Rekonstruktion zufälliger Landschaften

Matthias Löwe
Eurandom
P.O. Box 513
5600 MB Eindhoven
The Netherlands
email: lowe@eurandom.tue.nl

Zusammenfassung

Unter einer zufälligen m-farbigen Landschaft in d Dimensionen verstehen wir eine zufällige Färbung des d-dimensionalen Gitters $\mathbb{Z}^d$ mit m Farben. Eine neue, interessante Fragestellung in der Wahrscheinlichkeitstheorie beschäftigt sich mit dem Problem, aus den Beobachtungen, die eine Irrfahrt auf einer zufälligen m-farbigen Landschaft macht, nichttriviale Schlussfolgerungen auf die Landschaft zu ziehen.
Wir geben hier einen Überblick über die Fortschritte in diesem Themengebiet in den vergangenen Jahren, stellen einige grundlegende Techniken vor und beleuchten auch kurz das verwandte Problem des Werfens verschiedener Münzen in zufälliger Reihenfolge.

1 Einleitung

Um einen Eindruck von der Klasse von Problemen zu bekommen, die in dieser Arbeit zusammenfassend beleuchtet werden soll, stelle man sich vor, man habe einen Freund, O. Rakel, der auf rein zufällige Weise eine Stadt erkundet, deren Häuser in verschieden möglichen Farben, sagen wir blau und rot, gestrichen sind. Vor jedem Haus bleibt unser Freund stehen, ruft uns an (was aufgrund der modernen Technik möglich ist) und teilt uns die Farbe des Gebäudes mit. Dann bewegt er sich zufällig weiter ohne eine der möglichen Richtungen zu bevorzugen. Nehmen wir ferner an, wir wüssten, dass sich unser Freund nur in einer von zwei verschiedenen Städten, sagen wir Hic oder Ibi befindet, deren Stadtpläne (inklusive der Häuserfarben) wir kennen. Werden wir, ohne weitere Information, allein aufgrund der Farbabfolge, die wir mitgeteilt bekommen, entscheiden können, ob sich O.Rakel in Hic oder Ibi befindet? Können wir vielleicht, wenn wir selbst die Stadtpläne nicht haben, mit O. Rakels Angaben einen Stadtplan der Stadt erstellen, in der er sich befindet?

Diese ebenso seltsam wie hoffnungslos anmutenden Fragen haben ihre Wurzeln in der Ergodentheorie, genauer in der sogenannten $T-T^{-1}$–Vermutung, die 1981 von Kalikow gelöst wurde (siehe [4]). Als eine eigeständige "statistische" oder probabilistische Fragestellung wurde sie unabhängig von Benjamini und Weiss und von den Hollander und Keane [3] entdeckt. Sie erinnert an die schwierigen *inversen Probleme* aus der Statistik, wo es darauf ankommt, aus Beobachtungen einer der Realisierung der Faltung zweier Maße Rückschlüsse auf die Marginalverteilunegn zu ziehen. Versuche, die sich ergebenden Probleme zu lösen, haben im vergangenen Jahrzehnt zu erstaunlichen Antworten geführt.

Aus mathematischer Sicht brauchen wir für die Beschreibung des Problems zwei Zutaten: zunächst eine Färbung

$$\xi : \mathbb{Z}^d \to \{1, \ldots, m\}$$

des d-dimensionalen, ganzzahligen Gitters mit m Farben. Das Problem könnte allgemeiner für jeden beliebigen Graphen $\mathcal{G}$ gestellt werden, da sich aber alle verfügbaren Resultate auf den Fall $\mathcal{G} = \mathbb{Z}^d$ beziehen, verzichten wir auf die überflüssige Allgemeinheit. ξ heiße Landschaft und wird in der Folge meist zufällig gewählt sein, wobei wir später auf den Zufallsmechanismus eingehen werden.

Die zweite Zutat des Problems ist eine symmetrische Nächste-Nachbarschafts-Irrfahrt auf dem Gitter $\mathbb{Z}^d$, $(S_n)_{n\in\mathbb{N}_0}$, die in $x_0 = 0$ startet (obschon das eher sekundär ist) und sich ohne anzuhalten bewegt. Mathematisch bedeutet das $S_0 = x_0 = 0$ and

$$P(S_{n+1} = x + e_i | S_n = x) = P(S_{n+1} = x - e_i | S_n = x) = \frac{1}{2d}$$

für $x \in \mathbb{Z}^d$, $n \in \mathbb{N}_0$ und $i = 1, \ldots, d$. Hierbei bezeichnet e_i den i'ten Einheitsvektor in $\mathbb{Z}^d$. Die Schwierigkeit des Problems rührt nun daher, dass man ξ selbst nicht beobachten kann. Die gesamte Kenntnis der Landschaft stammt vielmehr von den Beobachtungen, die die Irrfahrt auf ihr macht; genauer besteht sie aus dem *Farbverlauf* $\chi := (\chi_n)_{n\in\mathbb{N}_0}$ von $(S_n)_{n\in\mathbb{N}_0}$ auf ξ, d.h.

$$\chi_n = \xi(S_n) \quad \text{für alle } n.$$

Grob gesprochen lässt sich das Problem der zufälligen Landschaften nun skizzieren als die Aufgabe, Schlussfolgerungen von χ auf die Landschaft ξ zu ziehen. Genauer unterteilt sich die Fragestellung in drei Richtungen:

1. Lassen sich zwei gegebene Landschaften ξ und η anhand des auf ihnen produzierten Farbverlaufs unterscheiden, d.h. kann man anhand von χ erkennen, ob die zugrunde liegende Landschaft ξ oder η ist?

2. Gibt es Landschaften, die sich nicht aufgrund des Farbverlaufes unterscheiden lassen?

3. Lässt sich ein unbekanntes ξ anhand von χ rekonstruieren?

Im zweiten Kapitel dieser kleinen Übersichtsarbeit sollen die Probleme, die im Zusammenhang mit den Fragen 1 und 2 auftreten, beleuchtet werden und Resultate zu diesem Themenkreis vorgestellt werden. Das dritte Kapitel stellt den Schwerpunkt dieses Artikels dar. Hier wird die Frage der Rekonstruierbarkeit von Landschaften diskutiert. Einige grundlegende Ideen werden angerissen und darauf aufbauende Sätze dargestellt. Im vierten Kapitel schließlich stellen wir ein Themengebiet vor, das der Rekonstruktion von Landschaften eng verwandt ist, aber dennoch eigene Techniken verlangt und überraschende Resultate bietet.

2 Unterscheidbarkeit zufälliger Landschaften

In diesem Abschnitt sollen die Fragen 1 und 2 aus der Einleitung diskutiert werden. Die erste Bemerkung hierzu ist beinahe trivial, soll aber dennoch gemacht werden, da sie den Ausgangspunkt für eine Vermutung bildet, deren Untersuchung mehr als ein Jahrzehnt in Anspruch genommen hat:
Wenn sich zwei Landschaften ξ und ξ' nur um Symmetrien des Gitters $\mathbb{Z}^d$ unterscheiden, dann ist nicht entscheidbar, ob ein Farbverlauf χ auf ξ oder auf ξ' produziert wurde. Um dies einzusehen, genügt es, sich vor Augen zu halten, dass man ja schon im ersten Schritt nicht weiß, in welche Richtung die Irrfahrt läuft und dass – unterscheiden sich ξ und ξ' nur durch Gittersymmetrien – S_n auf ξ den gleichen Farbverlauf erzeugen kann wie auf ξ', wenn die Irrfahrt einem entsprechenden symmetrischen Pfad folgt.
Ähnlich lässt sich einsehen, dass für $d = 1, 2$ und ein $v \in \mathbb{Z}^d$ mit

