth shortest cycle length. For the common cases we provide examples with $r = 5$. The difference between the longest and the shortest cycles is based on counting forwards or backwards in a particular cycle structure.

For the longest cycle length we detect the first position from behind which is unequal to 0. Corresponding probabilities with identical results must be added.
The following figure presents the distributions of the length of the longest cycle for permutations of length $n = 50$ with different parameter choice $\theta_1 = 0.5, \theta_2 = 1$ and $\theta_3 = 1.5$.

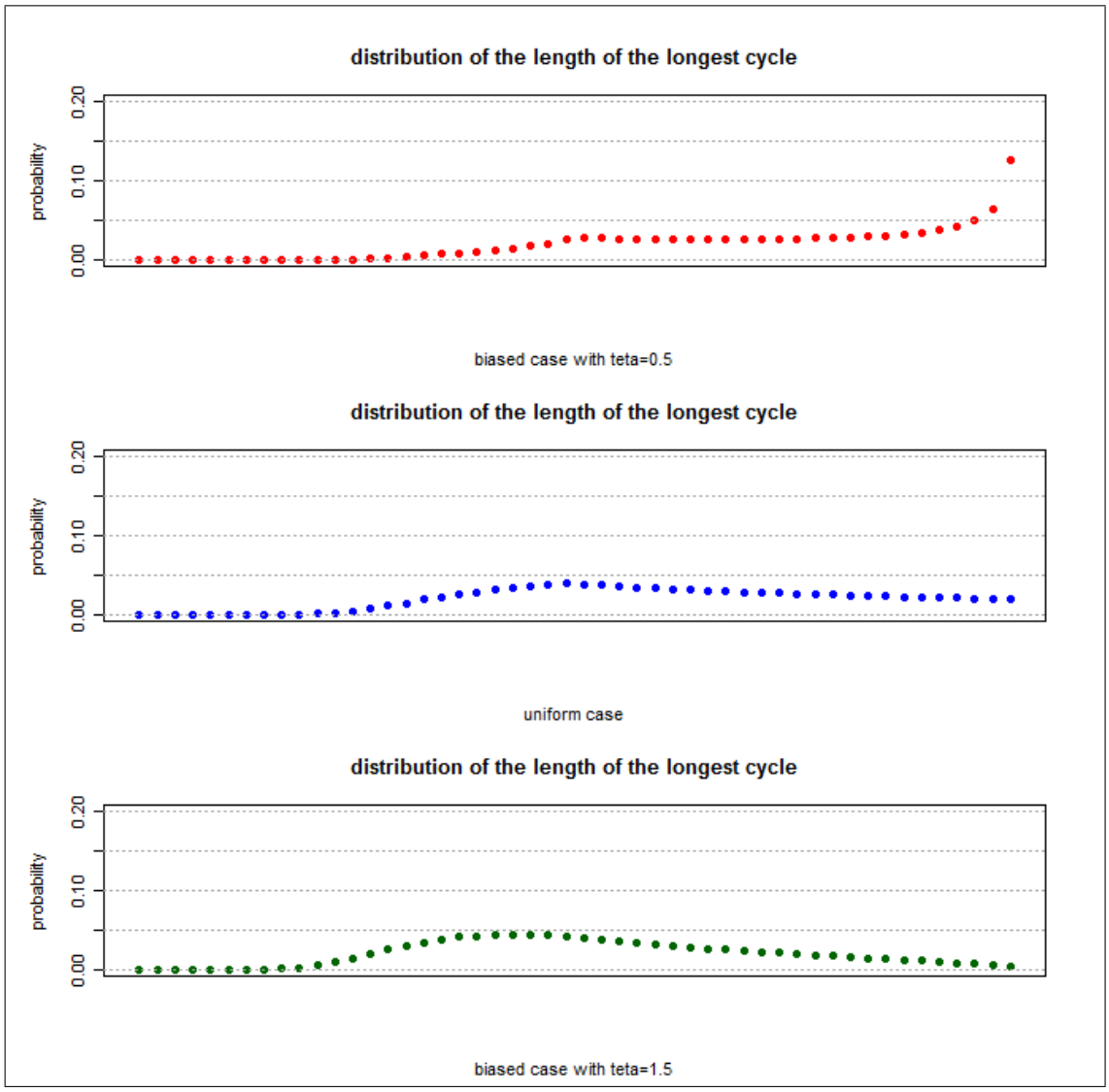


Figure 7: Explicit distribution for the length of the longest cycle with $n = 50$

For $\theta = 0.5$ almost all mass belongs to large cycles, where as in the other cases a bow can be recognized. The center of $\theta = 1.5$ is on the left side of the one of $\theta = 1$. This is consistent with (9). Note that there are little waves in the bows.

For the length of the rth longest cycle we build up accumulative cycle structures from behind. By going backwards the first position must be detected where the entry is greater or equal to r. Of course it is possible that a cycle structure does not contain r cycles. In this case the length of the rth cycle is defined equal to 0. In the following figures (rth longest and shortest case) this is missed out because of the scaling. This means for these figures and their points p_i for $i = 1, \ldots, n$ the probability $1 - \sum p_i$ has to be allocated to length 0. The following figure presents the distributions of the length of the 5th longest cycle for permutations of length $n = 40$ with different parameter choice $\theta_1 = 0.5, \theta_2 = 1$ and $\theta_3 = 1.5$.

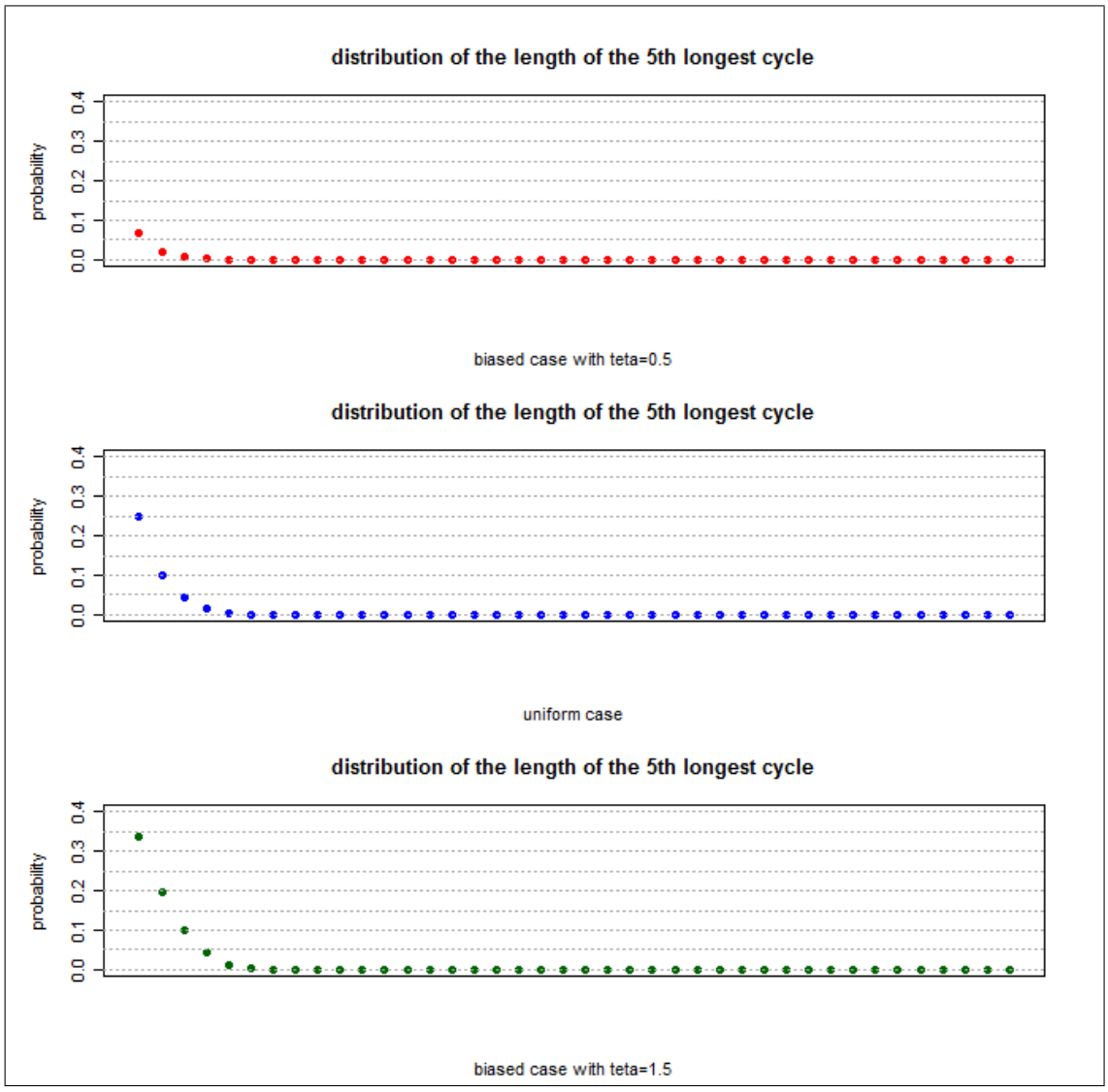


Figure 8: Explicit distribution for the length of the 5th longest cycle with $n = 40$

Again the results are consistent with the θ-constellation. In the case of $\theta = 0.5$ all most all mass is on length equal to 0. This probability is not mentioned in the figures above and corresponds to the fact that there do not exist at least 5 cycles in the cycle structure. It has been checked that the probabilities for this are the same as the one in the case of the 5th shortest cycle. For $\theta = 0.5$ the probability for this must be the greatest since we have the mass on structures with few cycles. With increasing θ this phenomenon is damped.

For the length of the shortest cycle we just detect the position of the first entry unequal to 0. The following figure presents the distributions of the length of the shortest cycle for permutations of length $n = 50$ with different parameter choice $\theta_1 = 0.5, \theta_2 = 1$ and $\theta_3 = 1.5$.

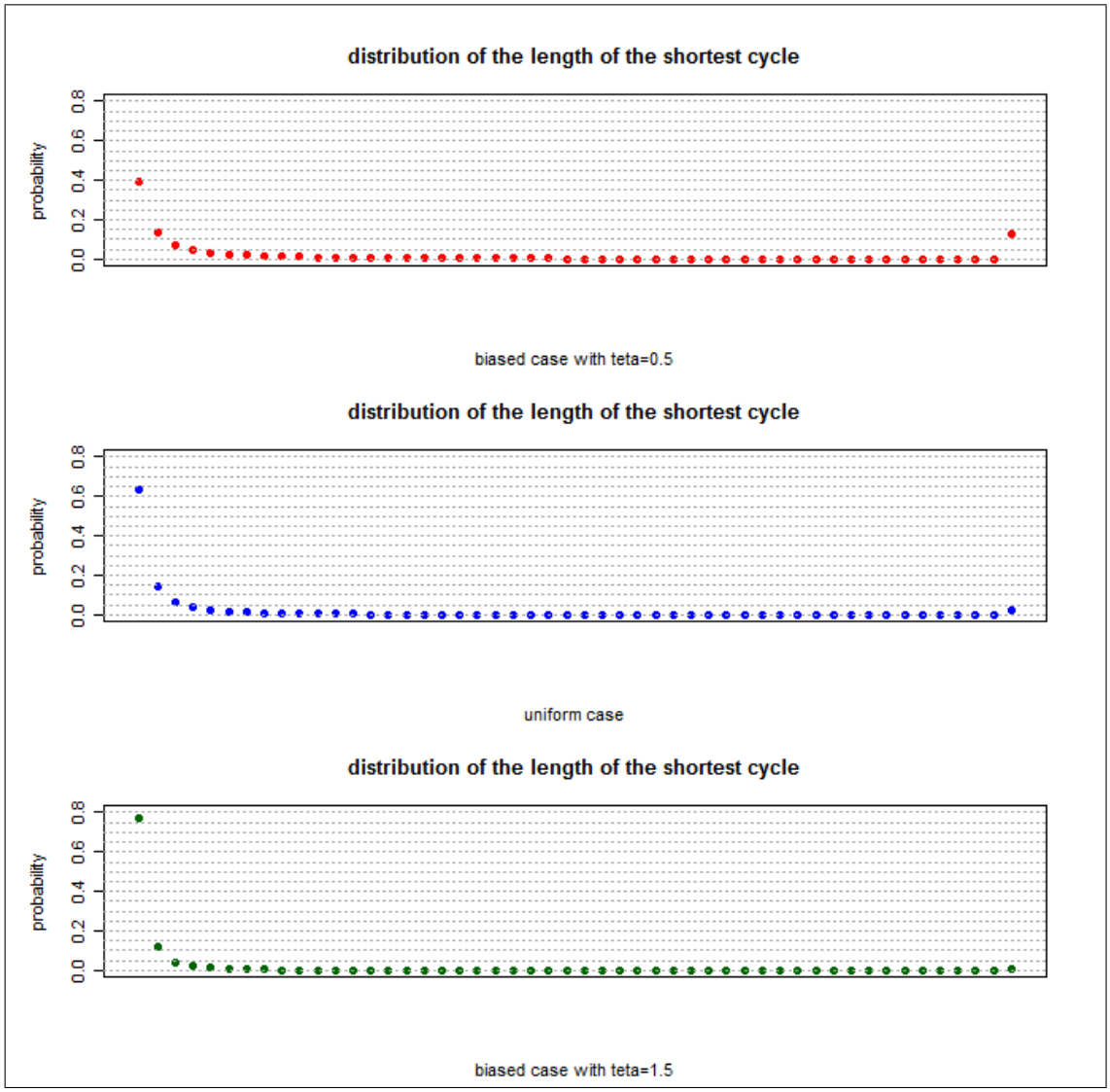


Figure 9: Explicit distribution for the length of the shortest cycle with $n = 50$

Where as the cases $\theta = 1$ and $\theta = 1.5$ are quite similar to the probability density functions of exponential distributions, $\theta = 0.5$ has so much mass on the cycle structure $[0, \ldots, 0, 1]$ that the probability after getting down to 0 and taking a long run there increases again at the end.

For the length of the rth shortest cycle we build up accumulative cycle structures. By going forwards the first position must be detected where the entry is greater or equal to the parameter r. Again it is possible that a certain cycle structure has not r cycles. For this see the description in the section about the rth longest cycle.
The following figure presents the distributions of the length of the 5th shortest cycle for permutations of length $n = 50$ with different parameter choice $\theta_1 = 0.5, \theta_2 = 1$ and $\theta_3 = 1.5$.

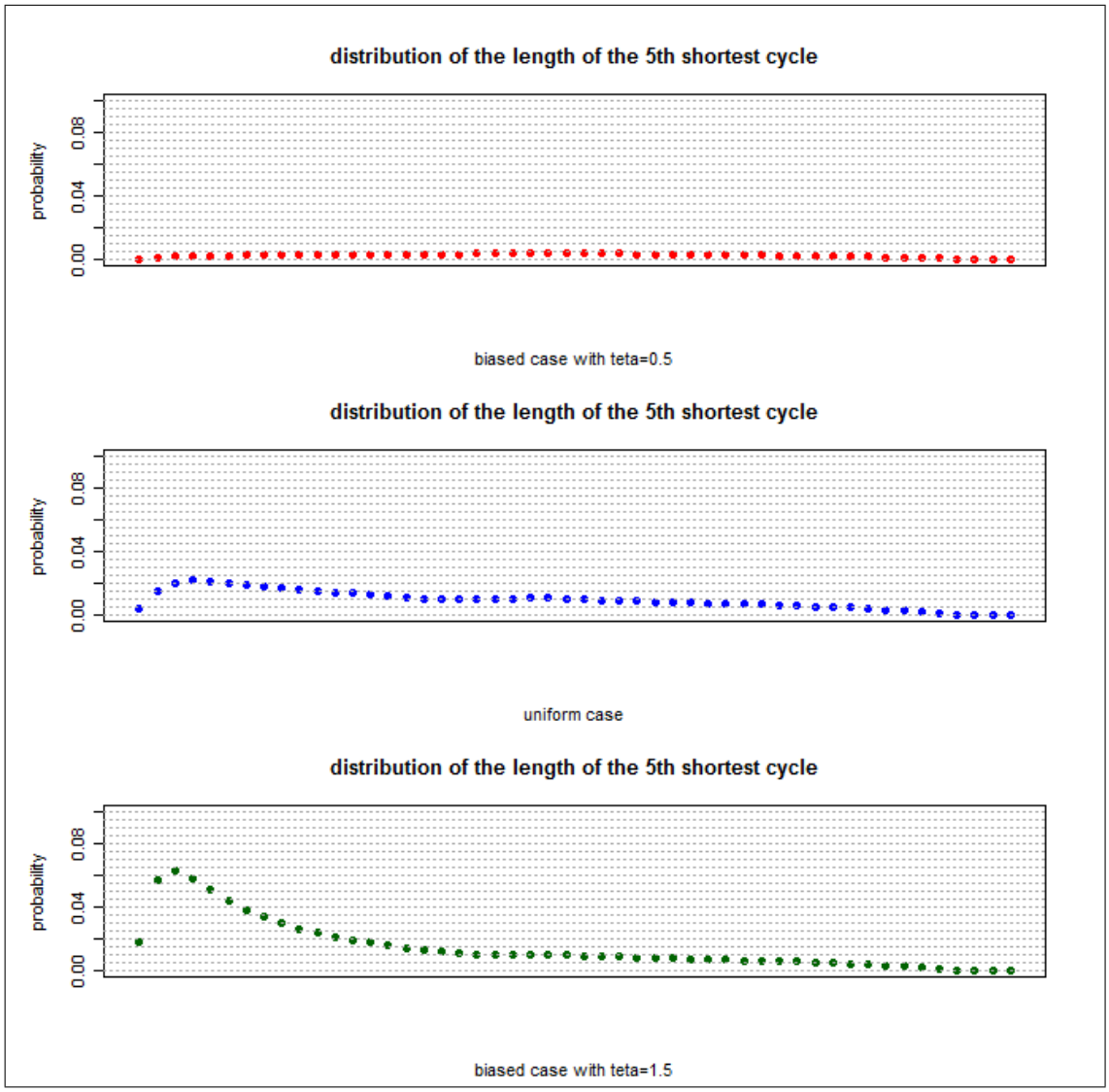


Figure 10: Explicit distribution for the length of the 5th shortest cycle with $n = 50$

For $\theta = 0.5$ we have almost all mass on large cycles. The corresponding structures do not contain many cycles respectively here 5 cycles. That is why almost all mass is on length equal to 0. Again little waves can be seen. The movement of the major mass from the right to the left with increasing θ is consistent with (9).

3.2.3 The distribution of the length of a random chosen cycle

For the length of a random chosen cycle each cycle in the structure is uniformly distributed. So each length in the cycle structure gets its probability by multiplying the structure probability with the number of existing cycles divided by the total number of cycles in the structure. The sum of the probabilities of a certain length over all structures delivers the distribution of the length of a random chosen cycle.
The following figure presents the distributions of the length of a random chosen cycle for permutations of length $n = 50$ with different parameter choice $\theta_1 = 0.5, \theta_2 = 1$ and $\theta_3 = 1.5$.

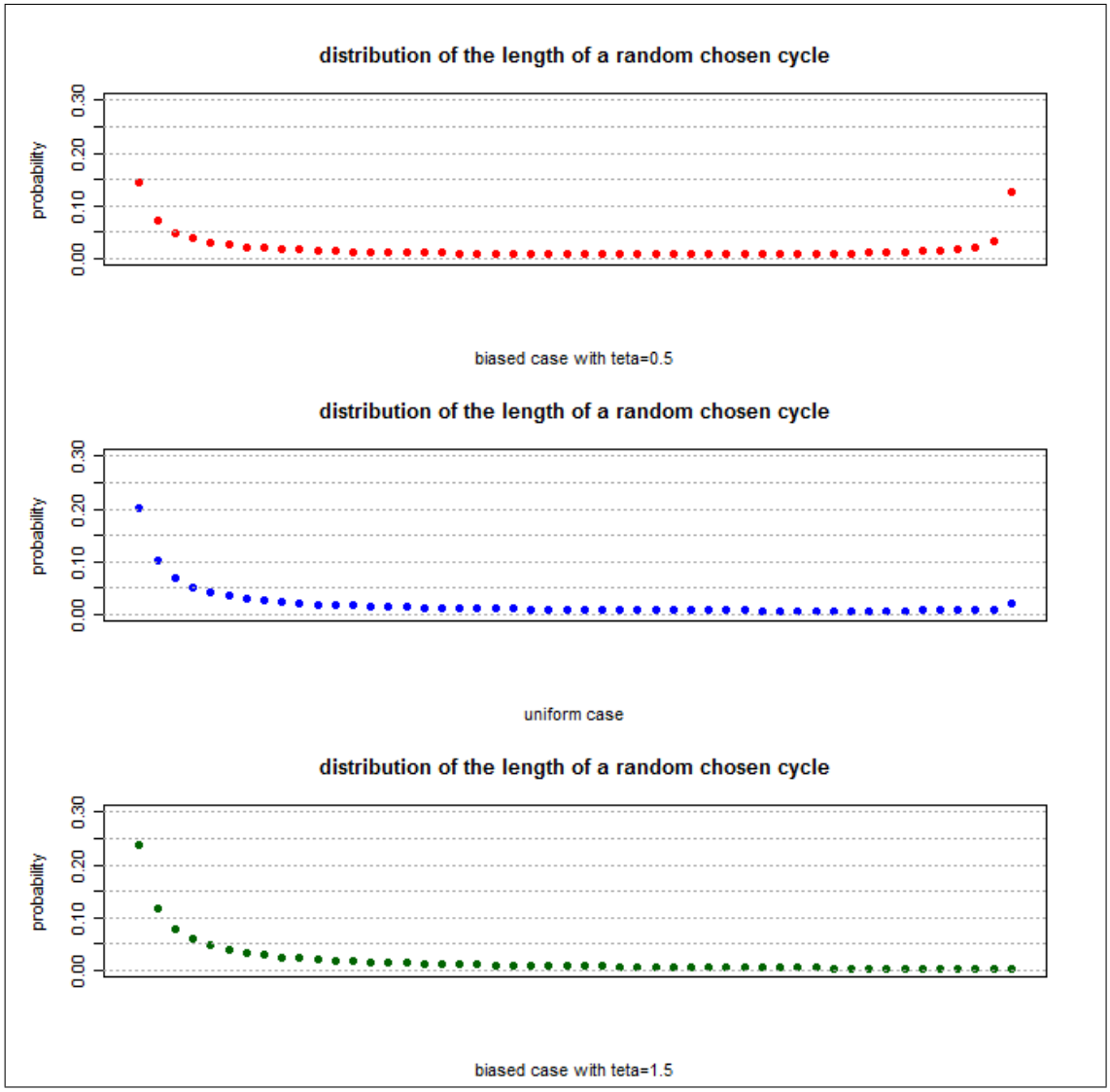


Figure 11: Explicit distribution for the length of a random chosen cycle with $n = 50$

Most of the mass is on small cycles since here only higher number of cycles exist. Again (9) can be recovered since the most mass on short cycles is given for $\theta = 1.5$. With $\theta = 1$ a small rising of the probability can be seen at the end. This is based on the high probability for the structure $[0, \ldots, 0, 1]$ even in the uniform case. This development strongly continues so that the mapping for $\theta = 0.5$ is quite similar to a stretched 'U'. Consider the scaling when comparing with the distribution of the shortest cycle length.

