

## Kapitel 4

# Markov-Basen für Kontingenztests

Zu den Grundaufgaben der Statistik gehört auch die Frage nach dem Zusammenhang zwischen zwei oder mehreren Zufallsvariablen. Um beispielsweise zu testen, ob ein Medikament besser ist als ein anderes, teilt man die Probanden zufallsgesteuert ein in zwei Kontrollgruppen, die mit je einem der beiden Medikamente behandelt werden, und mißt den Erfolg. Falls die beiden Zufallsvariablen *Medikament* und *Erfolg* voneinander unabhängig sind, haben beide Medikamente gleich viel (oder wenig) Erfolg, andernfalls läßt sich mit einer gewissen Wahrscheinlichkeit eines der beiden als das bessere identifizieren.

### §1: Kontingenztafeln

Für zwei unabhängige Zufallsvariablen  $X$  und  $Y$  mit Werten in  $\{1, \dots, r\}$  bzw.  $\{1, \dots, c\}$  ist  $p(X = x, Y = y) = p(X = x)p(Y = y)$  für alle  $(x, y) \in \{1, \dots, r\} \times \{1, \dots, c\}$ . Wegen der Beziehungen

$$p(X = x) = \sum_{y=1}^c p(X = x, Y = y)$$

und

$$p(Y = y) = \sum_{x=1}^r p(X = x, Y = y)$$

können wir das auch so ausdrücken, daß  $p(X = x, Y = y)$  nur abhängt von den beiden Summen

$$\sum_{y=1}^c p(X = x, Y = y) \quad \text{und} \quad \sum_{x=1}^r p(X = x, Y = y).$$

Um die Unabhängigkeit der beiden Variablen experimentell zu prüfen, wählen wir eine natürliche Zahl  $n$  und betrachten  $n$  Werte  $(x_k, y_k)$  der Zufallsvariablen  $X \times Y$ . Für die zugehörige Kontingenztafel definieren wir eine neue Zufallsvariable  $U = (U_{ij})_{\substack{i=1,\dots,r \\ j=1,\dots,c}}$ , wobei  $U_{ij}$  die Anzahl von Paaren  $(x_k, y_k)$  mit  $x_k = i$  und  $y_k = j$  bezeichnet. Außerdem betrachten wir noch die Zufallsvariablen

$$U_{i+} = \sum_{j=1}^c U_{ij} \quad \text{und} \quad U_{+j} = \sum_{i=1}^r U_{ij}$$

für  $i = 1, \dots, r$  und  $j = 1, \dots, c$ . Die Erwartungswerte der Zufallsvariablen  $U_{ij}$  sind dann

$$\mathbb{E}(U_{ij}) = np(X = i, Y = j) = n \cdot \sum_{j=1}^c p(X = i, Y = j) \cdot \sum_{i=1}^r p(X = i, Y = j)$$

Entsprechend sind die Erwartungswerte von  $U_{i+}$  und  $U_{+j}$  gleich

$$\mathbb{E}(U_{i+}) = n \cdot \sum_{j=1}^c p(X = i, Y = j) \quad \text{und} \quad \mathbb{E}(U_{+j}) = n \cdot \sum_{i=1}^r p(X = i, Y = j).$$

Durch Vergleich der Erwartungswerte folgen die Beziehungen

$$\mathbb{E}(U_{ij}) = \frac{\mathbb{E}(U_{i+})\mathbb{E}(U_{+j})}{n},$$

$$\mathbb{E}(U_{i+}) = \sum_{j=1}^c \mathbb{E}(U_{ij}) \quad \text{und} \quad \mathbb{E}(U_{+j}) = \sum_{i=1}^r \mathbb{E}(U_{ij}).$$

Natürlich können wir schon wegen der Ganzzahligkeit der  $u_{ij}$  nicht erwarten, daß für einen beobachteten Wert  $u$  von  $U$  auch  $u_{ij} = u_{i+}u_{+j}/n$  ist, aber die Abweichung zwischen den beiden Seiten sollte mit großer Wahrscheinlichkeit klein sein. Als Maß der Abweichung wählen wir die Zahl

$$\chi^2(u) = \sum_{i=1}^r \sum_{j=1}^c \frac{(u_{ij} - \hat{u}_{ij})^2}{\hat{u}_{ij}} \quad \text{mit} \quad \hat{u}_{ij} = \frac{u_{i+}u_{+j}}{n}.$$

In der Statistik zeigt man, daß die Verteilung der Zufallsvariablen  $\chi^2(U)$  asymptotisch gegen eine  $\chi^2$ -Verteilung mit  $(r-1)(c-1)$  Freiheitsgraden konvergiert, falls alle  $u_{ij}$  gegen unendlich gehen.

