



# BOOLESCHE NETZWERKE MIT ZUFÄLLIGEN REGELN: EIN REGELBASIERTES MODELL MIT ZUFÄLLIGEN REGELN FÜR REGULATORISCHE GENETISCHE NETZWERKE

AUSARBEITUNG

Julia Kroos

Münster, 21. Juni 2012

Bachelor- und Masterseminar zur WT: Mathematische Biologie  
SoSe 2012

Dozent: Prof. Dr. Matthias Löwe, Dr. Felipe Torres Tapia

Fachbereich Mathematik und Informatik  
Institut für Mathematische Statistik  
Westfälische Wilhelms-Universität Münster

# Inhaltsverzeichnis

<b>1</b>	<b>Biologische Motivation</b>	<b>1</b>
<b>2</b>	<b>Boolesche Netzwerke als Model für regulatorische genetische Netzwerke</b>	<b>2</b>
2.1	Ein Beispiel . . . . .	4
2.2	Dynamik von Booleschen Netzwerken . . . . .	5
<b>3</b>	<b>Boolesche Netzwerke mit zufälligen Regeln</b>	<b>7</b>
3.1	Ein Beispiel . . . . .	9
3.2	Dynamik von Booleschen Netzwerken mit zufälligen Regeln . . . . .	11
<b>A</b>	<b>Beweis der Gleichung (7)</b>	<b>13</b>

# 1 Biologische Motivation

In vielzelligen Organismen mit spezialisierten Zellen und Geweben enthält jede Zelle das gleiche genetische Material. Damit die Entwicklung normal verläuft und jede Zelle ihre korrekte spezialisierte Funktion übernimmt und aufrecht erhält, müssen bestimmte Proteine zum richtigen Zeitpunkt und in den richtigen Zellen synthetisiert werden. Anders als die DNA-Replikation, die in jeder Zelle allgemein nach einem Alles-oder-Nichts-Prinzip erfolgt, ist die Genexpression hochgradig selektiv.[1]

Ein bekanntes Beispiel hierfür ist die Zell-Zyklus-Regulation. Während des Zellzyklus durchläuft die Zelle verschiedene Phasen, die Verdopplung und Teilung der Erbinformationen und die zufällige Aufteilung der übrigen Zellbestandteile, um zwei genetisch identische Tochterzellen zu bilden. Damit die Zelle von der G1-Phase in die S-Phase wechselt, in der das genetische Material verdoppelt wird, muss das Retinoblastoma (Rb) Protein gehemmt werden. Das Rb Protein wird durch Cyclin E und Cyclin Dependent Kinase 2 (cdk2) phospharesiert und somit inaktiviert. Cdk2 und Cyklin E werden wiederum durch die folgenden Komplexe reguliert: Cdk Activating Kinase (CAK) hat eine aktivierende Wirkung und ist ein Komplex, der aus Cyclin H und cdk7 besteht. Diese beiden Proteine aktivieren somit cdk2 und Cyclin E. Die Komplexe p21 und WAF1 hemmen cdk2 und Cyclin E und werden selbst durch p53 aktiviert. Das Protein p53 wiederum kann Cyclin H inaktivieren. Abbildung 1 illustriert die Effekte der einzelnen Komplexe auf die DNA-Synthese schematisch.

Hierbei wird bereits deutlich, dass negative Regulation ein wichtiger Abwehrmechanismus in der Zelle ist. Wenn beispielsweise Zellen Mutagenen wie Strahlung oder hohen Temperaturen ausgesetzt sind, kann die DNA beschädigt werden. Dann ist es sinnvoll, den Schaden zu beheben, bevor die DNA repliziert wird und das beschädigte genetische Material in die nächste Generation weitergegeben wird.

In dem obigen Beispiel löst defekte DNA eine Kaskade aus, die p53 aktiviert, welches wiederum p21 und WAF1 aktiviert. Die Proteine p21 und WAF1 hemmen cdk2 und Cyclin E, Rb wird aktiviert und die DNA Synthese somit gestoppt.

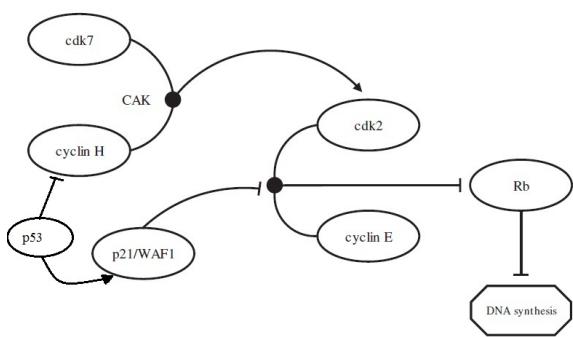


Abbildung 1: Die Zell-Zyklus-Regulation, wobei das stumpfe Ende eine negative Regulation und der Pfeil eine positive Regulation impliziert

## 2 Boolesche Netzwerke als Model für regulatorische genetische Netzwerke

**Definition 1. Boolesche Algebra (nach Huntington)**

Eine Boolesche Algebra ist eine Menge  $\mathcal{A}$  mit Verknüpfungen auf  $\mathcal{A}$ , so dass für alle Elemente  $a, b, c \in \mathcal{A}$  gilt

- (i) Kommutativität:  $a \wedge b = b \wedge a$
- (ii) Distributivität:  $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$
- (iii) es ex. neutrale Elemente:  $\exists 0 \in \mathcal{A}, 1 \in \mathcal{A}: a \wedge 1 = a, a \vee 0 = a$
- (iv) es ex. Komplemente:  $\forall a \in \mathcal{A} \exists \neg a \in \mathcal{A}: a \wedge \neg a = 0$

Eine Boolesche Algebra ist also einen algebraischen Struktur, welche die Eigenschaften der logischen Operatoren  $\wedge$ ,  $\vee$  und  $\neg$ , sowie die Eigenschaften der mengentheoretischen Verknüpfungen  $\cap$ ,  $\cup$  und  $\setminus$  verallgemeinert.