3.3 Run time of the total algorithm

Target of this section is a rough run time estimation of the MCIT algorithm. It will become clear that even for permutations of length 50 which are analyzed in this work high computational performance or rather high run time is needed. Since we are dealing with distributions concerning the cycle notation of permutations it is not necessary to run through all permutations ($50! \approx 3 \cdot 10^{64}$), but through all possible cycle structures also depending on the length of the permutations. These are given by the sequence $(a_n)_{n\in\mathbb{N}}$ (the so called A000041), where a_n stands for the number of partitions of n fulfilling the condition $b + 2c + 3d + \ldots = n$ with $b, c, \ldots \in \mathbb{N}_0$. The following figure shows the number of partitions up to $n = 45$.

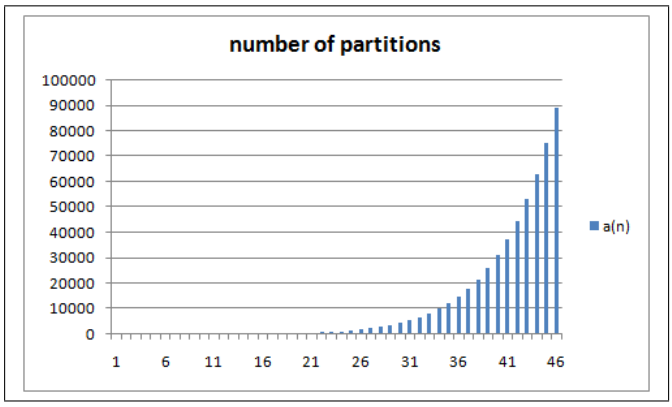


Figure 12: Sequence of the number of partitions

Obviously a_n increases exponentially. In order to generate the distributions the transition probabilities of section 3.1 are essential. There we had for example

$$P\left(Y_t = K' | Y_{t-1} = K\right) = \frac{k_i i}{t} \quad \forall i \in \{1, \ldots, t-1\} \quad \text{with} \quad k_i > 0,$$

where K stands for one of the partitions. To check each k_i and calculate, if possible, the transition probability at least $c\,(t-1)$ elementary operations, such as multiplication and devision, are necessary. For the term above there would be $c = 2$. The run time for these elementary operations depends on the power of the computer, e.g. 10^{-6} seconds. Putting parts together the total run time of the algorithm can be estimated by

$$\begin{aligned} \text{run time} &= c_1 \text{length}\; e^{c_2 \text{length}} \\ \Leftrightarrow \log\left(\frac{\text{run time}}{\text{length}}\right) &= \log c_1 + c_2 \text{length}. \end{aligned} \tag{11}$$

Now constants c_1 and c_2 may be estimated by ordinary least squares method (OLS). Strictly speaking there would have to be some corrections to do the best, but for a rough estimation (11) is good enough. Matlab also allows to measure cpu time directly by

$$\begin{aligned} t_1 &= \text{cputime}; \\ &\text{Zyklen10(50,1,5)}; \\ t_2 &= \text{cputime} - t_1. \end{aligned}$$

This leads to the following run times on a 8 gigabyte ram computer:

length	runtime
10	0.03 seconds
20	2.13 seconds
30	3.83 minutes
40	4.25 hours
50	10.72 days

With this and (11) constants c_1 and c_2 can be estimated. This was done for the lengths $10 - 40$, which leads to $c_1 = 5.18 \cdot 10^{-5}$ and $c_2 = 0.3954$. Calculating the run time for 50 then delivers 11.56 days, which is pretty near to the true cpu time. On the basis of that it becomes clear that the practical use of the MCIT algorithm in appendix C is limited for permutations far-off 50. However there might be potential for further improvements. For example note that the algorithm contains distributions mentioned in section 3 all in one.

4 Selected distributions in the context of permutations

In this section we present explicit formulas for the former distributions, if such formulas exist. Beside whether they exist or not we show approximations for them by considering the case where the permutation length converges to ∞, i.e. we examine the so called limit distributions.

4.1 The distribution of permutations according to their cycle structure

4.1.1 Explicit distribution

Uniform Case
Since $n!$ permutations of length n exist the challenge in the uniform case consists of the derivation of the number of permutations $N(n, y)$ which have cycle structure $y = (y_1, \ldots, y_n) \in \mathcal{N}_0^n$ where y_j for $1 \leq j \leq n$ stands for the number of cycles of length i. One has to prove

$$N(n, y) = 1_{\{\sum_{j=1}^{n} j y_j = n\}} n! \prod_{j=1}^{n} \left(\frac{1}{j}\right)^{y_j} \frac{1}{y_j!}. \tag{12}$$

As a first step it is useful to consider the canonical cycle notation und to neglect the brackets within. That leads to the total number of permutations and explains where the term $n!$ comes from in equation (12). Obviously one is over counting in this setting. To detect the degree of over counting different cases must be taken into account. Consider that

$$(12)(34) \Leftrightarrow (34)(12).$$

This means one has to correct the influence of the order between identical cycles since this is irrelevant for the cycle structure. In the example above there are $2 = 2!$ possibilities for the same structure of cycles of length 2. Consequently there are $y_j!$ possibilities for the same structure of cycles of length j. That is why in equation (12) a division with the product of the $\frac{1}{y_j!}$ for $1 \leq j \leq n$ is necessary. Next consider that

$$(12)(34) \Leftrightarrow (12)(43) \Leftrightarrow (21)(34) \Leftrightarrow (21)(43).$$

This means one has to correct the influence of the order within the cycles when this is irrelevant for the cycle structure. In the example above there are $2^2 = 4$ possibilities for the same structure of cycles of length 2. Note that here the order between the cycles has not been changed. Since there are j possibilities for a cycle of length j to write down the same cycle (each of the j elements can be written at the beginning) and y_j of such cycles exist there are j^{y_j} possibilities for the same structure of cycles of length j. That is why in equation (12) a division with the product of j^{y_j} for $1 \leq j \leq n$ is necessary. The constraint in the indicator function above should be clear since the total of the y_j cycles multiplied by their length j for $1 \leq j \leq n$ can only be filled up by n elements. Since the probability of a particular permutation in the uniform setting is given by $\frac{1}{n!}$ the demanded probability follows by

$$P(Y_n = y) = 1_{\{\sum_{j=1}^{n} j y_j = n\}} \prod_{j=1}^{n} \left(\frac{1}{j}\right)^{y_j} \frac{1}{y_j!}. \tag{13}$$

The formulas above also deliver the base for the biased case.

Biased Case
The distribution of θ-biased permutations according to their cycle structure is determined by the so called Ewens Sampling Formula ESF(θ). Compare [1]. As it is discussed in detail in section 4.2 the distribution of a particular θ-biased permutation π with k cycles is given by

$$P_\theta(\pi) = \frac{\theta^k}{\theta^{(n)}} = \frac{\prod_{j=1}^{n} \theta^{y_j}}{\theta^{(n)}},$$

where y_j corresponds again to the number of cycles of length j and the denominator contains the raising factorial. That is why (13) must be changed into

$$P(Y_n = y) = 1_{\{\sum_{j=1}^{n} j y_j = n\}} \frac{n!}{\theta^{(n)}} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{y_j} \frac{1}{y_j!}, \tag{14}$$

which displays ESF(θ).

4.1.2 Application of previous formulas

One could ask why it is a good idea to apply Markov chain imbedding technique, i.e. to determine the probability that certain cycle structure occurs by transition probabilities of a Markov chain, when the formulas mentioned in the previous section exist. The answer must be divided into two parts. First: The formulas given under the title 'Explicit distribution' can only be applied when the cycle structure is known. This means an algorithm generating numerical values has to create all possible cycle structures of a certain length at any rate, before it makes sense to apply the formulas. Second: The fact that for the uniform and θ-biased case such formulas exist does not ensure that this is also the case for biased transitions from n to $n+1$, which follow a more complicated way. For the second point see section 5.

Obviously the distribution of the cycle structure, i.e. the one of conjugacy classes in the symmetric group S_n, can also be determined for a certain n by the following two step mechanism

- use the Matlab - Code of appendix C to generate all the cycle structures of order n
- apply the formulas for uniform and biased case.

Exactly this was done, especially in order to confirm the results of MCIT. The results are presented in appendix B. Like in the algorithm of appendix C all other distributions presented in the following sections can be generated out of the results which are provided by the examination of the cycle structure. Nevertheless it is useful to analyze the other distributions without this indirection.

4.2 The distribution of the number of cycles

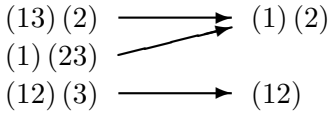
4.2.1 Explicit distribution

Uniform Case
Following the structure of the previous chapters one has to differ the uniform and the biased case. Starting with the uniform case it is crucial to look at unsigned Stirling numbers of the first kind $|s(n,k)|$, for which it is said that they represent the number of permutation with n elements and k disjoint cycles $\#\{\pi(n,k)\}$. We formally want to derive why this is true:

For this we have to prove the following recurrence relation

$$\#\{\pi(n+1,k)\} = n\#\{\pi(n,k)\} + \#\{\pi(n,k-1)\}.$$

It is clear that there are only two possibilities to create a permutation of length $n+1$ with k disjoint cycles out of a permutation of length n. First of all if the original permutation also has k cycles and no new cycle is added. Next if the original permutation has $k-1$ cycles and a new cycle is added. That is why the formula above contains two summands. For the first term on the right side of the equation one has to consider a particular permutation with its k cycles. It is possible to write next to the right of each of the n elements a newcomer, which extends the permutation to the length of $n+1$ without changing its total number of cycles. For the second term it is clear that for each of the permutations there is only one possibility to reach the necessary number of cycles by inserting a new cycle which contains the newcomer. The following example for the number of permutations of length 3 with 2 disjoint cycles illustrates the implicit mapping where the permutations are mapped on their origin:



Stirling numbers of the first kind are defined by

$$(x)_n = \sum_{k=0}^{n} s(n,k)x^k, \tag{15}$$

where $(x)_n = x\,(x-1)\dots(x-n+1)$ denotes the falling factorial. Then the unsigned Stirling number of the first kind is defined by $|s(n,k)| = s(n,k)(-1)^{-(n+k)}$. In order to show that $\#\{\pi(n,k)\} = |s(n,k)|$ induction is used. Thereby $\#\{\pi(n,k)\}$ is defined equal

to 0 for $k \leq 0$.

$$induction - beginning \qquad n = 1$$

$$\sum_{k=0}^{1} \#\{\pi(1,k)\}(-1)^{-(1+k)} x^k = x = (x)_1$$

$$induction - step \qquad n \to n+1$$

$$\sum_{k=0}^{n+1} \#\{\pi(n+1,k)\}(-1)^{-(n+1+k)} x^k \underbrace{=}_{recurrence}$$

$$\sum_{k=0}^{n+1} n\#\{\pi(n,k)\}(-1)^{-(n+k)}(-1)x^k + \#\{\pi(n,k-1)\}(-1)^{-(n+1+k)} x^k =$$

$$-n\,(x)_n + x \sum_{k=-1}^{n} \#\{\pi(n,k)\}(-1)^{-(n+k)}(-1)^{-2} x^k =$$

$$-n\,(x)_n + x \sum_{k=0}^{n} \#\{\pi(n,k)\}(-1)^{-(n+k)} x^k =$$

$$(x-n)\,(x)_n = (x)_{n+1}\,.$$

After the number of permutations with k cycles is determined it is easy to write down the distribution of the number of cycles N since the total number of permutations of length n is given by $n!$:

$$P(N = k) = \frac{|s(n,k)|}{n!}.$$

Biased Case

The number of cycles N of a θ-biased permutation of length n can also be explicitly determined by using the so called Feller coupling [1]. For that we introduce independent variables D_i for $i = 1, \ldots, n$ which describe the i-way choice between closing and extending a cycle. If the first cycle with the 1 at position 1 is written down in canonical cycle notation, i.e. (1..., there will be n possibilities to continue. First of all the 1 can stand in a single cycle, i.e. (1)... . That shall be captured by the event $D_1 = 1$. Next the cycle can be extended with one of the remaining $n-1$ numbers, i.e. (12...,(13...,(.... That shall be captured by the events $D_1 = j$ with $2 \leq j \leq n$.

With this construction a particular permutation in its canonical cycle notation can be presented by the D_i's. We give an example for a permutation of length 5:

$$(12)(34)(5) \Leftrightarrow D_1 = 2, D_2 = 1, D_3 = 4, D_4 = 1, D_5 = 1.$$

Looking at the equivalence above there are in general 5 possibilities for D_1, i.e.(1),(12,(13, (14,(15. After (12 has been chosen there are 4 possibilities for D_2, i.e. (12),(123,(124,(125 and so on until the last cycle has to be closed by D_5. Like in section 3.1 we can determine the probabilities for closing and updating the cycles:

$$P_\theta\,(D_i = 1) = \frac{\theta}{\theta + i - 1} \qquad P_\theta\,(D_i = j) = \frac{1}{\theta + i - 1}, \quad j \neq 1.$$

Stochastic independence is given since the choice of closing the actual cycle is not influenced by the fact, how long the cycle is, i.e. when the last cycle has been closed. So it is easy to infer from the distribution of the D_i's to the one of a particular permutation because of the independence:

$$P_\theta\,(\pi) = \frac{\theta^k \quad 1^{n-k}}{(\theta+1-1)\,(\theta+2-1)\ldots(\theta+n-1)} = \frac{\theta^k}{\theta^{(n)}}, \tag{16}$$

where $\pi \in \mathcal{S}_n$. k is equal to the number of cycles in π and $(\cdot)^{(\cdot)}$ stands for the Pochhammer symbol of the raising factorial. Next the distribution of the number of cycles N can directly be deviated from the number of permutations with k cycles:

$$P(N=k) = \sum_{\pi:\#cycles=k} P(\pi) = \frac{\theta^k}{\theta^{(n)}} \sum_{\pi:\#cycles=k} 1 = \frac{\theta^k |s(n,k)|}{\theta^{(n)}}, \tag{17}$$

where $k = 1, \ldots, n$ and the unsigned Stirling number of the first kind $|s(n,k)|$ again identifies the number of permutations in $\mathcal{S}_n$ having k disjoint cycles.
Since explicit formulas for the distribution of the number of cycles exist calculations described in section 3 can be checked by these. We obtained the same results. Next we turn to the limit distributions.

4.2.2 Local limit distribution

Uniform Case
Since we used the unsigned Stirling number of the first kind to quantify the probability of the number of cycles being equal to k we try the following approximation for $s(n,k)$

$$|s(n,k)| \sim \frac{(n-1)!}{(k-1)!} (\gamma + \log n)^{k-1}. \tag{18}$$

To prove that as true we largely follow the way in [12]. The Stirling number was introduced in (15). Rewriting of the left side leads to

$$(x-0)(x-1)\ldots(x-n+1).$$

This looks quite similar to Viète's formulas. So let $s(n,k)$ stand for the coefficients in the following equation

$$x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 = (x-u_1)\ldots(x-u_n),$$

where $u_i = i-1$ for $1 \leq i \leq n$ identify the roots for the case that the equation is put equal to 0. Then Viète's formulas tell us that

$$s(n,k) = (-1)^{n-k} \sum_{1 \leq i_1 < i_2 < \ldots < i_{n-k} \leq n} u_{i_1} \ldots u_{i_{n-k}} \quad for \quad 0 \leq k \leq n-1.$$

Consider that the sum runs over all combinations of $n-k$ elements without repetition and permutation, where the u_i are integers up to $n-1$. That is why dividing by $(n-1)!$ leads to $k-1$ elements remaining in the denominator, i.e.

$$s(n,k) = (-1)^{n-k}(n-1)! \sum_{1 < i_1 < \ldots < i_{k-1} \leq n} \frac{1}{u_{i_1} \ldots u_{i_{k-1}}} \quad for \quad 2 \leq k \leq n-1. \tag{19}$$

Note that by $1 < i_1$ the case $u_i = 0$ is excluded in the last sum. Furthermore the last equation does not capture the cases $s(n,0)$ and $s(n,1)$. The first one can be easily included by $\mathbb{1}_{\{n=0\}}$ and the second one by definition of $u_{i_0} := 1$. We mention that the sign of the Stirling number of the first kind is determined by $(-1)^{n-k}$, because this was used in the section about the explicit formulas. Momentarily let us have a look at the term below,

$$\sum_{u_{i_1}=1}^{n-1} \cdots \sum_{u_{i_{k-1}}=1}^{n-1} \frac{1}{u_{i_1}} \cdots \frac{1}{u_{i_{k-1}}}, \tag{20}$$

which has no restrictions regarding repetition and permutation. The target is to correct the sum in such a way that it is identical to the one in (19). We start with an example for $n = 4$. Viètes formulas deliver

$$\begin{aligned} -a_3 &= u_1 + u_2 + u_3 + u_4 = 1 + 2 + 3 = 6 \\ a_2 &= u_1u_2 + u_1u_3 + u_1u_4 + u_2u_3 + u_2u_4 + u_3u_4 = 1 \cdot 3 + 2 \cdot 3 + 1 \cdot 2 = 11 \\ -a_1 &= u_1u_2u_3 + u_1u_2u_4 + u_2u_3u_4 = 1 \cdot 2 \cdot 3 = 6 \\ a_0 &= u_1u_2u_3u_4 = 0. \end{aligned}$$

Equation (19) delivers

$$\begin{aligned} -s\left(n,3\right) &= (n-1)!\left(\frac{1}{u_2u_3} + \frac{1}{u_3u_4} + \frac{1}{u_2u_4}\right) = 3!\left(\frac{1}{3} + \frac{1}{6} + \frac{1}{2}\right) = 6 \\ s\left(n,2\right) &= (n-1)!\left(\frac{1}{u_2} + \frac{1}{u_3} + \frac{1}{u_4}\right) = 3!\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3}\right) = 11 \\ -s\left(n,1\right) &= (n-1)!\frac{1}{u_{i_0}} = 6 \\ s\left(n,0\right) &= 0. \end{aligned}$$

Calculation with (20) for the cases $k = 2, 3$ delivers

$$\begin{aligned} k = 3: \quad \left(\frac{1}{u_2} + \frac{1}{u_3} + \frac{1}{u_4}\right)^{k-1} &= \frac{1}{u_2^2} + \frac{2}{u_2u_3} + \frac{2}{u_2u_4} + \frac{1}{u_3^2} + \frac{2}{u_3u_4} + \frac{1}{u_4^2} \\ k = 2: \quad \left(\frac{1}{u_2} + \frac{1}{u_3} + \frac{1}{u_4}\right)^{k-1} &= \frac{1}{u_2} + \frac{1}{u_3} + \frac{1}{u_4}. \end{aligned}$$

With this example the structure of the correction mentioned above becomes clear. Repetition in the case $k = 3$ can be corrected by subtracting $\sum_{u_i=1}^{n-1}\left(\frac{1}{u_i}\right)^2$ and permutation can be corrected by dividing through $(k-1)!$. For the case $k = 2$ no correction regarding repetition is necessary and applying the correction with $(k-1)! = 1$ does not matter. Putting $\sigma_l = \sum_{u_{i_j}=1}^{n-1}\left(\frac{1}{u_{i_j}}\right)^l$ for $1 \leq j \leq k-1$ is helpful for a general rewriting of (19) to a sum, where repetition is allowed. For example we had in the case $(n,k) = (4,3)$ analogy between (19) and $\sigma_1^2 - \sigma_2$. Another example would be $(n,k) = (5,4)$. Since $u_{i_1} = u_{i_2}$,$u_{i_1} = u_{i_3}$ and $u_{i_2} = u_{i_3}$ is possible, the correction by $\sigma_1\sigma_2$ must take place three times. But with this the case $u_{i_1} = u_{i_2} = u_{i_3}$ has been corrected two times too often. That is why (19) is here equal to $\sigma_1^3 - 3\sigma_1\sigma_2 + 2\sigma_3$. Taking this into account and remembering the correction of permutation finally leads to

$$s\left(n,k\right) = \left(-1^{n-k}\right)\frac{(n-1)!}{(k-1)!}\sum a_{(\lambda_1,\ldots,\lambda_{k-1})}\sigma_1^{\lambda_1}\ldots\sigma_{k-1}^{\lambda_{k-1}},$$

where $\lambda_l = 0, 1, \ldots$ for $1 \leq l \leq k-1$. Out of the construction it is clear that $\sum_{l=1}^{k-1} l\lambda_l = k-1$ since the total degree of each of the involved fractions is equal to $k-1$. Here the explicit values of $a_{(\lambda_1,\ldots,\lambda_{k-1})}$ need not to be examined. We just treat it as a numerical constant when n is going to infinity. Now we are ready to look at the limit

$$\lim_{n\to\infty}\frac{\left|s\left(n,k\right)\right|}{(n-1)!\left(\gamma+\log n\right)^{k-1}} =$$

$$\lim_{n\to\infty}\frac{1}{\left(\gamma+\log n\right)^{k-1}}\left(-1\right)^{-(n-k)}\left(-1\right)^{n-k}\frac{(n-1)!}{(n-1)!\,(k-1)!}\sum a_{(\lambda_1,\ldots,\lambda_{k-1})}\sigma_1^{\lambda_1}\ldots\sigma_{k-1}^{\lambda_{k-1}} =$$

$$\lim_{n\to\infty}\frac{1}{(k-1)!}\left[\underbrace{\frac{\sigma_1^{k-1}}{\left(\gamma+\log n\right)^{k-1}}}_{\to 1} + \ldots\right].$$

The limit above follows for large n because of $\sigma_1 = \sum_{u_i=1}^{n-1} \frac{1}{u_i} \approx \log(n) + \gamma$. Next we examine the additional terms not written down above. Obviously

$$\prod_{l=1}^{k-1} \sigma_l^{\lambda_l} \leq \prod_{l=1}^{k-1} \sigma_1^{\lambda_l} = \sigma_1^{\sum \lambda_l}.$$

Since $\sum_{l=1}^{k-1} \lambda_l \stackrel{!}{=} \sum_{l=1}^{k-1} l\lambda_l = k-1$ is only true for $k = 2$ and this is covered by the case where no additional term exists it can be concluded $\sum \lambda_l \leq k-2$. Since σ_1^{k-2} can be approximated by $(\gamma + \log(n))^{k-2}$ we finally get for all the additional terms

$$\lim_{n\to\infty} \frac{(\gamma + \log n)^{k-2}}{(\gamma + \log n)^{k-1}} = 0.$$

Consider that all this was done for fixed k. Finally (18) is proved since

$$\lim_{n\to\infty} \frac{|s(n,k)|}{(n-1)!\,(\gamma + \log n)^{k-1}} = \frac{1}{(k-1)!} \Rightarrow$$
$$|s(n,k)| \sim \frac{(n-1)!}{(k-1)!} (\gamma + \log n)^{k-1}.$$

With this the corresponding distribution is given directly by

$$P(N = k) = \frac{s(n,k)}{n!} \sim \frac{(\gamma + \log n)^{k-1}}{n\,(k-1)!}.$$

The following figure compares for permutations of length 50 the results generated in section 3.2.1 with those of the local limit distribution. Note that here the local limit approach leads to probabilities that add up to more than 1.