Leider können aus praktischen Gründen oft nur Experimente realisiert werden, bei denen zumindest einige der  $u_{ij}$  recht klein sind. Eine Faustregel besagt, daß man definitiv nicht mit der  $\chi^2$ -Verteilung arbeiten sollte, wenn nicht alle  $u_{ij}$  mindestens gleich fünf sind.

## §2: Fishers exakter Test

Das erste Verfahren, die Wahrscheinlichkeit für die Zufälligkeit der Abweichung der Werte von  $U_{ij}$  und  $U_{i+}U_{+j}/n$  auch im Falle kleiner Werte einiger  $u_{ij}$  zu schätzen, war FISHERS exakter Test. Er betrachtet zu einer gegebenen Kontingenztafel  $(u_{ij})$  alle Tafeln  $(v_{ij})$  mit  $v_{ij} \in \mathbb{N}_0$  und  $v_{i+} = u_{i+}$  für alle  $i$  sowie  $v_{+j} = u_{+j}$  für alle  $j$ . Für jede dieser Tafeln berechnet er das zugehörige  $\chi^2$  und schätzt die Wahrscheinlichkeit für die Zufälligkeit der Abweichung als den Anteil aller Tafeln, die zu keinem größeren  $\chi^2$  führen als die gegebene Tabelle  $(u_{ij})$ . Im Falle von Vierfeldertests ist das auch noch bei moderat großen Zeilen- und Spaltensummen praktikabel, denn offensichtlich sind durch diese Summen alle  $v_{ij}$  eindeutig festgelegt, sobald man einen dieser vier Werte kennt. Mit wachsender Zahl der Freiheitsgrade steigt allerdings auch der Aufwand für FISHERS exakten Test dramatisch an, so daß die Betrachtung aller Tabellen mit denselben Randverteilungen wie die gegebene Tabelle nicht mehr mit realistischem Aufwand möglich ist.

## §3: Log-lineare Modelle

Bevor wir uns überlegen, wie wir in solchen Fällen vorgehen können, wollen wir zunächst die betrachtete Situation etwas verallgemeinern. Bei der Untersuchung auf Unabhängigkeit sollten die Erwartungswerte der  $U_{ij}$  nur abhängen von denen der  $U_{i+}$  und der  $U_{+j}$ . Auch im Falle abhängiger Größen kann es sein, daß es eine begrenzte Zahl von Linearkombinationen der  $U_{ij}$  gibt mit der Eigenschaft, daß die Erwartungswerte aller  $U_{ij}$  aus denen dieser Linearkombinationen berechnet werden können. Solche Situationen formalisiert der Begriff eines log-linearen Modells:

**Definition:**  $X_1, \dots, X_m$  seien Zufallsvariablen, und  $X_\ell$  nehme Werte

aus der Menge  $\{1, \dots, r_\ell\}$  an. Für

$$i = (i_1, \dots, i_m) \in \mathcal{R} \stackrel{\text{def}}{=} \{1, \dots, r_1\} \times \dots \times \{1, \dots, r_m\}$$

sei  $p_i = p(X_1 = i_1, \dots, X_m = i_m)$ . Weiter sei  $A \in \mathbb{Z}^{d \times \#\mathcal{R}}$  eine Matrix, deren Spalten alle die gleiche Summe haben. Das log-lineare Modell  $\mathcal{M}_A$  zur Matrix  $A$  ist die Menge aller Wahrscheinlichkeitstabellen  $(p_i)_{i \in \mathcal{R}}$  mit der Eigenschaft, daß der Zeilenvektor  $(\log p_i)_{i \in \mathcal{R}}$  im von den Zeilenvektoren von  $A$  aufgespannten Untervektorraum von  $\mathbb{R}^{\#\mathcal{R}}$  liegt.

Daß wir hier die Logarithmen der Wahrscheinlichkeiten betrachten und nicht diese selbst, liegt natürlich daran, daß bei den hier interessierenden Modellen viele Produkte von Wahrscheinlichkeiten eine Rolle spielen. Durch Übergang zu Logarithmen können wir daraus Summen machen und damit insbesondere auch Methoden aus der Linearen Algebra anwenden.