**Definition 2. Boolesche Funktion**

Boolesche Funktionen sind von der Form  $F: \mathcal{A}^m \rightarrow \mathcal{A}^1$ , wobei  $\mathcal{A}$  eine Boolesche Algebra ist.

Boolesche Funktionen sind also in Ausdrücke der Booleschen Algebra einsetzbar und können wie Variablen behandelt werden.

**Definition 3. Boolesches Netzwerk**

Ein Boolesches Netzwerk  $G(V, F)$  ist definiert durch eine Menge von Knoten  $V = \{x_1, \dots, x_n\}$  und eine Liste Boolescher Funktionen  $F = \{f_1, \dots, f_n\}$ . Jedes  $x_i \in \{0, 1\}$  mit  $i = 1, \dots, n$  ist eine Binärvariable, deren Wert nach einem Zeitschritt nur durch die Werte der anderen Knoten  $x_{i_1}, \dots, x_{i_k}$ ,  $1 \leq k \leq n$  und eine Boolesche Funktion  $f_i \in F$  bestimmt wird, d.h.

$$x_i(t+1) = f_i(x_{i_1}(t), \dots, x_{i_k}(t)) \quad \text{für } i = 1, \dots, n.$$

Im Folgenden nehmen wir zur Vereinfachung an, dass der Zustand eines Genes von den Zuständen aller anderen Gene abhängt, also  $k \equiv n$  gilt.

**Definition 4. fictitious, essentiell**

Für eine Funktion  $f$  heißt die Variable  $x_i$  fictitious, falls

$$f(x_i, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) = f(x_i, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n)$$

gilt. Eine Variable, die nicht fictitious ist, heißt essentiell.

Mit der obigen Definition repräsentiert jeder Knoten  $x_i$  den Zustand von Gen  $i$  mit

$$x_i = 1 \Leftrightarrow \text{Gen } i \text{ wird exprimiert}$$

$$x_i = 0 \Leftrightarrow \text{Gen } i \text{ wird nicht exprimiert.}$$

Die Liste der Booleschen Funktionen repräsentiert die Regeln der Interaktionen zwischen den Genen.

Um die dynamischen Eigenschaften dieser Netzwerke zu verstehen, ist es sinnvoll so genannte *Schaltpläne*  $G'(V', F')$  zu betrachten. Dazu konstruieren wir nun  $n$  zusätzliche Knoten  $x'_1, \dots, x'_n$ . Für jedes  $i = 1, \dots, n$  können wir eine Verbindung von  $x_j$  zu  $x'_i$  zeichnen für alle  $j = 1, \dots, n$ . Damit wird die Menge der Knoten zu  $V' = \{x_1, \dots, x_n, x'_1, \dots, x'_n\}$  und  $F'$  entspricht  $F$ , wobei allerdings die Funktionen den Punkten  $x'_1, \dots, x'_n$  zugeordnet sind. Die Funktionen in  $F'$ , die den  $x_1, \dots, x_n$  zugeordnet sind, sind die trivialen Identitätsfunktionen, also  $f(x_i) = x_i$ . Mathematisch ausgedrückt

$$x'_i = f_i(x_1, \dots, x_n)$$

Für  $n = 3$  lässt sich der Übergang von Zeitpunkt  $t$  zu  $t + 1$  der Zustände der Gene wie folgt in einem Schaltplan darstellen.

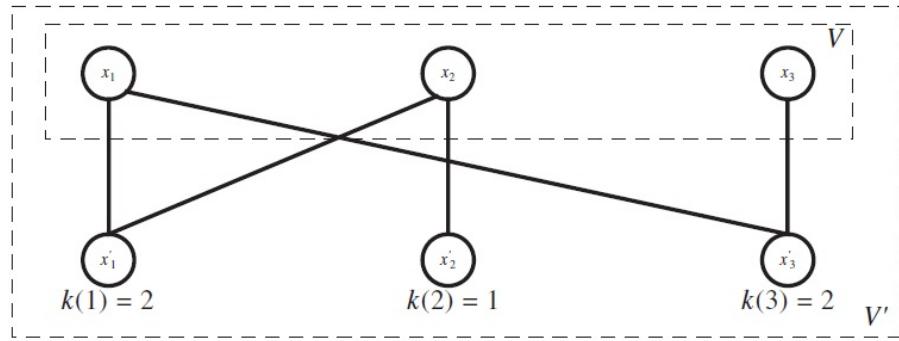


Abbildung 2: Ein Beispiel für einen Schaltplan für  $n = 3$

Insgesamt bilden die Zustände der einzelnen Gene des Genoms zu einem bestimmten Zeitpunkt die so genannte *Genexpressionsanalyse*.

