Combinatorial Characterization of Transducers with Bounded Variance

Sara Kropf

Alpen-Adria-Universität Klagenfurt

Joint work with Clemens Heuberger and Stephan Wagner

AofA
Strobl, June 12, 2015
Theorem (Hwang’s Quasi-Power-Theorem)

Let \( \Omega_n \) be a sequence of real random variables. Suppose the moment generating function satisfies

\[
\mathbb{E}(e^{\Omega_n s}) = e^{u(s)\Phi(n)+\nu(s)}(1 + \mathcal{O}(\kappa_n^{-1}))
\]

under some conditions.

Then

\[
\mathbb{E}\Omega_n = u'(0)\Phi(n) + \mathcal{O}(1),
\]

\[
\nabla\Omega_n = u''(0)\Phi(n) + \mathcal{O}(1).
\]

If \( \sigma^2 := u''(0) \neq 0 \), then \( \frac{\Omega_n - \mathbb{E}\Omega_n}{\sqrt{\nabla\Omega_n}} \) is asymptotically normally distributed.

When is the variance bounded?
Transducers

- transducer \( \mathcal{T} \) with a finite number of states

\[
\text{Output}(X_n) = \sum \text{output random word } X_n \in A^n \text{ as input today: equidistribution on } A^n \text{ read from right to left}
\]

Example with \( X_n = 11001 \) input: 11001

Output(11001) = 4
Transducers

- transducer $T$ with a finite number of states
- $\text{Output}(X_n) = \text{sum of the output}$
- random word $X_n \in A^n$ as input
- today: equidistribution on $A^n$
- read from right to left
Transducers

- transducer $\mathcal{T}$ with a finite number of states
- Output($X_n$) = sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left

Example with $X_n = 11001$

input: 11001
output: 11001

Output(11001) =
Transducers

- transducer $\mathcal{T}$ with a finite number of states
- Output($X_n$) = sum of the output
- random word $X_n \in A^n$ as input
- today: equidistribution on $A^n$
- read from right to left

Example with $X_n = 11001$

<table>
<thead>
<tr>
<th>input:</th>
<th>11001</th>
<th>output:</th>
<th>1</th>
</tr>
</thead>
</table>

Output(11001) = 1
Transducers

- transducer $T$ with a finite number of states
- Output($X_n$) = sum of the output
- random word $X_n \in A^n$ as input
- today: equidistribution on $A^n$
- read from right to left

Example with $X_n = 11001$

<table>
<thead>
<tr>
<th>input:</th>
<th>11001</th>
</tr>
</thead>
<tbody>
<tr>
<td>output:</td>
<td>11</td>
</tr>
</tbody>
</table>

Output(11001) =
Transducers

- transducer $T$ with a finite number of states
- $Output(X_n) = \text{sum of the output}$
- random word $X_n \in A^n$ as input
- today: equidistribution on $A^n$
- read from right to left

Example with $X_n = 11001$

<table>
<thead>
<tr>
<th>input:</th>
<th>11001</th>
</tr>
</thead>
<tbody>
<tr>
<td>output:</td>
<td>011</td>
</tr>
</tbody>
</table>

$Output(11001) =$
Transducers

- transducer $\mathcal{T}$ with a finite number of states
- $\text{Output}(X_n) = \text{sum of the output}$
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left

Example with $X_n = 11001$

<table>
<thead>
<tr>
<th>input:</th>
<th>11001</th>
<th>output:</th>
<th>1011</th>
</tr>
</thead>
</table>

Output(11001) =
Transducers

- transducer $\mathcal{T}$ with a finite number of states
- Output($X_n$) = sum of the output
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left

Example with $X_n = 11001$

<table>
<thead>
<tr>
<th>Input:</th>
<th>11001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output:</td>
<td>01011</td>
</tr>
</tbody>
</table>

Output(11001) =
Transducers

- transducer $\mathcal{T}$ with a finite number of states
- $\text{Output}(X_n) = \text{sum of the output}$
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left

