*Discrete Mathematics and Theoretical Computer Science* DMTCS vol. **18**:3, 2016, #11



# *Variance and Covariance of Several Simultaneous Outputs of a Markov Chain*

Sara Kropf

*Institut für Mathematik, Alpen-Adria-Universität Klagenfurt*

*received 1st Dec. 2015, accepted 17th June 2016.*

The partial sum of the states of a Markov chain or more generally a Markov source is asymptotically normally distributed under suitable conditions. One of these conditions is that the variance is unbounded. A simple combinatorial characterization of Markov sources which satisfy this condition is given in terms of cycles of the underlying graph of the Markov chain. Also Markov sources with higher dimensional alphabets are considered.

Furthermore, the case of an unbounded covariance between two coordinates of the Markov source is combinatorically characterized. If the covariance is bounded, then the two coordinates are asymptotically independent.

The results are illustrated by several examples, like the number of specific blocks in 0-1-sequences and the Hamming weight of the width-$w$ non-adjacent form.

**Keywords:** Markov source, variance, covariance, independence, Hamming weight, Matrix-Tree Theorem, transducer, central limit theorem

## 1 Introduction

We investigate the random vector defined as the $n$-th partial sum of a Markov source over a higher dimensional alphabet. Under suitable conditions, this random variable is asymptotically jointly normally distributed. Its mean and variance-covariance matrix is linear in the number of summands (cf. [6, Theorem 2.22]). On the one hand, these conditions include irreducibility and aperiodicity of the underlying graph of the Markov chain, which can be checked easily for a given Markov chain. On the other hand, we also have to check that the variance-covariance matrix is regular, which requires technical computations. In this article, we give a simple combinatorial characterization of Markov sources whose corresponding variance-covariance matrix is singular.

The covariance between two coordinates of this random vector is also of interest: If it is bounded, then these two coordinates are asymptotically independent because of the joint normal distribution. We give a combinatorial characterization of this case.

These characterizations are given in terms of subgraphs of the underlying graph of the Markov chain: For the variance-covariance matrix, we only have to consider all cycles. A regular variance-covariance

The author is supported by the Austrian Science Fund (FWF): P 24644-N26.
Email-address: sara.kropf@aau.at

ISSN 1365–8050 

matrix will be proven to be equivalent to the linear independence of certain functions of cycles of the underlying graph of the Markov chain. For the characterization of an unbounded covariance, we have to consider functional digraphs. This result is proven using an extension of the Matrix-Tree Theorem in [5, 20].

As Markov sources are closely related to automata and transducers, our results can also be used for the asymptotic analysis of sequences which can be computed by transducers. This includes the Hamming weight of many syntactically defined digit expansions as performed in [11, 16, 15, 13, 14]. Furthermore, occurrences of digits or subwords can also be computed by transducers. Their variance (and covariance) is analyzed in [12, 2, 19, 3, 22, 8, 10].

In [18], the variance of the output of a transducer as well as the covariance between the input and the output were analyzed. In this article, we consider the more general setting of Markov chains. The proofs are similar as those in [18], but the results are valid in a broader context and can be formulated more clearly. In contrast to [18], we allow the input sequence of the transducer to be generated by a Markov source. This allows us to model an input sequence for a transducer whose letters do not occur with equal probabilities and/or have dependencies between the letters. The precise relation between the setting of this article and that of [18] is given in Section 3.

As an example, we prove that the Hamming weight of the so-called width-$w$ non-adjacent form is asymptotically jointly normally distributed for two different values of $w \geq 2$. The width-$w$ non-adjacent form is a binary digit expansion with digits in $\{0, \pm 1, \pm 3, \dots, \pm(2^{w-1} - 1)\}$ and the syntactical rule that at most one of any $w$ adjacent digits is non-zero. This digit expansion exists and is unique for every integer (cf. [21, 1]). Furthermore, it has minimal Hamming weight among all digit expansions with this base and digit set.

The outline of this article is as follows: In Section 2, we define our setting and the types of graphs we use to state the combinatorial characterization of independent output sums and singular variance-covariance matrices. These characterizations are given in Section 3 and examples are given in Section 4. In Section 5, we finally prove the results of Section 3.

## 2 Preliminaries

In this article, a *finite Markov chain* consists of a finite state space $\{1, \dots, M\}$, a finite set of transitions $\mathcal{E}$ between the states, each with a positive transition probability, and a unique[(i)] initial state $1$. We denote the transition probability for a transition $e$ by $p_e$. Then we have

$$\sum_{\substack{e \in \mathcal{E} \\ e \text{ starts in } i}} p_e = 1$$

for all states $i$. Note that for all transitions $e \in \mathcal{E}$, we require $p_e > 0$. Further note that there may be multiple transitions between two states but always only a finite number of them. This may be useful for different outputs later on.

The transition probabilities induce a probability distribution on the paths of length $n$ starting in the initial state $1$. Let $X_n$ be a random path of length $n$ according to this model.

[(i)] This is no restriction as we can always add an additional state and the transitions starting in this state with probabilities corresponding to the non-degenerate initial distribution. The output functions are then extended by mapping these transitions to $0$.

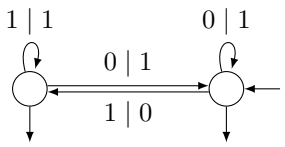



Fig. 1: A small example of a transducer.

All states of the underlying digraph of the Markov chain are assumed to be accessible from the initial state. Contracting each strongly connected component of the underlying digraph gives an acyclic digraph, the so-called condensation. We assume that this condensation has only one leaf (i.e., one vertex with out-degree $0$). The strongly connected component corresponding to this leaf is called *final component*. We assume that the period (i.e., the greatest common divisor of the lengths of all cycles) of this final component is $1$. We call such Markov chains *finally connected* and *finally aperiodic*.

Additionally we use *output functions* $k\colon \mathcal{E} \to \mathbb{R}$. The corresponding random variable $K_n$ is the sum of all values of $k$ along a random path $X_n$. We call $K_n$ the *output sum* of the Markov chain with respect to $k$. We use several output functions $k_1, \ldots, k_m$ and the corresponding random variables $K_n^{(1)}, \ldots, K_n^{(m)}$ simultaneously for one Markov chain.