Betrachten wir als Beispiel das obige Unabhängigkeitsmodell. Hier ist  $\mathcal{R} = \{1, \dots, r\} \times \{1, \dots, c\}$  und

$$p_{(i,j)} = p(X = i, Y = j) = p(X = i) \cdot p(Y = j).$$

Somit ist  $\log p_{(i,j)} = \log p(X = i) + \log p(Y = j)$ .

Ist  $A$  die  $(r+c) \times rc$ -Matrix, deren Spalte mit Index  $(i, j)$  in den Komponenten  $i$  und  $r+j$  eine Eins stehen hat und sonst lauter Nullen, so ist

$$(\log p_{(1,1)}, \dots, \log p_{(r,c)}) = \sum_{i=1}^r \log p(X = i) a^{(1)} + \sum_{j=1}^c \log p(Y = j) a^{(r+j)},$$

wobei  $a^{(\ell)}$  für den  $\ell$ -ten Zeilenvektor von  $A$  steht. Für  $r = 3$  und  $c = 2$  etwa ist

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

und  $(\log p_{(1,1)} \quad \log p_{(1,2)} \quad \log p_{(2,1)} \quad \log p_{(2,2)} \quad \log p_{(3,1)} \quad \log p_{(3,2)})$  ist gleich

$$\begin{aligned}
 & \log p(X = 1) \quad \times \quad (1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0) \\
 + & \log p(X = 2) \quad \times \quad (0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0) \\
 + & \log p(X = 3) \quad \times \quad (0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1) \\
 + & \log p(Y = 1) \quad \times \quad (1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0) \\
 + & \log p(Y = 2) \quad \times \quad (0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1).
 \end{aligned}$$

Wenn wir eine Kontingenztafel wie gerade eben als einen Vektor  $u$  auffassen, gibt uns das Matrixprodukt  $Au$  die sämtlichen Zeilen- und Spaltensummen; beim Unabhängigkeitsmodell hängen die Erwartungswerte der Variablen  $U_{ij}$  nur von diesen ab. Entsprechend erwarten wir bei einem beliebigen log-linearen Modell, daß auch dort die Erwartungswerte der Zufallsvariablen  $U_i$ ,  $i \in \mathcal{R}$ , nur vom Zufallsvariablenvektor  $AU$  abhängen sollten. Sobald wir wissen, wie sich die Erwartungswerte der  $U_i$  aus denen von  $AU$  berechnen lassen, können wir FISHERS exakten Test auch auf diese Situation verallgemeinern, indem wir zu einer gegebenen Tafel  $u$  aller Tafeln  $v = (v_i)_{i \in \mathcal{R}}$  betrachten, für die  $Av = Au$  ist.

**Definition:** Die *Faser*  $\mathcal{F}(u)$  zu einer gegebenen Tafel  $(u_i)_{i \in \mathcal{R}}$  bezüglich des Modells  $\mathcal{M}_A$  ist die Menge aller  $v \in \mathbb{N}_0^{\#\mathcal{R}}$  mit  $Av = Au$ .

Auch hier wird die Faser oft zu groß sein als daß wir  $\chi^2$  für jedes einzelne Element ausrechnen können, so daß wir uns mit einer Stichprobe begnügen müssen, die wir uns über eine Irrfahrt verschaffen wollen. Die einzelnen Schritte der Irrfahrt sind durch Elemente einer sogenannten MARKOV-Basis gegeben:

**Definition:** Eine MARKOV-Basis des log-linearen Modells  $\mathcal{M}_A$  ist eine endliche Menge  $\mathcal{B}$  von Tafeln  $b \in \mathbb{Z}^{\#\mathcal{R}}$  mit  $Ab = 0$ , für die gilt: Für jede Tafel  $u \in \mathbb{N}_0^{\#\mathcal{R}}$  und je zwei Elemente  $v, v' \in \mathcal{F}(u)$  gibt es eine Folge von Elementen  $b_1, \dots, b_L \in \mathcal{B}$ , so daß gilt:

- 1.) Für  $\ell = 1, \dots, L$  ist  $v + b_1 + \dots + b_\ell \in \mathbb{N}_0^{\#\mathcal{R}}$
- 2.)  $v + b_1 + \dots + b_L = v'$

Im Falle des Unabhängigkeitsmodells für zwei Zufallsvariablen können wir leicht eine MARKOV-Basis finden: Wir nehmen einfach alle Tafeln  $b^{(ijk\ell)}$ ,  $i, k = 1, \dots, r$  und  $j, \ell = i, \dots, c$ , deren Einträge allesamt verschwinden mit Ausnahme von

$$b_{ij}^{(ijk\ell)} = b_{k\ell}^{(ijk\ell)} = 1 \quad \text{und} \quad b_{i\ell}^{(ijk\ell)} = b_{kj}^{(ijk\ell)} = -1.$$

Im nächsten Paragraphen wollen wir uns überlegen, daß es für jedes log-lineare Modell eine MARKOV-Basis gibt.