Betrachten wir den Zustandsraum eines Booleschen Netzwerkes mit  $n$  Genen, so gibt es  $2^n$  mögliche Genexpressionsanalysen. Jedes dieser Genexpressionsanalysen hat eine Nachfolger Genexpressionsanalyse im nächsten Zeitschritt, das durch die Regeln der Booleschen Funktionen bestimmt ist.

Häufig ist die Dynamik solcher Systeme gerichtet, d.h. auf lange Sicht geht das System in ein *Einzugsgebiet* oder auch einen *Attraktor* über. Wechselt das System periodisch

zwischen verschiedenen Einzugsgebieten, so spricht man von *Grenzzyklen*.

Die Menge aller Zustände des Systems, die in das selbe Einzugsgebiet übergehen, bilden das so genannten *Bassin des Attraktors*.

Diese große Menge an möglichen Genexpressionsanalysen schließt eine computerbasierte Analyse allerdings nicht aus, da die meisten Booleschen Funktionen nur wenige essentielle Variablen haben und somit nur eine geringe Zahl an Genexpressionsanalysen zu dem Einzugsgebiet korrespondiert.

## 2.1 Ein Beispiel

Wir betrachten ein Boolesches Netzwerk mit zwei Genen, also  $V = \{x_1, x_2\}$  und der Liste der Booleschen Funktionen  $F = \{f_1, f_2\}$ . Diese Funktionen sind definiert durch

$$f_1(x) = x_1 \wedge x_2 \quad \text{und} \quad f_2(x) = x_1$$

mit  $x = (x_1, x_2) \in \{0, 1\}^2$ . Weiter sei die anfängliche Verteilung  $D^0 = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$  gegeben. Der Zustandsraum dieses Beispiels ist  $\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ .

Wir wollen nun eine Übergangsmatrix  $\mathbf{A} = (\mathbf{A}_{ij})_{i,j=1,\dots,4} \in M^{4 \times 4}$  bestimmen, so dass

$$D^1 = D^0 \cdot \mathbf{A}$$

gilt. Betrachten wir die Funktion  $f: \{0, 1\}^2 \rightarrow \{0, 1\}^2$  mit  $x \mapsto (f_1(x), f_2(x))$ , so ergibt sich

$$\begin{aligned} f((0, 0)) &= (0, 0) & 1 &\mapsto 1 \\ f((0, 1)) &= (0, 1) & 2 &\mapsto 2 \\ f((1, 0)) &= (0, 0) & 3 &\mapsto 1 \\ f((1, 1)) &= (1, 1) & 4 &\mapsto 4. \end{aligned}$$

Numerieren wir nun die möglichen Zustände durch, so führt die Funktion  $f$  zum Übergang beispielsweise von Zustand 1 zu 1. Dies führt dann zur Übergangsmatrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Mit Hilfe der Matrix  $\mathbf{A}$  kann nun die Wahrscheinlichkeitsverteilung nach einem Zeitschritt bestimmt werden, d.h.

$$D^1 = D^0 \cdot \mathbf{A} = \left[ \frac{1}{2}, \frac{1}{4}, 0, \frac{1}{4} \right].$$

Also tritt nach einem Zeitschritt mit Wahrscheinlichkeit 0.5 der Zustand  $(0, 0)$  ein und wir erhalten das folgende Übergangsdiagramm, in dem jeder Übergang mit Wahrscheinlichkeit 1 auftritt.

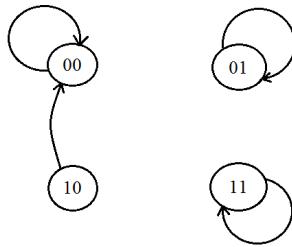


Abbildung 3: Das entsprechende Übergangsdiagramm

## 2.2 Dynamik von Booleschen Netzwerken

Da Boolesche Netzwerke rein deterministisch sind, ist auch ihre Dynamik rein deterministisch. Der einzige zufällige Anteil ist die Wahl der Anfangswerte.

Wir wollen nun die Dynamik von Booleschen Netzwerken probabilistisch beschreiben. Dazu wählen wir ein Boolesches Netzwerk  $G(V, F)$  mit  $n$  Genen bzw. Knoten  $x_1, \dots, x_n$  und eine anfängliche multivariate Verteilung  $D(x)$  mit  $x \in \{0, 1\}^n$ , welche die Wahrscheinlichkeit für jeden möglichen Zustand angibt.

Wir wollen nun die multivariate Verteilung nach einem Zeitschritt bestimmen. Für diese gilt:

$$\mathbb{P}(f_1(x) = i_1, \dots, f_n(x) = i_n) = \sum_{x \in \{0, 1\}^n : f_k(x) = i_k, k=1, \dots, n} D(x) \quad (1)$$

mit  $x = (x_1, \dots, x_n)$  und  $i_k \in \{0, 1\}^n$  für  $k = 1, \dots, n$ . Die Gleichung (1) kann iterativ benutzt werden, um die Verteilung  $\mathbb{P}(f_k(x) = i_k, k = 1, \dots, n)$  für beliebige Zeitschritte zu bestimmen. Damit erhalten wir als iteratives System

$$D^{t+1} = \Psi(D^t) \quad (2)$$

mit der Abbildung  $\Psi: [0, 1]^{2^n} \rightarrow [0, 1]^{2^n}$ , die durch die Gleichung (1) definiert ist.