$$P(S'_l - S''_l = v) > 0$$

für ein $l \in \mathbb{N}$ (hierbei sind (S'_l) und (S''_l) unabhängige Kopien von (S_n)) die Landschaften ξ und $\theta_v\xi$ nicht anhand von χ zu unterscheiden sind. Hierbei bezeichnet $\theta_v\xi$ die um $v \in \mathbb{Z}^d$ verschobene Landschaft ξ.
Somit lässt sich Frage 1 aus Abschnitt 1 folgendermaßen präzisieren:
Lassen sich zwei gegebene Landschaften ξ und η, wobei ξ und η nicht äquivalent sind, anhand des auf ihnen produzierten Farbverlaufs unterscheiden? Dabei bezeichnen wir ξ und η als äquivalent, falls man η aus ξ durch Verschiebung des Ursprungs und eine Drehung oder Spiegelung, die das Gitter invariant lässt, erhalten kann. Die naheliegende Antwort auf diese Frage ist – zumindest in einer Dimension – "Ja", d.h., wann immer zwei Landschaften ξ und η nicht äquivalent sind, kann man anhand von χ erkennen, ob die dem Farbverlauf zugrunde liegende Landschaft ξ oder η ist.
Um diese Vermutung eingehender diskutieren zu können, machen wir uns zunächst Gedanken über eine genauere Fassung des Begriffs der Unterscheidbarkeit. Es gibt nämlich Landschaften, die trivialerweise schon nach wenigen Schritten der Irrfahrt unterschieden werden können. Ist beispielsweise der Ursprung 0 durch ξ anders gefärbt als durch η, so ist eine Unterscheidung von ξ und η anhand von χ schon nach einer einzigen Beobachtung möglich, sind alle Nachbarn des Ursprungs unter ξ anders gefärbt als unter η, genügen zwei Beobachtungen usf. Solche trivialen Auswege sollen ausgeschlossen werden. Daher verlangen wir, dass zwei Landschaften nur dann unterscheidbar heißen sollen, wenn man sie unterscheiden kann, egal, wo man in den Beobachtungen beginnt. Genauer sei Q^l_ξ das Maß auf χ, das bei festem ξ durch $(\xi(S_n))_{n\geq l}$ induziert wird. Wir nennen nun zwei Landschaften ξ und η genau dann *unterscheidbar*, wenn die Maße Q^l_ξ und Q^l_η für jedes $l \in \mathbb{N}_0$ orthogonal (gegenseitig singulär) sind.

Mit dieser Definition lässt sich ein erster Schritt in Richtung der obigen Vermutung gehen, die sich schon in den anfänglichen Arbeiten über das Problem zufälliger Landschaften findet. Das wohl fundamentalste Resultat zur Unterscheidbarkeit von Landschaften geht dabei auf eine grundlegende Arbeit von Benjamini und Kesten [1] zurück. Dort werden die Landschaften (zumindest teilweise) zufällig gewählt. Dabei, sowie in den meisten anderen Arbeiten, die sich mit zufälligen Landschaften befassen, ist der Zufallsmechanismus dergestalt, dass die Farben der einzelnen Knoten des $\mathbb{Z}^d$ statistisch unabhänging sind. Genauer ist die Wahrscheinlichkeitsverteilung ρ einer Landschaft η gegeben durch

$$\rho(\eta) = \prod_{i \in \mathbb{Z}^d} \hat{\rho}(\eta_i).$$

$\hat{\rho}$ ist in diesem Zusammenhang eine Wahrscheinlichkeit auf $\{1, \ldots, m\}$, die allen Farben positive Wahrscheinlichkeit zumisst und η_i ist die i'te Komponente von η. Für solche Landschaften erhielten Benjamini und Kesten [1] das folgende Unterscheidbarkeitsresultat:

Theorem 2.1 *(Benjamini und Kesten (1996) [1])*
Es sei $d = 1$ oder $d = 2$. Dann ist jede feste zweifarbige Landschaft (d.h. $m = 2$) unterscheidbar von ρ-fast jeder zweifarbigen Landschaft η.

Wie Kesten in [7] betont ist der Beweis des obigen Satzes ohne weitere Schwierigkeiten erweiterbar auf den Fall beliebiger m. Weiterhin ist auch die Annahme, dass (S_n) eine nächste Nachbarschafts-Irrfahrt ist, nicht wesentlich. Wie Kesten [7] weiter ausführt, genügt die weit schwächere Voraussetzung an S_n, dass die Sprünge Erwartungswert 0 haben und beschränkt sind. Die Einschränkung an die Dimension $d = 1$ oder $d = 2$ hingegen ist essentiell. In der Tat ist leicht einzusehen, dass die Aufgabe zwei gegebene Landschaften zu unterscheiden ab Dimension drei immer schwieriger wird, da die Irrfahrt in diesen Dimensionen nicht mehr rekurrent ist und einen immer kleiner werdenden Teil des Gitters besucht. Dass das Problem allerdings nicht einzig auf die Rekurrenz der Irrfahrt zurückgespielt werden kann, belegt eindrucksvoll der folgende Satz, der ebenfalls auf Benjamini und Ketsen zurückgeht:

Theorem 2.2 *(Benjamini und Kesten (1996) [1])*
Es sei $d \geq 3$ fest. Dann gibt es ein $m_0 = m_0(d)$, so dass für $m \geq m_0$ jede feste m-farbige Landschaft ξ unterscheidbar ist von ρ-fast jeder m-farbigen Landschaft η.

Die Frage, ob es möglich ist, dass Theorem 2.1 in großen Dimensionen falsch wird und ab welcher Dimension und mit welcher Zahl von Farben dies geschieht, ist vielleicht eine der schwierigsten offenen Fragen des gesamten Themengebietes (in der Tat vermuten Benjamini und Kesten in [1] sogar, dass sich für genügend große Dimensionen $\rho \times \rho$-fast sicher kein zufällig gewähltes Paar von Landschaften mehr unterscheiden lässt).

Auf den Beweis der Theoreme 2.1 und 2.2, der ebenso subtil die Statistik einer Irrfahrt im $\mathbb{Z}^d$ ausnutzt wie die Statistik einer zufällig gewählten Landschaft, soll hier nicht detaillierter eingegangen werden. In einer Dimension oder bei genügend vielen Farben folgt er auch aus den Rekonstruktionsresultaten im folgenden Abschnitt.
Es sei nur darauf hingewiesen, dass ein Satz wie Theorem 2.1 nicht vollständig überraschend ist, wenn man bedenkt, dass – wenn $\hat{\rho}$ beispielsweise die Gleichverteilung auf allen Farben ist – eine zufällig gewählte Landschaft η sich von einer fest vorgegebenen Landschaft ξ fast sicher in mindestens der Hälfte aller Knoten unterscheidet. Da die Irrfahrt (S_n) in ein und zwei Dimensionen jeden Knoten des $\mathbb{Z}^d$ unendlich oft besucht, ist somit ein Resultat wie Theorem 2.1 zumindest im Bereich des Vorstellbaren. Noch sehr viel erstaunlicher ist vor diesem Hintergrund das folgende Ergebnis von Kesten [6],

das – in Anwesenheit von genügend vielen Farben – sogar die Unterscheidbarkeit zweier Landschaften behauptet, die sich nur in der Farbe *eines einzigen Knotens* unterscheiden: Genauer

Theorem 2.3 *(Kesten (1995) [6])*
Es sei $d = 1$ und $m \geq 5$. Dann lässt sich ρ-fast jede Landschaft ξ unterscheiden von einer Landschaft ξ' mit

$$\xi'_i = \begin{cases} \xi_i & i \neq 0 \\ \neq \xi_0 & i = 0. \end{cases}$$

Es sei bemerkt, dass Theorem 2.3 nicht trivial ist, weil unsere Definiton der Unterscheidbarkeit von Landschaften verlangt, dass Q^l_ξ und $Q^l_{\xi'}$ für *jedes* l orthogonal sind. Insofern spielt es auch keine Rolle, welchen Knoten wir unterschiedlich färben. Wir werden es ebenso wie Teile von Theorem 2.1 als Konsequenz der Rekonstruktionsresultate in Abschnitt 3 erhalten.