*Remark* 2.1. Usually, one is interested in a function evaluated at the sequence of random states of the Markov chain. This is equivalent to this setting with an output function of the transitions: For the one direction, the restriction of the output function to the outgoing transitions of one state is constant for every state. For the other direction, we use the standard construction of the Markov chain with state space $\{(i,j) \mid 1 \leq i,j \leq M\}$.

Thus, our setting can be seen as a Markov source with a finite set of $m$-dimensional vectors as alphabet.

We are interested in the joint distribution of the random variables $K_n^{(1)}, \ldots, K_n^{(m)}$. For one coordinate, we will prove that the expected value of $K_n^{(i)}$ is $e_i n + \mathcal{O}(1)$ for constants $e_i$. The variance-covariance matrix of $K_n^{(1)}, \ldots, K_n^{(m)}$ will turn out to be $\Sigma n + \mathcal{O}(1)$ for a matrix $\Sigma$. We call $\Sigma$ the asymptotic variance-covariance matrix and its entries the asymptotic variances and covariances.

We will combinatorically characterize Markov chains with output functions such that the variance-covariance matrix is regular. Furthermore, we give a combinatorial characterization of the case that the asymptotic covariance is zero. As this is only influenced by two output functions, we restrict ourselves to $K_n^{(1)}$ and $K_n^{(2)}$ in this case.

*Remark* 2.2. Markov chains with output functions are closely related to transducers with a probability distribution for the input: A transducer is defined to consist of a finite set of states, an initial state, a set of final states, an input alphabet, an output alphabet and a finite set of transitions, where a transition starts in one state, leads to another state and has an input and an output label from the corresponding alphabets. See [4, Chapter 1] for a more formal definition. An example of a transducer is given in Figure 1. We label the transitions with "input label | output label". The initial state is marked by an ingoing arrow starting at no other state and the final states are marked by outgoing arrows leading to no other state.

A Markov chain with one output function can be obtained by a transducer with additional probability distributions for the outgoing transitions of each state and by deleting the input labels of the transducer.

If we have two transducers where only the outputs of the transitions are different, we can choose probability distributions for the outgoing transitions of each state. Then we obtain a Markov chain with

two output functions. Thus, we can use our results for two output functions (see Examples 4.2 and 4.3).

*Remark* 2.3. We can additionally have *final output functions* $f\colon \{1,\dots,M\} \to \mathbb{R}$ for each output function $k$ and redefine the random variable $K_n$ as the sum of the values of the output function $k$ along a random path $X_n$ plus the final output $f$ of the final state of this path. We will see that this does not change the main terms of the asymptotic behavior. Thus, the results in Section 3 are still valid (see also Remark 5.5).

*Remark* 2.4. The Parry measure are probabilities $p_e$ such that every path of length $n$ has the same weight up to a constant factor (cf. [24, 23]). If we are interested in probabilities such that every path of length $n$ starting in the initial state $1$ has exactly the same weight, we have to use the Parry measure with additional *exit weights*: Each path is additionally weighted by these exit weights according to the final state of the path (cf. [17, Lemma 4.1]).

However, the sum of the weights of all paths of length $n$ is no longer normalized: It differs from $1$ by an exponentially small error term for $n \to \infty$. This gives an approximate equidistribution of all paths of length $n$. As we are interested in the asymptotic behavior for $n \to \infty$, the expected value and the variance of the corresponding measurable function $K_n$ can still be defined as usual.

If we use these exit weights $w_s$ in our setting, the main terms of the asymptotic behavior are not changed. Thus, the theorems in Section 3 are still valid (see also Remark 5.5).

These exit weights can also be used to simulate final and non-final states of a transducer by setting the weights of non-final states to $0$. However, not all exit weights of the final component are allowed to be zero.

Next, we define some subgraphs of the underlying graph of the final component and extend the probabilities and the output functions to these subgraphs.

**Definition 2.5.** We define the following types of directed graphs as subgraphs of the final component of the Markov chain.

- A *rooted tree* is a weakly connected digraph with one vertex which has out-degree $0$, while all other vertices have out-degree $1$. The vertex with out-degree $0$ is called the *root* of the tree.
- A *functional digraph* is a digraph whose vertices have out-degree $1$. Each component of a functional digraph consists of a directed cycle and some trees rooted at vertices of the cycle. For a functional digraph $D$, let $\mathcal{C}_D$ be the set of all cycles of $D$.

The probabilities $p_e$ can be multiplicatively extended to a weight function for arbitrary subgraphs of the Markov chain: Let $D$ be any subgraph of the underlying graph of the Markov chain, then define the weight of $D$ by

$$p_D = \prod_{e \in D} p_e.$$

For a path $P$ of length $n$, this is exactly the probability $\mathbb{P}(X_n = P)$.

However, the output function $k$ is additively extended to cycles $C$ of the underlying graph of the Markov chain by

$$k(C) = \sum_{e \in C} k(e).$$

This can further be extended to functional digraphs:

**Definition 2.6.** Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be the sets of all spanning subgraphs of the final component of the Markov chain $\mathcal{M}$ which are functional digraphs and have one and two components, respectively.

For functions $g$ and $h\colon \mathcal{E} \to \mathbb{R}$, we define

$$\begin{aligned} g(\mathcal{D}_1) &= \sum_{D\in\mathcal{D}_1} p_D \sum_{C\in\mathcal{C}_D} g(C), \\ (g,h)(\mathcal{D}_1) &= \sum_{D\in\mathcal{D}_1} p_D \sum_{C\in\mathcal{C}_D} g(C)h(C), \\ (g,h)(\mathcal{D}_2) &= \sum_{D\in\mathcal{D}_2} p_D \sum_{C_1\in\mathcal{C}_D} \sum_{\substack{C_2\in\mathcal{C}_D \\ C_2\neq C_1}} g(C_1)h(C_2). \end{aligned}$$

As functions $g$ and $h$, we use the output functions $k_1, \ldots, k_m$ and the constant function $\mathbb{1}(e) = 1$.

# 3 Main Results

In this section, we present the combinatorial characterization of output functions of Markov chains which are asymptotically independent and of Markov chains with output functions with a singular variance-covariance matrix. The proofs can be found in Section 5.

If the underlying directed graph of the Markov chain is $j$-regular, every transition has probability $1/j$, we only have two output functions and the first output function $k_1\colon \mathcal{E} \to \{0, 1, \ldots, j-1\}$ is such that the restrictions of $k_1$ to the outgoing transitions of one state is bijective for every state, then these results are stated in [18] (see also Remark 2.2).