Wir wollen nun anstelle der Abbildung  $\Psi$  eine Matrix  $A$  verwenden.

Seien  $D^t$  und  $D^{t+1}$  dargestellt durch zwei  $1 \times 2^n$  Vektoren, welche die multivariate

Verteilung enthalten und  $A$  sei eine  $2^n \times 2^n$  Binärmatrix, definiert durch

$$A_{i,j} = \begin{cases} 1 & , \text{ wenn } \exists x \in \{0,1\}^n : C(f_1(x), \dots, f_n(x)) = j, C(x_1, \dots, x_n) = i \\ 0 & , \text{ sonst} \end{cases}$$

mit  $C(i_1, \dots, i_n) = 1 + \sum_{j=1}^n 2^{n-j} \cdot i_j$  und  $i_j \in \{0,1\}$ .

Hierbei ordnet  $C$  jedem Binärvektor eindeutig eine ganze Zahl zu und der Eintrag  $A_{i,j}$  der Matrix  $A$  gibt an, ob ein Übergang von  $(x_1, \dots, x_n)$  zu  $(f_1(x), \dots, f_n(x))$  statt findet, wenn  $(x_1, \dots, x_n)$  der Zahl  $i$  und  $(f_1(x), \dots, f_n(x))$  der Zahl  $j$  zugeordnet ist. Diese Darstellung ist sehr effizient, da die Matrix  $A$  genau einen Eintrag pro Reihe enthält, der ungleich 0 ist.

Somit kann die Gleichung (2) auch geschrieben werden als

$$\begin{aligned} D^{t+1} &= D^t \cdot A \\ &= D^0 \cdot A^{t+1} \end{aligned} \tag{3}$$

mit anfänglicher Verteilung  $D^0 = D(x)$ .

Gleichung (3) ist die Darstellung einer Markovkette, wobei die Übergangsmatrix  $A$  eine Binärmatrix ist. Dies ist sinnvoll, da der Übergang von einem Zustand in den anderen nur durch die Booleschen Funktionen bestimmt wird und die Übergangswahrscheinlichkeiten entweder 0 oder 1 sind.

### 3 Boolesche Netzwerke mit zufälligen Regeln

Da es viele potentielle Funktionen für ein Gen gibt, kann nicht einfach eine dieser Funktionen ausgewählt werden. Die Lösung dieses Problems ist es, das Konzept des Booleschen Netzwerkes um eine wahrscheinlichkeitstheoretische Komponente zu erweitern. Die Idee hierbei ist es mehr als eine mögliche Funktion für jeden Knoten zu zulassen.

So beschreibt nun

$$F_i = \{f_j^{(i)}\}_{j=1,\dots,l(i)}$$

die Menge der möglichen Funktionen, die den Wert des Gens  $i$  bestimmen, wobei  $l(i)$  die Anzahl der möglichen Funktionen für das Gen  $i$  ist.

Die Funktionen  $f_j^{(i)}$  werden meist als *Prädiktor* bezeichnet, da diese Funktionen durch Messungen oder Schätzungen des minimalen Fehlers bestimmt werden.

Eine Realisierung des Booleschen Netzwerkes mit zufälligen Regeln zu einer bestimmten Zeit wird durch einen Vektor von Booleschen Funktionen definiert. Für  $N$  mögliche Realisierungen bei  $n$  Genen gibt es  $N$  Vektorfunktionen  $\mathbf{f}_1, \dots, \mathbf{f}_N$  der Form  $\mathbf{f}_k = (f_{k_1}^{(1)}, \dots, f_{k_n}^{(n)})$  mit  $k = 1, \dots, N$ ,  $1 \leq k_i \leq l(i)$  und  $f_{k_i}^{(i)} \in F_i$  für  $i = 1, \dots, n$ .  $\mathbf{f}_k : \{0, 1\}^n \rightarrow \{0, 1\}^n$  ist also eine Übergangsfunktion, die eine mögliche Realisierung des gesamten Booleschen Netzwerkes mit zufälligen Regeln angibt.

Bei gegebenen Werten aller Gene  $(x_1, \dots, x_n)$  gibt  $\mathbf{f}_k(x_1, \dots, x_n) = (x'_1, \dots, x'_n)$  die Werte der Gene nach einem Zeitschritt an.

#### **Definition 5. zugrundeliegender Maß-Raum und Markovketten**

Der zugrundeliegende Maß-Raum ist  $(\Omega, \mathcal{B})$  mit  $\Omega = \{0, 1\}^n$  und  $\mathcal{B} = \mathcal{P}(\{0, 1\}^n)$ .

Sei nun  $X = (X_1, \dots, X_n)$  eine Folge von Zufallsvariablen, welche die einzelnen Gene  $1, \dots, n$  repräsentieren und  $\Omega$  ist der Zustandsraum.