Die obigen sehr beeindruckenden Resultate nähren den Verdacht, dass die eingangs erwähnte Vermutung, dass je zwei nicht äquivalente Landschaften unterscheidbar sind, wahr sein sollte. Diese Vermutung stellte sich vor knapp 2 Jahren überraschenderweise als falsch heraus. In einer bemerkenswerten Arbeit zeigte Lindenstrauss [8] in Dimension eins die Existenz einer überabzählbaren Menge von ununterscheidbaren Landschaften. Er konstruierte dieses Beispiel für Landschaften mit zwei Farben, aber wie Kesten in [7] ausführt, ist dieselbe Kontruktion auch mit unendlich vielen Farben möglich. Grob gesprochen besteht die Grundidee dieser Konstruktion darin, Landschaften mit großen monochromatischen Inseln zu bauen. Auf diesen Inseln erzeugt die Irrfahrt einen Farbverlauf, der keinerlei Rückschlüsse darauf zulässt, wo sich die Irrfahrt befindet.
Ganz offenbar impliziert Lindenstrauss' Resultat eine negative Antwort auf die eingangs dieses Abschnitts erwähnte Vermutung: es gibt ununterscheidbare Landschaften auf $\mathbb{Z}$, die sich nicht nur durch Verschiebung des Ursprungs und eine Spieglung unterscheiden. Dieses Ergebnis, das auch in höheren Dimensionen wahr sein sollte (schließlich werden die Schwierigkeiten bei der Unterscheidung von Landschaften in zunehmender Dimension nur größer), beantwortet auch die in der Einleitung angerissene Frage 2.

3 Rekonstruktion zufälliger Landschaften

Das Ergebnis von Lindenstrauss [8] legt nahe, dass die Rekonstruktion von Landschaften ein extrem schwieriges, vielleicht sogar unlösbares Problem ist. Bei genauerem Hinsehen jedoch realisiert man schnell, dass, obschon überabzählbar viele Landschaften nicht unterscheidbar sind und somit natürlich auch nicht rekonstruierbar, dies trotz allem möglicherweise nur eine Nullmenge in der Menge aller möglichen Landschaften ist.
Wie wir im letzten Abschnitt schon diskutiert haben, lassen sich Lanschaften, die durch Verschiebungen des Ursprungs und Drehungen und Spiegelungen des $\mathbb{Z}^d$ auseinander hervorgehen, nicht durch den Farbverlauf χ der Irrfahrt (S_n) unterscheiden. Selbstverständlich kann man dann auch nur auf die Rekonstruierbarkeit einer Landschaft ξ bis auf Äquivalenz hoffen. Hierbei heissen zwei Landschaften ξ und η äquivalent (in Symbolen $\xi \sim \eta$), falls es eine Drehung oder Spiegelung M des $\mathbb{Z}^d$ gibt und einen Vektor $a \in \mathbb{Z}^d$,

so dass

$$\xi(z) = \eta(a + Mx) \qquad \text{für alle } z \in \mathbb{Z}^d$$

gilt.

Offenbar lassen sich nun die in Frage 1 angeschnittenen Probleme der Unterscheidung zweier Landschaften am einfachsten klären, wenn man eine Landschaft ξ (bis auf Äquivalenz) aus dem Farbverlauf χ der Irrfahrt rekonstruieren kann. Ein außerordentlich überraschendes Resultat von Matzinger [13],[14] besagt nun, dass dies in einer Dimension und in generischen Situationen bei einer gemäß ρ zufällig gewählten Landschaft stets möglich ist.

Theorem 3.1 *(Matzinger (1998, 1999) [13],[14], siehe auch Löwe, Matzinger [10])* *Es sei $d = 1$ und $2 \leq m < \infty$. Dann gibt es eine bezüglich der kanonischen Projektionen messbare Funktion*

$$\mathcal{A} : \{0, \ldots, m-1\}^{\mathbb{N}} \to \{0, \ldots, m-1\}^{\mathbb{Z}}$$

von der Menge aller Farbverläufe in die Menge aller Landschaften, so dass mit Wahrscheinlichkeit eins bezüglich ρ gilt

$$P(\mathcal{A}(\chi) \sim \xi) = 1. \tag{1}$$

Hierbei bezeichnet P die Verteilung der Irrfahrt.

Wir wollen nun ein wenig die grundlegenden Beweisideen von 2.2 kennenlernen.

Beweisideen von Theorem 3.1: Um zu verstehen, wie die Rekonstruktion einer Landschaft prinzipiell vonstatten geht, stellen wir uns vor, es gäbe in der Landschaft zwei zusätzliche Farben, sagen wir

$$A, B \notin \{1, \ldots, m\},$$

die nur an genau einer Stelle vorkommen. Ferner nehmen wir an, dass wir wissen, dass die Knoten x_1 und x_2, die mit A bzw. B gefärbt sind, genau L Schritte auseinanderliegen. Um dann die Landschaft zwischen x_1 und x_2 (bis auf Spiegelsymmetrie) zu rekonstruieren, würden wir warten, bis die Irrfahrt einmal auf direktem Wege von x_1 nach x_2 oder umgekehrt läuft. Die dabei gelesenen Farben sind offenbar exakt die Färbung der Landschaft zwischen x_1 und x_2. Um nun festzustellen, wann die Irrfahrt in x_1 bzw. x_2 ist, benutzen wir die Farben A und B. Wann immer wir also im Farbverlauf A lesen und L Schritte später ein B, kann die Irrfahrt keinen Umweg gemacht haben und wissen wir, dass wir auf direktem Wege von x_1 nach x_2 gegangen sind.

Nun haben wir zwei Annahmen gemacht, nämlich dass es zwei einzigartige Farben gibt und dass wir deren Abstand kennen. Diese Annahmen sind natürlich für eine "echte" Landschaft nicht erfüllt. Nehmen wir nun an, wir wüssten zwar weiterhin um die Existenz der besonderen Farben A und B, aber nicht um deren Abstand L. Da wir nun mit dem Farbverlauf χ eine unendliche Folge von Beobachtungen aus der Landschaft haben und (S_n) auf $\mathbb{Z}$ rekurrent ist, können wir einen guten Schätzer für L konstruieren. Da (S_n) nämlich unendlich oft nach x_1 zurückkehrt, hat die Irrfahrt unendlich viele Chancen, um auf direktem von x_1 nach x_2 zu laufen. Da dieses Ereignis positive P-Wahrscheinlichkeit

(genauer Wahrscheinlichkeit 2^{-L}) hat, wird es mit Wahrscheinlichkeit eins irgendwann eintreten. Nehmen wir also die kürzeste Distanz zwischen zwei Beobachtungen in χ, von denen eine A ist und die andere B, dann ist dies ein Schätzer für L, der mit Wahrscheinlichkeit eins L richtig schätzt.
Die Frage, wie man mit den fehlenden besonderen Farben umgeht, ist weit heikler. Wir werden hier einen Zugang vorstellen, der der vielleicht natürlichste (eine Art "Proof from the book") ist und gewissermaßen ein Entropie-Argument benutzt. Er funktioniert für den Fall $m \geq 3$ und basiert, grob gesprochen, auf der Idee, dass die Information, die man mit dem Lesen einer Farbe bekommt, die Entropie der Irffahrt überwiegt, falls die Anzahl der Farben größer ist als die Anzahl der neuen Punkte pro Schritt der Irrfahrt. In der Tat läßt sich aber auch bei einer kleineren Anzahl von Farben immer noch rekonstruieren. Auf die Idee dort, soll am Schluss dieser Beweisskizze kurz eingegangen werden.

Für den Moment nehmen wir an, dass $m \geq 3$ ist und wir betrachten den m-regulären Baum ohne Wurzel, T_m (also einen Baum ohne Wurzel, in dem jeder Knoten Grad m hat). Wir wählen nun irgendeinen Knoten o des T_m und nennen ihn den Ursprung. Nun färben wir die Knoten des T_m folgendermaßen:
o bekommt dieselbe Farbe wie der Urspung $0 \in \mathbb{Z}$ unter ξ. Weiter gelte die Regel, dass jeder Knoten v in T_m Nachbarn jeder Farbe aus $\{1, \ldots, m\}$ hat. Da es genau m Farben gibt und jeder Knoten genau m Nachbarn hat, legen diese beiden Regeln die Färbung der Knoten von T_m bis auf Isomorphie eindeutig fest.