Example with $X_n = 11001$

<table>
<thead>
<tr>
<th>input: 11001</th>
<th>output: 101011</th>
<th>Output(11001) =</th>
</tr>
</thead>
</table>

Diagram of a transducer with states and transitions.
Transducers

- transducer $\mathcal{T}$ with a finite number of states
- $\text{Output}(X_n) = \text{sum of the output}$
- random word $X_n \in \mathcal{A}^n$ as input
- today: equidistribution on $\mathcal{A}^n$
- read from right to left

Example with $X_n = 11001$

<table>
<thead>
<tr>
<th>input:</th>
<th>11001</th>
</tr>
</thead>
<tbody>
<tr>
<td>output:</td>
<td>101011</td>
</tr>
</tbody>
</table>

$\text{Output}(11001) = 4$
Other Probability Model and Several Outputs

All results also possible for:
- inputs coming from a Markov chain
- for every transition a probability
- sum of probabilities of output transitions is 1

Some results are independent of the choice of this Markov chain.

Several simultaneous outputs.
Applications

- algorithms with finite memory usage
- many digit expansions:
  - Hamming weight
  - sum of digits function, . . .
- many recursions
- motifs
Applications

- algorithms with finite memory usage
- many digit expansions:
  - Hamming weight
  - sum of digits function, ... 
- many recursions
- motifs

- completely $q$-additive functions
- digital sequences
- $q$-automatic sequences
Applications

- digit sum of binary expansion
Applications

- digit sum of binary expansion
- Hamming weight of non-adjacent form (NAF):
  - digits \{0, \pm 1\}, base 2
  - at least one of any two adjacent digits is 0
Applications

- digit sum of binary expansion
- Hamming weight of non-adjacent form (NAF):
  - digits \{0, \pm 1\}, base 2
  - at least one of any two adjacent digits is 0
- Hamming weight of width-\(w\) NAF:
  - digits \{0, \pm 1, \pm 3, \ldots, \pm(2^{w-1} - 1)\}, base 2
  - at least \(w - 1\) of \(w\) consecutive digits are 0
Variability Condition

Theorem (Hwang’s Quasi-Power-Theorem)

Let $\Omega_n$ be a sequence of real random variables. Suppose the moment generating function satisfies

$$\mathbb{E}(e^{\Omega_ns}) = e^{u(s)\Phi(n) + v(s)}(1 + O(\kappa_n^{-1}))$$

under some conditions.

Then

$$\mathbb{E}\Omega_n = u'(0)\Phi(n) + O(1),$$

$$\nabla\Omega_n = u''(0)\Phi(n) + O(1).$$

If $\sigma^2 := u''(0) \neq 0$, then $\frac{\Omega_n - \mathbb{E}\Omega_n}{\sqrt{\nabla\Omega_n}}$ is asymptotically normally distributed.

Assume that $\mathcal{T}$ is strongly connected.

Output($X_n$) satisfies all assumptions, except maybe the variability condition $\sigma^2 \neq 0$. 
Theorem (Heuberger–K.–Wagner 2015)

Let \( T \) be strongly connected. Then the following assertions are equivalent:

1. The asymptotic variance \( \sigma^2 \) is 0.
2. There is a constant \( k \) such that the average output of every cycle is \( k \).
3. There is a constant \( k \) such that \( \text{Output}(X_n) = kn + O(1) \).
Bounded Variance

**Theorem (Heuberger–K.–Wagner 2015)**

Let $\mathcal{T}$ be strongly connected. Then the following assertions are equivalent:

1. The asymptotic variance $\sigma^2$ is 0.
2. There is a constant $k$ such that the average output of every cycle is $k$.
3. There is a constant $k$ such that $\text{Output}(X_n) = kn + \mathcal{O}(1)$.