Wir machen die Annahme, dass  $X$  eine Markovkette ist, also gilt

$$\begin{aligned} \mathbb{P}\{X_{t+1} = (x'_1, \dots, x'_n) | X_0 = a_0, \dots, X_{t-1} = a_{t-1}, X_t = (x_1, \dots, x_n)\} = \\ \mathbb{P}\{X_{t+1} = (x'_1, \dots, x'_n) | X_t = (x_1, \dots, x_n)\} \end{aligned}$$

mit  $a_0, \dots, a_{t-1} \in \{0, 1\}$ .

Sei von nun an  $\mathbf{f} = (f^{(1)}, \dots, f^{(n)})$  ein beliebiger Vektor.  $\mathbf{f}$  kann also alle möglichen Realisierungen des Booleschen Netzwerkes mit zufälligen Regeln annehmen.

Dann ist die Wahrscheinlichkeit, dass  $f_j^{(i)}$  benutzt wird, um Gen  $i$  zu bestimmen, gegeben durch

$$c_j^{(i)} = \mathbb{P}(f^{(i)} = f_j^{(i)}) = \sum_{k: f_{k_i}^{(i)} = f_j^{(i)}} \mathbb{P}(\mathbf{f} = \mathbf{f}_k). \quad (4)$$

Da die  $c_j^{(i)}$  Wahrscheinlichkeiten sind, muss gelten:

$$\sum_{j=1}^{l(i)} c_j^{(i)} = 1. \quad (5)$$

Generell muss die Auswahl der Booleschen Funktionen nicht unabhängig sein. Allerdings nehmen wir dies der Einfachheit halber hier an, so dass

$$\mathbb{P}(f^{(i)} = f_j^{(i)}, f^{(h)} = f_k^{(h)}) = \mathbb{P}(f^{(i)} = f_j^{(i)}) \cdot \mathbb{P}(f^{(h)} = f_k^{(h)})$$

für  $i, h = 1, \dots, n$  und  $j = 1, \dots, l(i)$  und  $k = 1, \dots, l(h)$  gilt. Unter der Annahme der Unabhängigkeit ist dann die Anzahl der möglichen Realisierungen gegeben durch  $N = \prod_{i=1}^n l(i)$ .

Ist  $l(i) = 1$  für alle  $i = 1, \dots, n$ , so ist auch  $N = 1$  und das Boolesche Netzwerke mit zufälligen Regeln wird zu einem Standard Booleschen Netzwerk.

Im Grunde ist die Dynamik von Booleschen Netzwerken mit zufälligen Regeln die gleiche, wie die der Booleschen Netzwerke, außer dass der Wert eines Knotens durch einen möglichen Prädiktor bestimmt wird, der nach seiner zugehörigen Wahrscheinlichkeit ausgewählt wird. In anderen Worten haben wir zu jedem Zeitpunkt eins aus  $N$  möglichen Netzwerken.

Um wie in 2.2 vorgehen zu können, müssen wir die Wahrscheinlichkeit bestimmen, dass ein bestimmtes Netzwerk ausgewählt wird. Dies ist mit der Matrix  $\mathbf{K}$  möglich, die alle möglichen Kombinationen der  $f_j^{(i)}$  enthält:

$$\mathbf{K} = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 1 & l(n) \\ \vdots & \vdots & \ddots & 2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 2 & l(n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ l(1) & l(2) & \dots & l(n-1) & l(n) \end{pmatrix} \in \mathbf{M}^{(N \times n)}.$$

Hierbei korrespondiert jede Reihe von  $\mathbf{K}$  zu einer möglichen Netzstruktur und der Eintrag  $\mathbf{K}_{ij}$  besagt, dass  $f_j^{(i)}$  für das Gen  $i$  benutzt wird.

Damit lässt sich die Wahrscheinlichkeit, dass Netzwerk  $i$  ausgewählt wird, schreiben

als

$$\begin{aligned} P_i &= \mathbb{P}(\text{Netzwerk } i \text{ wird gewählt}) \\ &= \prod_{j=1}^n c_{\mathbf{K}_{ij}}^{(j)}. \end{aligned} \tag{6}$$

Auch hierbei gilt  $\sum_{i=1}^N P_i = 1$ , denn

$$\sum_{i=1}^N P_i \stackrel{(6)}{=} \sum_{i=1}^N \prod_{j=1}^n c_{\mathbf{K}_{ij}}^{(j)} = \prod_{j=1}^n \sum_{i=1}^{l(j)} c_i^{(j)} \stackrel{(5)}{=} 1. \tag{7}$$

Nun wollen wir den Zustandsraum des Booleschen Netzwerkes mit zufälligen Regeln betrachten, der aus  $2^n$  Zuständen besteht. Dazu wollen wir ähnlich wie in 2.2 die Übergangsmatrix  $\mathbf{A}$  bestimmen. Im Gegensatz zu den Booleschen Netzwerken, kann das Netzwerk nun von einem Zustand in eine Zahl möglicher Zustände übergehen, beinhaltet also einen zufälligen Prozess.