The next definition describes a sequence of random variables whose difference from its expected value is bounded for all elements.

**Definition 3.1.** The output sum $K_n$ of a Markov chain is called *quasi-deterministic* if there is a constant $a \in \mathbb{R}$ such that

$$K_n = an + \mathcal{O}(1)$$

holds for all $n$.

Next we give the combinatorial characterization of output sums with bounded variance in the case of a not necessarily independent identically distributed input sequence.

**Theorem 1.** *For a finite, finally connected and finally aperiodic Markov chain $\mathcal{M}$ with an output function $k$, the following assertions are equivalent:*

*(a) The asymptotic variance $v$ of the output sum is $0$.*

*(b) There exists a state $s$ of the final component and a constant $a \in \mathbb{R}$ such that*

$$k(C) = a\mathbb{1}(C)$$

*holds for every closed walk $C$ of the final component visiting the state $s$ exactly once.*

*(c) There exists a constant $a \in \mathbb{R}$ such that*

$$k(C) = a\mathbb{1}(C)$$

*holds for every directed cycle $C$ of the final component of $\mathcal{M}$.*

*In that case, $an + \mathcal{O}(1)$ is the expected value of the output sum and Statement (b) holds for all states $s$ of the final component.*

*If $\mathcal{M}$ is furthermore strongly connected, the following assertion is also equivalent:*

*(d) The random variable $K_n$ is quasi-deterministic with constant $a$.*

In the case that the value of the output function is 0 or 1 for each transition, there are only two trivial output functions with asymptotic variance zero.

**Corollary 3.2.** *Let $k\colon \mathcal{E} \to \{0,1\}$. Then the asymptotic variance $v$ is zero if and only if the output function $k$ is constant on the final component.*

The next theorem extends Theorem 1 to the joint distribution of several simultaneous output sums by combinatorically describing the case of a singular variance-covariance matrix.

**Theorem 2.** *Let $\mathcal{M}$ be a finite, finally connected, finally aperiodic Markov chain with $m$ output functions $k_1, \dots, k_m$. Then the variance-covariance matrix $\Sigma$ is regular if and only if the functions $\mathbb{1}, k_1, \dots, k_m$ are linearly independent as functions from the vector space of cycles of the final component to the real numbers, i.e. there do not exist real constants $a_0, \dots, a_m$, not all zero, such that*

$$a_0\mathbb{1}(C) + a_1k_1(C) + \cdots + a_mk_m(C) = 0 \tag{1}$$

*holds for all cycles (or equivalently, for all closed walks) $C$ of the final component.*

*The random variables $K_n^{(1)}, \dots, K_n^{(m)}$ are asymptotically jointly normally distributed if and only if $\Sigma$ is regular.*

*Remark* 3.3. Theorems 1 and 2 and Corollary 3.2 are independent of the choice of the probabilities of the transitions. Only the structure of the underlying graph of the Markov chain and the output functions influence the result. Note, however, that according to our general assumptions, all transitions have *positive* probability.

The next theorem gives a combinatorial characterization of output functions of a Markov chain which are asymptotically independent. As this characterization is given by the covariance, we can restrict ourselves to two output functions without loss of generality.

**Theorem 3.** *Let $\mathcal{M}$ be a finite, finally connected, finally aperiodic Markov chain with two output functions $k_1$ and $k_2$.*

*Then the random variable $K_n^{(i)}$ has the expected value $e_in + \mathcal{O}(1)$ and the variance $v_in + \mathcal{O}(1)$ where the constants are*

$$\begin{aligned} e_i &= \frac{k_i(\mathcal{D}_1)}{\mathbb{1}(\mathcal{D}_1)}, \\ v_i &= \frac{1}{\mathbb{1}(\mathcal{D}_1)}\big((k_i - e_i\mathbb{1}, k_i - e_i\mathbb{1})(\mathcal{D}_1) - (k_i - e_i\mathbb{1}, k_i - e_i\mathbb{1})(\mathcal{D}_2)\big) \end{aligned} \tag{2}$$

*for $i = 1, 2$.*

*The covariance of $K_n^{(1)}$ and $K_n^{(2)}$ is $cn + \mathcal{O}(1)$ with the constant*

$$c = \frac{1}{\mathbb{1}(\mathcal{D}_1)}\big((k_1 - e_1\mathbb{1}, k_2 - e_2\mathbb{1})(\mathcal{D}_1) - (k_1 - e_1\mathbb{1}, k_2 - e_2\mathbb{1})(\mathcal{D}_2)\big).$$

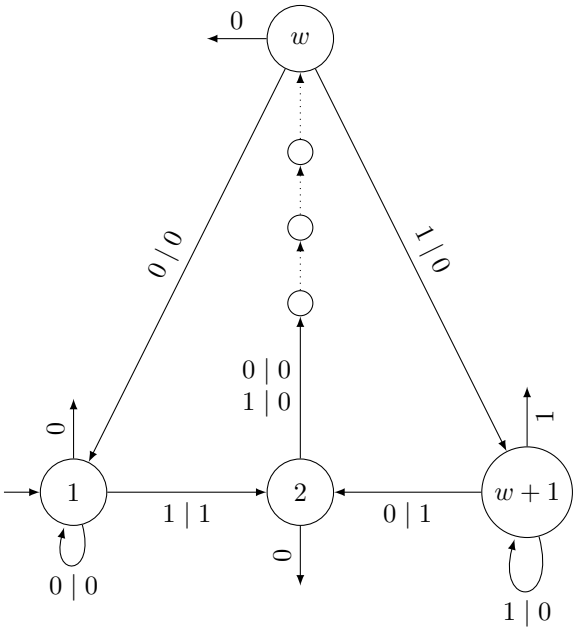



Fig. 2: Transducer $\mathcal{T}(w)$ to compute the Hamming weight of the width-$w$ non-adjacent form.