**Corollary (Heuberger–K.–Wagner 201)**

Let $\mathcal{T}$ be strongly connected, aperiodic with output alphabet $\{0, 1\}$. Then the asymptotic variance $\sigma^2$ is 0 if and only if all output letters are the same.
Small Example

\[ \sigma^2 = \frac{432}{2197} \]

\[ \sim \text{ asymptotic variance } \neq 0 \]
Small Example

\[ \exists \text{ asymptotic variance} \neq 0 \]

Sage: \( \sigma^2 = \frac{432}{2197} \)
Example: \( \tau \)-adic Digit Expansion

- algebraic integer \( \tau \)
- joint expansion of \( d \)-dimensional vectors in \( \mathbb{Z}[\tau]^d \)
- redundant digit set \( \mathcal{D} \) which satisfies
  - \( \mathcal{D} \cap \tau \mathbb{Z}^d = \{0\} \)
  - a subadditivity condition

Theorem (Heigl–Heuberger 2012)

If the asymptotic variance \( \sigma^2 \) of the minimal Hamming weight with digit set \( \mathcal{D} \) is \( \neq 0 \), then the minimal Hamming weight is asymptotically normally distributed.
Example: \( \tau \)-adic Digit Expansion

- algebraic integer \( \tau \)
- joint expansion of \( d \)-dimensional vectors in \( \mathbb{Z}[\tau]^d \)
- redundant digit set \( \mathcal{D} \) which satisfies
  - \( \mathcal{D} \cap \tau \mathbb{Z}^d = \{0\} \)
  - a subadditivity condition
- input: \( \tau \)-adic expansions with the irredundant digit set \( \mathcal{A} \) of length \( \leq n \) with equidistribution
Example: $\tau$-adic Digit Expansion

- algebraic integer $\tau$
- joint expansion of $d$-dimensional vectors in $\mathbb{Z}[\tau]^d$
- redundant digit set $\mathcal{D}$ which satisfies
  - $\mathcal{D} \cap \tau \mathbb{Z}^d = \{0\}$
  - a subadditivity condition
- input: $\tau$-adic expansions with the irredundant digit set $\mathcal{A}$ of length $\leq n$ with equidistribution

**Theorem (Heigl–Heuberger 2012)**

If the asymptotic variance $\sigma^2$ of the minimal Hamming weight with digit set $\mathcal{D}$ is $\neq 0$, then the minimal Hamming weight is asymptotically normally distributed.
Example: $\tau$-adic Digit Expansion

Heigl–Heuberger construct a transducer for each $\tau$ and $\mathcal{D}$:

- cycle with average output 0
Example: \( \tau \)-adic Digit Expansion

Heigl–Heuberger construct a transducer for each \( \tau \) and \( D \):

- Cycle with average output 0
- But not all minimal weights are 0
- \( 0 \cdots 0 \) always leads to the initial state
- \( \rightsquigarrow \) cycle with average output \( \neq 0 \)
Heigl–Heuberger construct a transducer for each $\tau$ and $D$:

- Cycle with average output $0$
- But not all minimal weights are $0$
- $0\cdots0$ always leads to the initial state
- $\rightsquigarrow$ Cycle with average output $\neq 0$
- Variability condition is satisfied
- $\rightsquigarrow$ Asymptotic normality

Example: $\tau$-adic Digit Expansion
Theorem (Heuberger–K.–Wagner 2015)

Let $T$ be strongly connected. Then the following assertions are equivalent:

1. The asymptotic variance $\sigma^2$ is 0.
2. There is a constant $k$ such that the average output of every cycle is $k$.
3. There is a constant $k$ such that $\text{Output}(X_n) = kn + O(1)$.
Idea of the Proof of the Theorem

1 ⇔ 2:

- Assume: asymptotic expected value of Output($X_n$) is 0
- Probability generating function

\[ A(y, z) = \sum_{l \in \mathbb{R}} \sum_{n=0}^{\infty} a_{ln} K^{-n} y^l z^n \]

with \( K = |A| \) and \( a_{ln} \) = number of input words of length \( n \) with output sum \( l \)