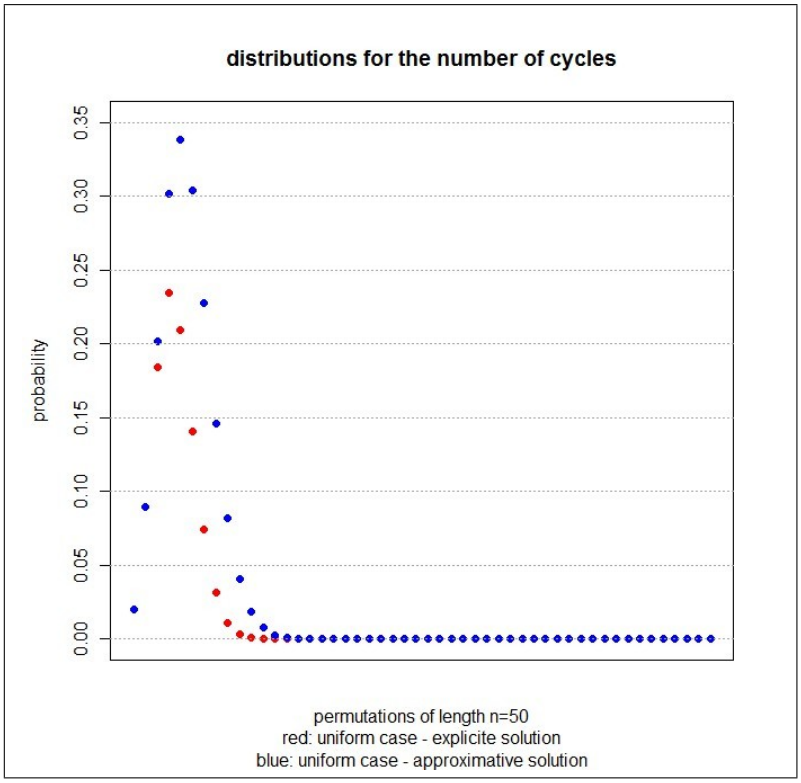


Figure 13: Explicit and local limit distribution for the number of cycles - uniform case

Biased Case
Since the local limit distribution depends on an approximative presentation of the unsigned Stirling numbers of the first kind it follows immediately by what has been derived for the uniform case that

$$P_\theta\left(N=k\right) \sim \frac{\theta^k\left(n-1\right)!\left(\gamma+\log n\right)^{k-1}}{(k-1)!\theta^{(n)}}.$$

The following figure compares for θ-biased permutations of length 50 the results generated in section 3.2.1 with those of the local limit distribution. Here $\theta = 1.5$. Note that again no probability measure is given.

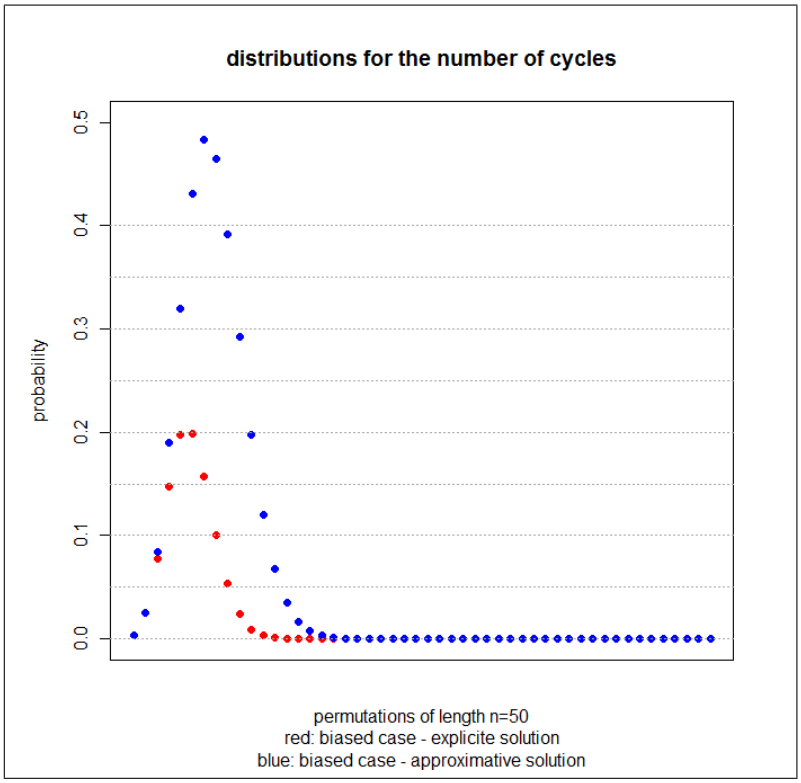


Figure 14: Explicit and local limit distribution for the number of cycles - biased case

After all the local limit method is comfortable to compute, but delivers only weak results. There are other similar local limit approaches in literature, e.g. in [13]. As a conclusion of the results here they can be seen critically.

4.2.3 Limit distribution

Uniform Case
The derivation of the limit distribution of the number of cycles $N = \sum_j Y_j^n$ with Y_j^n standing for the number of cycles of length j of an permutation of length n largely follows the approach that was presented by Shepp and Lloyd (1966) in [22]. First of all we want to connect equation (13) with a conditional distribution of independent Poisson - distributed random variables α_j. This connection also plays an important role in Arratia et al.'s book of Logarithmic Combinatorial Structures (2003) [1] where it is called Conditioning Relation:

$$P\left(Y_1^n = y_1, \ldots, Y_n^n = y_n\right) = P\left(\alpha_1 = a_1, \ldots, \alpha_n = a_n | \sum_{j=1}^{n} j a_j\right). \tag{21}$$

The upper index n determining the length of the permutation is in the following neglected to simplify matters. According to (13) we have to show the following in order to prove (21):

$$P\left(\alpha_1 = a_1, \ldots, \alpha_n = a_n | \sum_{j=1}^{n} j a_j\right) = 1_{\{\sum_{j=1}^{n} j\alpha_j = n\}} \prod_{j=1}^{n} \left(\frac{1}{j}\right)^{y_j} \frac{1}{y_j!}. \tag{22}$$

Let the α's be mutually independent and α_j Poisson distributed with parameter $\frac{z^j}{j}$, where $0 \leq z \leq 1$ i.e.

$$P(\alpha_j = a) = e^{\left(-\frac{z^j}{j}\right)} \frac{\left(\frac{z^j}{j}\right)^a}{a!}, \quad a = 0, 1, \ldots. \tag{23}$$

Since $0 \leq z \leq 1$ and j stands for a positive integer we have $-\frac{z^j}{j} > ln\left(1 - \frac{z^j}{j}\right)$ and can easily estimate

$$e^{-\frac{z^j}{j}} > 1 - \frac{z^j}{j} \Leftrightarrow 1 - e^{-\frac{z^j}{j}} < \frac{z^j}{j} \Rightarrow$$
$$P(\alpha_j \neq 0) = 1 - e^{-\frac{z^j}{j}} < \frac{z^j}{j}, \quad j = 1, 2, \ldots.$$

Since $\sum_j^\infty \frac{z^j}{j} = -\log(1-z)$ it is clear that $\sum_{j=1}^\infty P(\alpha_j \neq 0) < \infty$. Next Lemma of Borel-Cantelli tells us that if the sum of the probabilities of particular incidents is finite the probability that infinitely many of the incidents occur will be equal to 0, i.e. $P(\alpha_j \neq 0 \quad for \quad \#j = \infty) = 0$. What follows is $P\left(\sum_j^\infty \alpha_j < \infty\right) = 1$ and so $\sum_j^\infty j\alpha_j$ is also finite with probability 1. Next we derive the joint distribution of the α's

$$\begin{aligned} P(\alpha_1 = a_1, \alpha_2 = a_2, \ldots) = & \quad \prod_{j=1}^\infty e^{\left(-\frac{z^j}{j}\right)} \frac{\left(\frac{z^j}{j}\right)^{a_j}}{a_j!} \\ = & \quad \underbrace{e^{\left(-\sum_{j=1}^\infty \frac{z^j}{j}\right)}}_{e^{\log(1-z)}} z^{\sum_{j=1}^\infty j a_j} \prod_{j=1}^\infty \frac{\left(\frac{1}{j}\right)^{a_j}}{a_j!} \\ = & \quad (1-z)\, z^{v(a)} \prod_{j=1}^\infty \frac{\left(\frac{1}{j}\right)^{a_j}}{a_j!}, \end{aligned}$$

where $v(\alpha) = \sum_{j=1}^\infty j\alpha_j$. Now we want to determine the probability of this random variable. For that we take the way along the probability generating function

$$G(t) = E\left(t^{v(\alpha)}\right) = \prod_{j=1}^\infty E\left(t^{j\alpha_j}\right)$$

$$\begin{aligned} E\left(t^{j\alpha_j}\right) &= \sum_{k=0}^\infty t^{jk} P(\alpha_j = k) = \sum_{k=0}^\infty t^{jk} \frac{\left(\frac{z^j}{j}\right)^k}{k!} e^{-\left(\frac{z^j}{j}\right)} \\ &= e^{\frac{(tz)^j}{j}} e^{-\left(\frac{z^j}{j}\right)} \\ &= e^{\frac{z^j}{j}\left(t^j - 1\right)} \Rightarrow \\ \prod_{j=1}^\infty E\left(t^{j\alpha_j}\right) &= \prod_{j=1}^\infty e^{\frac{z^j}{j}\left(t^j-1\right)} \\ &= e^{ln\left(\frac{1}{1-zt}\right) - ln\left(\frac{1}{1-z}\right)} \\ &= \frac{1-z}{1-zt}. \end{aligned}$$

Now we are using that $P(v(\alpha)=n)=G^{(n)}(0)\frac{1}{n!}$. Differentiating leads to

$$\frac{\partial^n \frac{1-z}{1-zt}}{(\partial t)^n}=\frac{1\ldots n\,(1-z)\,z^n}{(1-zt)^{n+1}}.$$

Finally putting $t=0$ and dividing by $n!$ delivers $P(v(\alpha)=n)=(1-z)\,z^n$. Now we are ready to determine the conditional distribution mentioned above

$$\begin{aligned}P(\alpha_1=a_1,\alpha_2=a_2,\ldots|v(\alpha)=n)&=\frac{P(v(\alpha)=n,\alpha_1=a_1,\alpha_2=a_2,\ldots)}{P(v(\alpha)=n)}\\&=\frac{\prod_{j=1}^{\infty}\frac{\left(\frac{1}{j}\right)^{a_j}}{a_j!}(1-z)\,z^n}{(1-z)\,z^n}\\&=\prod_{j=1}^{\infty}\frac{\left(\frac{1}{j}\right)^{a_j}}{a_j!}.\end{aligned}$$

So with replacing y_j by a_j we are almost back at the probability of (13) and have shown the Conditioning Relation. Let us put $\alpha_{n+1}(\pi)=\alpha_{n+2}(\pi)=\ldots=0$ for a given permutation π so that we can write

$$\mathcal{L}(Y_1^n,\ldots,Y_n^n)=\mathcal{L}\left((\alpha_1^n,\ldots,\alpha_n^n)\,|\sum_{j=1}^{n}j\alpha_j=n\right),$$

where $\mathcal{L}$ stands for the discrete distribution of the appearing random variables. Now the Conditioning Relation can be used to connect the expectation of $N=\sum_{j=1}^{n}Y_j$ with the one of $N_\alpha=\sum_{j=1}^{\infty}\alpha_j$, i.e.

$$E(N_\alpha)=E(E(N_\alpha|v(\alpha)))= \tag{24}$$

$$\sum_{n=0}^{\infty}P(v(\alpha)=n)\,E(N_\alpha|v(\alpha))=\sum_{n=0}^{\infty}(1-z)\,z^nE(N). \tag{25}$$

In the equation above a conditional expectation $E(X|Y)$ can be written as a $\sigma(Y)$-measurable function of Y. This is the so called 'Faktorisierungslemma'. Compare Lemma 8.3 in [18]. That is why the distribution of the random variable $E(X|Y)$ is determined by the one of Y. Instead of looking directly at the distribution of N we first consider the one of N_α. For that we take the way along the characteristic function using the independence

of the Poisson distributed α_j, i.e.

$$\begin{aligned}
\phi_{N_\alpha}(t) &= E\left(e^{itN_\alpha}\right) = E\left(e^{it\sum_{j=1}^{\infty}\alpha_j}\right) \\
&= E\left(e^{it\alpha_1}e^{it\alpha_2}\ldots\right) = \prod_{j=1}^{\infty} E\left(e^{it\alpha_j}\right) \\
E\left(e^{it\alpha_j}\right) &= \sum_{k=0}^{\infty} e^{itk}\frac{\left(\frac{z^j}{j}\right)^k}{k!}e^{-\left(\frac{z^j}{j}\right)} \\
&= \sum_{k=0}^{\infty}\frac{\left(\frac{e^{it}z^j}{j}\right)^k}{k!}e^{-\left(\frac{z^j}{j}\right)} \\
&= e^{\frac{z^j}{j}\left(e^{it}-1\right)} \Rightarrow \\
E\left(e^{itN_\alpha}\right) &= \prod_{j=1}^{\infty} e^{\frac{z^j}{j}\left(e^{it}-1\right)} \\
&= \left(e^{\sum_{j=1}^{\infty}\frac{z^j}{j}}\right)^{\left(e^{it}-1\right)} \\
&= \left(\frac{1}{1-z}\right)^{e^{it}-1}, \quad -\infty \leq t \leq \infty.
\end{aligned}$$

According to (25) $E\left(e^{itN}\right)$ represents in $\frac{E\left(e^{itN_\alpha}\right)}{1-z} = \left(\frac{1}{1-z}\right)^{e^{it}} = \sum_{n=0}^{\infty} E\left(e^{itN}\right) z^n$ a binomial coefficient. In fact there is an analytical form of $E\left(e^{itN}\right)$. In order to derivate this e^{it} is replaced by a and the so called hyper geometric function is used, i.e.

$$\begin{aligned}
\sum_{n=0}^{\infty}\frac{\Gamma(n+a)}{\Gamma(a)\,n!}z^n = \\
\frac{\Gamma(a)}{\Gamma(a)} + \frac{\Gamma(1+a)}{\Gamma(a)}z + \frac{\Gamma(2+a)}{2\Gamma(a)}z^2 + \frac{\Gamma(3+a)}{6\Gamma(a)} + \ldots = \\
1 + \frac{a\Gamma(a)}{\Gamma(a)}z + \frac{a(1+a)\Gamma(a)}{2\Gamma(a)}z^2 + \frac{a(1+a)(2+a)\Gamma(a)}{6\Gamma(a)} + \ldots = \\
1 - a(-z) + \frac{a(1+a)(-z)^2}{2} - \frac{a(1+a)(2+a)(-z)^3}{6} + \ldots = \\
1 - a(-z) - a(-a-1)\frac{1}{2}(-z)^2 - a(-a-1)(-a-2)\frac{1}{6}(-z)^3 = \\
\sum_{n=0}^{\infty}\binom{-a}{n}(-z)^n = \\
(1-z)^{-a},
\end{aligned}$$

where $\binom{a}{n} = \frac{a(a-1)(a-2)\ldots(a-k+1)}{n!}$ for $k > 0$ stands for the generalized binomial coefficient and the last step corresponds to the generalized form of the binomial theorem. Consider that $\binom{a}{n} = 1$ through definition. More details for Γ-function are given in appendix A. Returning from a to e^{it} then delivers $\sum_{n=0}^{\infty}\frac{\Gamma(n+a)}{\Gamma(a)n!}z^n = \left(\frac{1}{1-z}\right)^{e^{it}}$. That is why

$$\phi_N(t) = E\left(e^{itN}\right) = \frac{\Gamma\left(n+e^{it}\right)}{\Gamma\left(e^{it}\right)n!}. \tag{26}$$

After determining $\phi_N(t)$ the moments of N follow by $E\left(N^k\right) = \frac{\phi_N^k(0)}{i^k}$. We start with the expectation

$$\begin{aligned}
\frac{\partial \frac{\Gamma\left(n+e^{it}\right)}{\Gamma\left(e^{it}\right)}}{\partial t} = \\
\frac{D(\Gamma)\left(n+e^{it}\right) ie^{it}\Gamma\left(e^{it}\right) - \Gamma\left(n+e^{it}\right) D(\Gamma)\left(e^{it}\right)}{\Gamma\left(e^{it}\right)^2 n!} \Rightarrow \\
E(N) = \frac{D(\Gamma)(n+1)}{\Gamma(1)\, n!} - \frac{\Gamma(n+1)\, D(\Gamma)(1)}{\Gamma(1)^2\, n!} = \\
\frac{\Psi(n+1)\,\Gamma(n+1) + \Gamma(n+1)\,\gamma}{n!} = \\
\Psi(n+1) + \gamma = \sum_{k=1}^{n} \frac{1}{k} = \\
\log(n) + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \ldots = \\
\log(n) + O(1),
\end{aligned}$$

where big O - notation is used to describe the error, $E(N) - log(n)$, being smaller in absolute value than some real constant times 1, when n is near enough to infinity. After all the residual can be neglected in the limit. In this calculation several properties of Γ- and the Ψ-function (gamma and digamma function) were used, which are mentioned in the appendix A. Next we turn to the variance. Similar calculation as done above then delivers

$$E\left(N^2\right) = \frac{1}{i^2}\phi_N^2(0) = \Psi(1, n+1) + \Psi(n+1)^2 + \Psi(n+1)\,\gamma + \gamma^2 - \frac{\Pi^2}{6} + \gamma,$$

where $\Psi(1, n+1) = \sum_{k=0}^{\infty} \frac{1}{n+1+k}$ stands for trigamma function. Appendix A delivers more information for that. Since $E(N)^2 = \Psi(n+1)^2 + 2\Psi(n+1)\,\gamma + \gamma^2$ is already uncovered we finally obtain

$$\begin{aligned}
Var(N) &= E\left(N^2\right) - E^2(N) \\
&= \sum_{k=1}^{n} \frac{k-1}{k^2} \\
&= \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k^2} \\
&= \log(n) + \gamma - \ldots \\
&= \log(n) + O(1).
\end{aligned}$$

Now we are ready to deduce the limiting distribution. Consider that it is not directly possible to obtain standard normal distribution for the normalized random variable $\frac{N-\log(n)}{\sqrt{\log(n)}}$ by central limit distribution since the assumption that the Y_j are independent and identically distributed does not hold. Instead of this we look back at the Feller coupling used in the section about the explicit distribution in biased permutations. There we constructed permutations with their canonical cycle notation by introducing independent random variables

$$D_i = \begin{cases} 1 & \text{if actual cycle is closed} \\ j & \text{if actual cycle is extended} \end{cases}, \quad 1 \le i \le n, \quad j \ne 1.$$

It is easy to turn the D_is into independent Bernoulli distributed random variables

$$\xi = \mathbb{1}_{\{cycle-closing\}}.$$

Obviously the number of cycles N in such a permutation is identical to $\sum \xi_i$. Because of the given independence of the ξ_is and the existing variance N will be approximatively normal distributed if the Lindeberg's condition holds, i.e. $\forall \epsilon > 0 \; \forall i$

$$\lim_{n\to\infty} \frac{1}{\sum \sigma_{\xi_i}^2} \sum_{i=1}^{n} \int\limits_{\{|\xi_i - E(\xi_i)| > \epsilon \sigma_N\}} (\xi_i - E(\xi_i))^2 \, dP = 0. \tag{27}$$

Remember that there was only one possibility for the cycle to be closed among several to be extended. The total number of possibilities was depending on i. For D_1 we had a n-way choice, for D_2 we had a $n-1$-way choice and so on. This leads directly to $P(\xi_i = 1) = \frac{1}{n-i+1}$ and $E(\xi_i) = \frac{1}{n-i+1}$. Of course the expectation is obtained directly by

$$E(N) = \sum_{i=1}^{n} E(\xi_i) = \sum_{i=1}^{n} \frac{1}{n-i+1} = \sum_{i=1}^{n} \frac{1}{i} \approx \log n,$$

but parts of the derivation above are needed at other places. Furthermore $\sum \sigma_{\xi_i}^2 = \sigma_{\sum \xi_i}^2$ because of independence. So (27) can be transferred to

$$\lim_{n\to\infty} \frac{1}{\log(n)} \sum_{i=1}^{n} \sum_{a_i : a_i - \mu_i > \epsilon \sigma_N} \left(a_i - \frac{1}{n-i+1} \right)^2 P(\xi_i = a_i) \tag{28}$$

$$\leq \lim_{n\to\infty} \frac{1}{\log(n)} \sum_{i=1}^{n} \left(1 - \frac{1}{n-i+1} \right)^2 + \left(-\frac{1}{n-i+1} \right)^2 = 0. \tag{29}$$

Analysis tells us that the last equation holds so that Lindeberg-Feller central limit theorem can be applied, i.e.

$$\lim_{n\to\infty} P\left(\frac{N - \log(n)}{\sqrt{log(n)}} \leq x \right) = \int_{-\infty}^{x} \frac{1}{2\pi} e^{\left(\frac{t^2}{2}\right)} dt.$$

In order to get an idea how well this convergence is working we compare for permutations of length 50 the results generated in section (3) with those of a normal distributed random variable $X \sim N\left(\log 50, \sqrt{\log 50}\right)$.

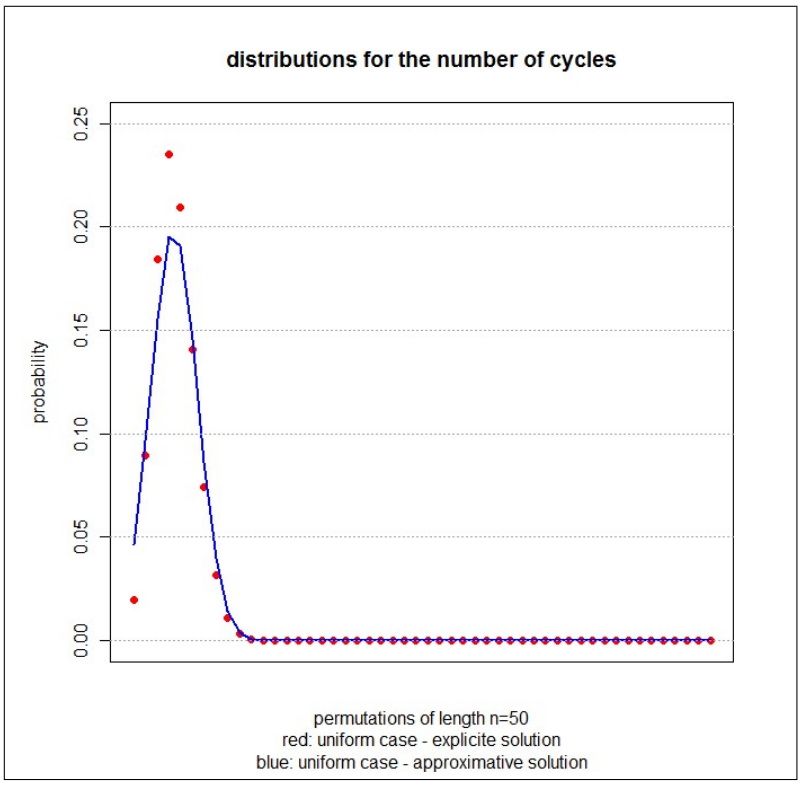


Figure 15: Explicit and limit distribution for the number of cycles - uniform case

Biased Case

Consider (17), where $P(N_\theta)$ was derived. With this the probability generating function follows immediately, i.e.

$$\begin{aligned} G(t) = E\left(t^{N_\theta}\right) &= \sum_{k=1}^{n} t^k P(N_\theta = k) \\ &= \sum_{k=1}^{n} t^k \frac{\theta^k |s(n,k)|}{\theta} = \frac{(t\theta)^{(n)}}{\theta^{(n)}} \\ &= \frac{\Gamma(n+\theta t)\,\Gamma(\theta)}{\Gamma(\theta t)\,\Gamma(n+\theta)}. \end{aligned}$$

Note that the unsigned Stirling number of the first kind turns the falling factorial in (15) into the rising one. For the computation of the expectation $E(N_\theta) = \lim_{t\to 1} G'(t)$ can be used, i.e.

$$G'(t) = \frac{\Psi(n+t\theta)\,\Gamma(n+t\theta)\,\theta\Gamma(\theta)}{\Gamma(\theta+n)\,\Gamma(t\theta)} - \frac{\Gamma(n+t\theta)\,\Gamma(\theta)\,\Psi(t\theta)\,\theta}{\Gamma(\theta+n)\,\Gamma(t\theta)} \Rightarrow$$

$$\lim_{t\to 1} G'(t) = \theta\left(\Psi(\theta+n) - \Psi(\theta)\right) = \sum_{j=1}^{n} \frac{\theta}{\theta+j-1} \approx \theta \log n.$$

A similar calculation as done above delivers

$$Var(N_\theta) = \lim_{\theta\to 1} G''(t) + \lim_{\theta\to 1} G'(t) - \left(\lim_{\theta\to 1} G'(t)\right)^2 \Leftrightarrow$$

$$Var(N_\theta) = \sum_{j=1}^{n} \frac{\theta(j-1)}{(\theta+j-1)^2} \approx \theta \log n.$$

Concerning section 3.1 probabilities for the independent Bernoulli random variables ξ_i^θ are given by

$$P\left(\xi_i^\theta = 1\right) = \frac{\theta}{n+\theta-i+1} \qquad P\left(\xi_i^\theta = 0\right) = \frac{n-i+1}{n+\theta-i+1}.$$

This does not matter for the validity of (29). So Lindeberg's condition holds again and so

$$\lim_{n\to\infty} P\left(\frac{N_\theta - \theta\log(n)}{\sqrt{\theta\log(n)}} \leq x\right) = \int_{-\infty}^{x} \frac{1}{2\pi} e^{\left(\frac{t^2}{2}\right)} dt.$$

The following figure compares for θ-biased permutations of length 50 the results generated in section (3) with those of a normal distributed random variable $X \sim N\left(1.5\log 50, \sqrt{1.5\log 50}\right)$. Here again $\theta = 1.5$.