*The random variables $K_n^{(1)}$ and $K_n^{(2)}$ are asymptotically independent if and only if*

$$(k_1 - e_1\mathbb{1}, k_2 - e_2\mathbb{1})(\mathcal{D}_1) = (k_1 - e_1\mathbb{1}, k_2 - e_2\mathbb{1})(\mathcal{D}_2).$$

In the case that the expected values of $K_n^{(1)}$ and $K_n^{(2)}$ are both bounded, i.e. $e_1 = e_2 = 0$, these random variables are asymptotically independent if and only if

$$(k_1, k_2)(\mathcal{D}_1) = (k_1, k_2)(\mathcal{D}_2).$$

# 4 Examples

In this section, we first prove the asymptotic joint normal distribution of the Hamming weights of two different digit expansions by using Theorem 2. Then we investigate the independence of length 2 blocks of 0-1-sequences by using Theorem 3. In both cases we start with two transducers to construct a Markov chain with two output functions, once as a Cartesian product, once via Remark 2.2.

*Example* 4.1 (Width-$w$ non-adjacent forms). Let $2 \leq w_1 < w_2$ be integers. We consider the asymptotic joint distribution of the Hamming weight of the width-$w_1$ non-adjacent form ($w_1$-NAF) and the Hamming weight of the $w_2$-NAF. The width-$w$ non-adjacent form is a binary digit expansion with digit set $\{0, \pm 1, \pm 3, \ldots, \pm(2^{w-1} - 1)\}$ and the syntactical rule that at most one of any $w$ adjacent digits is non-zero.

It will turn out that this distribution is normal if and only if the variance-covariance matrix is regular. Using Theorem 2, we have to find closed walks in the corresponding Markov chain such that all coefficients in (1) have to be zero.

The transducer $\mathcal{T}(w)$ in Figure 2 computes the Hamming weight of the $w$-NAF of the integer $n$ when the input is the binary expansion of $n$ (cf. [15]). It has $w+1$ states. Next, we construct the Cartesian product of the transducers for $w_1$ and $w_2$ and choose any non-degenerate probability distribution, i.e. with all probabilities non-zero, for the outgoing transitions of a state. Thus, we obtain a Markov chain $\mathcal{M}$ with $(w_1+1)(w_2+1)$ states with two different output functions $h_1$ and $h_2$ corresponding to the outputs of the transducers for $w_1$ and $w_2$, respectively. We can now use Theorem 2 to prove that these two Hamming weights are asymptotically jointly normally distributed.

The Cartesian product of two closed walks in $\mathcal{T}(w_1)$ and $\mathcal{T}(w_2)$ with the same input sequence is a closed walk in $\mathcal{M}$. We construct three different closed walks and prove that all three coefficients in (1) have to be zero. For brevity, we denote a closed walk in the Cartesian product $\mathcal{M}$ and its projections to $\mathcal{T}(w_1)$ and $\mathcal{T}(w_2)$ by the same letter.

First, we choose the closed walk $C_1$ starting in state 1 with input sequence 0. We obtain $h_1(C_1)=0$ in $\mathcal{T}(w_1)$, $h_2(C_1)=0$ in $\mathcal{T}(w_2)$ and $\mathbb{1}(C_1)=1$. Second, we choose the closed walk $C_2$ starting in 1 with input sequence $10^{w_2-1}$. Because $w_1<w_2$ and the loop at state 1, $C_2$ is a closed walk in $\mathcal{T}(w_1)$ and $\mathcal{T}(w_2)$. We obtain $h_1(C_2)=1$ in $\mathcal{T}(w_1)$, $h_2(C_2)=1$ in $\mathcal{T}(w_2)$ and $\mathbb{1}(C_2)=w_2$. The third choice depends on whether $w_1=w_2-1$ or not:

- $w_1\neq w_2-1$: We choose the closed walk $C_3$ starting in 1 with input sequence $10^{w_1-1}10^{w_1-1}0^{\alpha}$ where $\alpha=\max(w_2-2w_1,0)$. On the one hand, this is a closed walk in $\mathcal{T}(w_1)$ consisting of two times the cycle $1\to w_1\to 1$ and $\alpha$ times the loop at state 1. On the other hand, this is a closed walk in $\mathcal{T}(w_2)$ consisting of the cycle $1\to w_2\to 1$ and the correct number of loops at state 1. We obtain $h_1(C_3)=2$ in $\mathcal{T}(w_1)$, $h_2(C_3)=1$ in $\mathcal{T}(w_2)$ and $\mathbb{1}(C_3)=\max(w_2,2w_1)$.
- $w_1=w_2-1$: We choose the closed walk $C_3$ starting in 1 with input sequence $10^{w_1-1}10^{w_1-1}10^{w_1-1}$. On the one hand, this is a closed walk in $\mathcal{T}(w_1)$ consisting of three times the cycle $1\to w_1\to 1$. On the other hand, this is a closed walk in $\mathcal{T}(w_2)$ consisting of the closed walk $1\to w_2\to w_2+1\to w_2\to 1$ and the correct number of loops at state 1. We obtain $h_1(C_3)=3$ in $\mathcal{T}(w_1)$, $h_2(C_3)=2$ in $\mathcal{T}(w_2)$ and $\mathbb{1}(C_3)=3w_1$.

This yields a system of linear equations for the coefficients $a_0$, $a_1$ and $a_2$ with coefficient matrix

$$\begin{pmatrix} 1 & 0 & 0\\ w_2 & 1 & 1\\ \max(w_2,2w_1) & 2 & 1\end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 0 & 0\\ w_2 & 1 & 1\\ 3w_1 & 3 & 2\end{pmatrix},$$

which only has the trivial solution. Thus, the Hamming weights of the $w_1$-NAF and the $w_2$-NAF are asymptotically jointly normally distributed, independently of the choice of the distributions for the Markov chain.

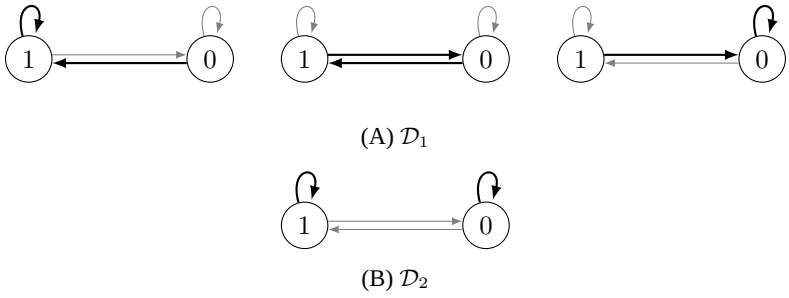
The next two examples investigate the asymptotic independence of length two blocks of 0-1-sequences.