- \( A(1, z) \) has a simple dominant pole at \( z = 1 \)
Idea of the Proof of the Theorem

1 ⇔ 2:

• assume: asymptotic expected value of Output($X_n$) is 0

• probability generating function

$$A(y, z) = \sum_{l \in \mathbb{R}} \sum_{n=0}^{\infty} a_{ln} K^{-n} y^l z^n$$

with $K = |A|$ and $a_{ln} =$ number of input words of length $n$

with output sum $l$

• $A(1, z)$ has a simple dominant pole at $z = 1$

$$\mathbb{E}(\text{Output}(X_n)) = [z^n]A_y(1, z) = \mathcal{O}(1)$$

$$\nabla\text{Output}(X_n)) = [z^n]A_{yy}(1, z) + \mathcal{O}(1)$$
Idea of the Proof of the Theorem

Decomposition:

\[ A(y, z) = 1 - C(y, z) P(y, z) \]

\[ P(1, z) \text{ is analytic in } |z| < 1 + \varepsilon \]

\[ 1 - C(1, z) = (1 - z) g(z) \] with \[ g(1) \neq 0 \]
Idea of the Proof of the Theorem

Decomposition:

- probability generating functions

\[ C(y, z), \ P(y, z) \]
Idea of the Proof of the Theorem

Decomposition:

- probability generating functions $C(y, z), P(y, z)$

- by the symbolic method:

$$A(y, z) = \frac{1}{1 - C(y, z)} P(y, z)$$
Idea of the Proof of the Theorem

Decomposition:

- probability generating functions \( C(y, z), P(y, z) \)
- by the symbolic method:
  \[
  A(y, z) = \frac{1}{1 - C(y, z)} P(y, z)
  \]
- \( P(1, z) \) is analytic in \(|z| < 1 + \varepsilon\)
- \( P(1, 1) \neq 0 \)
- \( 1 - C(1, z) = (1 - z)g(z) \) with \( g(1) \neq 0 \)
Idea of the Proof of the Theorem

- Singularity Analysis \( \mapsto \)

\[
\nabla (\text{Output}(X_n)) = P(1, 1)g(1)^{-2} C_{yy}(1, 1)n + O(1)
\]

thus,

\[
\nabla (\text{Output}(X_n)) = O(1) \iff C_{yy}(1, 1) = 0 \iff \sum_{C \in \mathcal{C}} \text{Output}(C)^2 K^{-\text{Length}(C)} = 0 \iff \forall C \in \mathcal{C} : \text{Output}(C) = 0
\]
Consider $m$ different outputs $k_1, \ldots, k_m$ of a transducer instead of Output.

Using a multi-dimensional Quasi-Power-Theorem:

**Theorem (K. 2015+)**

The $m$ output sums are asymptotically jointly normally distributed, if and only if:

$$a_0 \text{Length}(C) + a_1 k_1(C) + \cdots + a_m k_m(C) = 0$$

holding for all cycles $C$ implies that $a_0 = \cdots = a_m = 0$. 
Bounded Covariance

- random variable \((\text{Input}(X_n), \text{Output}(X_n))\)
- 2-dimensional version of the Quasi-Power-Theorem
- \(\sim\) asymptotic normal distribution
random variable \((\text{Input}(X_n), \text{Output}(X_n))\)
2-dimensional version of the Quasi-Power-Theorem
\(\rightsquigarrow\) asymptotic normal distribution
When is the covariance bounded?
covariance bounded \(\iff\) components of the asymptotic random variable are independent

**Definition**
An independent transducer is a transducer which has a bounded covariance of \((\text{Input}(X_n), \text{Output}(X_n))\).
A functional digraph is a directed graph where every vertex has out-degree 1. This is a map from a finite set into itself.
A functional digraph is a directed graph where every vertex has out-degree 1.

This is a map from a finite set into itself.