#### §4: Markov-Basen und Ideale

Wir führen für jedes  $i \in \mathcal{R}$  eine Variable  $p_i$  ein und betrachten den Polynomring  $k[p_i | i \in \mathcal{R}]$ . Jeder Tabelle  $u \in \mathbb{N}_0^{\#\mathcal{R}}$  ordnen wir das Monom  $p^u \stackrel{\text{def}}{=} \prod_{i \in \mathcal{R}} p_i^{u_i} \in k[p_i | i \in \mathcal{R}]$  zu.

**Definition:** a) Eine nichtleere Teilmenge  $\mathcal{L} \subset \mathbb{Z}^{\#\mathcal{R}}$  heißt *Gitter*, wenn für je zwei Elemente  $u, v \in \mathcal{L}$  auch  $u + v$  und  $-u$  in  $\mathcal{L}$  liegen.

b) Eine endliche Teilmenge  $\mathcal{B}$  eines Gitters  $\mathcal{L}$  heißt *MARKOV-Basis*, wenn es zu je zwei Elementen  $v, v' \in \mathbb{N}_0^{\#\mathcal{R}}$  mit  $v' - v \in \mathcal{L}$  eine Folge von Elementen  $b_1, \dots, b_L \in \mathcal{B}$  gibt, so daß gilt:

1.) Für  $\ell = 1, \dots, L$  ist  $v + b_1 + \dots + b_\ell \in \mathbb{N}_0^{\#\mathcal{R}}$

2.)  $v + b_1 + \dots + b_L = v'$

c) Das *Gitterideal* zu  $\mathcal{L} \subset \mathbb{Z}^{\mathcal{R}}$  ist

$$I_{\mathcal{L}} \stackrel{\text{def}}{=} (p^u - p^v \mid u, v \in \mathbb{N}_0^{\mathcal{R}} \setminus \{0\}, u - v \in \mathcal{L}).$$

Offensichtlich ist für ein log-lineares Modell  $\mathcal{M}_A$  die Menge  $\mathcal{L}$  aller  $u \in \mathbb{Z}^{\#\mathcal{R}}$  mit  $Au = 0$  ein Gitter, und eine MARKOV-Basis davon ist genau das, was im vorigen Paragraphen definiert wurde.

Für  $u \in \mathbb{Z}^{\mathcal{R}}$  definieren wir  $u^+, u^- \in \mathbb{N}_0^{\mathcal{R}}$  durch

$$u_i^+ = \max(u_i, 0) \quad \text{und} \quad u_i^- = \max(-u_i, 0).$$

Dann ist  $u = u^+ - u^-$ , und dies ist die einzige Darstellung von  $u$  als Differenz zweier Elemente  $v, w \in \mathbb{N}_0^{\mathcal{R}}$ , bei der für kein  $i \in \mathcal{R}$  sowohl  $u_i$  als auch  $v_i$  von Null verschieden sind.

**Satz:** Eine Menge  $\mathcal{B} \subset \mathcal{L}$  ist genau dann eine MARKOV-Basis von  $\mathcal{L}$ , wenn die Polynome  $p^{b^+} - p^{b^-}$  mit  $b \in \mathcal{B}$  das Ideal  $I_{\mathcal{L}}$  erzeugen.

Für den Beweis sei auf §4.4 des Buchs

SATOSHI AOIKI, HISAYUKI HARA, AKIMICHI TAKEMURA: Markov Bases in Algebraic Statistics, *Springer*, 2012

verwiesen.

Da jedes Ideal eines Polynomrings eine endliche GRÖBNER-Basis hat, zeigt dieser Satz die Existenz von MARKOV-Basen.