Die nächste Beobachtung ist die, dass sich die Landschaft ξ als ein zufälliger Pfad (sogar als eine Irrfahrt) auf diesem gefärbten T_m darstellen lässt. Dazu starten wir in $o \in T_m$ und wandern nun mit einem Nächsten-Nachbar-Pfad durch T_m, indem wir sukzessiv die Farben abtragen, die die Knoten in $\mathbb{Z}$ unter ξ tragen, wenn man schrittweise vom Ursprung nach rechts geht. Anschließend verfahren wir identisch, indem wir wieder in $o \in T_m$ starten und die Farben links vom Ursprung in $\mathbb{Z}$ abtragen. Dieses erzeugt einen Pfad φ_ξ auf T_m, der für den Fall, dass die Landschaft ξ gemäß ρ gewählt ist, einer Irrfahrt auf T_m entspricht. Diese Irrfahrt ist transient, da $m \geq 3$ ist. Da außerdem jede Farbe positive Wahrscheinlichkeit besitzt, gibt es einen Knoten v^*, so dass von v^* zwei verschieden Zweige ausgehen und φ_ξ auf jedem dieser Zweige unendlich viele Knoten besitzt. Kännte man nun den Pfad φ_ξ, so kännte man bis auf Verschiebung des Ursprungs und Spiegelungen auch die Landschaft ξ, da die Konstruktion, die von ξ zu φ_ξ führt, umkehrbar ist.

Nun ist aber φ_ξ ebenso wenig bekannt wie ξ. Allerdings kann man φ_ξ stückweise durch eine weitere Irrfahrt erkunden. In der Tat induziert ja auch die Irrfahrt (S_n) eine Irrfahrt (R_n) auf T_m, wenn wir sie mit Hilfe derselben Abbildung wie ξ auf T_m transportieren. Offenbar ist der Pfad von (R_n) im Pfad von φ_ξ enthalten. Außerdem ist (R_n) rekurrent, da (S_n) auf $\mathbb{Z}$ rekurrent ist. Die Schlüsselidee des Beweises ist es nun auszunutzen, dass wir zwei Irrfahrten φ_ξ und (R_n) haben, die auf den gleichen Knoten laufen, von denen aber die eine transient ist, die andere rekurrent. (R_n) folgt dabei dem Pfad φ_ξ, "tanzt" aber auf diesem möglicherweise vor und zurück.

Genauer funktioniert das wie folgt: Wir fixieren zwei Knoten $v, w \in T_m$, so dass v und w beide in der Knotenmenge des Pfades φ_ξ liegen. Da v und w in der Knotenmenge von φ_ξ sind, gibt es Punkte $x_1, x_2 \in \mathbb{Z}$, so dass $\xi(x_1)$ dieselbe Farbe hat wie v, $\xi(x_2)$ dieselbe

Farbe wie w und das Stück Landschaft, wenn man von x_1 nach x_2 geht genauso gefärbt ist wie der Pfad φ_ξ, wenn man ihm von v nach w folgt. Prinzipiell würde es also genügen, den Pfad φ_ξ zwischen v und w mit Hilfe von (R_n) zu erkunden.

Hierbei gibt es zwei prinzipielle Hindernisse. Folgt man (R_n), während die Irrfahrt von v nach w läuft, so kann es sein, dass sich (R_n) auf diesem Weg vor- und zurückbewegt. Dies kann zwei mögliche Ursachen haben. Entweder bewegt sich auf der Pfad φ_ξ auf dieser Strecke auch vor und zurück (was geschieht, wenn beispielsweise drei aufeinanderfolgende Punkte in $\mathbb{Z}$ zwischen x_1 und x_2 mit den Farben 1, 2 und dann wieder 1 gefärbt sind), so dass (R_n) zwangsläufig dieser Bewegung folgt. Umgekehrt kann es auch geschehen, dass sich (R_n) auf dem Weg von v nach w vor und zurück bewegt, weil die Irrfahrt (S_n) auf $\mathbb{Z}$ vorwärts und rückwärts geht. In diesem Fall wäre die Farbfolge, die von (R_n) gelesen wird, nicht identisch mit der Farbfolge in φ_ξ. Um zwischen diesen beiden Fällen zu unterscheiden, erinnern wir uns an einen Trick, den wir schon in der obigen Skizze kennen gelernt haben. Wenn (R_n) von v nach w geht und dabei exakt dem Pfad von φ_ξ folgt, dann ist dies die kürzeste aller Möglichkeiten für (R_n) um von v nach w zu gehen. Da nun (R_n) rekurrent ist, hat diese Irrfahrt unendlich viele Möglichkeiten, diese kürzeste Durchquerung der Strecke von v nach w zu realisieren. Da eine solche Durchquerung positive Wahrscheinlichkeit hat, wird diese kürzeste Durchquerung der Strecke von v nach w mit Wahrscheinlichkeit eins irgendwann tatsächlich vorkommen. Somit genügt es, unter allen Beobachtungen diejenige herauszusuchen, die eine kürzeste Durchquerung der Strecke von v nach w liefert. Mit Wahrscheinlichkeit eins wird die entsprechende Farbsequenz die Färbung der Landschaft zwischen x_1 und x_2 liefern.

Die zweite Schwierigkeit bei dieser Rekonstruktionsmethode ist die Tatsache, dass der Pfad φ_ξ die Strecke zwischen v und w möglicherweise mehrfach durchquert und möglicherweise sogar dergestalt, dass es mehrere kürzeste Durchquerungen mit unterschiedlichen Farbverläufen gibt. In diesem Falle würde unsere Rekonstruktionsmethode kein eindeutiges (und somit gar kein) Ergebnis liefern. Diese Möglichkeit lässt sich (je nach Landschaft ξ) für ein festes Paar v, w im der Knotenmenge von φ_ξ nicht ausschließen. Allerdings hilft uns die Transienz von φ_ξ aus diesem Dilemma. Diese liefert nämlich die Existenz eines Paares v, w, sagen wir v in dem einem von v^* ausgehenden Zweig, w in dem anderen, so dass die Strecke von v nach w nur genau einmal von dem Pfad φ_ξ durchquert wird. Andernfalls nämlich würde der Punkt v^* unendlich oft besucht im Widerspruch zur Transienz von φ_ξ.

Somit funktioniert der Rekonstruktionsalgorithmus wie folgt: Wir nehmen eine Folge von Punktepaaren (v_t, w_t), so dass v_t und w_t Knoten von φ_ξ sind und mit $d(v_t, w_t) \to \infty$, wenn $t \to \infty$ (hier bezeichnet $d(\cdot, \cdot)$ die Distanz im Graphen T_m) und suchen dann nach einer kürzesten Durchquerung der Strecke von v_t nach w_t durch die Irrfahrt (R_n) (im Bezug auf alle beobachteten Durchquerungen). Die dabei gelesenen Farben interpretieren wir als ein Stück rekonstruierter Landschaft. Aus den oben aufgeführten Gründen wird dieser Rekonstruktionsalgorithmus schließlich konstant, in dem Sinne als dass er auf gleichen endlichen Stücken von $\mathbb{Z}$ eine gleiche Landschaft liefert. Lässt man also $t \to \infty$ gehen, ergibt sich somit mit Wahrscheinlichkeit eins eine zu ξ äquivalente Landschaft.

Der hier vorgestellte Rekonstruktionsalgorithmus wurde erstmal von Matzinger in [13] für den Fall gemäß ρ-verteilter Landschaften vorgestellt. Gerade da das ursprüngliche Problem der Unterscheidung und Rekonstruktion von Landschaften seine Wurzeln in der Ergodentheorie hat, ist es naheliegend zu fragen, welche ergodischen Eigenschaften eigentlich von der Landschaft gefordert werden müssen, um obigen Algorithmus erfolgreich anwenden zu können. Diese Fragen wurden in [10] erschöpfend analysiert.