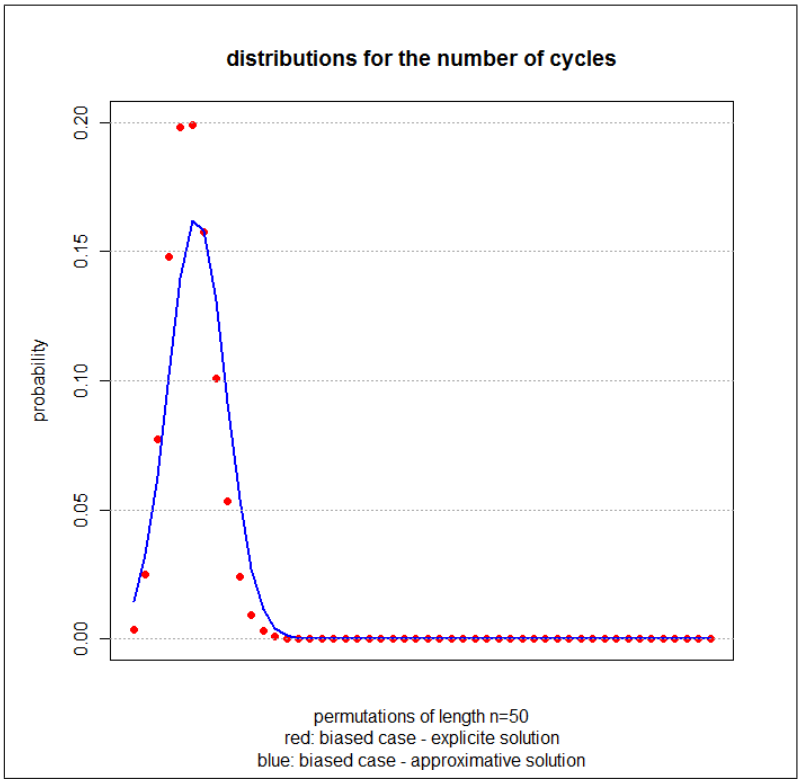


Figure 16: Explicit and limit distribution for the number of cycles - biased case

4.3 The distribution of the rth longest and shortest cycle

4.3.1 Uniform Case

First of all it must be mentioned that in the appropriate literature no approach was found to present the distribution for the length of the longest cycle without approximation. Our first target in this section is to comprehend the following result given by Shepp and Lloyd [22]:

$$\frac{1}{n}L_r^n \xrightarrow{d} L_r, \tag{30}$$

where L_r^n stands for the length of the r-longest cycle and L_r is a distribution that is determined by the Dickman function. It is sometimes called Dickman distribution. We get to know later how it is build up explicitly.

As a first step a repetition of (21) gives a hint that it is useful to consider independent Poisson distributed random variables again. So let there be a Poisson process on $T = -\infty < t < \infty$. Next T is interrupted by the time points $t_j(z) = \sum_{k=1}^{j-1} \frac{z^k}{k}$ for $j = 2, 3 \ldots$. Defining $t_1(z) = 0$ this leads to time points $t_1(z) = 0, t_2(z) = z, t_3(z) = z + \frac{z^2}{2}, \ldots, \sum_{k=1}^{\infty} \frac{z^k}{k} = \log\left(\frac{1}{1-z}\right)$, where the j-th interval $\{t : t_j(z) \leq t < t_{j+1}(z)\}$ for $j = 1, 2, \ldots$ has length $\frac{z^j}{j}$. Simultaneously a step function $\lambda_z(t)$ with $\lambda_z(t) = j$ for t in the j-th interval is given. In this setting no smoother time units are considered, i.e. the expected number of jumps in the j-th interval is directly given by the length of the interval. In other words the jumps α_j of the j-th interval are Poisson-distributed with probability

$$P(\alpha_j = a) = e^{-\frac{z^j}{j}} \frac{\left(\frac{z^j}{j}\right)^a}{a!}.$$

In the definition of the Poisson process the α_j are independent. So we are back at (23). Consider a particular cycle structure $y = (y_1, \ldots, y_n)$ of a permutation with length n. Then each cycle is represented by a jump and the number of cycles with length j correspond to the number of jumps in the time interval j. If for each jump in the time interval j the value of λ_j is allocated, we will have $\sum jy_j = \sum \lambda_j = n$.

Meanwhile we consider the rth shortest cycle. The length of the rth shortest cycle in permutation of cycle class α shall be given by $S_r = S_r(\alpha)$. It is clear that the rth shortest cycle can only exist, if the total cycle class contains at least r cycles, i.e. $S_r = 0 \Leftrightarrow \sum \alpha_j < r$. Let T_1 be the time up to the first jump, T_2 the time between the first and the second jump, T_3 the time between the second and the third jump and so on. In fact these are stop times. As it is known in a Poisson process these times are independently exponential distributed with the same parameter used for the Poisson distributed increments, i.e. here the length of the interval.

Now we are interested in a sum of these times since $P(S_r = j) = P(\sum_{k=1}^{r} T_k \in [t_j, t_{j+1}[)$. Obviously the rth longest cycle has length j only if the total time up to the rth jump is within the jth interval. Consider that in this setting everything is starting in time point 0. Changing into a continuous setting we must ask which density function has to be used.

Proposition
The density for the sum of the independently exponential distributed random variables $T_1, \ldots, T_r$ is given by

$$f(t) = \begin{cases} \frac{t^{r-1}}{\Gamma(r)} e^{-t} & : \quad t \geq 0 \\ 0 & : \quad t < 0 \end{cases}$$

The corresponding distribution is called Erlang distribution, when like here parameter r is an integer. Otherwise it is the Γ-distribution.

Proof
Via induction on r. Induction starts with $r = 1$:

$$P(T_1 \leq x) = \int_0^x e^{-t} dt = \int_0^x \frac{t^{1-1}}{(1-1)!} e^{-t} dt.$$

For $r \to (r+1)$ convolution, for example given in [15], is used:

$$\begin{aligned} F(t) = P\left(\sum_{k=1}^{r} T_k + T_{r+1} < t\right) = \\ \int_{]0,\infty[} f_{T_{r+1}} * \ldots * f_{T_1} d(u, s^r) = \\ \int_0^\infty \int_0^{t-s} f_{T_{r+1}} du \quad f_{\sum_{k=1}^r T_k} ds = \\ \int_0^t \left(1 - e^{t-s}\right) \frac{s^{r-1}}{(r-1)!} e^{-s} ds = \\ 1 - e^{-t} \sum_{i=0}^{r} \frac{t^i}{i!}. \end{aligned}$$

Then the proof is completed with

$$f(t) = \frac{\partial F(t)}{\partial t} = \frac{e^{-t} t^r}{r!}.$$

Now we are ready for a first determination of the probability of S_r,i.e.

$$P(S_r = j) = \int_{t_j(z)}^{t_{j+1}(z)} \frac{t^{r-1}}{(r-1)!} e^{-t} dt \quad \text{for} \quad j = 1, 2, \ldots.$$

Substituting t by $t_\infty(z) - t$ delivers an analogon for the length of the rth longest cycle,i.e.

$$P(S_r = j) = \int_{t_j(z)}^{t_{j+1}(z)} \frac{(t_\infty(z) - t)^{r-1}}{(r-1)!} e^{-(t_\infty(z)-t)} dt \quad \text{for} \quad j = 1, 2, \ldots. \tag{31}$$

As a next step consequences for $t_j(z)$ are analyzed, when $z = e^{-s}$ with $0 < s < \infty$ is chosen. First of all we have

$$t_\infty\left(e^{-s}\right) - t_j\left(e^{-s}\right) = \sum_{k=j}^{\infty} \frac{e^{-ks}}{k}.$$

Consider that ks is positive and given $ks < y < (k+1)s$ it can be concluded

$$\frac{e^{-ks}}{ks} > \frac{e^{-y}}{y} > \frac{e^{-(k+1)s}}{(k+1)\,s} \Rightarrow$$

$$\int_{ks}^{(k+1)s} \frac{e^{-ks}}{ks}dy > \int_{ks}^{(k+1)s} \frac{e^{-y}}{y}dy > \int_{ks}^{(k+1)s} \frac{e^{-ks}}{(k+1)\,s}dy \Leftrightarrow$$

$$\frac{e^{-ks}}{k} > \int_{ks}^{(k+1)s} \frac{e^{-y}}{y}dy > \frac{e^{-(k+1)s}}{k+1} \Rightarrow$$

$$\sum_{k=j}^{\infty} \int_{(k-1)s}^{ks} \frac{e^{-y}}{y}dy > \sum_{k=j}^{\infty} \frac{e^{-ks}}{k} > \sum_{k=j}^{\infty} \int_{ks}^{(k+1)s} \frac{e^{-y}}{y}dy \Leftrightarrow$$

$$Exp\left((j-1)\,s\right) > \sum_{k=j}^{\infty} \frac{e^{-ks}}{k} > Exp\left(js\right),$$

where $Exp\left(x\right)$ stands for the exponential integral, i.e. $Exp\left(x\right) = \int_{x}^{\infty} \frac{e^{-y}}{y}dy$ for $0 < x < \infty$. Consider now $f : x \to \int_{x}^{\infty} \frac{e^{-y}}{y}dy$.

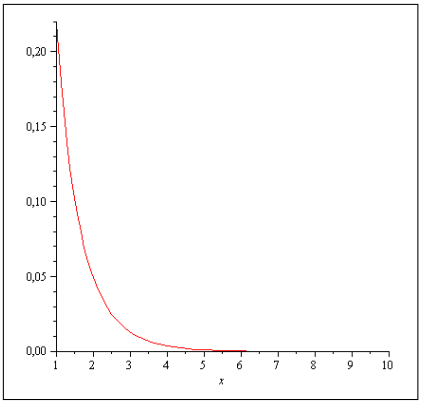


Figure 17: Mapping on exponential integral for $x = 1, \ldots, 10$

Obviously function ϕ is invertible and its inverse ϕ^{-1} also continuous (homeomorphism). What follows is

$$\lim_{\sum_{k=j}^{\infty} \frac{e^{-ks}}{k} \to \int_{x_j(s)}^{\infty} \frac{e^{-y}}{y}dy} \phi^{-1}\left(\sum_{k=j}^{\infty} \frac{e^{-ks}}{k}\right) = x_j\left(s\right).$$

In other words equation

$$\sum_{j}^{\infty} \frac{e^{-ks}}{k} = E\left(x_j\left(s\right)\right)$$

has an unique solution $x_j\left(s\right)$ for each $1 < j < \infty$ and each $0 < s < \infty$. It is also clear that

$$(j-1)\,s < x_j\left(s\right) < js < x_{j+1}\left(s\right) < (j+1)\,s, \tag{32}$$

because of the lower limit in the integrals. Connecting the last result with (31) the mth moment of the length of the rth longest cycle is given by

$$E\left((L_r^n)^m\right) = \sum_{j=1}^{\infty} j^m \int_{x_{j-1}(s)}^{x_j(s)} \frac{(Exp\,(x))^{r-1}}{(r-1)!} e^{-Exp(x)} \frac{e^{-x}}{x} dx.$$

Here integration by substitution with $\left(t_\infty\left(e^{-s}\right) - t\right) = Exp\,(x)$ and

$$\phi^{-1}\left(t_j\left(e^{-s}\right)\right) = \phi^{-1}\left(t_\infty\left(e^{-s}\right) - Exp\left(x_j\left(s\right)\right)\right) = x_j\left(s\right)$$

was applied. Upper and lower bounds for $s^m E\left((L_r^n)^m\right)$ can easily be given by using (32), i.e.

$$\text{upper bound:} \qquad \sum_{j=1}^{\infty} \left(x_{j+1}\left(s\right)\right)^m \int_{x_j(s)}^{x_{j+1}(s)} d\mu\left(x\right)$$

$$\text{lower bound:} \qquad \sum_{j=1}^{\infty} \left(x_j\left(s\right)\right)^m \int_{x_j(s)}^{x_{j+1}(s)} d\mu\left(x\right),$$

where the following is used

$$d\mu = \frac{\left(E\left(x\right)\right)^{r-1}}{(r-1)!} e^{-E(x)} \frac{e^{-x}}{x} dx.$$

Now consider $\int_{x_j(s)}^{x_{j+1}(s)} d\mu\left(x\right)$ in the sense of $\left(x_{j+1}\left(s\right) - x_j\left(s\right)\right)$ as acuteness so that the upper and lower bound mentioned above are the one of the integral

$$\int_{x_1(s)}^{\infty} x^m d\mu\left(x\right).$$

So the integral will be reached, if the mesh in upper and lower sum vanishes. Note that because of (32) this is the case with $s \to 0$. Putting $s = (1 - \hat{z})$ this is equivalent with $\hat{z} \to 1$. So it can be concluded

$$\lim_{\hat{z}\to 1} \frac{(1-\hat{z})^m}{m!} E\left((L_r)^m\right) = \int_0^{\infty} \frac{x^{m-1}}{m!} \frac{Exp\left(x\right)^{r-1}}{(r-1)!} e^{(-Exp(x)-x)} dx.$$

Up to now we considered independent Poisson distributed random variables. Although the conditioning relation is given in (21) these are not the true cycle variables. So the L_r mentioned above is also a quite similar, but artificial one. According to Shepp and Lloyd [22] Karamata-Hardy-Littlewood Tauberian theorem delivers for the true L_r

$$E\left((L_r)^m\right) \sim \int_0^{\infty} \frac{x^{m-1}}{m!} \frac{Exp\left(x\right)^{r-1}}{(r-1)!} e^{(-Exp(x)-x)} dx. \tag{33}$$

We do not follow the argumentation of Shepp and Lloyd in [22] any longer. Instead of this we take another way, but (33) should be kept in mind. In one of the next sections concerning the biased case we will repeat a result for the joint distribution of the length of the r longest **cycles** given in [1], i.e.

$$\frac{1}{n}\left(L_1^n, \ldots, L_r^n\right) \xrightarrow{d} \left(L_1, \ldots, L_r\right).$$

Thereby the random vector $(L_1, \ldots, L_r)$ has density

$$f_\theta^r(x_1, \ldots, x_r) = \frac{\theta^r x_r^{\theta-1}}{x_1 \ldots x_r} p_\theta\left(\frac{1 - x_1 - \ldots - x_r}{x_r}\right), \tag{34}$$

with $p_\theta(x) =$

$$x^{\theta-1}\left(1 + \sum_{k=1}^{\infty} \frac{(-\theta)^k}{k!} \int \ldots \int_{\substack{xy_1 > 1 \ldots xy_k > 1 \\ y_1 + \ldots + y_k < 1}} \left(1 - \sum_{j=1}^{k} y_j\right)^{\theta-1} \frac{dy_1 \ldots dy_k}{y_1 \ldots y_k}\right).$$

The distribution of the random vector is the Poisson-Dirichlet one, which will be introduced in detail in the next section starting from a Poisson point process. There (L_r) is given by

$$L_r = \frac{X_{[r]}}{S_\theta},$$

with $S_\theta = X_{[1]} + X_{[2]} + \ldots$ being independent from L_r and $\Gamma_{1,\theta}$ distributed. (In detail put $S_\theta = M_\theta$ and independence follows from (40). The density of $X_{[r]}$ is not self-explanatory. It has to be derived from the corresponding Poisson point process in section 4.3.2). See also [11]. $X_{[r]}$ has density

$$f_{X_{[r]}}(x) = \frac{\theta e^{-x}}{x} \frac{(\theta Exp(x))^{r-1}}{(r-1)!} e^{-\theta Exp(x)}.$$

So the mth moment of the single variable can be computed by

$$E\left(X_{[r]}^m\right) = E\left(L_r^m S_\theta^m\right) = E\left(L_r^m\right) E\left(S_\theta^m\right) \Leftrightarrow$$
$$\int_0^\infty x^m \frac{\theta e^{-x}}{x} \frac{(\theta Exp(x))^{r-1}}{(r-1)!} e^{-\theta Exp(x)} dx = E\left(L_r^m\right) \theta^r \frac{\Gamma(1+m)}{\Gamma(1)}.$$

Putting $\theta = 1$ leads to

$$\int_0^\infty \frac{x^m}{m!} \frac{e^{-x}}{x} \frac{(Exp(x))^{r-1}}{(r-1)!} e^{-Exp(x)} dx = E\left(L_r^m\right),$$

which is exactly the same than (33). Assuming that the distribution is sufficiently described by its moments (method of moments) we take the marginal distribution of (34) as the one of L_r. Putting in $\theta = 1$ and emphasizing the first element there leads to a density

$$f_1^1 = \frac{1}{x} p_1\left(\frac{1-x}{x}\right) \tag{35}$$

with

$$p_1(x) = \left(1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int \ldots \int_{\substack{xy_1 > 1 \ldots xy_k > 1 \\ y_1 + \ldots + y_k < 1}} \frac{dy_1 \ldots dy_k}{y_1 \ldots y_k}\right).$$

This is the so called Dickman distribution determined by the Dickman function p_1. Note that the taken approach is appropriate as long as (34) is not derived from (33). See section 4.3.3. For further details of Dickman function see section 4.3.4 and appendix E. The following figure compares for uniform permutations of length 50 the results generated in section 3.2.2 with those of (35).

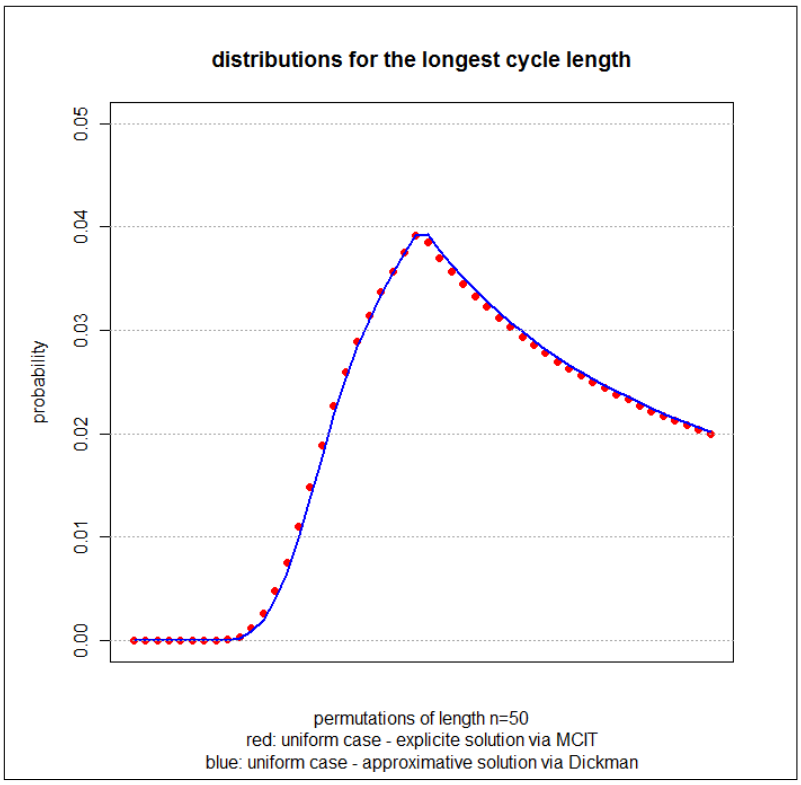


Figure 18: Explicit and limit distribution for the longest cycle length - uniform case

The asymptotic result is really good. A great advantage of this is that the result is obtained without knowing the corresponding cycle structures. It follows directly from computation of the Dickman function. For that see appendix E.

4.3.2 Poisson-Dirichlet distribution

Before one is able to introduce the Poisson-Dirichlet distribution preliminaries are necessary given for example in [15]. These are recalled here to complete the framework.

Definition
A random measure on $(E, \mathcal{E})$ is a family of random variables $X = \{X(A), A \in \mathcal{E}\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ with $P(X \in \mathcal{M}(E))=1$, i.e. X is almost surely a Radon-measure on $(E, \mathcal{E})$. Thereby as usually E stands for the basic set and $\mathcal{E}$ for an appropriate σ-algebra. However it is not enough to take the standard Borel sets under all circumstances. There must be constraints such as boundary when using a locally compact polish space for E. For further details in the corresponding measure theory see [15]. We assume the intensity measure to be σ-finite.

The intensity measure $\mu \in \mathcal{M}(E)$ of X is determined $\forall A \in \mathcal{E}$ by the function

$$E(X): \quad \mathcal{E} \longrightarrow [0, \infty] \qquad A \longmapsto E(X(A)) = \mu(A).$$

A random measure with intensity measure μ is called a Poisson point process if

$$\forall A \in \mathcal{E}: \qquad P(X(A) = k) = \frac{\mu(A)^k}{k!} e^{-\mu(A)} \quad \text{for} \quad k = 0, 1, \ldots.$$

and if for mutually disjoint $A_1, \ldots, A_n$ the random variables $X(A_1), \ldots, X(A_n)$ are stochastically independent.

Definition
The Laplace transform of a random measure X is defined by

$$\mathcal{L}_X(f) = E\left(e^{-\int f dX}\right),$$

where f is a $\mathcal{E}$-measurable mapping.

Although it is not shown here explicitly it is crucial to consider that the distribution of a random measure is sufficiently determined by the values of its Laplace transform mentioned above. For the details see [15] page 511. Another basic result not explicitly shown here is that for each $\mu \in \mathcal{M}(E)$ a Poisson point process exists such that μ is the intensity measure of the Poisson point process.

Proposition
Let $\mu \in \mathcal{M}(E)$ and X a Poisson point process with intensity measure μ. Then the corresponding Laplace transform is given by

$$\mathcal{L}(f) = e^{\int \left(e^{-f(x)}-1\right)\mu(dx)}.$$

Proof
The proof is given for elementary functions, i.e. $f = \sum_{l=1}^{n} \alpha_l \mathbb{1}_{A_l}$ with complex numbers $\alpha_1, \ldots, \alpha_n$ and mutually disjoint sets $A_1, \ldots, A_n \in \mathcal{E}$. It can be completed by algebraic induction. Using the definition of a Laplace transform for a random measure delivers

$$\begin{aligned} E\left(e^{-\sum_{l=1}^{n} \alpha_l X(A_l)}\right) &= \prod_{l=1}^{n} E\left(e^{-\alpha_l X(A_l)}\right) \\ &= \prod_{l=1}^{n} e^{\mu(A_l)\left(e^{-\alpha_l}-1\right)} \\ &= e^{\sum_{l=1}^{n} \mu(A_l)\left(e^{-\alpha_l}-1\right)} \\ &= e^{\int \mu(dx)\left(e^{-f(x)}-1\right)}. \end{aligned}$$

The second step uses the known form of the characteristic function for the Poisson distribution and the last step the integral definition.