*Example* 4.2 (10- and 11-blocks). The two transducers in Figure 3 count the number of 10- and 11-blocks in 0-1-sequences. After deleting the outputs, both transducers are the same. Thus, any non-degenerate probability distribution on the outgoing edges of the states gives a Markov chain with two output functions $k_{10}$ (for the 10-blocks) and $k_{11}$ (for the 11-blocks).

Because of the two loops and the cycle $0\to 1\to 0$, Theorem 2 implies that the number of 10- and 11-blocks is asymptotically normally distributed.


(A) 10-blocks (B) 11-blocks

Fig. 3: Transducers to compute the number of 10- and 11-blocks.




(A) $\mathcal{D}_1$

(B) $\mathcal{D}_2$

Fig. 4: Functional digraphs of the transducers of Examples 4.2 and 4.3.

The next question is: For which choices of probability distributions is the number of 10- and 11-blocks asymptotically independent? All functional digraphs with one or two components are given in Figure 4. Using Theorem 3, we obtain the following system of equations for the values of the probabilities such that the numbers of 11-blocks and 10-blocks are asymptotically independent: first by definition

$$\begin{aligned} 1 &= p_{0\to0} + p_{0\to1}, \\ 1 &= p_{1\to0} + p_{1\to1}, \end{aligned}$$

then by (2)

$$\begin{aligned} e_{10} &= \frac{p_{0\to1}p_{1\to0}}{p_{0\to1}p_{1\to1} + 2p_{0\to1}p_{1\to0} + p_{0\to0}p_{1\to0}}, \\ e_{11} &= \frac{p_{0\to1}p_{1\to1}}{p_{0\to1}p_{1\to1} + 2p_{0\to1}p_{1\to0} + p_{0\to0}p_{1\to0}}, \end{aligned}$$

and finally for the independence

$$\begin{aligned} &p_{0\to1}p_{1\to1}(-e_{10})(1-e_{11}) + p_{0\to1}p_{1\to0}(1-2e_{10})(-2e_{11}) + p_{0\to0}p_{1\to0}(-e_{10})(-e_{11}) \\ &\quad = p_{0\to0}p_{1\to1}(-e_{10})(-e_{11}) + p_{0\to0}p_{1\to1}(-e_{10})(1-e_{11}). \end{aligned}$$

This system has non-trivial real solutions, i.e. solutions where all probabilities are non-zero, with

$$p_{0\to0} = -\frac{1}{2}p_{1\to1} + 2 - \frac{1}{2}\sqrt{p_{1\to1}^2 - 8p_{1\to1} + 8}$$

for all $0 < p_{1\to1} < 1$. Then we have $2 - \sqrt{2} < p_{0\to0} < 1$.

Thus, for these transition probabilities, the number of 10-blocks and the number of 11-blocks are asymptotically independent.

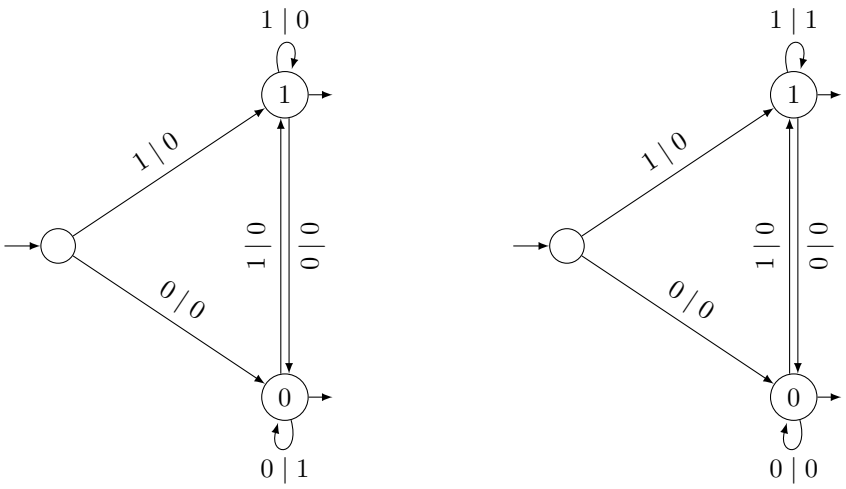
One such example of a non-trivial solution is $p_{1\to1} = p_{1\to0} = 0.5$, $p_{0\to0} \approx 0.7192$ and $p_{0\to1} \approx 0.2808$. Note that for the symmetric distributions $p_{0\to0} = p_{0\to1} = p_{1\to1} = p_{1\to0} = 0.5$, we obtain asymptotic dependence of the number of 10- and 11-blocks.

*Example* 4.3 (00- and 11-blocks). The two transducers in Figure 5 count the number of 00- and 11-blocks in 0-1-sequences. They have the same underlying graph and the same input labels. Thus, choosing any non-degenerate probability distribution of the outgoing edges of the states yields a Markov chain with two output functions.

Because of the two loops and the cycle $0 \to 1 \to 0$, Theorem 2 implies that the number of 00- and 11-blocks is asymptotically normally distributed.

The next question is: For which choices of probability distributions is the number of 00- and 11-blocks asymptotically independent? The functional digraphs of the final component are the same as in Example 4.2, see again Figure 4. By Theorem 3, the system of equations for the transition probabilities $p_e$ such that the two output functions are asymptotically independent are: first by definition

$$\begin{aligned} 1 &= p_{0\to0} + p_{0\to1}, \\ 1 &= p_{1\to0} + p_{1\to1}, \end{aligned}$$




(A) 00-blocks (B) 11-blocks

Fig. 5: Transducers to compute the number of 00- and 11-blocks.

then by (2)

$$e_{00} = \frac{p_{0\to0}p_{1\to0}}{p_{0\to1}p_{1\to1} + 2p_{0\to1}p_{1\to0} + p_{0\to0}p_{1\to0}},$$
$$e_{11} = \frac{p_{0\to1}p_{1\to1}}{p_{0\to1}p_{1\to1} + 2p_{0\to1}p_{1\to0} + p_{0\to0}p_{1\to0}},$$

and finally for the independence

$$\begin{aligned} &p_{0\to1}p_{1\to1}(-e_{00})(1-e_{11}) + p_{0\to1}p_{1\to0}(-2e_{00})(-2e_{11}) + p_{0\to0}p_{1\to0}(1-e_{00})(-e_{11}) \\ &\quad = p_{0\to0}p_{1\to1}(1-e_{00})(1-e_{11}) + p_{0\to0}p_{1\to1}(-e_{00})(-e_{11}). \end{aligned}$$

These equations have no solution with $0 < p_e < 1$ for all transitions $e$. Thus, the numbers of 00- and 11-blocks are asymptotically dependent for all choices of the input distributions, as expected.