Schon die Aussage des Theorems 3.1 besagt, dass die Rekonstruktion in einer Dimension allerdings nicht auf den Fall von drei Farben oder mehr beschränkt bleibt, der oben vorgestellte Algorithmus allerdings schon. Der Fall von zwei Farben ($m = 2$) ist, insbesondere für den Fall, dass man auch Irrfahrten zulässt, die eine positive Haltewahrscheinlichkeit haben, technisch ungleich mühsamer. Die Grundidee ist wieder die am obigen Beispiel des Falles mit besonderen Farben illustrierte. Allerdings müssen im Fall $m = 2$ neue Ideen gefunden werden die fehlenden "besonderen Farben" zu ersetzen. Der Grundgedanke hierbei ist der folgende. Für jedes $k \in \mathbb{N}$ treten mit positiver Wahrscheinlichkeit 2^{-k} monochrome Blöcke der Länge k auf. Diese sind aufgrund ihres seltenen Auftretens typisch in ihrer Umgebung. Auf diesen Blöcken erzeugt die Irrfahrt (S_n) typischerweise eine monochrome Beobachtung der Länge k^2. Umgekehrt ist es auch so, dass eine Beobachtung der Länge k^2 mit überwältigender Wahrscheinlichkeit von einem Block der Länge mindestens k kommt und nicht von einem um Größenordnungen kleinerem. Somit lassen sich solche monochromen Blöcke der Länge k auf Stücken der Landschaft ξ, die nicht länger sind als 2^k, gut als eine Approximation für besondere Farben nehmen. Der technische Aufwand bei der Umsetzung dieser Idee (die zudem eine Induktion über k beinhaltet) ist erheblich und würde den Rahmen dieses Übersichtsartikels sprengen.

$\square$

Wie schon Kesten in [7] bemerkte, sind die oben illustrierten Techniken, insbesondere die oben genauer gezeigte Übersetzung des Problems in eine Fragestellung über Irrfahrten auf einem gefärbten Baum, sehr stark davon abhängig, dass es sich bei (S_n) um eine Nächste-Nachbarschafts-Irrfahrt ohne Sprünge handelt. Die Frage, ob eine Rekonstruktion auch bei Irrfahrten mit möglicher längerer Reichweite möglich ist, liegt auf der Hand. Eine Antwort darauf konnten wir in [11] geben. Sie stellt sich wie folgt dar:

Theorem 3.2 *(Löwe, Matzinger 2000 [11])*
Es sei $d = 1$ und ξ eine eindimensionale zufällige Landschaft mit m Farben. Der Zufallsmechanismus ρ hierbei ist dergestalt, dass die $\xi(z)$, $z \in \mathbb{Z}$ unabhängige und identisch verteilte Zufallsvariablen sind mit

$$\rho(\xi(0) = i) = \frac{1}{m}$$

für alle Farben $i \in \{1, \ldots, m\}$.
Weiter sei $(S_n)_{n\in\mathbb{N}}$ eine Irrfahrt auf $\mathbb{Z}$ mit Start im Ursprung 0 und den folgenden Eigenschaften

- *(S_n) hat unabhängige, identisch verteilte Zuwächse, d.h.*

$$S_{n+1} - S_n = X_n$$

  wobei (X_n) eine Folge unabhängiger, identisch verteilter Zufallsvariablen ist.

- *X_1 is $\mathbb{Z}$-wertig und hat endliche Reichweite, d.h.*

$$supp(X_1) := \{x \in \mathbb{Z} : P(X_1 = x) \neq 0\}$$

ist eine endliche Menge.

- *(S_n) is aperiodisch in dem Sinne, dass $ggt\, supp(X_1) = 1$ gilt (um sicherzustellen, dass jeder Punkt überhaupt mit positiver Wahrscheinlichkeit durch (S_n) erreicht wird).*

- *(S_n) ist ohne Drift, d.h.*

$$E(X_1) = 0.$$

Nun sei $m \geq |\, supp(X_1)| + 1$ und wir nehmen ferner an, dass die Irrfahrt $(S_n)_{n\geq 0}$ und die Landschaft ξ die obigen Bedingungen erfüllen und unabhängig voneinander sind. Dann kann man aus einer Realisierung des Prozesses $(\xi(S_n))_{n\geq 0}$ fast sicher (mit Bezug auf ρ und P) ξ bis auf Äquivalenz rekonstruieren. Mit anderen Worten: es gibt eine messbare Funktion (bezüglich der kanonischen σ-Algebren)

$$\mathcal{A} : \{1, \ldots, m\}^{\mathbb{N}} \to \{1, \ldots, m\}^{\mathbb{Z}}$$

so dass für ρ-fast alle Landschaften ξ gilt

$$P(\mathcal{A}(\chi) \sim \xi) = 1. \tag{2}$$

Die zentrale Idee beim Beweis von Theorem 3.2 ist eine Verfeinerung der Ideen des Beweises von Theorem 3.1 für den Fall von $m = 2$ Farben. Zusätzlich zu den ohnehin erheblichen technischen Schwierigkeiten, erheben sich hier die folgenden Probleme: Zum einen ist es durch die Sprünge der Irrfahrt möglich lange monochrome Sequenzen auch dann zu lesen, wenn sich die Irrfahrt nicht auf monochromen Teilstücken der Landschaft ξ befindet. Dies erfordert zusätzliche Abschätzungen, um zu zeigen, dass dennoch typischerweise lange monochrome Sequenzen in den Beobachtungen χ von relativ langen monochromen Teilstücken der Landschaft ξ herrühren und so unseren Ersatz für die "besonderen Farben" retten.

Die zweite wesentliche Schwierigkeit beim Beweis von Theorem 3.2 besteht darin, dass durch die Sprünge eine schnellste Durchquerung eines Stückes Landschaft von einer monochromen "Insel" zur nächsten nicht direkt ein Stück Landschaft abliest, sondern nur die Farben von Knoten im Abstand maximaler Sprunglänge. Wir werden also im Allgemeinen mehrere leiterartige Wörter lesen, die verschiedenen solcher kürzester Durchquerungen der Distanz zwischen monochromen Inseln entsprechen. Eine entscheidende Aufgabe beim Beweis des obigen Theorems, ist es eine Technik zu finden, diese verschiedenen "Leitern" zu einer Landschaft zusammenzustecken. In [11] wird dieses Problem induktiv gelöst, indem man eine gewisse Überlappung mit schon rekonstruierten Landschaftsstücken fordert. Diese ordnen dann die "Leitern" in der richtigen Reihenfolge. Für detaillierte Ausführungen sei auf [11] verwiesen.

Eine mindestens ebenso naheliegende Feststellung ist die, dass die unter der Beweisskizze für Theorem 3.1 aufgeführten Ideen sehr stark den eindimensionalen Rahmen des Problems benutzen. Diese Bemerkung traf schon Kesten in [7]. Er vermutete, dass in zwei Dimensionen eine Rekonstruktion von Landschaften nicht möglich sei. Um die Beweggründe für diese Vermutung besser zu verstehen, wollen wir ein wenig die prinzipiellen Probleme in zwei Dimensionen erläutern.
Offensichtlich hinkt die oben vorgeführte Technik der Übersetzung des Problems in eine Fragestellung über Irrfahrten auf Bäumen daran, dass schon $\mathbb{Z}^2$ kein Baum ist und somit das Problem nicht auf Bäume übersetzt werden kann.
Hinter diesen Schwierigkeiten steckt ein fundamentaleres Problem. Wie schon weiter oben aufgeführt, besteht in einer Dimension eine grundsätzliche Idee darin, auf eine kürzeste Durchquerung einer Strecke durch die Irrfahrt (S_n) zu warten. So eine kürzeste Durchquerung ist überschneidungsfrei und liest damit direkt die Landschaft zwischen den beiden Endpunkten der Strecke eindeutig ab. Dies stimmt auch in zwei Dimensionen, nur gibt es hier das Problem, dass es sogar zwischen zwei gegebenen Endpunkten einer Strecke (über die man i.a. nicht verfügt) mehr als eine kürzeste überschneidungsfreie Durchquerung gibt und man nicht weiss, welche davon gerade von der Irrfahrt gelesen wird.