Proposition

1. Let E and F be locally compact and polish spaces. $\phi : E \to F$ shall be measurable. Let $\mu \in \mathcal{M}(E)$ with $\mu \circ \phi^{-1} \in \mathcal{M}(F)$. Let X be a Poisson point process on E with intensity measure μ. Then $X \circ \phi^{-1}$ identifies a Poisson point process on F with intensity measure $\mu \circ \phi^{-1}$.
2. Let $\mu \in \mathcal{M}((0,\infty))$ and X be a Poisson point process on $(0,\infty)$ with intensity measure μ. Put $M := \int xX(dx)$. Is $\int (1 \wedge x)\, \mu(dx) < \infty$ valid, then M is a divisible non negative random variable with Lévy measure μ. Thereby $\wedge$ stands for the minimum.

Proof

1. For a measurable f on F the Laplace transform of the Poisson point process can directly be determined by transformation formula, i.e.

$$\begin{aligned}\mathcal{L}_{X\circ\phi^{-1}}(f) &= \mathcal{L}_X(f\circ\phi)\\ &= e^{\int\left(e^{-f(\phi(x))}-1\right)\mu(dx)}\\ &= e^{\int\left(e^{-f(y)}-1\right)\left(\mu\circ\phi^{-1}\right)(dx)}.\end{aligned}$$

   The proof is completed by the fact that the Laplace transform is characteristic for a Poisson point process.

2. First a definition must be given: A measure $\mu\in\mathcal{M}((0,\infty))$ is called divisible if for each $n\in\mathbb{N}$ a measure $\mu_n\in\mathcal{M}((0,\infty))$ exists such that convolution done $(n-1)$-times delivers the original measure itself, i.e.

$$\mu=\mu_n*\ldots*\mu_n.$$

The definition can be transferred to random variables. Then the statement above follows by famous **Lévy Khinchin formula** and putting $\alpha=0$:

Let $\nu\in\mathcal{M}([0,\infty))$ and let u be the $-\log$-Laplace transform of ν. Then ν is divisible if and only if an unique $\alpha\geq 0$ and an unique σ-finite measure $\mu\in\mathcal{M}((0,\infty))$ exist with $\int(1\wedge x)\,\nu(dx)<\infty$ such that

$$u(t)=\alpha t+\int\left(1-e^{-tx}\right)\mu(dx)\quad\text{for}\quad t\geq 0.$$

The measure μ is then called the Lévy measure of ν. We pass on a proof of Lévy Khinchin formula for example given in [15] page 321.

Remark
Let X be a Poisson point process on $(0,\infty)\times[0,\infty)$ with intensity measure $\mu\otimes\lambda$, where λ stands for the Lebesgue measure. Put $M_0=0$ and

$$M_t:=\int\limits_{(0,\infty)\times(0,t\,]}xX(d(x,s)). \tag{36}$$

According to the proposition above $X(\cdot\times(s,t\,])$ is also a Poisson point process with intensity measure $(t-s)\,\mu$. Furthermore

$$M_t-M_s=\int\limits_{(0,\infty)\times(s,t\,]}xX(d(x,s)),$$

then has Lévy measure $(t-s)\,\mu$. This construction (of a subordinator) will be used to introduce the Poisson-Dirichlet distribution. Now we are ready with our review on Poisson point process and turn to this target.

Since the Poisson-Dirichlet distribution plays an important role in section 4.3 it deserves an extensive introduction. The distribution was originally introduced by Kingman, e.g.

[14]. The introduction here follows the one given in [15]. As a first brick consider the Dirichlet distribution $Dir_{\theta_1,\dots,\theta_n}$ defined on the $n-1$ dimensional simplex, i.e.

$$\Delta_n := \{(x_1,\dots,x_n) \in [0,1]^n : x_n = 1 - \sum_{i=1}^{n-1} x_i\}. \tag{37}$$

The probability density function is given by

$$f_{\theta_1,\dots,\theta_n}(x_1,\dots,x_n) = \frac{\Gamma\left(\sum_{i=1}^n \theta_i\right)}{\prod_{i=1}^n \Gamma(\theta_i)} x_1^{\theta_1-1} \dots x_n^{\theta_n-1}.$$

Note that (37) ensures that a probability measure is given, i.e.

$$\int\limits_{[0,1]^{n-1}} \frac{\Gamma\left(\sum_{i=1}^n \theta_i\right)}{\prod_{i=1}^n \Gamma(\theta_i)} x_1^{\theta_1-1} \dots x_{n-1}^{\theta_{n-1}-1} \left(1 - \sum_{i=1}^{n-1} x_i\right)^{\theta_n-1} dx_1 \dots dx_{n-1} = 1. \tag{38}$$

The next brick represents Moran-Gamma-Subordinator. This is a stochastic process $(M_t)_{t\geq 0}$ with from the right continuous and monotonically increasing paths. Its incrementals are independent, stationary and Γ-distributed, i.e.

$$M_t - M_s \sim \Gamma_{1,t-s} \qquad \text{for} \qquad t > s \geq 0.$$

The following proposition explains the distribution of a compound random vector.

Proposition
Suppose $X \sim Dir_{\theta_1,\dots,\theta_n}$ and $Z \sim \Gamma_{1,\sum_{i=1}^n \theta_i}$ are independent random variables. Then the random variables $S_i := ZX_i$ for $i = 1,\dots,n$ and $n \in \mathbb{N}$ are independent and also Gamma distributed with $S_i \sim \Gamma_{1,\theta_i}$.

Proof
Put $s := \sum_{i=1}^n s_i$ and $\Delta_n^1 := \{x_1,\dots,x_{n-1} : x_i > 0 \quad \text{and} \quad \sum_{i=1}^{n-1} x_i < 1\}$. Considering the independence the random vector $(X_1,\dots,X_{n-1},Z)$ with $x \in \Delta_n^1$ and $z \in (0,\infty)$ has density f equal to

$$\left(\frac{\Gamma\left(\sum_{i=1}^n \theta_i\right)}{\prod_{i=1}^n \Gamma(\theta_i)} \prod_{i=1}^n x_i^{\theta_i-1}\right)\left(\frac{1}{\Gamma\left(\sum_{i=1}^n \theta_i\right)} z^{\sum_{i=1}^n \theta_i - 1} e^{-z}\right) = \prod_{i=1}^n \frac{x_i^{\theta_i-1}}{\Gamma(\theta_i)} z^{\sum_{i=1}^n \theta_i - 1} e^{-z}. \tag{39}$$

The mapping $F : \Delta_{n-1}^1 \times (0,\infty) \longrightarrow (0,\infty)^n$, $(x_1,\dots,x_{n-1},z) \longmapsto (zx_1,\dots,zx_n)$ is easily invertible. Putting $z = s$, which makes sense as it is mentioned below, leads to

$$F^{-1} : (s_1,\dots,s_n) \longmapsto \left(\frac{s_1}{s},\dots,\frac{s_{n-1}}{s},s\right).$$

In order to determine the density g of $(S_1,\dots,S_n)$ transformation formula for densities can be used, i.e.

$$g(s_1,\dots,s_n) = \frac{f\left(F^{-1}(s_1,\dots,s_n)\right)}{|det\left(F'\left(F^{-1}(s_1,\dots,s_n)\right)\right)|}.$$

It can be shown that the Jacobian matrix of $F(x_1,\dots,x_{n-1},z)$ is equal to z^{n-1}. That is why $|det\left(F'\left(\frac{s_1}{s},\dots,\frac{s_{n-1}}{s},s\right)\right)| = s^{n-1}$. The density of the numerator is known out of

(39). Putting the parts together delivers

$$\begin{aligned} g(s_1,\ldots,s_n) &= \prod_{i=1}^{n}\left(\frac{s_i}{s}\right)^{\theta_i-1}\frac{1}{\Gamma(\theta_i)}\frac{s^{\sum_{i=1}^{n}\theta_i-1}e^{-s}}{s^{n-1}} \\ &= \prod_{i=1}^{n} s_i^{\theta_i-1}\frac{e^{-s_i}}{\Gamma(\theta_i)}\frac{s^{\sum_{i=1}^{n}\theta_i-1}}{s^{\sum_{i=1}^{n}\theta_i-n}s^{n-1}} \\ &= \prod_{i=1}^{n} s_i^{\theta_i-1}\frac{e^{-s_i}}{\Gamma(\theta_i)}. \end{aligned}$$

Then $g(s_1,\ldots,s_n)$ identifies the density of independent Gamma-distributed random variables.

Coming back to the Moran-Gamma-Subordinator and defining $t_0 = 0$ and $t_i := \sum_{j=1}^{i}\theta_j$ it is straightforward to determine

$$\begin{aligned} M_{t_i} - M_{t_{i-1}} &\sim \Gamma_{1,(t_i-t_{i-1})} = \Gamma_{1,\theta_i} \\ M_{t_n} - M_{t_0} &\sim \Gamma_{1,\sum_{i=1}^{n}\theta_i}. \end{aligned}$$

Putting $S_i := \left(M_{t_i} - M_{t_{i-1}}\right)$ and $S := M_{t_n}$ it follows that

$$X = \left(\frac{M_{t_1}}{M_{t_n}},\ldots,\frac{M_{t_n} - M_{t_{n-1}}}{M_{t_n}}\right) \sim Dir_{\theta_1,\ldots,\theta_n} \tag{40}$$

and X and S are independent random variables.

The marginal expectation of a Dirichlet distributed vector $(X_1^n,\ldots,X_n^n)$ can be easily calculated, i.e.

$$E(X_i^n) = \frac{\theta_i}{\sum_{i=1}^{n}\theta_i} \quad \text{for} \quad i = 1,\ldots,n.$$

According to Kingman in [14] we consider a simple, but intuitive situation, where $\theta_1 = \ldots = \theta_n = \frac{\theta}{n}$ for a real $\theta > 0$. The marginal expectation turns into $E(X_i^n) = \frac{\theta/n}{\theta} = \frac{1}{n}$ corresponding to

$$\lim_{n\to\infty} E\left(|X_i^n - 0|^1\right) = 0.$$

With this also convergence in distribution follows and writing this for k of the variables we have

$$(X_1^n,\ldots,X_k^n) \xrightarrow{d} (0,\ldots,0),$$

where $k \leq n \in \mathbb{N}$. In contrast to this non-exciting limit the Poisson-Dirichlet distribution occurs as a limit distribution, where the X_i^n are submitted to an order,i.e.

$$X_1^n \geq X_2^n \geq \ldots,$$

and where they represent the normalized jumps of the Moran-Gamma-subordinator according to (40). As it is often done in this section we use the setting given in [15].

Definition
Let $\theta > 0$ be a parameter and $(M_t)_{t\in[0,\theta]}$ be the Moran-Gamma-subordinator. Let

$$\hat{m_1} = \frac{m_1}{M_\theta} \geq \hat{m_2} = \frac{m_2}{M_\theta} \geq \ldots$$

be its ordered and normalized up throws on $\Delta_O = \{x_1 \geq x_2 \geq \ldots \geq 0 : \sum_{i=1}^{\infty} x_i = 1\}$. Then the distribution of the random vector $(\hat{m_1},\hat{m_2},\ldots)$ is called the **Poisson-Dirichlet**

distribution with parameter θ (PD_θ).

The definition is only complete with showing $\sum_{i=1}^{\infty} \hat{m} = 1$. This guarantees like in the case of the Dirichlet distribution that a probability measure is given with Poisson-Dirichlet distribution. In order to show that all preliminaries about the Poisson point process are needed:

Let X be a Poisson point process on $(0,\infty) \times (0,\theta]$ with intensity measure $\mu \otimes \lambda$, where λ stands for the Lebesgue measure. According to the Moran-Gamma-subordinator μ must be the Lévy measure of the Γ-distribution, which is in fact divisible. See below. Concerning the fact that a jump, which is allowed to have different heights, can always occur it is useful to consider an elementary time span. That is why μ has to be the Lévy measure of $\Gamma_{1,1}$-distribution, i.e.

$$\mu(dx) = \frac{e^{-x}}{x} dx.$$

In this setting it can be shown that $X(\{x,s\})$ with $x \in (0,\infty)$ and $s \in (0,\theta]$ can either be 1 or 0. So (36) can be manipulated to

$$M_t = \sum_{(x,s):X(\{x,s\})=1, s \leq t} x.$$

Then it is also clear that

$$\begin{aligned} m_1 &= \sup\{x \in (0,\infty) : X(\{x\} \times (0,\theta]) \\ m_2 &= \sup\{x \leq m_1 : X(\{x\} \times (0,\theta]) \\ &\vdots \end{aligned}$$

Now by adding the $\hat{m}$ up to infinity numerator and denominator will be equal to M_θ.

Remark
Indeed the Γ-distribution is divisible since the sum of independent Γ_{θ_i}-distributed random variables is also Γ-distributed with parameter $\sum \theta_i$. One of the quickest way to confirm this is by checking tables of Laplace transform:

$$\begin{aligned} f(t) = \frac{e^{-t} t^{\theta_i - 1}}{\Gamma(\theta_i)} &\overset{Laplace\ transform}{\longrightarrow} F(s) = \frac{1}{(1+s)^{\theta_i}} \qquad \text{and} \\ n - \text{times convolution} &\overset{Laplace\ transform}{\longrightarrow} F_1(s) \dots F_n(s) \qquad \Leftrightarrow \\ f(t) = \frac{e^{-t} t^{\sum \theta_i - 1}}{\Gamma(\sum \theta_i)} &\overset{Laplace\ transform}{\longrightarrow} F(s) = \frac{1}{(1+s)^{\sum \theta_i}}. \end{aligned}$$

4.3.3 Biased Case

In this section we provide a detailed presentation of results given in [1]. This also covers the approach in section 4.3.1, where we considered the uniform case as a special case of the biased one. Therefore some preliminaries are needed, where Poi_λ stands for a Poisson distributed random variable with parameter λ. As a comparison between (13) and (14) suggests it is useful to consider the Conditioning Relation, i.e. (21) with $Poi_{\theta/i}$ distributed random variables Z_i.

Proposition

For a random variable $\sum_{i=1}^{n} iZ_i$ with $Z_i \sim Poi_{\theta/i}$ the following equation holds:

$$kP\left(\sum_{i=1}^{n} iZ_i = k\right) = \theta \sum_{j=1}^{n} P\left(\sum_{i=1}^{n} iZ_i = k-j\right) \quad \text{for} \quad k = 1, 2, \ldots. \tag{41}$$

Proof
The size biased variable X^* of a non negative random variable X with finite mean $\mu > 0$ and distribution is introduced by the equivalent conditions:

$$F^*(dx) = \frac{xF(dx)}{\mu} \quad \text{for} \quad x > 0$$
$$E(g(X^*)) = \frac{E(Xg(X))}{E(X)} \quad \forall g \in M_b,$$

where M_b stands for the space of all bounded and measurable functions with values in $\mathbb{R}$. The size biased variable of a Poi_λ distributed X is $X+1$ distributed with the same parameter:

$$\begin{aligned} E(g(X+1)) &= \sum_{j=0}^{\infty} g(j+1) \frac{\lambda^j e^{-\lambda}}{j!} \\ &= \sum_{j=0}^{\infty} g(j+1) \frac{\lambda^{j+1} e^{-\lambda}}{(j+1)!} \frac{j+1}{\lambda} \\ &= \frac{1}{\lambda} \sum_{j=1}^{\infty} jg(j) \frac{\lambda^j e^{-\lambda}}{j!} \\ &= \frac{1}{\lambda} \sum_{j=0}^{\infty} jg(j) \frac{\lambda^j e^{-\lambda}}{j!} \\ &= \frac{E(Xg(X))}{E(X)}. \end{aligned}$$

It can be shown that for independent nonnegative random variables $X_1, \ldots, X_n$ with $E(\sum_{i=1}^{n} X_i) = \mu > 0$ the following equation holds in distribution, i.e.

$$\mathcal{L}\left(\left(\sum_{i=1}^{n} X_i\right)^*\right) = \sum_{j=1}^{n} \frac{\mu_j}{\mu} \mathcal{L}\left(\sum_{i=1, i\neq j}^{n} X_i + X_j^*\right).$$

Applying this for our variable of interest and using the multiplicative property of the size biasing delivers

$$\begin{aligned} \sum_{i=1, i\not\sim j}^{n} iZ_i + (jZ_j)^* &= \sum_{i=1, i\not\sim j}^{n} iZ_i + jZ_j^* \\ &= \sum_{i=1, i\not\sim j}^{n} iZ_i + j(Z_j + 1) \\ &= \sum_{i=1}^{n} iZ_i + j. \end{aligned}$$

What follows is

$$\begin{aligned} P\left(\left(\sum_{i=1}^{n} iZ_i\right)^* = k\right) &= \sum_{j=1}^{n} \frac{E(jZ_j)}{E(\sum_{i=1}^{n} iZ_i)} P\left(\sum_{i=1}^{n} iZ_i + j = k\right) \\ &= \sum_{j=1}^{n} \frac{\theta}{E(\sum_{i=1}^{n} iZ_i)} P\left(\sum_{i=1}^{n} iZ_i = k-j\right). \end{aligned}$$

Then with the definition of size biasing,i.e.

$$k\left(P\left(\sum_{i=1}^{n} iZ_i = k\right)\right) = E\left(\sum_{i=1}^{n} iZ_i\right) P\left(\left(\sum_{i=1}^{n} iZ_i\right)^* = k\right)$$
$$= \theta \sum_{j=1}^{n} P\left(\sum_{i=1}^{n} iZ_i = k - j\right).$$

So (41) is given. As a next step results from Arratia et al. [1] are given in order to introduce a θ-dependent Dickman function p_θ as asymptotic for $\sum_{i=1}^{n} iZ_i$. This is conforming to section 4.3.1.

Theorem
Let $Z_i \sim Poi_{\theta/i}$ with $i = 1, \ldots, n$.

1. Then as $n \to \infty$ the random variable $\frac{1}{n}\sum_{i=1}^{n} iZ_i$ converges in distribution to a random variable X_θ, whose distribution is characterized by the following Laplace transform
$$E\left(e^{-sX_\theta}\right) = e^{\left(-\int_0^1 \left(1-e^{-sx}\right)\frac{\theta}{x}dx\right)}. \tag{42}$$

2. The limit variable X_θ has density $p_\theta(x)$ for $x > 0$, where
$$p_\theta(x) = \frac{e^{-\gamma\theta}x^{\theta-1}}{\Gamma(\theta)} A \tag{43}$$
$$A = \left(1 + \sum_{k=1}^{\infty} \frac{(-\theta)^k}{k!} \int \cdots \int_{\substack{xy_1 > 1 \ldots xy_k > 1 \\ y_1 + \ldots + y_k < 1}} \left(1 - \sum_{j=1}^{k} y_j\right)^{\theta-1} \frac{dy_1 \ldots dy_k}{y_1 \ldots y_k}\right). \tag{44}$$

Proof

1. The Laplace transform of $\frac{1}{n}\sum_{i=1}^{n} iZ_i$ is given by
$$E\left(e^{-\frac{s}{n}\sum_{i=1}^{n} iZ_i}\right) = e^{-\sum_{i=1}^{n} \frac{\theta}{i}\left(1-e^{-\frac{si}{n}}\right)}.$$
Using the Lebesgue measure to have a measure for the points $\frac{1}{i}$ for $i = 1, \ldots, n$ on $(0, 1)$ the last sum can be written as an integral, i.e.
$$\lim_{n\to\infty} E\left(e^{-\frac{s}{n}\sum_{i=1}^{n} iZ_i}\right) = e^{-\int_0^1 \frac{\theta}{x}\left(1-e^{-sx}\right)dx}.$$

2. Substituting $x = \frac{t}{s}$ leads to
$$\int_0^1 \frac{(1-e^{-sx})}{x}dx = \int_0^s \frac{(1-e^{-t})}{t}dt.$$

Furthermore it is known that

$$\int_0^s \frac{\left(1-e^{-t}\right)}{t} dt = Exp\left(s\right) + \log s + \gamma,$$

where Exp again stands for the exponential integral. Putting the last result into (42) delivers

$$E\left(e^{-sX_\theta}\right) = e^{-\gamma\theta} s^\theta e^{-Exp(s)}.$$

Using the exponential progression and writing down the exponential integral leads to

$$E\left(e^{-sX_\theta}\right) = e^{-\theta\gamma} s^{-\theta} \sum_{k=0}^{\infty} \frac{(-\theta)^k}{k!} \left(\int_s^\infty \frac{e^{-y}}{y} dy \right)^k. \tag{45}$$

By checking tables of Laplace transform one obtains that $s^{-\theta} e^{-s\sum_{i=1}^k v_i}$ is the Laplace transform of $\frac{\left(x-\sum_{i=1}^k v_i\right)^{\theta-1}}{\Gamma(\theta)}$. Applying the definition of Laplace transform delivers

$$s^{-\theta} e^{-s\sum_{i=1}^k v_i} = \int_{\sum_{i=1}^k v_i}^\infty e^{-sx} \frac{\left(x-\sum_{i=1}^k v_i\right)^{\theta-1}}{\Gamma\left(\theta\right)} dx. \tag{46}$$

Using the last result and applying Fubini's theorem leads to

$$\begin{aligned}
s^{-\theta} \left(\int_s^\infty \frac{e^{-y}}{y} dy \right)^k = \\
s^{-\theta} \prod_{i=1}^k \left(\int_1^\infty \frac{e^{-sv_i}}{v_i} dv_i \right) = \\
\int_1^\infty \dots \int_1^\infty \frac{s^{-\theta} e^{-s(v_1+\dots+v_k)}}{v_1 \dots v_k} dv_1 \dots dv_k = \\
\int_1^\infty \dots \int_1^\infty \left(\int_{\sum_{i=1}^k v_i}^\infty e^{-sx} \frac{\left(x-\sum_{i=1}^k v_i\right)^{\theta-1}}{\Gamma\left(\theta\right)} dx \right) \frac{dv_1 \dots dv_k}{v_1 \dots v_k} = \\
\int_0^\infty e^{-sx} \left(\int_1^\infty \dots \int_1^\infty \mathbb{1}_{\sum_{i=1}^k v_i < x} \frac{\left(x-\sum_{i=1}^k v_i\right)^{\theta-1}}{\Gamma\left(\theta\right)} \frac{dv_1 \dots dv_k}{v_1 \dots v_k} \right) dx = \\
\int_0^\infty e^{-sx} \underbrace{\frac{x^{\theta-1}}{\Gamma\left(\theta\right)} \left(\int \dots \int_{\substack{xy_1>1 \dots xy_k>1 \\ y_1+\dots+y_k<1}} \left(1-\sum_{i=1}^k y_i\right)^{\theta-1} \frac{dy_1 \dots dy_k}{y_1 \dots y_k} \right)}_{(*)} dx.
\end{aligned}$$

Note that this computation was done from Arratia et al. [1] and that the total expression represents the Laplace transform of (*). Completing (*) with $e^{-\theta\gamma} \sum_{k=0}^\infty \frac{(-\theta)^k}{k!} \dots$ from equation (45) finally delivers the result. Now we are able to connect the results obtained above. This is done in [1], too.