## §5: Markov-Ketten

Ein häufiger zitiertes Vorbild einer Irrfahrt ist der Heimweg eines Betrunknen über einen Platz mit vielen Laternen. Jedesmal, wenn er sich den Kopf an einer Laterne anschlägt, geht er (mit gleicher Wahrscheinlichkeit) nach rechts oder nach links. Seine Entscheidungen an den verschiedenen Laternen sind unabhängig voneinander, hängen also insbesondere nicht davon ab, *wie* er an die aktuelle Laterne gekommen ist. MARKOV-Ketten sind stochastische Prozesse mit entsprechenden Eigenschaften:

**Definition:** a) Ein stochastischer Prozess ist eine Folge  $X^{(1)}, X^{(2)}, \dots$  von Zufallsvariablen mit Werten in einer festen Menge  $A$ . Wir betrachten hier nur den Fall einer endlichen Menge  $A = \{a_1, \dots, a_m\}$ .

b) Ein solcher Prozess heißt MARKOV-Prozess oder MARKOV-Kette, wenn für alle  $n \in \mathbb{N}$  gilt

$$\begin{aligned} p(X^{(n+1)} = x^{(n+1)} \mid X^{(1)} = x^{(1)}, \dots, X^{(n)} = x^{(n)}) \\ = p(X^{(n+1)} = x^{(n+1)} \mid X^{(n)} = x^{(n)}). \end{aligned}$$

c) Eine MARKOV-Kette heißt *zeitinvariant*, wenn die bedingte Wahrscheinlichkeit  $p(X^{(n+1)} = y \mid X^{(n)} = x)$  nicht von  $n$  abhängt. In diesem Fall setzen wir  $p_{ij} = p(X^{(n+1)} = a_j \mid X_n = a_i)$  und bezeichnen die  $m \times m$ -Matrix  $P$  mit Einträgen  $p_{ij}$  als die *Übergangsmatrix* des Prozesses.



Der russische Mathematiker ANDREĬ ANDREEVIČ MARKOV (Андре́й Андре́евич Марко́в, 1856–1922) studierte in Sankt Petersburg, wo er später auch Professor wurde. Er beschäftigte sich zunächst hauptsächlich mit Zahlentheorie und Analysis; erst später folgen die wahrscheinlichkeitstheoretischen Arbeiten, für die er heute vor allem bekannt ist. Der Name Марков wird in lateinischen Buchstaben verschieden transkribiert; MARKOVs französische Arbeiten erschienen mit der Schreibweise MARKOFF; nach den klassischen deutschen Transkriptionsregeln müßte man MARKOW schreiben. Die Schreibweise MARKOV entspricht den englischen Regeln und scheint sich mittlerweile in der Mathematik ziemlich durchgesetzt zu haben.

Wir betrachten im folgenden nur zeitinvariante MARKOV-Ketten. Zeitinvarianz muß selbstverständlich nicht bedeuten, daß alle Zufallsvariablen  $X^{(n)}$  dieselbe Verteilung haben, sie sind allerdings auch nicht unabhängig voneinander: Bezeichnet  $p^{(n)} \in \Delta_{m-1}$  die Verteilung von  $X^{(n)}$ , so ist

$$p_j^{(n+1)} = p(X^{(n+1)} = a_j) = \sum_{i=1}^m p(X^{(n+1)} = a_j \mid X^{(n)} = a_i) = \sum_{i=1}^m p_i^{(n)} p_{ij};$$

wenn wir die  $p^{(n)}$  als Zeilenvektoren schreiben, ist also  $p^{(n+1)} = p^{(n)} P$ .

**Definition:** a) Eine MARKOV-Kette heißt *stationär*, wenn alle  $X^{(n)}$  dieselbe Wahrscheinlichkeitsverteilung haben.

b) Eine MARKOV-Kette heißt *reversibel*, wenn für alle  $i, j, n$  gilt:

$$p_i^{(n)} p_{ij} = p_j^{(n)} p_{ji}.$$

c) Eine MARKOV-Kette heißt *irreduzibel*, wenn es für je zwei mögliche Werte  $a_i, a_j$  stets ein  $r \in \mathbb{N}$  gibt, so daß  $p(X^{(r+1)} = a_j \mid X^{(1)} = a_i) > 0$  ist.

d) Der Wert  $a_i$  heißt *aperiodisch*, wenn der ggT aller natürlicher Zahlen  $r$  mit  $p(X^{(r+1)} = a_i \mid X^{(1)} = a_i) > 0$  gleich eins ist.