Die Wahrscheinlichkeit für einen Übergang von  $(x_1, \dots, x_n)$  zu  $(x'_1, \dots, x'_n)$  ist dann gegeben durch

$$\begin{aligned} \mathbf{P}\{(x_1, \dots, x_n) \rightarrow (x'_1, \dots, x'_n)\} &= \\ \sum_{i: f_{\mathbf{K}_{i1}}^{(1)}(x_1, \dots, x_n) = x'_1, \dots, f_{\mathbf{K}_{in}}^{(n)}(x_1, \dots, x_n) = x'_n} P_i &= \\ \sum_{i=1}^N P_i \left[ \prod_{j=1}^n \left( 1 - |f_{\mathbf{K}_{ij}}^{(j)}(x_1, \dots, x_n) - x'_j| \right) \right], \end{aligned} \tag{8}$$

wobei die Werte von  $f_j^{(i)}$  und  $x'_j$  als Dezimalzahlen benutzt werden. Benutzen wir die Gleichungen (7) und (8) und die Tatsache, dass für jedes  $(x_1, \dots, x_n)$  ein  $(x'_1, \dots, x'_n)$  und ein  $i$  existieren, so dass  $\prod_{j=1}^n \left( 1 - |f_{\mathbf{K}_{ij}}^{(j)}(x_1, \dots, x_n) - x'_j| \right) = 1$  gilt, so erhalten wir insgesamt

$$\sum_{j=1}^{2^n} \mathbf{A}_{ij} = 1$$

für jedes  $i = 1, \dots, 2^n$ . Also ist  $\mathbf{A}$  eine Markovmatrix und das Boolesche Netzwerk mit zufälligen Regeln ist ein homogener Markovprozess. Dies bedeutet, dass sich die Übergangswahrscheinlichkeiten mit der Zeit nicht ändern.

### 3.1 Ein Beispiel

Wir betrachten ein Boolesches Netzwerk mit zufälligen Regeln mit drei Genen, also  $V = \{x_1, x_2, x_3\}$ , und einer Funktionenmenge  $F = (F_1, F_2, F_3)$  mit  $F_1 = \{f_1^{(1)}, f_2^{(1)}\}$ ,

$F_2 = \{f_1^{(2)}\}$  und  $F_3 = \{f_1^{(3)}, f_2^{(3)}\}$ . Die einzelnen Funktionen sind definiert durch die folgende Wahrheitstabelle

$x_1x_2x_3$	$f_1^{(1)}$	$f_2^{(1)}$	$f_1^{(2)}$	$f_1^{(3)}$	$f_2^{(3)}$
000	0	0	0	0	0
001	1	1	1	0	0
010	1	1	1	0	0
011	1	0	0	1	0
100	0	0	1	0	0
101	1	1	1	1	0
110	1	1	0	1	0
111	1	1	1	1	1
$c_j^{(i)}$	0.6	0.4	1	0.5	0.5

Tabelle 1: Die Werte der Booleschen Funktionen in Form einer Wahrheitstabelle

Da wir zwei Funktionen für  $x_1$ , eine Funktion für  $x_2$  und zwei Funktionen für  $x_3$  haben, gibt es  $N = 2 \cdot 1 \cdot 2 = 4$  mögliche Kombinationen dieser Funktionen und somit vier mögliche Netzwerke. Die Matrix  $\mathbf{K}$  ist dann gegeben durch:

$$\mathbf{K} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \end{pmatrix}.$$

Beispielsweise die zweite Zeile der Matrix  $\mathbf{K}$  bedeutet, dass die Prädiktoren  $(f_1^{(1)}, f_1^{(2)}, f_2^{(3)})$  benutzt werden. Mit Hilfe von (8) kann nun die Matrix  $\mathbf{A}$  bestimmt werden als:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ P_4 & P_3 & 0 & 0 & P_2 & P_1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & P_2 + P_4 & P_1 + P_3 \\ 0 & 0 & 0 & 0 & P_2 + P_4 & P_1 + P_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Um die Konstruktion der Übergangsmatrix  $\mathbf{A}$  zu verdeutlichen betrachten wir beispielsweise die Übergangswahrscheinlichkeit  $\mathbb{P}\{(1, 1, 0) \rightarrow (1, 0, 0)\}$ , die zum Eintrag  $\mathbf{A}_{7,5}$  korrespondiert, da  $(1, 1, 0)$  der siebte und  $(1, 0, 0)$  der fünfte Eintrag in der obigen Tabelle ist. Um nun  $\mathbf{A}_{7,5}$  zu bestimmen, betrachten wir  $(x_1, x_2, x_3) = (1, 1, 0)$  in der

obigen Tabelle und suchen nach möglichen Kombinationen der Prädiktoren, die den Wert  $(1, 0, 0)$  liefern. Die beiden Möglichkeiten  $(f_1^{(1)}, f_1^{(2)}, f_2^{(3)})$  und  $(f_2^{(1)}, f_1^{(2)}, f_2^{(3)})$ , die dies erfüllen, korrespondieren zu der zweiten und zu der vierten Zeile der Matrix  $\mathbf{K}$ . Also gilt

$$\mathbb{P}\{(1, 1, 0) \rightarrow (1, 0, 0)\} = P_2 + P_4.$$

Analog folgen alle anderen Werte für  $\mathbf{A}$ . Insgesamt erhalten wir nun das folgende Übergangsdiagramm