# 5 Proofs

In this section, we prove the results from Section 3. Most of the proofs follow along the same ideas as in [18]. The main differences are that one has to replace "complete transducer" by "Markov chain" and the input sum by the output sum $K_n^{(1)}$.

We first prove Theorem 3 with the help of two lemmas. For one of these lemmas, we use a version of the Matrix-Tree Theorem for weighted directed forests proved in [5, 20]. At the end of this section, we prove Theorems 1 and 2.

**Definition 5.1.** Let $A$, $B \subseteq \{1, \dots, N\}$. Let $\mathcal{F}_{A,B}$ be the set of all forests which are spanning subgraphs of the final component of the Markov chain $\mathcal{M}$ with $|A|$ trees such that every tree is rooted at some vertex $a \in A$ and contains exactly one vertex $b \in B$.

Let $A = \{i_1, \dots, i_n\}$ and $B = \{j_1, \dots, j_n\}$ with $i_1 < \dots < i_n$ and $j_1 < \dots < j_n$. For $F \in \mathcal{F}_{A,B}$, we define a function $g\colon B \to A$ by $g(j) = i$ if $j$ is in the tree of $F$ which is rooted in vertex $i$. We further define the function $h\colon A \to B$ by $h(i_k) = j_k$ for $k = 1, \dots, n$. The composition $g \circ h\colon A \to A$ is a permutation of $A$. We define $\operatorname{sign} F = \operatorname{sign} g \circ h$.

If $|A| \neq |B|$, then $\mathcal{F}_{A,B} = \emptyset$. If $|A| = |B| = 1$, then $\operatorname{sign} F = 1$ and $\mathcal{F}_{A,B}$ consists of all spanning trees rooted in $a \in A$.

**Theorem** (All-Minors-Matrix-Tree Theorem [5, 20])**.** *For a directed, weighted graph with loops and multiple edges, let $L = (l_{ij})_{1 \le i,j \le N}$ be the Laplacian matrix, that is $\sum_{j=1}^{N} l_{ij} = 0$ for every $i = 1, \dots, N$ and $-l_{ij}$ is the sum of the weights $p_e$ of all edges $e$ from $i$ to $j$ for $i \neq j$. Then, for $|A| = |B|$, the minor $\det L_{A,B}$ satisfies*

$$\det L_{A,B} = (-1)^{\sum_{i \in A} i + \sum_{j \in B} j} \sum_{F \in \mathcal{F}_{A,B}} p_F \operatorname{sign} F$$

*where $L_{A,B}$ is the matrix $L$ whose rows with index in $A$ and columns with index in $B$ are deleted.*

The All-Minors-Matrix-Tree Theorem is still valid for $|A| \neq |B|$ if we assume that the determinant of a non-square matrix is $0$. For notational simplicity, we use this convention in the rest of this section.

**Definition 5.2.** The transition matrix $W(x_1, \dots, x_m)$ of a Markov chain with $M$ states and $m$ output functions $k_1, \dots, k_m$ is a $M \times M$ matrix whose $(i,j)$-th entry is

$$\sum_{e\colon i \to j} p_e x_1^{k_1(e)} \cdots x_m^{k_m(e)}$$

where $p_e$ is the probability of the transition $e$.

Let $A(x_1, \dots, x_m)$ be the $N \times N$ transition matrix of the final component of the Markov chain. Let the order of the states be such that the transition matrix of the whole Markov chain $W(x_1, \dots, x_m)$ has the block structure

$$W(x_1, \dots, x_m) = \begin{pmatrix} * & * \\ 0 & A(x_1, \dots, x_m) \end{pmatrix} \tag{3}$$

where $*$ denotes any matrix. If the Markov chain is strongly connected, the matrices $*$ are not present (they have $0$ rows).

We first use the All-Minors-Matrix-Tree Theorem to connect the derivatives of the characteristic polynomial of the transition matrix with a sum of weighted digraphs in the next lemma.

**Lemma 5.3.** *For $f(x_1, x_2, z) = \det(I - zA(x_1, x_2))$, we have*

$$\begin{aligned} f_{x_i}(1,1,1) &= -k_i(\mathcal{D}_1), & f_{x_1x_2}(1,1,1) &= (k_1,k_2)(\mathcal{D}_2) - (k_1,k_2)(\mathcal{D}_1), \\ f_z(1,1,1) &= -\mathbb{1}(\mathcal{D}_1), & f_{x_iz}(1,1,1) &= (k_i,\mathbb{1})(\mathcal{D}_2) - (k_i,\mathbb{1})(\mathcal{D}_1), \\ f_{x_ix_i}(1,1,1) + f_{x_i}(1,1,1) &= (k_i,k_i)(\mathcal{D}_2) - (k_i,k_i)(\mathcal{D}_1), \\ f_{zz}(1,1,1) + f_z(1,1,1) &= (\mathbb{1},\mathbb{1})(\mathcal{D}_2) - (\mathbb{1},\mathbb{1})(\mathcal{D}_1) \end{aligned}$$

*for $i = 1, 2$.*

This lemma can be proven in the same way as [18, Lemma 5.3] using the All-Minors-Matrix-Tree Theorem [5, 20].

The following lemma will be used for $m \geq 2$ output functions later on.