$D_1$ and $D_2$ are the sets of functional digraphs with one respectively two components which are subgraphs of the given transducer.
Bounded Covariance

\[
\text{InputOutput}(\mathcal{D}_1) = \sum_{D \in \mathcal{D}_1} \text{Input}(\text{cycle})\text{Output}(\text{cycle}),
\]

\[
\text{InputOutput}(\mathcal{D}_2) = \sum_{D \in \mathcal{D}_2} \text{Input}(\text{one cycle})\text{Output}(\text{other cycle})
\]

Theorem (Heuberger–K.–Wagner 2015)

Suppose the asymptotic expected value of \(\text{Input}(X_n), \text{Output}(X_n)\) is \((0, 0)\). Then the transducer is independent if and only if \(\text{InputOutput}(\mathcal{D}_2) = \text{InputOutput}(\mathcal{D}_1)\). Also possible: 2 outputs, Markov chain.
Bounded Covariance

\[
\text{InputOutput}(D_1) = \sum_{D \in D_1} \text{Input(cycle)}\text{Output(cycle)},
\]

\[
\text{InputOutput}(D_2) = \sum_{D \in D_2} \text{Input(one cycle)}\text{Output(other cycle)}
\]

**Theorem (Heuberger–K.–Wagner 2015)**

Suppose the asymptotic expected value of \((\text{Input}(X_n), \text{Output}(X_n))\) is \((0, 0)\).

Then the transducer is independent if and only if

\[
\text{InputOutput}(D_2) = \text{InputOutput}(D_1).
\]

Also possible: 2 outputs, Markov chain
Width-\(w\) Non-Adjacent Form

- asymptotic covariance = 0
- arbitrarily large independent transducers
- Hamming weight of binary expansion and Hamming weight of \(w\)-NAF are independent
- \(w = 2\): NAF (Heuberger–Prodinger 2007)
Width-$w$ Non-Adjacent Form

2 $\leq w_1 < w_2$ with $w_1 \neq w_2 - 1$:
- closed walk with input 0
- closed walk with input $10^{w_2-1}$
- closed walk with input $10^{w_1-1}10^{w_1-1}0\cdots0$

\[ \begin{pmatrix} 1 & 0 & 0 \\ * & 1 & 1 \\ * & 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = 0 \]

$\Rightarrow$ asymptotic normal distribution
Gray Code

First values:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>6</th>
<th>101</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>11</td>
<td>8</td>
<td>1100</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>9</td>
<td>1101</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>110</td>
<td>10</td>
<td>1111</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>111</td>
<td>11</td>
<td>1110</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>111</td>
<td>11</td>
<td>1110</td>
<td></td>
</tr>
</tbody>
</table>

starting transitions unimportant

asymptotic covariance = 0

independent transducer

Hamming weight of binary expansion

and Hamming weight of Gray code are independent
Gray Code

First values:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>6</td>
<td>101</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>8</td>
<td>1100</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>9</td>
<td>1101</td>
</tr>
<tr>
<td>4</td>
<td>110</td>
<td>10</td>
<td>1111</td>
</tr>
<tr>
<td>5</td>
<td>111</td>
<td>11</td>
<td>1110</td>
</tr>
</tbody>
</table>

- starting transitions unimportant
Gray Code

First values:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>6</th>
<th>101</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>7</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>8</td>
<td>1100</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>9</td>
<td>1101</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>110</td>
<td>10</td>
<td>1111</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>111</td>
<td>11</td>
<td>1110</td>
<td></td>
</tr>
</tbody>
</table>

- starting transitions unimportant
- asymptotic covariance = 0
- independent transducer
- Hamming weight of binary expansion and Hamming weight of Gray code are independent
Conclusion

- combinatorial description for transducers with
  - bounded variance
  - singular variance-covariance matrix
  - bounded covariance

- \( \sim \) asymptotically normally distributed
- can be checked
  - without long computations
  - in general settings