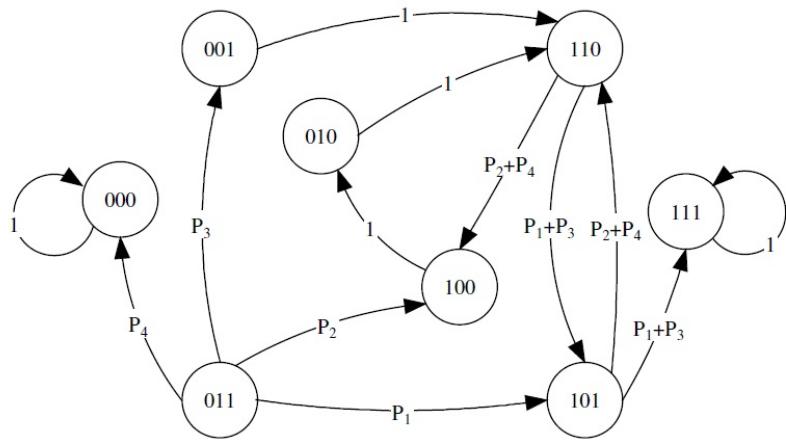


Abbildung 4: Das entsprechende Übergangsdiagramm

### 3.2 Dynamik von Booleschen Netzwerken mit zufälligen Regeln

Bis jetzt haben wir gesehen, dass das Verhalten von Booleschen Netzwerken mit zufälligen Regeln durch Markovketten beschrieben werden kann. Dabei wird die Übergangsmatrix  $\mathbf{A}$  eindeutig durch die gegebenen Booleschen Funktionen und ihre Wahrscheinlichkeiten definiert. Auf Grund dessen ist die Theorie über das Grenzverhalten von Markovketten anwendbar.

Um das Grenzverhalten von Booleschen Netzwerken mit zufälligen Regeln zu verdeutlichen, betrachten wir das Beispiel 3.1. Wie in Tabelle 1 ersichtlich ist, sind die Wahrscheinlichkeiten für die Prädiktoren  $c_1^{(1)} = 0.6$ ,  $c_2^{(1)} = 0.4$ ,  $c_1^{(2)} = 1$ ,  $c_1^{(3)} = 0.5$  und  $c_2^{(3)} = 0.5$ . Mit Hilfe der Gleichung (6) lassen sich die Wahrscheinlichkeiten für die vier verschiedenen Netzwerke bestimmen, beispielsweise  $P_2$  durch

$$P_2 = \mathbb{P}(\text{Netzwerk 2 wird gewählt}) = \prod_{j=1}^4 c_{\mathbf{K}_2,j}^{(j)} = c_1^{(1)} \cdot c_1^{(2)} \cdot c_2^{(3)} = 0.6 \cdot 1 \cdot 0.5 = 0.3$$

Analog ergeben sich die übrigen Wahrscheinlichkeiten  $P_1 = 0.3$ ,  $P_3 = 0.2$  und  $P_4 = 0.2$ . Mit der anfänglichen multivarianten Verteilung  $D^0 = [\frac{1}{8}, \dots, \frac{1}{8}]$  erhalten wir durch

Einsetzen der obigen Werte in die Matrix  $\mathbf{A}$  und iterieren von Gleichung (3) die folgende Grenzwahrscheinlichkeit

$$\pi = [0.15, 0, 0, 0, 0, 0, 0, 0.85]$$

Dies bedeutet, dass mit Wahrscheinlichkeit 0.15 alle drei Gene auf lange Sicht ausgeschaltet (000) sind, oder mit Wahrscheinlichkeit 0.85 alle aktiviert (111) sind.

Diese beiden Zustände werden als *absorbierend* bezeichnet.

Diese Grenzwahrscheinlichkeit lässt sich durch das Übergangsdiagramm in Abbildung 4 verifizieren, denn hier wird ersichtlich, dass wenn einer dieser beiden Zustände erreicht wird, dieser auch nicht mehr verlassen wird.

Diese Eigenschaft von Booleschen Netzwerken mit zufälligen Regeln korrespondiert zu dem Konzept des Einzugsbereichs bei Booleschen Netzwerken, ebenso wie die irreduziblen Markovketten zu den Grenzzyklen korrespondieren. Boolesche Netzwerke mit zufälligen Regeln besitzen also die gleichen dynamischen Eigenschaften, wie normale Boolesche Netzwerke, können allerdings mit Unsicherheiten umgehen.

Ob und welche Gleichgewichtsverteilung ein Boolesches Netzwerk mit zufälligen Regeln annimmt, hängt von der anfänglichen Multivarianten Verteilung  $D^0$  ab.

## A Beweis der Gleichung (7)

Zeige:  $\sum_{i=1}^N \prod_{j=1}^n c_{\mathbf{K}_{i,j}}^{(j)} = \prod_{j=1}^n \sum_{i=1}^{l(j)} c_i^{(j)}$  (\*)

Beweis per Induktion nach  $n$

IA:  $n = 1$

Definiere  $N := l(1)$ , dann gilt

$$\sum_{i=1}^N \prod_{j=1}^n c_{\mathbf{K}_{i,j}}^{(j)} = \sum_{i=1}^{l(1)} c_{\mathbf{K}_{i,1}}^{(1)} = \sum_{i=1}^{l(1)} c_i^{(1)} = \prod_{j=1}^n \sum_{i=1}^{l(1)} c_i^{(1)}.$$

IV: (\*) gilt für bel. festes  $n$  und wir haben

$$\mathbf{K} = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 1 & l(n) \\ \vdots & \vdots & \ddots & 2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 2 & l(n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ l(1) & l(2) & \dots & l(n-1) & l(n) \end{pmatrix} \in M^{N \times n}, N = \prod_{i=1}^n l(i)$$