**Lemma 5.4.** *Let $f(x_1,\dots,x_m,z)=\det(I-zA(x_1,\dots,x_m))$. Then there is a unique dominant root $z=\rho(x_1,\dots,x_m)$ of $f$ in a neighborhood of $(1,\dots,1)$.*

*The moment generating function of $(K_n^{(1)},\dots,K_n^{(m)})$ has the asymptotic expansion*

$$\mathbb{E}(\exp(s_1K_n^{(1)}+\cdots+s_mK_n^{(m)}))=e^{u(s_1,\dots,s_m)n+v(s_1,\dots,s_m)}(1+\mathcal{O}(\kappa^n))$$

*where $\kappa<1$,*

$$u(s_1,\dots,s_m)=-\log\rho(e^{s_1},\dots,e^{s_m}),$$

*and $v(s_1,\dots,s_m)$ are analytic functions in a small neighborhood of $(0,\dots,0)$.*

**Proof:** The moment generating function of $(K_n^{(1)},\dots,K_n^{(m)})$ is

$$\mathbb{E}(\exp(s_1K_n^{(1)}+\cdots+s_mK_n^{(m)}))=[z^n]v_1^t(I-zW(e^{s_1},\dots,e^{s_m}))^{-1}v_2(e^{s_1},\dots,e^{s_m})$$

for the initial vector $v_1$, and a vector $v_2(x_1,\dots,x_m)$ encoding all the final information of the states[(ii)] where we write $[z^n]b(z)$ for the coefficient of $z^n$ in the power series $b$. Because of the block structure of the transition matrix $W$ of the whole Markov chain in (3), we obtain

$$\begin{aligned}\mathbb{E}(x_1^{K_n^{(1)}}\cdots x_m^{K_n^{(m)}})&=[z^n]\frac{F_1(x_1,\dots,x_m,z)}{\det(I-zW(x_1,\dots,x_m))}\\&=[z^n]\frac{F_1(x_1,\dots,x_m,z)}{F_2(x_1,\dots,x_m,z)f(x_1,\dots,x_m,z)}\end{aligned}$$

for "polynomials" $F_1$ and $F_2$ , i.e. finite linear combinations of $x_1^{\alpha_1}\cdots x_m^{\alpha_m}z^\beta$ for $\alpha_i\in\mathbb{R}$ and $\beta$ a non-negative integer. The function $F_2$ corresponds to the determinant of the non-final part of the Markov chain.

We obtain the coefficient of $z^n$ by singularity analysis (cf. [7]): Since the final component of $\mathcal{M}$ is again a Markov chain, the dominant singularity of $1/f(1,\dots,1,z)$ is 1 by the theorem of Perron–Frobenius (cf. [9]). By the aperiodicity of the final component, this dominant singularity is unique and it is $\rho(1,\dots,1)=1$.

Next, we consider the non-final components of the Markov chain using the same arguments as in [18]. The corresponding non-final component $\mathcal{M}_0$ is not a Markov chain as the transition matrix is not stochastic. Let $\mathcal{M}_0^+$ be the Markov chain that is obtained from $\mathcal{M}_0$ by adding loops with the missing probabilities where necessary. The dominant eigenvalue of the transition matrix of $\mathcal{M}_0^+$ is 1. As the transition matrices of $\mathcal{M}_0$ and $\mathcal{M}_0^+$ satisfy element-wise inequalities but are not equal (at $(x_1,\dots,x_m)=(1,\dots,1)$), the theorem of Perron–Frobenius (cf. [9, Theorem 8.8.1]) implies that the dominant eigenvalues of $\mathcal{M}_0$ have absolute value less than 1. Thus, the dominant singularities of $F_2(1,\dots,1,z)^{-1}$ are at $|z|>1$.

As $A(1,\dots,1,z)=(1-z)^{-1}$, we obtain $F_1(1,\dots,1)\neq 0$.

Thus, there is a is the unique, dominant singularity of

$$\frac{F_1(1,\dots,1,z)}{F_2(1,\dots,1,z)f(1,\dots,1,z)},$$

(ii) This information is the final output (see Remark 2.3) and the exit weight (see Remark 2.4) included as $w_ix_1^{f_1(i)}\cdots x_m^{f_m(i)}$ in the $i$-th coordinate of $v_2(x_1,\dots,x_m)$. This does not change the asymptotic behavior (see Remark 5.5).

which is $\rho(1,\dots,1) = 1$. This also holds for $(x_1,\dots,x_m)$ in a small neighborhood of $(1,\dots,1)$ by the continuity of the eigenvalues of the transition matrices. Thus, $\rho(x_1,\dots,x_m)$ is this unique dominant singularity.

Now, singularity analysis (cf. [7]) implies the statement of this lemma. □

*Remark* 5.5. The main term of the asymptotic expansion of the moment generating function only depends on $\rho(x_1,\dots,x_m)$ and therefore on $f(x_1,\dots,x_m,z)$. It does not depend on the "polynomials" $F_1(x_1,\dots,x_m,z)$ and $F_2(x_1,\dots,x_m,z)$. Thus, only the final component influences the main term. Neither the states in the non-final part of the Markov chain nor the final outputs and exit weights influence the main term.

Now, we can use the previous two lemmas to prove Theorem 3.

**Proof of Theorem 3.:** By Lemma 5.4 for two output functions $k_1$ and $k_2$, the moment generating function satisfies the conditions of the Quasi-Power Theorem [18, Theorem 5.1], which yields the expected value

$$\mathbb{E}(K_n^{(1)}, K_n^{(2)}) = n \operatorname{grad} u(\mathbf{0}) + \mathcal{O}(1)$$

and the variance

$$\mathbb{V}(K_n^{(1)}, K_n^{(2)}) = n H_u(\mathbf{0}) + \mathcal{O}(1)$$

with $\operatorname{grad} u(\mathbf{0})$ and $H_u(\mathbf{0})$ the gradient and the Hessian of $u$ at $\mathbf{0}$, respectively. Furthermore, we obtain an asymptotic joint normal distribution of the standardized random vector if the Hessian is not singular by [18, Theorem 3.9]. Otherwise, the limiting random vector is either a pair of degenerate random variables, or a degenerate and normally distributed one, or a linear transformation thereof. Thus, the random variables $K_n^{(1)}$ and $K_n^{(2)}$ are asymptotically independent if and only if the covariance is zero.

By implicit differentiation, we obtain the following formulas for the constants of the moments in terms of the partial derivatives of $f$:

$$\begin{aligned}
e_i &= \frac{f_{x_i}}{f_z}\Big|_{\mathbf{1}},\\
v_i &= \frac{1}{f_z^3}(f_{x_i}^2(f_{zz}+f_z) + f_z^2(f_{x_ix_i}+f_{x_i}) - 2f_{x_i}f_zf_{x_iz})\Big|_{\mathbf{1}},\\
c &= \frac{1}{f_z^3}(f_{x_1}f_{x_2}(f_{zz}+f_z) + f_z^2 f_{x_1x_2} - f_{x_2}f_zf_{x_1z} - f_{x_1}f_zf_{x_2z})\Big|_{\mathbf{1}}
\end{aligned}$$

for $i = 1, 2$.