Das zweite Problem bei der zweidimensionalen Rekonstruktion ist noch prinzipiellerer Natur. Es betrifft die Rekurrenz der zweidimendsionalen Irrfahrt oder genauer ihre Rückkehrzeit. Selbstverständlich ist die Irrfahrt (S_n) in zwei Dimensionen ebenso rekurrent wie in einer. Somit besucht sie jeden Punkt mit Wahrscheinlichkeit eins, was offenbar eine unabdingbare Voraussetzung für die Rekonstruktion ist (daher ist auch die Frage, ob Rekonstruktion von Landschaften in drei Dimensionen oder mehr möglich ist, schnell beantwortet). Allerdings gibt es einen erheblichen Unterschied in der Häufigkeit der Besuche eines einzelnen Punktes. Während beispielsweise die Lokalzeit (die erwartete Anzahl von Besuchen) des Urspungs zur Zeit t sich in einer Dimension verhält wie $\sqrt{t}$, ist dieselbe Größe in zwei Dimensionen nur noch $\log t$, was im Vergleich zu $\sqrt{t}$ verschwindend wenig ist. Nun mag man sich auf den Standpunkt stellen, dass die Anzahl der Besuche im Ursprung (oder jedem anderen Knoten des Graphen) unerheblich ist, solange sie nur gegen Unendlich strebt, die Irrfahrt also rekurrent ist. Dass dies durchaus nicht notwendig ein allgemeines Prinzip zu sein braucht, zeigen die erstaunlichen Resultate von Harris und Keane [2] für ein unserem Rekonstruktionsproblem durchaus verwandten Problem. Diese sollen im folgenden Abschnitt dargestellt werden.

Die geringeren Lokalzeiten in zwei Dimensionen führen nun zu der folgenden Schwierigkeit: Das wesentliche Problem bei der Rekonstruktion von Landschaften ist nicht primär der Rekonstruktionsteil, sondern herauszufinden, wo die Rekonstruktion stattfindet. Je seltener nun die Irrfahrt an einen Punkt zurückkehrt, umso genauer müssen die Techniken sein, die gewährleisten, dass sie sich an der Stelle befindet, an der wir sie vermuten. Es ist plausibel, dass eine Erhöhung der Anzahl der zur Verfügung stehenden Farben die Lösung dieses Problems eher erleichtert. In der Tat hätte man (im Extremfall) unendlich viele Farben, so wäre jede Farbe charakteristisch für den Punkt, an dem sie auftritt. Eine Farbe zu lesen wäre somit gleichbedeutend mit vollständiger Information über den Ort, an dem sich die Irrfahrt aufhält (dennoch bliebe immer noch das eigentliche Rekonstruktionsproblem zu lösen). Es wird somit naheliegender, dass für den Fall hinreichend vieler Farben Kestens Vermutung tatsächlich widerlegt werden kann und eine Rekonstruktion möglich ist. Dies wurde in [9] von Matzinger und dem Autor gezeigt. Genauer ist die zentrale Aussage in [9] die folgende:

Theorem 3.3 *(Löwe, Matzinger (1999), [9])*
Es sei die Dimension $d = 2$ und (S_n) sei die einfache Nächste-Nachbarschafts-Irrfahrt in $\mathbb{Z}^2$ ohne Sprünge und mit Haltewahrscheinlichkeit null. ρ sei das Wahrscheinlichkeitsmaß auf der Menge aller zweidimensionalen Landschaften, das die Farben der einzelnen Knoten unabhängig macht und die Marginalverteilung

$$\rho(\xi(0) = i) = \frac{1}{m}$$

für alle Farben $i \in \{1, \ldots, m\}$ besitzt. Dann gibt es ein $m_0 \in \mathbb{N}$, so dass, wann immer $m \geq m_0$ ist, gilt: Es gibt eine messbare Funktion (bezüglich der kanonischen σ-Algebren)

$$\mathcal{A} : \{1, \ldots, m\}^{\mathbb{N}} \to \{1, \ldots, m\}^{\mathbb{Z}^2}$$

so dass für ρ-fast alle Landschaften ξ gilt

$$P(\mathcal{A}(\chi) \sim \xi) = 1. \tag{3}$$

Hierbei sind wieder zwei zweidimensionale Landschaften ξ und η äquivalent (symbolisch schreiben wir $\xi \sim \eta$, wenn sie sich nur durch Verschiebungen des Ursprungs und eine Spiegelung oder Drehung, die das Gitter invariant lässt, unterscheiden).

Beweisidee zu Theorem 3.3: Die oben angedeuteten Schwierigkeiten überwinden wir durch einen Induktionsbeweis. Die Induktion wird dabei über einen Parameter r geführt, wobei r den Radius euklidischer Bälle bezeichnet, auf denen die Landschaft sukzessiv rekonstruiert werden soll.
Aus Gründen der Übersichtlichkeit stellen wir dabei zunächst den Indukstionschritt vor. Wie schon einleitend erwähnt gibt es dabei zwei verschiedene Schritte: einen Teil, in dem wir sicherstellen, dass sich die Irrfahrt tatsächlich an dem Ort befindet, an dem die Rekonstruktion stattfinden soll und den eigentlichen Rekonstruktionsschritt.
Um festzustellen, dass die Irrfahrt sich tatsächlich im Ball um den Ursprung B^r vom Radius r befindet (und nicht sonstwo) benutzen wir die Kenntnis der Farben in B^{r-1}, über die wir nach Induktionsannahme verfügen, in folgender Weise. Wir definieren eine Menge von Signalwörtern, die aus sämtlichen nicht-überlappenden, horizontalen Farbsequenzen der Länge $c_1 \log r$ besteht, die innerhalb von B^{r-1} gelesen werden können. Hierbei sollte man sich c_1 als konstant und typischerweise sehr klein vorstellen. Liest die Irrfahrt nun in einen Zeitintervall der Länge r^2 mehr als $r^{2-\varepsilon_1}$ (hierbei ist ε_1 eine geeignete positive Zahl) dieser Signalwörter, so nehmen wir das als Anzeichen, dass sich die Irrfahrt in B^{r-1} befindet und wir versuchen können, die Landschaft auf dem Rand von B^{r-1} und somit auf B^r zu rekonstruieren. Hierbei ist zumindest die Größenordnung der Signalwörter von $\log r$ von entscheidender Wichtigkeit. Wählt man sie zu groß, so wird die Irrfahrt selbst dann keine Signalwörter lesen, wenn sie sich innerhalb von B^{r-1} befindet; ist sie zu klein, kann man überall beliebige Mengen an Signalwörtern lesen. Hingegen hat die Tatsache, dass wir die Signalwörter nur horizontal und nicht überlappend wählen,

vorwiegend den technischen Grund, dass dann die Ereignisse verschiedene solcher Wörter innerhalb von B^{r-1} zu lesen stochastisch unabhängig sind.
Tatsächlich lässt sich unter Ausnutzung der stochastischen Unabhängigkeit mit Hilfe einer exponentiellen Abschätzung beweisen, dass innerhalb eines Zeitintervalls der ungefähren Länge e^{r^α} mit sehr großer Wahrscheinlichkeit nur dann mehr als $r^{2-\varepsilon_1}$ Signalwörter in r^2 Schritten gelesen werden, wenn die Irrfahrt sich in B^{r-1} befindet. Hierbei kann α groß gewählt werden, wenn man die Anzahl der Farben groß genug wählt. Diese große Sicherheit beim Erkennen von B^{r-1} ist auch notwendig, denn in einem Zeitintervall der Länge e^{r^α} wird die Irrfahrt mit großer Wahrscheinlichkeit nicht öfter als r^α-mal in den Ursprung zurückkehren. Dies ist gerade ausreichend, um mit großer Wahscheinlichkeit jeden der $\mathcal{O}(r)$-Randpunkte genügend häufig zu besuchen, um dort eine Rekonstruktion vorzunehmen.