IS:  $n \mapsto n + 1$

Wir haben die Matrix  $\mathbf{K}'$ , die gegeben ist durch

$$\mathbf{K}' = \left( \begin{array}{c|c} K & \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \\ \hline K & \begin{matrix} 2 \\ \vdots \\ 2 \end{matrix} \\ \hline \vdots & \vdots \\ \hline K & \begin{matrix} l(n+1) \\ \vdots \\ l(n+1) \end{matrix} \end{array} \right) \in M^{N' \times (n+1)}$$

mit  $N' := N \cdot l(n + 1)$ , also

$$\mathbf{K}'_{i,j} = \begin{cases} \mathbf{K}_{m,j} & , \text{ wenn } j \leq n \text{ mit } m = i \pmod{N} \\ m & , \text{ wenn } j = n + 1 \text{ mit } i \in \{m \cdot 1, \dots, m \cdot N\} \end{cases}$$

Damit gilt insgesamt

$$\begin{aligned} \prod_{j=1}^{n+1} \sum_{i=1}^{l(j)} c_i^{(j)} &= \left( \sum_{i=1}^{l(n+1)} c_i^{(n+1)} \right) \left( \prod_{j=1}^n \sum_{i=1}^{l(j)} c_i^{(j)} \right) \\ &\stackrel{IV}{=} \left( \sum_{i=1}^{l(n+1)} c_i^{(n+1)} \right) \left( \sum_{k=1}^N \prod_{j=1}^n \underbrace{c_{\mathbf{K}'_{k,j}}^{(j)}}_{=c_{\mathbf{K}'_{k,j}}^{(j)}, \text{ da } j \leq n, k \in \{1, \dots, N\}} \right) \\ &= \sum_{i=1}^{l(n+1)} \sum_{k=1}^N c_i^{(n+1)} \prod_{j=1}^n c_{\mathbf{K}'_{k,j}}^{(j)} \\ &= \sum_{k=1}^N \underbrace{c_1^{(n+1)}}_{=c_{\mathbf{K}'_{k,n+1}}^{(n+1)} \text{ für } k=1, \dots, N} \prod_{j=1}^n c_{\mathbf{K}'_{k,j}}^{(j)} + \dots \\ &\quad + \sum_{k=1}^N \underbrace{c_{l(n+1)-1}^{(n+1)}}_{=c_{\mathbf{K}'_{k,n+1}}^{(n+1)} \text{ für } k=N \cdot (l(n+1)-2)+1, \dots, N \cdot (l(n+1)-1)} \prod_{j=1}^n c_{\mathbf{K}'_{k,j}}^{(j)} \\ &\quad + \sum_{k=1}^N \underbrace{c_{l(n+1)}^{(n+1)}}_{=c_{\mathbf{K}'_{k,n+1}}^{(n+1)} \text{ für } k=N \cdot (l(n+1)-1)+1, \dots, N \cdot l(n+1)} \prod_{j=1}^n c_{\mathbf{K}'_{k,j}}^{(j)} \\ &= \sum_{k=1}^N \prod_{j=1}^{n+1} c_{\mathbf{K}'_{k,j}}^{(j)} + \dots + \sum_{k=N(l(n+1)-2)+1}^{N(l(n+1)-1)} \prod_{j=1}^{n+1} c_{\mathbf{K}'_{k,j}}^{(j)} + \sum_{k=N(l(n+1)-1)+1}^{N \cdot l(n+1)} \prod_{j=1}^{n+1} c_{\mathbf{K}'_{k,j}}^{(j)} \\ &= \sum_{k=1}^{N'} \prod_{j=1}^{n+1} c_{\mathbf{K}'_{k,j}}^{(j)}, \end{aligned}$$

wobei die letzte Gleichung mit der Definition  $N' := N \cdot l(n + 1)$  folgt.  $\square$

## **Abbildungsverzeichnis**

1	Die Zell-Zyklus-Regulation, wobei das stumpfe Ende eine negative Regulation impliziert und der Pfeil eine positive Regulation, Quelle: [2] . . . . .	1
2	Ein Beispiel für einen Schaltplan für $n = 3$ , Quelle: [2] . . . . .	3
3	Übergangsdiagramm für das Beispiel 2.1, Quelle: eigene Darstellung mit Paint in Anlehnung an [2] . . . . .	5
4	Übergangsdiagramm für das Beispiel 3.1, Quelle: [2] . . . . .	11

## **Tabellenverzeichnis**

1	Die Werte der Booleschen Funktionen in Form einer Wahrheitstabelle für das Beispiel 3.1 . . . . .	10
---	---	----

## **Literatur**

- [1] W.K. Purves, D. Sadava, G.H. Orians, H.C. Heller: Biologie, Spektrum Verlag, 7. Auflage 2006
- [2] I. Shmulevich, E.R. Dougherty, S. Kim, W. Zhang: Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks, Bioinformatics, Vol. 18, Nr. 2 2002, S. 261-274
- [3] H. De Jong: Modeling and Simulation of Genetic Regulatory Systems: A Literature Review, Journal of Computational Biology, Vol. 9, Nr.2 2002, S. 67-103