Now, Lemma 5.3 implies the results as stated in the theorem. □

**Proof of Theorem 1:** This follows by the same arguments as in [18, Theorem 3.1]. □

**Proof of Corollary 3.2:** This follows by the same arguments as in [18, Corollary 3.6]. □

**Proof of Theorem 2:** WLOG, we assume that $\mathbb{E}K_n^{(i)} = \mathcal{O}(1)$ for $i = 1,\dots,m$ by subtracting the corresponding constant of the expected value from each output function. There exists a unitary matrix $T = (t_{ji})_{1\le j,i\le m}$ such that the variance-covariance matrix $\Sigma$ can be diagonalized as $T\Sigma T^\top = D$. The

diagonal matrix $D$ is the variance-covariance matrix of the linearly transformed random vector $\boldsymbol{Y}_n = T\boldsymbol{K}_n$.

Then $\Sigma$ is singular if and only if the diagonal matrix $D$ is singular. This is equivalent to

$$\mathbb{V}(t_{j1}K_n^{(1)} + \cdots + t_{jm}K_n^{(m)}) = \mathcal{O}(1) \tag{4}$$

holds for a $j \in \{1, \ldots, m\}$. Now consider the output function $t_{j1}k_1 + \cdots + t_{jm}k_m$. By Theorem 1, (4) is equivalent to

$$t_{j1}k_1(C) + \cdots + t_{jm}k_m(C) = 0$$

holding for all cycles of the final component (since the expected value of this output function is $\mathcal{O}(1)$).

If we shift back the output function such that the expected value is no longer bounded, we obtain an additional summand $a_0\mathbb{1}(C)$.

The asymptotic joint normal distribution follows from Lemma 5.4 and the multidimensional Quasi-Power Theorem [6, Theorem 2.22]. □

# References


[1] Roberto Avanzi, *A note on the signed sliding window integer recoding and a left-to-right analogue*, Selected Areas in Cryptography: 11th International Workshop, SAC 2004, Waterloo, Canada, August 9-10, 2004, Revised Selected Papers (H. Handschuh and A. Hasan, eds.), Lecture Notes in Comput. Sci., vol. 3357, Springer-Verlag, Berlin, 2005, pp. 130–143.

[2] Roberto Avanzi, Clemens Heuberger, and Helmut Prodinger, *Scalar multiplication on Koblitz curves. Using the Frobenius endomorphism and its combination with point halving: Extensions and mathematical analysis*, Algorithmica **46** (2006), 249–270.

[3] Edward A. Bender and Fred Kochman, *The distribution of subword counts is usually normal*, European J. Combin. **14** (1993), no. 4, 265–275.

[4] Valérie Berthé and Michel Rigo (eds.), *Combinatorics, automata and number theory*, Encyclopedia Math. Appl., vol. 135, Cambridge University Press, Cambridge, 2010.

[5] Seth Chaiken, *A combinatorial proof of the all minors matrix tree theorem*, SIAM J. Alg. Disc. Meth. **3** (1982), no. 3, 319–329.

[6] Michael Drmota, *Random trees*, SpringerWienNewYork, 2009.

[7] Philippe Flajolet and Robert Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.

[8] Philippe Flajolet, Wojciech Szpankowski, and Brigitte Vallée, *Hidden word statistics*, J. ACM **53** (2006), no. 1, 147–183.

[9] Chris D. Godsil and Gordon Royle, *Algebraic graph theory*, Graduate texts in mathematics, vol. 207, Springer Verlag (New York), 2001.

[10] Massimiliano Goldwurm and Roberto Radicioni, *Average value and variance of pattern statistics in rational models*, Implementation and Application of Automata (Jan Holub and Jan Žďárek, eds.), Lecture Notes in Comput. Sci., vol. 4783, Springer Berlin Heidelberg, 2007, pp. 62–72.

[11] Peter J. Grabner, Clemens Heuberger, and Helmut Prodinger, *Distribution results for low-weight binary representations for pairs of integers*, Theoret. Comput. Sci. **319** (2004), 307–331.

[12] Peter J. Grabner, Clemens Heuberger, Helmut Prodinger, and Jörg Thuswaldner, *Analysis of linear combination algorithms in cryptography*, ACM Trans. Algorithms **1** (2005), 123–142.

[13] Peter J. Grabner and Jörg M. Thuswaldner, *On the sum of digits function for number systems with negative bases*, Ramanujan J. **4** (2000), no. 2, 201–220.

[14] Florian Heigl and Clemens Heuberger, *Analysis of digital expansions of minimal weight*, 23rd Intern. Meeting on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms (AofA'12), DMTCS Proceedings, 2012, pp. 399–411.

[15] Clemens Heuberger and Sara Kropf, *Analysis of the binary asymmetric joint sparse form*, Combin. Probab. Comput. **23** (2014), 1087–1113.

[16] Clemens Heuberger, Sara Kropf, and Helmut Prodinger, *Output sum of transducers: Limiting distribution and periodic fluctuation*, Electron. J. Combin. **22** (2015), no. 2, 1–53.

[17] ———, *Analysis of carries in signed digit expansions*, Monatsh. Math. (2016), published online first, doi:10.1007/s00605-016-0917-x.

[18] Clemens Heuberger, Sara Kropf, and Stephan Wagner, *Variances and covariances in the central limit theorem for the output of a transducer*, European J. Combin. **49** (2015), 167–187.

[19] Clemens Heuberger and Helmut Prodinger, *Analysis of alternative digit sets for nonadjacent representations*, Monatsh. Math. **147** (2006), 219–248.

[20] John W. Moon, *Some determinant expansions and the matrix-tree theorem*, Discrete Math. **124** (1994), 163–171.

[21] James A. Muir and Douglas R. Stinson, *Minimality and other properties of the width-$w$ nonadjacent form*, Math. Comp. **75** (2006), 369–384.

[22] Pierre Nicodème, Bruno Salvy, and Philippe Flajolet, *Motif statistics*, Theoret. Comput. Sci. **287** (2002), no. 2, 593–617.

[23] William Parry, *Intrinsic Markov chains*, Trans. Amer. Math. Soc. **112** (1964), 55–66.

[24] Claude E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423.