Der eigentliche Rekonstruktionschritt findet dann nur während der Zeitintervalle der Länge r^2 statt, in denen sich die Irrfahrt mit großer Wahrscheinlichkeit in B^{r-1} aufhält. Während dieser Zeiten suchen wir Zeitabschnitte der Länge $c_2 \log r$ auf, in denen die Irrfahrt auf einer Geraden aus dem Inneren von B^{r-1} kommend durch einen Randpunkt von B^{r-1} läuft. Die Farbe des Randpunktes ist somit die erste unbekannte Farbe auf dieser Strecke. Eine wesentliche Schwierigkeit, die es bei dieser Methode zu überwinden gilt, ist es herauszufinden, ob die Irrfahrt tasächlich auf einer Geraden läuft (um sicherzustellen, dass man zum "richtigen" Randpunkt läuft). Die technische Bewältigung dieser Schwierigkeit ist ein wenig zu aufwendig um hier vorgeführt zu werden, aber der Grundgedanke ist erneut, dass Strecken der Länge $\log r$ innerhalb von B^{r-1} charakteristisch sind für den Platz, an dem sie gelesen werden.

Alle Wahrscheinlichkeitsaussagen in diesen Argumenten gelten mit summierbarer Wahrscheinlichkeit in r, so dass die entsprechenden Begründungen für alle r durchgehen bis auf eine endliche Ausnahmemenge.
Sei r_0 das maximale r dieser Ausnahmemenge. Um die Induktion vollständig zu machen, muss der Induktionsanfang also die Rekonstruktion der Landschaft auf B^{r_0} sein. Hierzu benötigen wir eine große Anzahl von Farben. Wir nehmen nämlich an, dass m so groß ist, dass zu einem Zeitpunkt $T(r_o)$, zu dem mit Wahrscheinlichkeit größer als $3/4$ jede mögliche Kante in B^{r_0} einmal passiert wurde, alle bis dahin besuchten Punkte mit Wahrscheinlichkeit größer $3/4$ unterschiedliche Farben haben. Mit Wahrscheinlichkeit größer als $1/2$ ist somit jeder der bis $T(r_0)$ besuchten Punkte von verschiedener Farbe und man kennt die komplette Einschritt-Nachbarschaft jedes Punktes in B^{r_0}. Die Landschaft in B^{r_0} läßt sich dann – beginnend mit der Farbe des Urspungs $0 \in \mathbb{Z}^2$ – gleich einem Puzzle zusammenstecken (für Details sei wieder auf [9] verwiesen).

Der aufmerksame Leser wird bemerkt haben, dass zwar der Induktionsschritt mit Wahrscheinlichkeit eins korrekt ausgeführt wird, der Induktionsanfang aber nur mit Wahrscheinlichkeit größer als $1/2$ durchgeführt werden kann. Dies scheint im Widerspruch zur Behauptung des Theorems 3.3 zu stehen. Die Rettung bietet uns das folgend Lemma:

Lemma 3.4 *Für jedes m gilt: Falls eine messbare Abbildung*

$$\overline{\mathcal{A}} : \{0, \ldots, m-1\}^{\mathbb{N}} \to \{0, \ldots, m-1\}^{\mathbb{Z}^2}$$

gibt, so dass

$$\mathbb{P}(\overline{\mathcal{A}}(\chi) \sim \xi) > 1/2$$

dann gibt es auch eine messbare Abbildung

$$\mathcal{A} : \{0, \ldots, m-1\}^{\mathbb{N}} \to \{0, \ldots, m-1\}^{\mathbb{Z}^2}$$

mit

$$\mathbb{P}(\mathcal{A}(\chi) \sim \xi) = 1.$$

Der Beweis des Lemmas beruht auf einer Anwendung des Ergodensatzes. Man wendet den Rekonstruktionsalgorithmus $\overline{\mathcal{A}}$ auf χ an, dann auf die Beobachtungen χ beginnend ab der zweiten Beobachtung, dann auf die Beobachtungen χ beginnend ab der dritten Beobachtung usf. Da die Wahrscheinlichkeit einer richtigen Rekonstruktion unter $\overline{\mathcal{A}}$ größer ist als die einer falschen Rekonstruktion, wird die Mehrzahl aller Rekonstruktionen bei diesem Verfahren zu der Orginal-Landschaft äquivalent sein.
In der Tat ist dies auch außerordentlich plausibel. Wenn der obige Rekonstruktionsalgorithmus fehlschlägt, dann deshalb, weil Landschaft um den Ursprung nicht den Anforderungen des Induktionsanfangs genügt. Da aber die Mehrzahl aller Bälle vom Radius r_0 durchaus eine Landschaft bietet, die den Ansprüchen des Induktionsanfangs genügt, reicht es aus zu warten, bis sich die Irrfahrt in eine solche Region begibt um eine korrekte Rekonstruktion zu erhalten.
Dies beendet die Skizze des Beweises von Theorem 3.3.

□

4 Ein verwandtes Problem

In diesem Abschnitt soll eine Fragestellung vorgestellt werden, die der Unterscheidung und Rekonstruktion von Landschaften artverwandt ist, insbesondere für den Fall, dass die Landschaft eine Periodizität aufweist. Dieser Problemkreis, der ein überraschendes Resultat birgt, wurde in einem Artikel von Harris und Keane [2] eröffnet. Er lässt sich folgendermaßen vorstellen:
Wir betrachten einen unendlichen, zusammenhängenden Graphen $\mathcal{G}$ (die wichtigsten Beispiele erhält man wieder für $\mathcal{G} = \mathbb{Z}$ oder $\mathcal{G} = \mathbb{Z}^2$) und eine Irrfahrt (S_n) auf $\mathcal{G}$ mit Start in einem Knoten o, den wir den Ursprung nennen (im Fall von $\mathcal{G} = \mathbb{Z}^d$ für ein d sei $o = 0$). In jedem Knoten v von $\mathcal{G}$ sei eine Münze mit Seiten $+1$ und -1 angehängt, die wir werfen, wenn die Irrfahrt den Knoten v erreicht. Hierbei seien sämtliche Münzen in Knoten v mit Ausnahme der Münze in o fair und stochastisch unabhängig, d.h. für die mit ihnen assoziierten Zufallsvariablen X_v mit Werten in $\{-1, +1\}$ gelte, dass sie unabhängig seien und

$$P(X_v = +1) = P(X_v = -1) = \frac{1}{2}$$

erfüllen. Zusätzlich sei auch X_o unabhängig von allen anderen X_v, habe aber möglicherweise einen Bias $\theta \in (-1, 1)$, d.h.

$$P(X_o = +1) = \frac{1}{2}(1 + \theta) \quad \text{und} \quad P(X_o = -1) = \frac{1}{2}(1 - \theta).$$

Zu jeden Zeitpunkt $n \in \mathbb{N}$ werfen wir nun die Münze am Knoten v mit $S_n = v$ und produzieren so eine Folge $(Y_n)_{n\in\mathbb{N}}$ von $+1$'en und -1'en. Die Frage, die wir beantworten wollen, ist: "Hat X_o wirklich einen Bias, d.h. ist $\theta \neq 0$ und falls ja, können wir θ verlässlich schätzen?"
Bezeichnen wir mit μ_θ die Verteilung der Folge der $(Y_n)_{n\in\mathbb{N}}$, wenn X_o den Bias θ hat. Offensichtlich lässt sich der Fall $\theta \neq 0$ vom Fall $\theta = 0$ mit Wahscheinlichkeit eins unterscheiden, wenn μ_θ und μ_0 orthogonal sind, in Symbolen $\mu_\theta \perp \mu_0$, und diese Unterscheidung ist mit Sicherheit nicht möglich, falls μ_θ absolut stetig ist bezüglich μ_0, also $\mu_\theta \ll \mu_0$ gilt. Sicherlich ist es für $\mu_\theta \perp \mu_0$ absolut notwendig, dass die Irrfahrt rekurrent ist, weil sonst X_o nur endlich oft geworfen wird und daher natürlich μ_0 auf jeden Fall absolut stetig ist bezüglich μ_θ. Interessanterweise ist die Rekurrenz von (S_n) nicht die relevante Bedingung um die Fälle $\mu_\theta \perp \mu_0$ und $\mu_\theta \ll \mu_0$ zu unterscheiden, wie folgendes Resultat von Harris und Keane zeigt. Um es formulieren zu können, müssen wir noch eine Notation einführen: u_k sei die Wahrscheinlichkeit, dass (S_n) zum Zeitpunkt k nach o zurückkehrt also

$$u_k := P(S_k = o).$$

Dann liest sich das Resultat von Harris und Keane [2] wie folgt:

Theorem 4.1 *(Harris, Keane (1997) [2])*
Mit den obigen Bezeichnungen gilt das folgende:

1. *Falls $\sum_{k=0}^\infty u_k^2 = \infty$ ist, dann gilt $\mu_\theta \perp \mu_0$.*
2. *Ist $\sum_{k=0}^\infty u_k^2 < 1$, so gilt $\mu_\theta \ll \mu_0$.*

Beachtenswert hierbei ist, dass es nicht etwa auf die Konvergenz der Reihe der Rückkehrzeiten ankommt, sondern auf die Konvergenz der Reihe ihrer Quadrate. Dies führt zu der folgenden witzigen Dichotomie:

Korollar 4.2 *(Harris, Keane (1997) [2])*
Falls $\mathcal{G} = \mathbb{Z}^d$ ist, so gilt:

1. *Falls $d = 1$ ist, gilt $\mu_\theta \perp \mu_0$.*
2. *Für $d \geq 2$ ist $\mu_\theta \ll \mu_0$.*

Der Grund, warum Korollar 4.2 eine direkte Konsequenz aus Theorem 4.1 ist, ist der, dass die Rückkehrwahrscheinlichkeiten u_k von S_n in einer Dimension $d = 1$ folgendes Verhalten aufweisen:

$$u_k = 0 \quad \text{falls } k \text{ gerade und} \quad u_{2k} = \binom{2k}{k} 2^{-2k} \sim \frac{1}{\sqrt{2k}},$$

während für $d = 2$ gilt

$$\sum_{k=0}^\infty u_k^2 < 2.$$

Theorem 4.1 kann überdies als ein Analogon für eine ganz spezielle Abhängigkeitsstruktur der folgenden berühmten Dichotomie von Kakutani [5] über Produktmaße aufgefasst werden.

Theorem 4.3 *(Kakutani (1943) [5])*
Es sei ν_0 die Verteilung des i.i.d. fairen Münzwurfes über $\Omega = \{-1,+1\}^{\mathbb{N}}$ und ν_ϑ die Verteilung des unabhängigen Münzwurfes über $\Omega = \{-1,+1\}^{\mathbb{N}}$ mit Bias $\vartheta := (\vartheta_n)_{n\in\mathbb{N}}$. Dann gilt:

1. *Divergiert $\sum \vartheta_n^2$, d.h. ist $\sum_{n=1}^{\infty} \vartheta_n^2 = \infty$, so gilt $\nu_0 \perp \nu_\vartheta$.*
2. *Konvergiert $\sum \vartheta_n^2$, d.h. ist $\sum_{n=1}^{\infty} \vartheta_n^2 < \infty$, so gilt $\nu_0 \ll \nu_\vartheta$ und $\nu_\vartheta \ll \nu_0$.*

Auf den ersten Blick mag schon Teil 1 von Korollar 4.2 überraschend erscheinen. In der Tat ist ja die eindimensionale Irrfahrt in $\mathbb{Z}$ zwar rekurrent, aber die Lokalzeit (erwartete Anzahl der Besuche) des Ursprungs 0 zur Zeit n ist nur ungefähr $\sqrt{n}$. Somit wird die Anzahl der Male, an denen die Münze in 0 geworfen wird, verschwindend klein werden verglichen mit der Gesamtanzahl von Münzwürfen. Bei genauerem Hinsehen bedeutet dies aber nur, dass ein Test, der auf dem Mittelwert der (Y_n) beruht, zu grob ist. In der Tat benutzt der Beweis von Theorem 4.1 durch Harris und Keane in [2] ein Martingalargument für die Dichte von μ_θ bezüglich μ_0 auf den ersten n Beobachtungen. Eine andere denkbare Methode zu testen, ob μ_0 vorliegt oder μ_θ, besteht darin, die Länge des längsten "Runs" in den ersten n Beobachtungen $Y_1, Y_2, \ldots, Y_n$ (somit der längsten Teilfolge mit $Y_i = Y_{i+1} = \ldots = Y_{i+k}$, wobei $i \geq 1$ und $i + k \leq n$ ist) zu beobachten. Die Verteilung dieser kann unter μ_0 sehr genau bestimmt werden. Unterschiedliche empirische Verteilungen des längsten Runs lassen somit auf ein $\theta \neq 0$ schliessen. Dieses Programm wurde kürzlich von Levin, Pemantle und Peres [12] in großer Meisterschaft durchgeführt. Ihnen gelang sogar die Rekonstruktion des Bias θ, falls dieser groß genug ist.

Eine sehr naheliegende Vermutung beim Betrachten von Theorem 4.1 und Theorem 4.3 ist, dass auch in Theorem 4.1 der Parameter θ keine Rolle spielen sollte, also dass stets gilt $\mu_\theta \ll \mu_0$. Überraschenderweise ist diese naheliegende Vermutung falsch. Vielmehr gilt das folgende Resultat von Levin, Pemantle und Peres [12]:
Notation: Für zwei Folgen (a_n) und (b_n) schreiben wir $a_n \approx b_n$, falls es $0 < C_1 \leq C_2 < \infty$ gibt, so dass $C_1 < \frac{a_n}{b_n} \leq C_2$ für alle $n \geq 1$

Theorem 4.4 *(Levin, Pemantle und Peres (1999) [12])*
Es sei $\frac{1}{2} < \gamma < 1$ und die Rückkehrwahrscheinlichkeiten (u_k) erfüllen $u_k \approx k^{-\gamma}$ und $u_{\max} := \max_k u_k > 2^{\gamma-1}$, dann gilt für jedes θ mit

$$\theta > \frac{2^\gamma}{u_{\max}} - 1,$$

dass $\mu_\theta \perp \mu_0$ ist.

Für einen Beweis verweisen wir auf den Artikel von Levin, Pemantle und Peres [12].

Literatur

[1] I. Benjamini, H. Kesten: *Distinguishing sceneries by observing the sceneries along a random walk path*, J. d'Anal. Math **69** (1996), 97–135.

[2] M. Harris, M. Keane: *Random coin tossing*, Probab. Theory Related Fields **109** (1997), 27–37.

[3] W. Th. F. den Hollander, M. Keane: *Ergodic properties of color records*, Physica **138A** (1986), 183–193.

[4] A. Kalikow: T, T^{-1} *transformation is not loosely Bernoulli*, Ann. Math. **115** (1982), 393-409.

[5] S. Kakutani: *On the equivalence of infinte product measures*, Ann. of Math **49** (1948), 214–224.

[6] H. Kesten: *Detecting a single defect form observing the scenery along a random walk path*, Ito's stochastic Calculus and Probability Theory, Springer, Tokyo, (1996), 171–183.

[7] H. Kesten: *Distinguishing and reconstructing sceneries from observations from random walk paths*, Microsurveys in discrete probability (D. Aldous and J. Propp, eds.), DIMACS Series in Discrete Mathematics and Theoretical Computer Sciences **41** (1998), 75–83.

[8] E. Lindenstrauss: *Indistinguishable Sceneries*, random Struct. Alg. **14** (1999), 71–86.

[9] M. Löwe and H. Matzinger, *Scenery Reconstruction in two dimensions with many colors*, Eurandom-Report 99-018 (1999), submitted.

[10] M. Löwe and H. Matzinger, *Reconstruction of Sceneries with Correlated Colors*, Eurandom-Report 99-032 (1999), submitted.

[11] M. Löwe and H. Matzinger, *Reconstructing a one dimensional Scenery form the Observation of a random Walk with Jumps* (1999), in preparation.

[12] D.A. Levin, R. Pemantle and Y. Peres; *A Phase Transition in Random Coin Tossing* (1999), Preprint.

[13] H. Matzinger; *Reconstructing a 3-color scenery by observing it along a simple random walk*, Random Structures Algorithms **15** (1999), 196–207.

[14] H. Matzinger; *Reconstruction of a one dimensional scenery seen along the path of a random walk with holding*, Ph. D. Thesis, Cornell University (1999).