**Satz:** Ist eine MARKOV-Kette reversibel, irreduzibel und aperiodisch, so ist sie stationär.



Zum *Beweis* sei auf Lehrbücher zur Theorie der MARKOV-Ketten verwiesen, zum Beispiel

ACHIM KLENKE: Wahrscheinlichkeitstheorie, *Springer*,<sup>2</sup>2008, Satz 18.18

Wir werden im folgenden, soweit nicht explizit etwas anderes gesagt ist, stets annehmen, daß unsere MARKOV-Ketten zeitinvariant sind. In diesem Fall können wir die Wahrscheinlichkeitsverteilungen aller Zufallsvariablen aus der von  $X_1$  und der Übergangsmatrix berechnen: Ist allgemein  $p_i^{(n)}$  die Wahrscheinlichkeit, mit der  $X_n$  dem Wert  $a_i$  annimmt, so ist

$$\begin{aligned} & p(X_n = a_{i_0}, X_{n+1} = a_{i_1}, \dots, X_{n+r} = a_{i_r}) \\ &= p(X_n = a_{i_0}) \prod_{\ell=1}^r p(X_{n+\ell} = a_{i_\ell} \mid X_{n+\ell-1} = a_{i_{\ell-1}}). \end{aligned}$$

Für  $r = 1$  wird das zu

$$p(X_n = a_i, X_{n+1} = a_j) = p(X_n = a_i)p(X_{n+1} = a_j \mid X_n = a_i) = p_i^{(n)} p_{ij},$$

was wir auch einfacher mit Matrizen und Vektoren formulieren können: Ist  $\mathbf{p}^{(n)} = (p_1^{(n)}, \dots, p_m^{(n)})^T$  der Spaltenvektor der Wahrscheinlichkeitsverteilung zu  $X_n$ , so ist  $\mathbf{p}^{(n+1)} = A^T \mathbf{p}^{(n)}$  und damit  $\mathbf{p}^{(n)} = (A^T)^{n-1} \mathbf{p}^{(1)}$ .

Somit bestimmen  $\mathbf{p}^{(1)}$  und die Übergangsmatrix  $A$  die Wahrscheinlichkeitsverteilungen aller  $X_n$  und erlauben damit auch die Berechnung der Wahrscheinlichkeiten aller Teiltupel, die von der MARKOV-Kette produziert werden.

## §6: Die Markovketten-Montecarlo Methode

Wir gehen aus einem log-linearen Modell  $\mathcal{M}_A$  und einer MARKOV-Basis  $\mathcal{B}$  von  $\text{Kern}_{\mathbb{Z}}(A)$ . Zu einer beobachteten Tabelle  $u$  betrachten wir eine Zufallsvariable  $U$  mit Werten in  $\mathcal{F}(u)$  und interessieren uns für die Wahrscheinlichkeit

$$p(\chi^2(U) \geq \chi^2(u)).$$

Der folgende Algorithmus von METROPOLIS liefert eine Schätzung dieser Wahrscheinlichkeit:

*Initialisierung:* Wähle ein beliebiges Element  $u^{(1)}$  aus  $\mathcal{F}(u)$ .

*t*-ter Iterationsschritt: Wähle zufällig ein Element  $b_t \in \mathcal{B}$ , wobei jedes Element mit gleicher Wahrscheinlichkeit auftreten kann, sowie ein Element  $\varepsilon_t \in \{+1, -1\}$ , wobei beide Möglichkeiten Wahrscheinlichkeit  $\frac{1}{2}$  haben. Dann ist

$$u^{(t+1)} = \begin{cases} u^{(t)} + \varepsilon_t b_t & \text{mit Wahrscheinlichkeit } q \\ u^{(t)} & \text{mit Wahrscheinlichkeit } 1 - q \end{cases}$$

$$\text{mit } q = \min \left( 1, \frac{p(U = u^{(t)} + \varepsilon_t b_t \mid U \in \mathcal{F}(u))}{p(U = u^{(t)} \mid U \in \mathcal{F}(u))} \right).$$

Zu dieser neuen Tabelle wird  $\chi^2$  berechnet, und anhand einer hinreichend langen Folge dieser Werte wird die obige Wahrscheinlichkeit geschätzt. Um die Schätzung möglichst unabhängig von  $u^{(1)}$  zu machen, werden dabei meist die ersten Glieder der Folge ignoriert.

Um zu sehen, daß wir so tatsächlich eine unverzerrte Schätzung der Wahrscheinlichkeit bekommen, müssen wir zeigen, daß die erzeugte Folge von  $\chi^2$ -Werten als erwartete Häufigkeit den gesuchten Anteil aller Elemente der Faser mit einem Wert von  $\chi^2$ , der den für die gegebene Tabelle von  $u$  nicht übersteigt. Siehe dazu das oben zitierte Buch von KLENKE, insbesondere die Paragraphen 18.2 und 18.3.