Theorem
If $k = k(n) \in \mathbb{Z}^+$ with $\lim_{n\to\infty} \frac{k}{n} = y \in (0, \infty)$, then

$$\lim_{n\to\infty} nP\left(\sum_{i=1}^{n} iZ_i = k\right) \sim p_\theta(y). \tag{47}$$

Proof
Equation (41) can be written as follows

$$kP\left(\sum_{i=1}^{n} iZ_i = k\right) = \theta P\left(k - n \leq \sum_{i=1}^{n} iZ_i < k\right).$$

Multiplying this by $\frac{n}{k}$ leads to

$$nP\left(\sum_{i=1}^{n} iZ_i = k\right) = \theta\frac{n}{k}P\left(\frac{k-n}{n} \leq \frac{1}{n}\sum_{i=1}^{n} iZ_i < \frac{k}{n}\right).$$

At last using the limit, i.e.

$$\lim_{n\to\infty} nP\left(\sum_{i=1}^{n} iZ_i = k\right) = \theta\frac{1}{y}P(y-1 \leq X_\theta < y) \sim p_\theta(y).$$

delivers the result. The last step works since p_θ (Dickman - function) has the following property

$$yp_\theta(y) = \theta \int_{y-1}^{y} p_\theta(x)\,dx. \tag{48}$$

Now we are ready to turn to the asymptotic distributions for the rth longest cycles in the biased case. Our proceeding is as follows: First of all we deal with the marginal distribution of the longest cycle. For this we compare the asymptotic results with the one generated by MCIT. After the marginal distribution we turn to the joint distribution of the vector that contains the lengths of the first r longest cycles. This will be the Poisson - Dirichlet distribution introduced in section 4.3.2 and used in 4.3.1 to derive the length of the longest cycle in the uniform case. The structure is chosen to emphasize how the single parts belong together. For the following see [1] again.

Lemma
The length of the longest cycle of a θ-biased permutation of length n converges in distribution to a random variable,i.e.

$$\frac{1}{n}L_1^{(n,\theta)} \to L_1^\theta,$$

which has distribution function

$$F_\theta(x) = e^{\gamma\theta}x^{\theta-1}\Gamma(\theta)\,p_\theta\left(\frac{1}{x}\right) \quad \text{for} \quad x > 0. \tag{49}$$

In the following the notation is often done without θ. However notice that this section is about the biased case and the terms are not the same as in section 4.3.1. p_θ is given by (44).

Proof
First of all it is useful to put $m := \lfloor nx \rfloor$ for $x \in (0,1]$ because of

$$P\left(\frac{1}{n}L_1^{(n,\theta)} \leq x\right) = P\left(L_1^{(n,\theta)} \leq m\right).$$

For simplification the floor function is neglected in the following computation. Correction of this would not bring any insights. Notice that if the longest cycle is lower or equal to m all $Y_i : \ i > m$ counting the number of cycles with length i must be equal to 0. Now the Conditioning Relation (21) and its $Poi_{\theta/i}$-distributed random variables can be used:

$$\begin{aligned}
P\left(\frac{1}{n}L_1^{(n,\theta)}\right) &= P\left(Y_{m+1}^{(n)} + \ldots + Y_n^{(n)} = 0\right) \\
&= P\left(Z_{m+1} + \ldots + Z_n = 0 | \sum_{i=1}^{n} iZ_i = n\right) \\
&= \frac{P\left(\sum_{m+1}^{n} Z_i = 0, \sum_{i=1}^{n} iZ_i = n\right)}{P\left(\sum_{i=1}^{n} iZ_i = n\right)} \\
&= P\left(Z_{m+1} = 0\right) \ldots P\left(Z_n = 0\right) \frac{P\left(\sum_{i=1}^{m} iZ_i = n\right)}{P\left(\sum_{i=1}^{n} iZ_i = n\right)} \\
&= \underbrace{e^{-\theta\left(\sum_{i=1}^{n}\frac{1}{i} - \sum_{i=1}^{m}\frac{1}{i}\right)}}_{(*)} \frac{1}{x} \underbrace{\frac{m}{n}\frac{P\left(\sum_{i=1}^{m} iZ_i = n\right)}{P\left(\sum_{i=1}^{n} iZ_i = n\right)}}_{(**)}.
\end{aligned}$$

Using the logarithm to approximate the harmonic number we obtain for (*)

$$e^{-\theta \log\left(\frac{n}{m}\right)} = \left(\frac{1}{x}\right)^{-\theta} = x^{\theta}.$$

Applying (47) with $y_1 = \frac{n}{m}$ and $y_2 = \frac{n}{n}$ delivers for (**):

$$\frac{m}{n}\frac{P\left(\sum_{i=1}^{m} iZ_i = n\right)}{P\left(\sum_{i=1}^{n} iZ_i = n\right)} \sim \frac{p_\theta\left(1/x\right)}{p_\theta\left(1\right)}.$$

Putting together the parts and computing $p_\theta\left(1\right) = \frac{e^{-\gamma\theta}}{\Gamma(\theta)}$ leads to the statement above

$$P\left(\frac{1}{n}L_1^{(n,\theta)} \leq x\right) \sim x^{\theta-1} e^{\gamma\theta} \Gamma\left(\theta\right) p_\theta\left(\frac{1}{x}\right).$$

The following figure compares for θ-biased permutations of length 50 the results generated in section 3.2.2 with those of a θ-dependent Dickman function. Thereby $\theta = 1.5$.

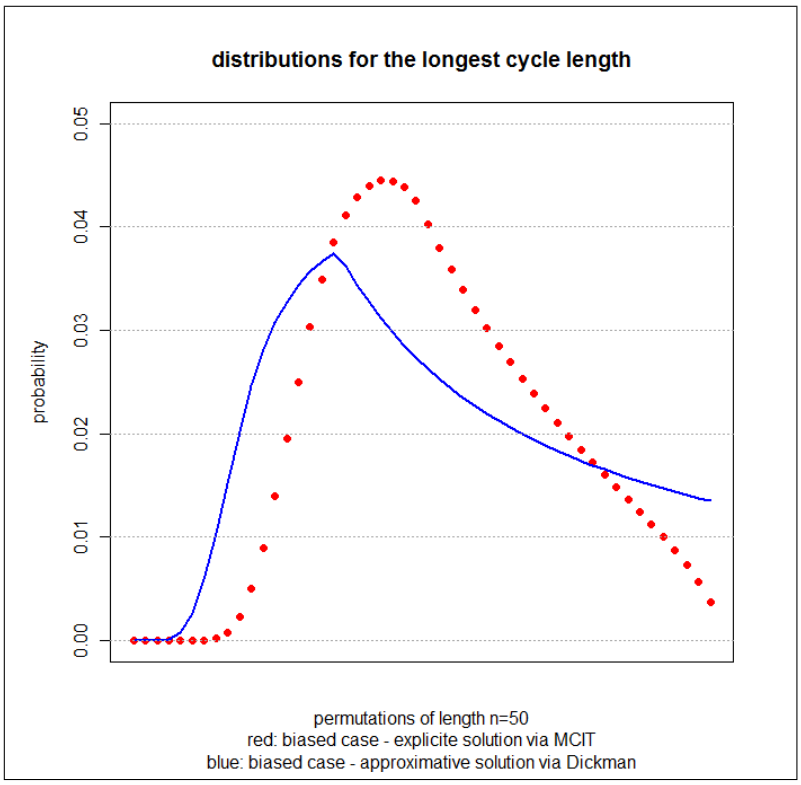


Figure 19: Explicit and limit distribution for the longest cycle length - biased case

Note that the asymptotic method in the figure above is not exactly (49). This is because it is not straightforward to obtain numerical values for (49). For details consider appendix E. So keep in mind that there might be numerical improvements which bring both methods, MCIT and the asymptotic one, even nearer together. A further analytical and graphical presentation of the rth marginal distribution is not given here. Such a marginal distribution can be derived from the joint distribution, which follows next. But a comparison with MCIT - values has to be taken carefully, since a numerical analysis of p_θ is only approximately possible. See appendix E. That is why it is not easy to differ which part of an error between the MCIT - values and the asymptotic approach has its origin in the numerical analysis and which in the difference between the methods. However such a comparison is possible. Next we turn to the joint distribution of the r largest cycle lengths. The following results that terminate this section are also given in [1].

Theorem
For $r \geq 2$ suppose that $0 < x_r < \ldots < x_1 < 1$ also satisfy $0 < \sum_{i=1}^{r} x_i < 1$. Then if integers $m_i = m_i(n)$ exist that satisfy $\frac{m_i}{n} \to x_i$ for $n \to \infty$ and $1 \leq i \leq r$ the following asymptotic is valid:

$$\begin{aligned} \lim_{n\to\infty} n^r P\left(L_1^{(n,\theta)} = m_1, \ldots, L_r^{(n,\theta)}\right) &= f_\theta^r(x_1, \ldots, x_r) \\ &= \frac{e^{\gamma\theta}\theta^r \Gamma(\theta)\, x_r^{\theta-1}}{x_1 \ldots x_r} p_\theta\left(\frac{1 - x_1 - \ldots - x_r}{x_r}\right). \end{aligned} \tag{50}$$

Notice that p_θ is defined in (44) and that $\frac{e^{\gamma\theta}\Gamma(\theta)}{e^{\gamma\theta}\Gamma(\theta)}$ can be reduced.

Proof
If $m_1, \ldots, m_r$ exist with $\frac{m_i}{n}$ they automatically satisfy the conditions

$$\begin{aligned} 1 \leq m_r < m_{r-1} < \ldots < m_1 < n \\ m_1 + \ldots + m_r \leq n. \end{aligned} \tag{51}$$

Let $y^n = \left[Y_n^n = 0, \ldots, Y_{m_1+1}^n = 0, Y_{m_1}^n = 1, Y_{m_1-1}^n = 0, \ldots\right]$ be the particular cycle structure, where all Y_{m_i} with $1 \leq i \leq r-1$ be equal to 1 and 0 elsewhere. With this the following joint probability can be written down:

$$P\left(L_1^n = m_1, \ldots, L_r^n = m_r\right) = P\left(Y^n = y^n, Y_{m_r}^n \geq 1\right),$$

where $Y_{m_r}^n \geq 1$ stands for the fact that the cycle structure from now on does not matter for the event on the left side. σ-additivity allows to split the last probability, i.e.

$$\underbrace{P\left(Y^n = y^n, Y_{m_r}^n = 1\right)}_{(*)} + \underbrace{\sum_{l \geq 2} P\left(Y^n = y^n, Y_{m_r}^n = l\right)}_{(**)}.$$

Applying the Conditioning Relation (*) can be written in terms of independent $Poi_{\theta/i}$ distributed random variables, i.e.

$$P\left(Y^n\left(Z\right) = y^n\left(Z\right), Z_{m_r} = 1 | \sum_{i=1}^n iZ_i = n\right).$$

Using the fact that the sum of the product of cycle length and its corresponding number from position m_r on up to the end is equal to $m := \sum_{i=1}^r m_i$ we obtain

$$\sum_{i=1}^{m_r - 1} iZ_i = n - m.$$

What follows for (*) is

$$\begin{aligned} P\left(Y^n\left(Z\right) = y^n\left(Z\right)\right) P\left(Z_{m_r} = 1\right) \frac{P\left(\sum_{i=1}^{m_r-1} iZ_i = n - m\right)}{P\left(\sum_{i=1}^n iZ_i = n\right)} = \\ \prod_{i=1}^r e^{-\frac{\theta}{m_i}} \left(\frac{\theta}{m_i}\right) \frac{n}{m_r - 1} \frac{m_r - 1}{n} \frac{P\left(\sum_{i=1}^{m_r-1} iZ_i = n - m\right)}{P\left(\sum_{i=1}^n iZ_i = n\right)} = \\ e^{\sum_{i=1}^n \left(\frac{1}{i} - \sum_{i=1}^n \frac{1}{i}\right)} \theta^r \frac{1}{m_1 \ldots m_r} \frac{1}{x_r} \frac{p_\theta\left(\frac{n-m}{m_r-1}\right)}{p_\theta\left(1\right)} = \\ \underbrace{e^{\log\left(\frac{n}{m_r-1}\right)}}_{\Delta} \theta^r \frac{1}{x_r} \underbrace{\frac{1}{m_1 \ldots m_r}}_{\Delta\Delta} e^{\gamma\theta} \Gamma\left(\theta\right) p_\theta\left(\frac{1 - x_1 - \ldots - x_r}{x_r}\right). \end{aligned}$$

Like in the last Lemma we were able to apply (47). Furthermore note that we treat $\frac{n}{m_r-1}$ as if $\frac{n}{m_r}$, but concerning the assumptions this can be done. Since $\Delta = x_r^\theta$ and combining $\Delta\Delta$ with n^r delivers $\left(\frac{1}{x_1 \ldots x_r}\right)$ the proof is complete, when $(**) = o\left(n\right)$,i.e.

$$\lim_{n \to \infty} \frac{1}{n^r} \sum_{l \geq 2} P\left(Y^n = y^n, Y_{m_r}^n = l\right) = 0.$$

We do not give a detailed argumentation for that here. It can be seen in [1] and depends on elementary estimates. Instead of this we attain the result mentioned in section 4.3.1.

For θ-biased random permutations the vector of the largest cycle lengths has the following asymptotic

$$\frac{1}{n}\left(L_1^n, L_2^n, \ldots\right) \xrightarrow{d} (L_1, L_2, \ldots)$$

on $\Delta = \{x \in \mathbb{R}_+^d : \ \sum_{i\leq 1} x_i = 1\} \subset [0,1]^d$. The density of $(L_1, \ldots, L_r)$ is then equal to f_θ^r.

At first glance there seems to be nothing to do, but in the theorem above no convergence in distribution is given. Arratia et al. wants to achieve this with Scheffé's theorem [21].

Theorem
If for a sequence $(p_n(x))$ of densities

$$\lim_{n\to\infty} p_n(x) = p(x) \quad \text{(point wise)}$$

for almost all x in $\mathbb{R}^d$ holds, then a sufficient condition that

$$\lim_{n\to\infty} \int_B p_n(x)\, dx = \int_B p(x)\, dx \quad \forall \quad \text{Borel sets } B \in \mathbb{R}^d \tag{52}$$

holds, is that $p(x)$ is a density. The proof of this is based on Lebesgue's theorem about dominated convergence. Obviously (52) identifies convergence in distribution, but the convergence in the theorem above is not point wise for densities, but point wise for probabilities. To correct this Arratia et al. introduce independent uniform random variables $U_1, \ldots, U_r$ on $(0,1)$ being also independent of $(L_1^n, \ldots, L_r^n)$. They define $\hat{L}_i^n := \frac{1}{n}\left(L_i^n + U_i\right)$ for $i = 1, \ldots, n$. Then indeed the density of $\left(\hat{L}_1^n, \ldots, \hat{L}_r^n\right)$ is given by

$$n^r P\left(L_1^n = \lfloor nx_1 \rfloor, \ldots, L_1^n = \lfloor nx_1 \rfloor\right),$$

and so the theorem above guarantees that the density of $\left(\hat{L}_1^n, \ldots, \hat{L}_r^n\right)$ converges point wise to f_θ^r. Assuming that f_θ^r is a density, namely the one of the Poisson Dirichlet (PD) distribution, and taking $(L_1, \ldots, L_r)$ as PD - distributed random vector Scheffés theorem delivers

$$\left(\hat{L}_1^n, \ldots, \hat{L}_r^n\right) \xrightarrow{d} (L_1, \ldots, L_r).$$

Simultaneously the following

$$\left(\hat{L}_1^n, \ldots, \hat{L}_r^n\right) - \frac{1}{n}\left(L_1^n, \ldots, L_r^n\right) = \frac{1}{n}\left(U_1, \ldots, U_r\right) \xrightarrow{d} (0, \ldots, 0)$$

holds so that $\frac{1}{n}\left(L_1^n, \ldots, L_r^n\right) \xrightarrow{d} (L_1, \ldots, L_r)$ can be concluded.

In the next section it will be said more about why f_θ^r is the density of Poisson-Dirichlet distribution and about the connection to the Dickman function and GEM distribution.

4.3.4 Overview

In section 4.3.1 we got to know a first impression how a limit distribution for the rth cycle length could be derived. Then in section 4.3.2 we technically introduced the Poisson Dirichlet distribution (PD) since we supposed it to be the right one. In section 4.3.3 we derived the marginal density f_r^θ as the one of the limit distribution of the first r cycle lengths.

What remains to do is to argue why f_r^θ is indeed the marginal density of PD. Why marginal density ? Let $(L_1, L_2, \ldots)$ be PD distributed. We only consider a finite subset of this random vector. That is why marginal is correct. Note that (44) can also be written as

$$p_\theta(x) = 1 + \sum_{k=1}^{[x]} \frac{(-\theta)^k}{k!} \int_{1/x}^{1} \ldots \int_{1/x}^{1} \frac{(1 - y_1 - \ldots - y_k)^{\theta-1}}{y_1 \ldots y_k} dy_1 \ldots dy_k.$$

See [11] and note that terms can be reduced. A first result will deal with (49). There we mainly had

$$P(L_1 \leq x) = p_\theta\left(\frac{1}{x}\right).$$

This means: If L_1 represents the first component of a Poisson - Dirichlet distributed random vector, it must be possible to show out of that

$$\begin{aligned} P(L_1 \leq x) &= p_\theta\left(\frac{1}{x}\right) \\ &= 1 + \sum_{k=1}^{[1/x]} \frac{(-\theta)^k}{k!} \int_{x}^{1} \ldots \int_{x}^{1} \frac{(1 - y_1 - \ldots - y_k)^{\theta-1}}{y_1 \ldots y_k} dy_1 \ldots dy_k. \end{aligned}$$

Remind yourself that in section 4.3.2 we introduced the PD as the limit one of Dirichlet distribution. That is why we start with the first Dirichlet component Z_1. Obviously we have

$$P(Z_1 \leq x) = 1 - P(Z_1 > x \cup \ldots \cup Z_n > x),$$

since the Z_i are ordered by their size. Now inclusion - exclusion formula can be applied. This leads to

$$1 - P(Z_1 > x \cup \ldots \cup Z_n > x) = 1 + \sum_{k=1}^{[1/x]} (-1)^k \binom{n}{k} P(Z_1 > x, \ldots, Z_k > x). \qquad (53)$$

Recall Dirichlet density from (38) and assume symmetry, i.e.

$$\theta_i = \frac{\theta}{n} \quad \forall i = 1, \ldots, k \quad \text{and} \sum_{i=1}^{n} \theta_i = \theta.$$

What follows is

$$(53) = 1 + \sum_{k=1}^{[1/x]} (-1)^k \binom{n}{k} \int_{x}^{1} \ldots \int_{x}^{1} \frac{\Gamma(\theta)(1 - y_1 - \ldots - y_k)^{\theta - \frac{k\theta}{n} - 1}}{\Gamma\left(\frac{\theta}{n}\right)^k \Gamma\left(\theta - \frac{k\theta}{n}\right) y_1^{1-\frac{\theta}{n}} \ldots y_k^{1-\frac{\theta}{n}}}.$$

Let $n \to \infty$ and consider

$$\frac{1}{\Gamma\left(\frac{\theta}{n}\right)^k} = \frac{\left(\frac{\theta}{n}\right)^k}{\left(\frac{\theta}{n}\Gamma\left(\frac{\theta}{n}\right)\right)^k} = \frac{\left(\frac{\theta}{n}\right)^k}{\Gamma\left(\frac{\theta}{n} + 1\right)^k}.$$

This leads to

$$\begin{aligned} P(Z_1 \leq x) &\to P(L_1 \leq x) \\ \frac{n!}{k!(n-k)!}\frac{\theta^k}{n^k} &\to \frac{\theta^k}{k!} \\ \frac{1}{\Gamma\left(\frac{\theta}{n}+1\right)^k} &\to 1. \end{aligned}$$

Putting parts together it follows that L_1 represents the first component of a Poisson Dirichlet distributed random vector. A similar calculation as above using Dirichlet distribution, symmetry and substitution shows that

$$\begin{aligned} f_{L_1 \ldots L_r}(x_1 \ldots x_r) &= P\left(L_1 \leq \frac{x_r}{1-x_1-\ldots-x_r}\right) \\ &= \frac{\theta^r x_r^{\theta-1}}{x_1 \ldots x_r} p_\theta\left(\frac{1-x_1-\ldots-x_r}{x_r}\right). \end{aligned} \tag{54}$$

For details see [11]. So in fact f_θ^r from (50) represents the density of PD. As a next step we want to derive the differential equation that typically characterizes Dickman's function and that is used for the numerical values generated by what is given in appendix E. Dickman's function is defined by

$$\begin{aligned} p_\theta(z) &= P\left(L_1 \leq \frac{1}{z}\right) \\ \text{with} \quad p_\theta(z) &= 1 \quad \text{for} \quad 0 < z < 1. \end{aligned}$$

For $z > 1$ we have

$$\begin{aligned} p'_\theta &= \frac{\partial P\left(L_1 < \frac{1}{z}\right)}{\partial z} \\ &= \frac{\partial F_{L_1}\left(\frac{1}{z}\right)}{\partial z} \\ &= f_{L_1}\left(\frac{1}{z}\right)\left(-\frac{1}{z^2}\right). \end{aligned}$$

By rearranging we have from (54)

$$f_{L_1}(x) = \frac{\theta(1-x)^{\theta-1}}{x} P\left(L_1 < \frac{x}{1-x}\right) \quad \text{for} 0 < x < 1.$$

Putting $x = \frac{1}{z}$ delivers

$$\begin{aligned} f_{L_1}\left(\frac{1}{z}\right) = \frac{\theta(z-1)^{\theta-1}}{z^{\theta-1}} z P\left(L_1 < \frac{\frac{1}{z}}{1-\frac{1}{z}}\right) \Rightarrow \\ -\frac{1}{z^2} f_{L_1}\left(\frac{1}{z}\right) = \frac{\theta(z-1)^{\theta-1}}{z^\theta} p_\theta(z-1). \end{aligned}$$

This leads to

$$\begin{aligned} p'_\theta(z) = -\frac{\theta(z-1)^{\theta-1}}{z^\theta} p_\theta(z-1) \Rightarrow \\ z^\theta p'_\theta(z) = -\theta(z-1)^{\theta-1} p_\theta(z-1). \end{aligned} \tag{55}$$

There also exists an integral equation

$$\begin{aligned}\left(z^{\theta}p_{\theta}(z)\right)' &= \theta z^{\theta-1}p_{\theta}(z) + z^{\theta}p'(z)\\ &= \theta z^{\theta-1}p_{\theta}(z) - \theta(z-1)^{\theta-1}p_{\theta}(z-1) \Leftrightarrow\\ z^{\theta}p_{\theta}(z) &= \int_{z-1}^{z} \theta y^{\theta-1}p_{\theta}(y)\,dy.\end{aligned} \tag{56}$$

for $z > 1$. With this we also have (48). Now we finally want to connect L_1 with a Beta distributed U_1. This is also given in [11].

Proposition
Let L_1 and U_1 be independent random variables with

$$P(L_1 \leq x) = p_{\theta}\left(\frac{1}{x}\right) \quad \text{and}$$
$$f_{U_1}(u) = \frac{1}{B(1,\theta)}u^0u^{\theta}(1-u)^{\theta-1}.$$

Then L_1 and $U_1 \vee (1-U_1)L_1$ have the same distribution.

Proof
Note that $B(1,\theta) = \int_0^1 (1-u)^{\theta-1}\,du$. By independence, substitution and (56) we have

$$P(U_1 \vee (1-U_1)L_1 \leq x) =$$
$$\int_0^x \theta(1-u)^{\theta-1}p_{\theta}\left(\frac{1-u}{x}\right)du \overset{y=(1-u)/x}{=}$$
$$-\int_{y(0)}^{y(x)} \theta x^{\theta-1}y^{\theta-1}p_{\theta}(y)\,x\,dy =$$
$$x^{\theta}\int_{\frac{1}{x}-1}^{\frac{1}{x}} \theta y^{\theta-1}p_{\theta}(y)\,dy \overset{(56)}{=}$$
$$x^{\theta}\left(\frac{1}{x}\right)^{\theta}p_{\theta}\left(\frac{1}{x}\right) = P(L_1 \leq x).$$

From PD to GEM
What has been said up to now our focus was on Poisson Dirichlet distribution (PD). Terminating this section we want to emphasize a detail already mentioned above. The vector $(L_1, L_2, \ldots)$ is naturally size-ordered, when L_i describes the length of the i longest cycle. In section 4.1 we discovered the Ewens Sampling Formula as essential for the distributions of the cycle structures. Here there are limit distributions, too. See [11]. Technical details are left out here.

Let $(\alpha_1, \ldots, \alpha_r)$. Size-ordering of the α_i leads in fact to

$$\left(\frac{\max_{i=1,\ldots,r}\alpha_i}{n}, \ldots, \frac{min_{i=1,\ldots,r}\alpha_i}{n}, 0, 0, \ldots\right) \to PD(\theta) \quad \text{for} \quad n \to \infty. \tag{57}$$

Now the question is uprising how the limit distribution looks like, if we do not have a size-ordering. This will be answered by $GEM(\theta)$, so called after Griffiths, Engen and McClosky.

Note that the definition of Dirichlet distributed random variables on a simplex is equivalent with a so called probability vector. Then random variables $V_1, V_2, \ldots$ are introduced such that

$$\begin{aligned} P(V_1 = i|L) &= L_i \qquad \text{for} \quad i = 1, 2, \ldots \\ P(V_2 = i|L, V_1) &= \frac{L_i}{1 - L_{V_1}} \qquad \text{for} \quad i = 1, 2, \ldots, i \neq V_1 \\ P(V_3 = i|L, V_1, V_2) &= \frac{L_i}{1 - L_{V_1} - L_{V_2}} \qquad \text{for} \quad i = 1, 2, \ldots, i \neq V_1, i \neq V_2 \\ &\vdots \end{aligned}$$

Instead of the probability vector $(L_1, L_2, \ldots)$ now consider the random probability vector $(L_{V_1}, L_{V_2}, \ldots)$. For example the probability of $L_{V_1} = L_i$ is determined by $V_1 = i$ conditional on L.

The random probability vector $L_V = (L_{V_1}, L_{V_2}, \ldots)$ is called the size-biased permutation of L. If $L = (L_1, L_2, \ldots)$ is $PD(\theta)$ and L_V is its size-biased permutation, L_V will have the same distribution as a vector of independent, identical Beta distributed random variables, i.e.

$$L_V \overset{d}{=} (U_1, (1 - U_1) U_2, (1 - U_1)(1 - U_2) U_3, \ldots).$$

As we already got to know, we have $U_i \sim \text{Beta}(1, \theta)$ with probability density function $f(x) = \theta (1 - x)^{\theta - 1}$. The distribution of L_V is called $GEM(\theta)$. Now we can give an analogon to (57) since it can be shown that an arbitrary cycle structure is a size-biased permutation of a size-ordered one, i.e.

$$\left(\frac{\alpha_1}{n}, \ldots, \frac{\alpha_r}{n}, 0, 0, \ldots\right) \to GEM(\theta) \qquad \text{for} \qquad n \to \infty.$$

4.3.5 The distribution of the rth shortest cycle

In contrast to the rth longest cycle we do not deal with the theory of the rth shortest cycle so extensively. Note that there are further approaches, e.g. in [1] and [22], which have a lot of reference points to what has been said in the sections before.

Again consider a particular cycle structure $[\alpha_1^n, \ldots, \alpha_n^n]$. As it is also mentioned in section 4.3.1 we have

$$S_r^n > l \Leftrightarrow \sum_{j=1}^{l} \alpha_j^n \leq r-1,$$

i.e. the rth shortest cycle can only be longer than l when the number of cycles up to the lth position is less than r. Substituting α_j by independent $Poi_{\frac{1}{j}}$ distributed random variables z_j for $j = 1, \ldots, n$ we take an heuristic way according to the Conditioning Relation, i.e.

$$\begin{aligned} P(S_r > l) &= P\left(\sum_{j=1}^{l} z_j < r\right) \\ &= Poi_{\sum_{j=1}^{l} \frac{1}{j}} \Rightarrow \\ P(S_r = l) &= P(S_r > l-1) - P(S_r > l) \\ &= Poi_{\sum_{j=1}^{l-1} \frac{1}{j}} - Poi_{\sum_{j=1}^{l} \frac{1}{j}}. \end{aligned} \tag{58}$$

The sum of independent Poisson distributed random variables is also Poisson distributed with the sum of the single parameters. This is easy to show by generating functions. So we have the right side in the equations above. With (58) it is possible to generate the probabilities recursively. The probability for length 1 is created by the difference up to 1. The following figure compares for permutations of length 50 the explicit distribution of the shortest cycle generated by MCIT with the one generated by (58). This is done for the uniform case as well as for the $\theta = 1.5$ case.

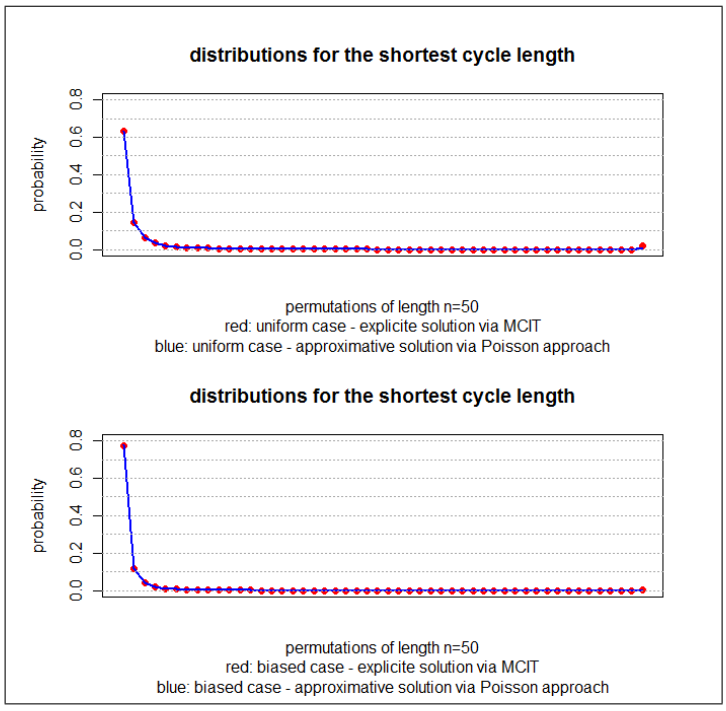


Figure 20: Explicit and approximative distributions for the shortest cycle length

For the biased case as mentioned before it is just enough to substitute $Poi_{\frac{1}{j}}$ by $Poi_{\frac{\theta}{j}}$ distributed random variables. The following figure compares for permutations of length 50 the explicit distribution of the 5 shortest cycle generated by MCIT with the one generated by the analogue terms of (58). This is again done for the uniform case and for $\theta = 1.5$ case as always in this section.

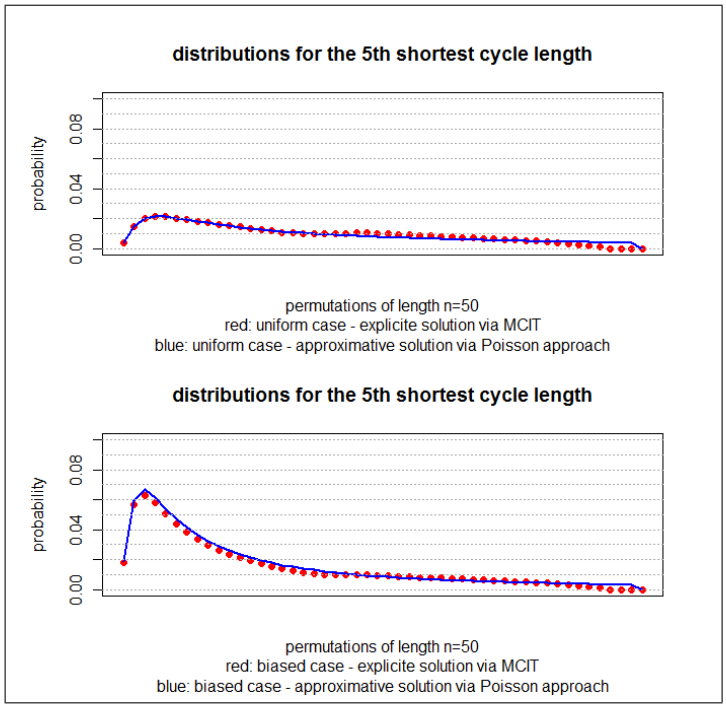


Figure 21: Explicit and approximative distributions for the 5th shortest cycle length

Note that the Poisson distribution in this setting suggest major probability on length 50. This is not true since there exists no 5th shortest cycle in structure $[0, \ldots, 0, 1]$. The mass must be on length 0. This has been corrected in the figures above. All in all the approximations are pretty good so that no further approaches are considered here.

4.4 The distribution of the length of a random chosen cycle

Here we can not provide a technically clean derivation of the asymptotic approach used below. Nevertheless we want to sketch how this is working. In [8] Goldschmidt and Martin deal with random recursive trees and show in their Theorem 3.1 that the following asymptotic holds

$$\frac{\log M_n}{\log n} \xrightarrow{d} U_{[0,1]}.$$

Thereby $U_{[0,1]}$ is an uniformly distributed random variable on $[0,1]$. M_n stands for 'the sum of the blocks not containing integer 1' (Goldschmidt and Martin [8]) before the last cutting reunites the random recursive tree to $(1,\ldots,n)$. For a deeper understanding of random recursive trees and a cutting procedure in them see [19].

Note that a random recursive tree on $(1,\ldots,n+1)$ corresponds to a random permutation on $(1,\ldots,n)$. This means we have to leave what we called 'cycle notation in a new form' in section 3.1, e.g. $[1,0,1,\ldots]$ and have to return to classical cycle notation, e.g. $(1)\,(234)\ldots$.

The cutting procedure works by picking an edge of the random recursive tree uniformly among all. This is quite similar than choosing a cycle uniformly among all. So the result of Goldschmidt and Martin before the last cutting is nothing else than an uniformly chosen cycle regarding the classical cycle notation. Taking the size of the blocks is then nothing else than taking the length of the cycle above.

Leaving technichal details out and considering the connection between recursive trees and permutations we obtain for X_n being the length of a random chosen cycle of permutations of length n

$$\begin{aligned} P\left(X_n \leq x\right) &= P\left(\frac{\log\left(X_n+1\right)}{\log\left(n+1\right)} \leq \frac{\log\left(x+1\right)}{\log\left(n+1\right)}\right) \\ &\sim P\left(U_{[0,1]} \leq \frac{\log\left(x+1\right)}{\log\left(n+1\right)}\right) \\ &= \frac{\log\left(x+1\right)}{\log\left(n+1\right)}. \end{aligned} \tag{59}$$

The following figure compares for uniformly distributed permutations the length of a random chosen cycle generated by MCIT with the one of (59)

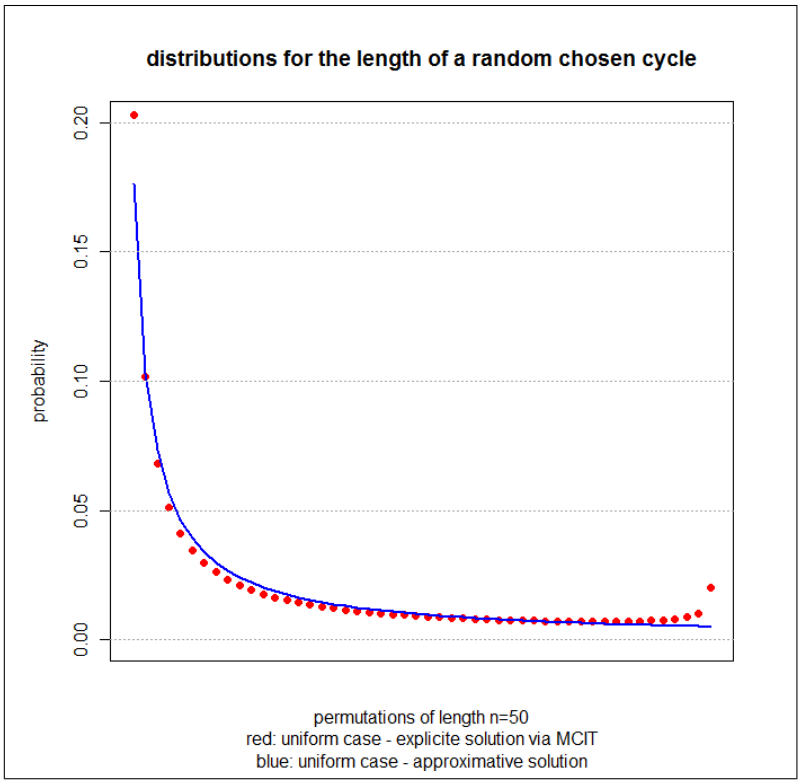


Figure 22: Explicit and approximative distributions for random chosen cycle length

Although the approximation seems to work quite well the following statements relativize this appreciation

1. The log terms automatically depress the values of the distribution. They are not able to capture the characteristic increasing of the probability at the end.

2. The presented setting does not fit to the modified cycle notation. That is it is not easily possible to compare the values given above with $\theta \neq 1$. distributed permutations.

5 Conclusion and perspective

According to the articles presented in section 2 and the use of Markov chain imbedding technique for the distributions concerning the cycle structure of permutations there are a lot of possibilities to apply this instrument. Of course it is ideal when the imbedding - such as in the case of success runs in a binomial trial - leads to a great saving of computational effort. Unfortunately this can not be achieved for the distributions of section 3. Nevertheless it is worthwhile to obtain explicit values for the distributions mentioned above. Approximations of section 4 vary in their quality. A detailed analysis of convergence rates is not topic of this thesis.

Because of the wide implementation range it is advised to consider Markov chain imbedding technique if an appropriate problem is given. Some experience is needed to estimate, how worthwhile an application is. Based on the distribution of permutations presented in this work we want to give some perspectives:

The mechanism of the biased case is indeed nothing else than the so called Chinese restaurant process [8] with occupying new tables instead of opening new cycles. To that effect a simple extension is given by inserting an additional real parameter α. This corresponds to the following probabilities:

$$\begin{aligned} P\left(\text{Sharing an existing table with } n_i \text{ other guests}\right) &= \frac{n_i - \alpha}{n + \theta} \\ P\left(\text{Occupying a new table}\right) &= \frac{\theta + k\alpha}{n + \theta}, \end{aligned}$$

with $i = 1, \ldots, k$ and k standing for the total number of occupied tables up to the moment. A probability measure is given since

$$\sum_{i=1}^{k} \frac{n_i - \alpha}{n + \theta} + \frac{\theta + k\alpha}{n + \theta} = 1.$$

Note that in order to avoid negative probabilities two constellations are possible

$$\begin{aligned} \alpha < 0 &\Rightarrow k\alpha < \theta \\ 0 \leq \alpha \leq 1 &\Rightarrow -\theta < \alpha. \end{aligned} \tag{60}$$

Obviously this could easily be used to generate the distributions above by other transition probabilities as for a single $\theta > 0$. Last but not least we want to widen the framework even more.

A lot of research has been done concerning coalescent theory of gene copies. For example Kingman showed that

$$P\left(k \text{ copies reduce to } (k-1) \text{ copies}\right) = \frac{k\,(k-1)}{4N},$$

where N is the population size. Goldschmidt and Martin argue in [8] that a random recursive tree on $[n+1]$ corresponds to a random permutation of $[n]$, where $[\cdot]$ stands for the set of integers up to $\cdot$. In section 3 and 4 we also had a hint on such trees. Furthermore they translate the results for trees into results for coalescents. Coalescent theory deals with trees that grow together. One of the essential question is which gene copies share a parent. In contrast to that the tree for permutations was rapidly expanding. But it is possible to reverse the structure and discover parallels. With slight modifications of the MCIT algorithm it is possible to generate results with totally other transition probabilities as long as the structure is not going too fast too big. This may also strike the areas of coalescent and random recursive trees.

Appendix

A Special functions in Combinatoric Analysis

In the descriptions above the following equations regarding Γ- and Ψ-function were used

$$\Gamma(n+1) = n\Gamma(n) \tag{61}$$

$$\Gamma(n+1) = n! \tag{62}$$

$$\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} \tag{63}$$

$$\Gamma(1) = 1 \tag{64}$$

$$\Gamma'(1) = -\gamma \tag{65}$$

$$\Psi(1) = -\gamma \tag{66}$$

$$\Psi(n+1) = \sum_{k=1}^{n} \frac{1}{k} - \gamma \tag{67}$$

$$\Psi(1,x) = \frac{\Psi'(x)}{\Psi(x)} \tag{68}$$

$$\Psi(1,1) = \frac{1}{6}\pi^2. \tag{69}$$

B Graphics for the distribution of permutations

According to the number of partitions a sequence (a_n) can be created, where a_n stands for the number of irreducible representations of the symmetric group S_n, when S_n is divided into equivalent classes. That is

$$\pi \sim \vartheta \Leftrightarrow \pi, \vartheta \text{ have the same cycle structure.}$$

Since (a_n) quickly leads to large values the comparison mentioned in section 4.1 is only done for permutations of length 10. Consider that this corresponds to 42 different cycle structures.

10	0	0	0	0	0	0	0	0	0		8	1	0	0	0	0	0	0	0	0
6	2	0	0	0	0	0	0	0	0		7	0	1	0	0	0	0	0	0	0
4	3	0	0	0	0	0	0	0	0		5	1	1	0	0	0	0	0	0	0
2	4	0	0	0	0	0	0	0	0		3	2	1	0	0	0	0	0	0	0
0	5	0	0	0	0	0	0	0	0		1	3	1	0	0	0	0	0	0	0
6	0	0	1	0	0	0	0	0	0		4	0	2	0	0	0	0	0	0	0
4	1	0	1	0	0	0	0	0	0		2	1	2	0	0	0	0	0	0	0
2	2	0	1	0	0	0	0	0	0		0	2	2	0	0	0	0	0	0	0
0	3	0	1	0	0	0	0	0	0		3	0	1	1	0	0	0	0	0	0
1	0	3	0	0	0	0	0	0	0		1	1	1	1	0	0	0	0	0	0
0	0	2	1	0	0	0	0	0	0		5	0	0	0	1	0	0	0	0	0
3	1	0	0	1	0	0	0	0	0		1	2	0	0	1	0	0	0	0	0
2	0	0	2	0	0	0	0	0	0		2	0	1	0	1	0	0	0	0	0
0	1	0	2	0	0	0	0	0	0		0	1	1	0	1	0	0	0	0	0
1	0	0	1	1	0	0	0	0	0		4	0	0	0	0	1	0	0	0	0
2	1	0	0	0	1	0	0	0	0		0	2	0	0	0	1	0	0	0	0
1	0	1	0	0	1	0	0	0	0		0	0	0	0	2	0	0	0	0	0
0	0	0	1	0	1	0	0	0	0		3	0	0	0	0	0	1	0	0	0
1	1	0	0	0	0	1	0	0	0		0	0	1	0	0	0	1	0	0	0
2	0	0	0	0	0	0	1	0	0		0	1	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	1

Figure 23: Complete cycle structure for permutations of length equal to 10

The problem with large values is that the probability on a single structure will mainly be equal to 0, where as depending on the choice of θ a few structures will have a much greater point mass. That is why a presentation for larger n is not as well-arranged as the following one.

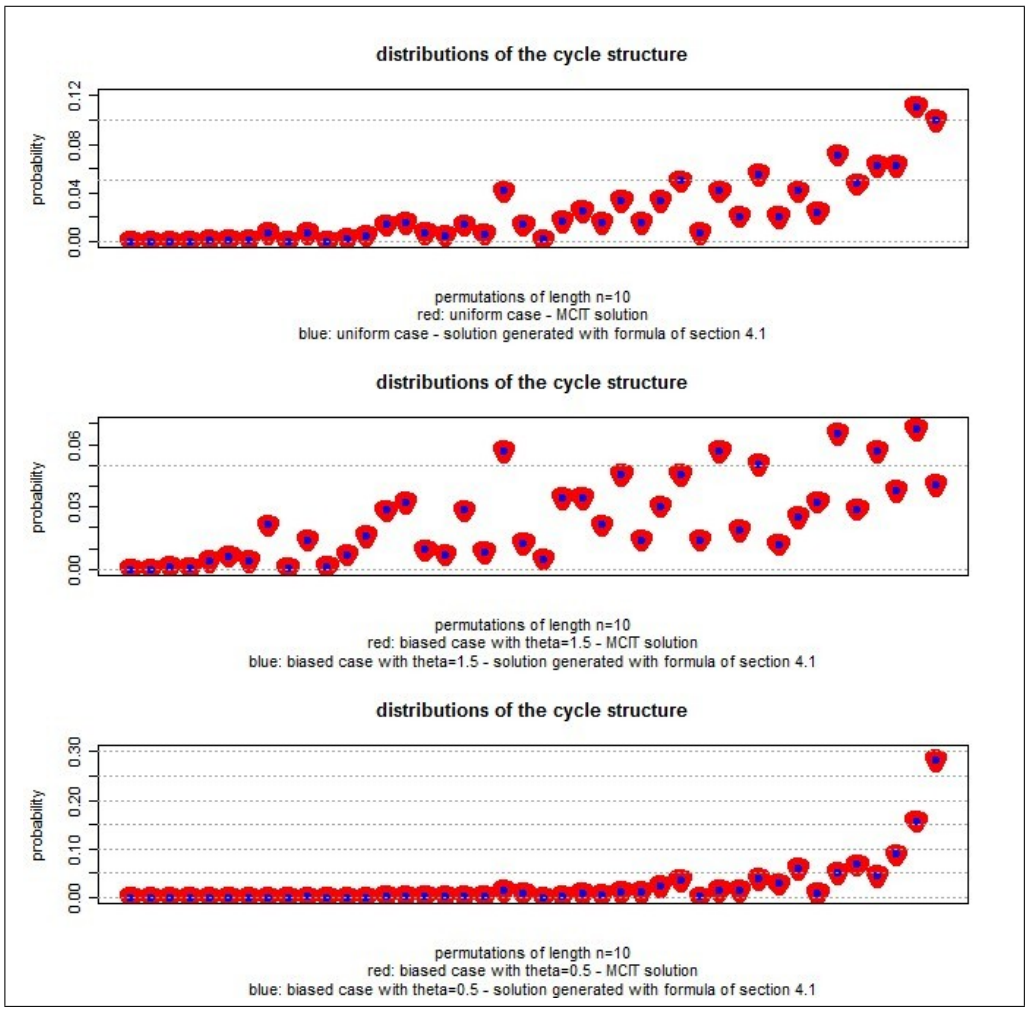


Figure 24: Comparison of MCIT - and analytic solution for the cycle structure

It can be seen that both methods lead exactly to the same results.

C Matlab - code for MCIT generated distributions

The following functions Zyklen10 and Zyklen11 compute in order of appearing the distribution of uniformly ($\theta = 1$) and biased ($\theta \neq 1$) distributed permutations of order n regarding their cycle structure. Functions is here used in the sense of an input file for Matlab and not mathematically. The functions use Markov chain imbedding technique. The Markov chain of interest is Y_t=[K_1,...,K_t] where K_i gives the number of cycles of length i. The final result are distributions of different characteristics such as

- the complete structure (X)
- the longest cycle (Y)
- the number of cycles (Y2)
- the shortest cycle (Y3,v=1)
- the length of a random taken cycle (Y4)
- the length of the fifth shortest cycle (Y5,v2=5)
- the length of the fifth longest cycle (Y6,v3=5)

For construction of function pbar see F. Ruskey [3]. p_bar(t,t) gives the total number of different permutations of order t regarding the cycle notation. p_bar(t,t-i) gives the number of order t regarding the cycle notation with cycles of length less or equal to t-i. The upper triangle matrix can be neglected. For its purpose here it could be substituted by loading values of A000041 from the Internet.

For construction of function nextPart see S. Grusea [10]. nextPart([k_1,...,k_t,t]) gives the next cycle realization in lexicographic order. It does not generate all permutations up to order t, but the ones of order t. It is exactly the Matlab counterpart to an algorithm implemented by ourselves in Maple.

```
function Zyklen10 = f(n,v,v2,v3);

p=pbar(n);
X=zeros(1,p(n,n));
X(1,1)=1;
K=[1];
for t=1:n-1
        X2=zeros(1,p(n,n));
        K=[t,zeros(1,t-1)];
        j=1;
        % first case
        K2=[K,0];
        K2(1)=K2(1)+1;
        Liste=K2;
        l=1;
        X2(1,l)=X(1,j)*1/(t+1);
        % second case
        K2=[K,0];
        K2(1)=K2(1)-1;
        K2(2)=K2(2)+1;
        Liste=[Liste;K2];
        l=2;
        X2(1,l)=X(1,j)*1*K(1)/(t+1);
        m=2;
        phead=pbar(t);
        for a=1:phead(t,t)-1
        Kaux=K;
        K=nextPart(Kaux,t);
        for g=1:p(t,t)
            if Liste2(g,:)==K
                j=g;
            end
        end
        %first case
        h=0;
        K2=[K,0];
        K2(1)=K2(1)+1;
        for i=1:m
            if Liste(i,:)==K2
              l=i;
              h=1;
            end
        end
        if h==0
            l=m+1;
            m=m+1;
            Liste=[Liste;K2];
        end
        X2(1,l)=X2(1,l)+X(1,j)*1/(t+1);
        %second case
        h=0;
        for k=1:t
            K2=[K,0];
            if K2(k)>0
                K2(k) = K2(k)-1;
                K2(k+1)=K2(k+1)+1;
                h=0;
                for i=1:m
```

```
                    if Liste(i,:)==K2
                        l=i;
                        h=1;
                    end
                end
                if h==0
                    l=m+1;
                    m=m+1;
                    Liste=[Liste;K2];
                end
                X2(1,l)=X2(1,l)+X(1,j)*k*K(k)/(t+1);
            end
        end
        end
X=X2;
Liste2=Liste;
end
Y=zeros(1,n);
for j=1:p(n,n)
    K=Liste(j,:);
    i=0;l=n;
    while K(n-i)==0
        i=i+1;
        l=n-i;
    end
    Y(1,l)=Y(1,l)+X(1,j);
end
Y2=zeros(1,n);
for j=1:p(n,n)
    K=Liste(j,:);
    l=sum(K);
    Y2(1,l)=Y2(1,l)+X(1,j);
end
Y3=zeros(1,n+1);
for j=1:p(n,n)
    w=0;
    K=zeros(1,n);
    for i =1:n
        w=w+Liste(j,i);
        K(1,i)=w;
    end
    if K(n) < v
        l=0;
    end
    i=0;
    while ((n-i>0) && (K(n-i) >= v))
        i=i+1;
        l=n-i+1;
    end
    Y3(1,l+1)=Y3(1,l+1)+X(1,j);
end
Y4=zeros(1,n);
for j=1:p(n,n)
    K=Liste(j,:);
    suu=sum(K);
    for i=1:n
        if Liste2(j,i) > 0
            Liste2(j,i)=X(1,j)*Liste(j,i)/suu;
```

```
        end
    end
end
for k=1:n
    L=Liste2(:,k);
    l=sum(L);
    Y4(1,k)=l;
end
Y5=zeros(1,n+1);
for j=1:p(n,n)
    w=0;
    K=zeros(1,n);
    for i =1:n
        w=w+Liste(j,i);
        K(1,i)=w;
    end
    if K(n) < v2
        l=0;
    end
    i=0;
    while ((n-i>0) && (K(n-i) >= v2))
        i=i+1;
        l=n-i+1;
    end
    Y5(1,l+1)=Y5(1,l+1)+X(1,j);
end
Y6=zeros(1,n+1);
for j=1:p(n,n)
    w=0;
    K=zeros(1,n);
    for i =n:-1:1
        w=w+Liste(j,i);
        K(1,i)=w;
    end
    i=0;
    while ((n-i>0)&&(K(n-i)<v3))
        i=i+1;
        l=n-i;
    end
    if K(1) < v3
        l=0;
    end
    Y6(1,l+1)=Y6(1,l+1)+X(1,j);
end

X,Y,Y2,Y3,Y4,Y5,Y6
```

```
function Zyklen11 = f(n,teta,v,v2,v3);

p=pbar(n);
X=zeros(1,p(n,n));
X(1,1)=1;
K=[1];
for t=1:n-1
        X2=zeros(1,p(n,n));
        K=[t,zeros(1,t-1)];
        j=1;
        % first case
        K2=[K,0];
        K2(1)=K2(1)+1;
        Liste=K2;
        l=1;
        X2(1,l)=X(1,j)*teta/(t+teta);
        % second case
        K2=[K,0];
        %K3=[K,0];
        K2(1)=K2(1)-1;
        K2(2)=K2(2)+1;
        %for z=1:t+1
            %K3(z)=K2(z)*z;
        %end;
        %if sum(K3)==t+1
        Liste=[Liste;K2];
        l=2;
        X2(1,l)=X(1,j)*1*K(1)/(t+teta);
        %end;
        m=2;
        phead=pbar(t);
        for a=1:phead(t,t)-1
        Kaux=K;
        K=nextPart(Kaux,t);
        for g=1:p(t,t)
            if Liste2(g,:)==K
                j=g;
            end
        end
        %first case
        h=0;
        K2=[K,0];
        K2(1)=K2(1)+1;
        for i=1:m
            if Liste(i,:)==K2
              l=i;
              h=1;
            end
        end
        if h==0
            l=m+1;
            m=m+1;
            Liste=[Liste;K2];
        end
        X2(1,l)=X2(1,l)+X(1,j)*teta/(t+teta);
        %second case
        h=0;
        for k=1:t
```

```
            K2=[K,0];
            if K2(k)>0
                K2(k) = K2(k)-1;
                K2(k+1)=K2(k+1)+1;
                %for z=1:t+1
                    %K3(z)=K2(z)*z;
                %end
                %if sum(K3)==t+1
                h=0;
                for i=1:m
                    if Liste(i,:)==K2
                        l=i;
                        h=1;
                    end
                end
                if h==0
                    l=m+1;
                    m=m+1;
                    Liste=[Liste;K2];
                end
                X2(1,l)=X2(1,l)+X(1,j)*k*K(k)/(t+teta);
                %end
            end
        end
        end
X=X2;
Liste2=Liste;
end
Y=zeros(1,n);
for j=1:p(n,n)
    K=Liste(j,:);
    i=0;l=n;
    while K(n-i)==0
        i=i+1;
        l=n-i;
    end
    Y(1,l)=Y(1,l)+X(1,j);
end
Y2=zeros(1,n);
for j=1:p(n,n)
    K=Liste(j,:);
    l=sum(K);
    Y2(1,l)=Y2(1,l)+X(1,j);
end
Y3=zeros(1,n+1);
for j=1:p(n,n)
    w=0;
    K=zeros(1,n);
    for i =1:n
        w=w+Liste(j,i);
        K(1,i)=w;
    end
    if K(n) < v
        l=0;
    end
    i=0;
    while ((n-i>0) && (K(n-i) >= v))
        i=i+1;
```

```
        l=n-i+1;
    end
    Y3(1,l+1)=Y3(1,l+1)+X(1,j);
end
Y4=zeros(1,n);
for j=1:p(n,n)
    K=Liste(j,:);
    suu=sum(K);
    for i=1:n
        if Liste2(j,i) > 0
            Liste2(j,i)=X(1,j)*Liste(j,i)/suu;
        end
    end
end
for k=1:n
    L=Liste2(:,k);
    l=sum(L);
    Y4(1,k)=l;
end
Y5=zeros(1,n+1);
for j=1:p(n,n)
    w=0;
    K=zeros(1,n);
    for i =1:n
        w=w+Liste(j,i);
        K(1,i)=w;
    end
    if K(n) < v2
        l=0;
    end
    i=0;
    while ((n-i>0) && (K(n-i) >= v2))
        i=i+1;
        l=n-i+1;
    end
    Y5(1,l+1)=Y5(1,l+1)+X(1,j);
end
Y6=zeros(1,n+1);
for j=1:p(n,n)
    w=0;
    K=zeros(1,n);
    for i =n:-1:1
        w=w+Liste(j,i);
        K(1,i)=w;
    end
    i=0;
    while ((n-i>0)&&(K(n-i)<v3))
        i=i+1;
        l=n-i;
    end
    if K(1) < v3
        l=0;
    end
    Y6(1,l+1)=Y6(1,l+1)+X(1,j);
end
X,Y,Y2,Y3,Y4,Y5,Y6
```

D Maple - code for the number of non overlapping k successes

```
> with(linalg):
> for z from 1 to 5 do
> p[z]:=1/(z+1);
> od;for z from 1 to 5 do
> q[z]:=1-p[z];
> od;
> mat[0]:=matrix([[1,0,0,0,0,0]]);
> for z from 1 to 5 do
>
mat[z]:=matrix([[q[z],p[z],0,0,0,0],[q[z],0,p[z],0,0,0],[0,0,q[z]
,p[z],0,0],[0,0,q[z],0,p[z],0],[0,0,0,0,q[z],p[z]],[0,0,0,0,0,1]]
);
> od;
```

$$p_1 := \frac{1}{2}$$
$$p_2 := \frac{1}{3}$$
$$p_3 := \frac{1}{4}$$
$$p_4 := \frac{1}{5}$$
$$p_5 := \frac{1}{6}$$
$$q_1 := \frac{1}{2}$$
$$q_2 := \frac{2}{3}$$
$$q_3 := \frac{3}{4}$$
$$q_4 := \frac{4}{5}$$
$$q_5 := \frac{5}{6}$$
$$mat_0 := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$mat_1 := \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$mat_2 := \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$
$$mat_3 := \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$
$$mat_4 := \begin{bmatrix} \frac{4}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 \\ \frac{4}{5} & 0 & \frac{1}{5} & 0 & 0 & 0 \\ 0 & 0 & \frac{4}{5} & \frac{1}{5} & 0 & 0 \\ 0 & 0 & \frac{4}{5} & 0 & \frac{1}{5} & 0 \\ 0 & 0 & 0 & 0 & \frac{4}{5} & \frac{1}{5} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$
$$mat_5 := \begin{bmatrix} \frac{5}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 \\ \frac{5}{6} & 0 & \frac{1}{6} & 0 & 0 & 0 \\ 0 & 0 & \frac{5}{6} & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{5}{6} & 0 & \frac{1}{6} & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

```
mat[7]:=matrix([[1],[1],[0],[0],[0],[0]]);
```

$$mat_7 := \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

```
> mat[8]:=matrix([[0],[0],[1],[1],[0],[0]]);
```

$$mat_8 := \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

```
> mat[9]:=matrix([[0],[0],[0],[0],[1],[1]]);
```

$$mat_9 := \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

```
> evalm(mat[1]&*mat[2]&*mat[3]&*mat[4]&*mat[5]);
```

$$\begin{bmatrix} \frac{91}{144} & \frac{19}{180} & \frac{31}{144} & \frac{1}{30} & \frac{1}{80} & \frac{1}{720} \\ \frac{53}{144} & \frac{11}{180} & \frac{103}{240} & \frac{5}{72} & \frac{1}{18} & \frac{1}{60} \\ 0 & 0 & \frac{91}{144} & \frac{19}{180} & \frac{23}{144} & \frac{37}{360} \\ 0 & 0 & \frac{53}{144} & \frac{11}{180} & \frac{53}{240} & \frac{7}{20} \\ 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{5}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

```
> mat[10]:=mat[1]&*mat[2]&*mat[3]&*mat[4]&*mat[5]:
>
evalm(mat[0]&*mat[10]&*mat[7]);evalm(mat[0]&*mat[10]&*mat[8]);eva
lm(mat[0]&*mat[10]&*mat[9]);
```

$$\begin{bmatrix} \frac{59}{80} \end{bmatrix}$$

$$\begin{bmatrix} \frac{179}{720} \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{72} \end{bmatrix}$$

```
> evalf(%%%);evalf(%%%);evalf(%%%);
```

$$\begin{bmatrix} 0.7375000000 \end{bmatrix}$$

$$\begin{bmatrix} 0.2486111111 \end{bmatrix}$$

$$\begin{bmatrix} 0.01388888889 \end{bmatrix}$$

```
> ### matrix correction because of non reachable states ###
> mat[4°]:=matrix([[4/5, 1/5, 0, 0, 0, 0], [4/5, 0, 1/5, 0, 0,
0], [0, 0, 4/5, 1/5, 0, 0], [0, 0, 4/5, 0, 1/5, 0], [0, 0, 0, 0,
1, 0], [0, 0, 0, 0, 0, 0]]);
```

$$mat_{4^\circ} := \begin{bmatrix} \frac{4}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 \\ \frac{4}{5} & 0 & \frac{1}{5} & 0 & 0 & 0 \\ 0 & 0 & \frac{4}{5} & \frac{1}{5} & 0 & 0 \\ 0 & 0 & \frac{4}{5} & 0 & \frac{1}{5} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

```
> mat[3°]:=matrix([[3/4, 1/4, 0, 0, 0, 0], [3/4, 0, 1/4, 0, 0,
0], [0, 0, 3/4, 1/4, 0, 0], [0, 0, 0, 1, 0, 0], [0, 0, 0, 0, 0,
0], [0, 0, 0, 0, 0, 0]]);
```

$$mat_{3^\circ} := \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

```
> mat[2°]:= matrix([[2/3, 1/3, 0, 0, 0, 0], [2/3, 0, 1/3, 0, 0,
0], [0, 0, 1, 0, 0, 0], [0, 0, 0, 0, 0, 0], [0, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 0]]);
```

$$mat_{2^\circ} := \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

```
> mat[1°]:=matrix([[1/2, 1/2, 0, 0, 0, 0], [0, 1, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 0], [0, 0, 0, 0, 0, 0], [0, 0, 0, 0, 0, 0], [0,
0, 0, 0, 0, 0]]);
```

$$mat_{1^\circ} := \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

```
> evalm(mat[1°]&*mat[2°]&*mat[3°]&*mat[4°]&*mat[5]);
```

$$\begin{bmatrix} \frac{91}{144} & \frac{19}{180} & \frac{31}{144} & \frac{1}{30} & \frac{1}{80} & \frac{1}{720} \\ \frac{19}{36} & \frac{4}{45} & \frac{37}{120} & \frac{1}{20} & \frac{1}{45} & \frac{1}{360} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

```
> mat[10°]:=(mat[1°]&*mat[2°]&*mat[3°]&*mat[4°]&*mat[5]):
>
evalm(mat[0]&*mat[10°]&*mat[7]);evalm(mat[0]&*mat[10°]&*mat[8]);e
valm(mat[0]&*mat[10°]&*mat[9]);
```

$$\begin{bmatrix} \frac{59}{80} \end{bmatrix}$$

$$\begin{bmatrix} \frac{179}{720} \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{72} \end{bmatrix}$$

```
> with(linalg):
> for z from 1 to 7 do
> p[z]:=1/(z+1):
> od:for z from 1 to 7 do
> q[z]:=1-p[z]:od:
> mat[0]:=matrix([[1,0,0,0,0,0,0,0]]):
> for z from 1 to 7 do
>
mat[z]:=matrix([[q[z],p[z],0,0,0,0,0,0],[q[z],0,p[z],0,0,0,0,
0],[q[z],0,0,p[z],0,0,0,0],[0,0,0,q[z],p[z],0,0,0],[0,0,0,q[z
],0,p[z],0,0,0],[0,0,0,q[z],0,0,p[z],0,0],[0,0,0,0,0,0,q[z],p[z],
0],[0,0,0,0,0,0,q[z],0,p[z]],[0,0,0,0,0,0,0,1]]):od:
> mat[8]:=matrix([[1],[1],[1],[0],[0],[0],[0],[0],[0]]):
> mat[9]:=matrix([[0],[0],[0],[1],[1],[1],[0],[0],[0]]):
> mat[10]:=matrix([[0],[0],[0],[0],[0],[0],[1],[1],[1]]):
>
mat[11]:=mat[1]&*mat[2]&*mat[3]&*mat[4]&*mat[5]&*mat[6]&*mat[7]:
>
evalm(mat[0]&*mat[11]&*mat[8]):evalm(mat[0]&*mat[11]&*mat[9]):eva
lm(mat[0]&*mat[11]&*mat[10]):evalf(%%%);evalf(%%%);evalf(%%%);
```

$$\begin{bmatrix} 0.9385912698 \end{bmatrix}$$

$$\begin{bmatrix} 0.06108630952 \end{bmatrix}$$

$$\begin{bmatrix} 0.0003224206349 \end{bmatrix}$$

E Mathematica - code for the Dickman - function

The following code by Francois Grieu [9] generates numerical values for Dickman's function. The algorithm is mainly based on step by step power series expansion of (56).

```
(Debug) In[1]:= Clear[DickmanF, Dickman$f, Dickm$pr, Dickm$ex, Dickm$dv, Dickm$ma];
DickmanF::usage = "DickmanF[x] is the proportion of integers N with \
factors below N^x";
Dickman$f::usage = "Dickman$f[y] is the proportion of integers N with \
factors below N^(1/y)"; Dickm$pr = 40 (*desired ABSOLUTE (rather than \
relative) accuracy*); Dickm$ex =
 Ceiling[Dickm$pr*Log[3, 10]] + 1 (*order of expansions*); Dickm$ma =
 Ceiling[Dickm$ex/
   Log[Dickm$ex]] (*domain limit*);(*Dickm$dv[n] approximates \
Dickman$f[t+n-1/2] for-1/2<=t<=1/2*)Dickm$dv[1] = N[1, Dickm$pr + 1];
Dickm$dv[n_] :=
 Dickm$dv[n] =
  Module[{u = Dickm$dv[n - 1], v},
   v = Integrate[(u + O[t]^Ceiling[Dickm$ex - n])/(t + n - 1/2),
     t]; (Normal[u] /. t -> 1/2) + (Normal[v] /. t -> -1/2) - v];
Dickman$f[y_?NumberQ] :=
 Which[y <= 0, 1, y > Dickm$ma, N[1, y Log[10., y]] - 1, True,
  Normal[Dickm$dv[Ceiling[y]]] /. t -> y - Ceiling[y] + 1/2];
DickmanF[x_?NumberQ] = If[x <= 0, 0, Dickman$f[1/x]];
(Debug) In[2]:= For[i = 1, i <= 50, i++, Print[DickmanF[(i/50)^(1.0)]]]
```

The values have been checked and the code has been slightly modified. Then it was applied for the results of section 4. Note that the results for the uniform case are pretty good, where as for the biased case a major simplification was accepted. Target of that was to obtain θ-dependent terms that minimize the gap to the explicit values. That can be achieved by taking the θ-root of Dickman's function's input. This was checked for $\theta = 0.5$ and for $\theta = 1.5$. Nevertheless it clearly should be said that the best would be to find an appropriate numerical solution for (49).

References

[1] Arratia, R., Barbour A.D., Tavaré S. (2003). Logarithmic Combinatorial Structures: A probalistic Approach. European Mathematical Society.

[2] Balakrishnan, N., Koutras M.V. (2002). Runs and Scans with Applications. John Wiley and Sons, Inc.

[3] Effler S., Ruskey F. (2003). A CAT Algorithm for listing permutations with a given number of inversions. Information Processing Letters 86/2. p. 107 - 112.

[4] Fu, J.C., Koutras, M.V. (1994). Distribution Theory of Runs: A Markov Chain Approach. Journal of the American Statistical Association, Vol. 89, No. 427, pp. 1050 - 1058.

[5] Fu, J.C., Spiring, F.A. (2002). On the average run lengths of quality control schemes using Markov chain approach. Statistics and Probability Letters 56, pp. 369 - 380.

[6] Fu, J.C., Chang Y.M. (2002). On Probability Generating Functions for Waiting Time Distributions of Compound Patterns in a Sequence of Multistate Trials. Journal of Applied Probability, Vol. 39, No. 1, pp. 70 - 80.

[7] Fu, J.C, Wendy Lou (2003). Distribution theory of runs and patterns and its applications. A finite Markov chain Imbedding approach. World Scientific Publishing Co. Pte. Ltd.

[8] Goldschmidt, C., Martin J. (2005). Random recursive trees and the Bolthausen-Sznitman coalescent. arXiv:math/0502263v2[mathPR].

[9] Grieu, F. (1999). Computing Dickman's function. viewed 29 September 2009. http://forums.wolfram.com/mathgroup/archive/1999/Dec/msg00377.html.

[10] Grusea, S. (2008). Applications du calcul des probabilités à la recherche de régions genomiques conservées. Dissertation.

[11] Holst, L. (2001). The Poisson - Dirichlet distribution and its relatives revisited. Paper for the Royal Institute of Technology.

[12] Jordan, C. (1965). Calculus of finite differences. Chelsea Publishing Company.

[13] Kemp, R. (1984). Fundamentals of the average case analysis of particular algorithms. Wiley-Teubner Series in Computer Science.

[14] Kingman, J.F.C. (1980). Poisson Processes. Oxford University Press.

[15] Klenke, A. (2008). Wahrscheinlichkeitstheorie. Springer Verlag.

[16] Koutras, M.V. (1997). Waiting Time Distributions Associated With Runs Of Fixed Length In Two State Markov Chains. Ann. Inst. Statist. Math., Vol. 49, No.1, pp. 123 - 139.

[17] Lou Wendy, W.Y. (1996). On Runs and Longest Run Tests: A Method of Finite Markov Chain Imbedding. Journal of the American Statistical Association, Vol. 91, No. 431, pp. 1595 - 1601.

[18] Meintrup,D., Schäffler, S. (2005). Stochastik: Theorie und Anwendungen. Springer Verlag.

[19] Meir, A., Moon J. W. (1974). Cutting down recursive trees. Math. Bioscience. 21: 173-181.

[20] Robin S., Daudin, J.J. (2001). Exact Distribution Of The Distances Between Any Occurrences Of A Set Of Words. Ann. Inst. Statist. Math., Vol. 53, No.4, pp. 895 - 905

[21] Scheffé, H. (1947). A useful convergence theorem for probability distributions. Annals of Mathematical Statistics. 18: 434-438.

[22] Shepp, L.A., Lloyd S.P. (1966). Ordered Cycle Lengths In A Random Permutation. Transactions of the American Mathematical Society.

List of used software

The following software packages are used in this thesis

1. LaTeX for creating this document.

2. Microsoft Office 2007 for intermediate steps.

3. R 2.8.1 for creating the graphics.

4. Matlab R2007b as a surface for the MCIT algorithm.

5. Maple 11 for intermediate steps and computation within a binomial trial.

6. Mathematica 6 for computation of Dickman function.