

# Ereignis- und Zeitreihenanalyse

Skriptum zu den Vorlesungen  
Evaluation und Forschungsmethoden (Statistik III + IV)<sup>1</sup>

U. Mortensen

Fachbereich Psychologie und Sportwissenschaften, Institut III  
Westfälische Wilhelms-Universität  
Fliegenerstr. 21

---

<sup>1</sup>Letzte Korrektur: 16. 07. 2014/Vorlesungen/katdata/survival

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung in die Ereignisanalyse</b>	<b>4</b>
1.1	Einführung . . . . .	4
1.2	Die Hazardfunktion und Verteilungen von Wartezeiten . . . . .	5
1.2.1	Die Definition der Hazardfunktion . . . . .	5
1.2.2	Die Exponentialverteilung und der Begriff der "reinen Zufälligkeit" . . . . .	7
1.2.3	Die Weibull-Verteilung . . . . .	13
1.2.4	Die Gompertz- bzw. Makeham-Verteilung . . . . .	18
1.3	Zur Abschätzung der Hazardfunktion . . . . .	20
1.3.1	Zensierungen . . . . .	20
1.3.2	Die Survivor-Funktion für diskrete Intervalle . . . . .	20
1.3.3	Die Sterbetafelmethode . . . . .	21
1.3.4	Der Kaplan-Meier-Schätzer der Survivor-Funktion . . . . .	22
1.3.5	Der log-Rang-Test . . . . .	23
1.4	Der Effekt exogener Variablen . . . . .	23
1.4.1	Das Exponentialmodell . . . . .	23
1.4.2	Das Weibull-Modell . . . . .	24
1.4.3	Das Proportional-Hazard-Modell von Cox . . . . .	25
1.4.4	Competing-Risk-Modelle . . . . .	27
1.4.5	Der Mehr-Episoden-Fall . . . . .	28
1.5	Markov-Ketten . . . . .	29
1.5.1	Grundlegende Definitionen . . . . .	29
1.5.2	Verweildauern und Semi-Markov-Prozesse . . . . .	34
1.5.3	Schätzungen für die Übergangswahrscheinlichkeiten . . . . .	36
1.5.4	Dyadische Interaktion . . . . .	37
<b>2</b>	<b>Zeitreihenanalyse</b>	<b>46</b>
2.1	Einführung . . . . .	46
2.2	ARIMA-Prozesse . . . . .	48
2.2.1	Autoregressive und Moving-Average Prozesse . . . . .	48

2.2.2	ARMA-Prozesse . . . . .	56
2.2.3	Integrierte Prozesse . . . . .	57

# Kapitel 1

## Einführung in die Ereignisanalyse

### 1.1 Einführung

Eine Reihe von Fragestellungen läßt sich über die Betrachtung von *Wartezeiten* bis zum Eintreffen bestimmter Ereignisse empirisch angehen:

1. Wie lange leben Menschen? Abgesehen davon, dass die Frage von allgemeinem Interesse ist, kann diese Frage Aufschluß über die Wirkung von Lebensgewohnheiten geben, so kann Stress, fett- und fuckerhaltige Nahrung, oder auch einfach das Geschlecht einen Einfluß auf die Lebensdauer haben.
2. Hat sich der Zeitpunkt der ersten Heiratsschließung verschoben? Viele sozio-ökonomische Entwicklungen wirken auf diesen Zeitpunkt ein, und die Betrachtung dieses Zeitpunktes kann Aufschluß über solche gesellschaftlichen Entwicklungen geben.
3. Ist eine neue Therapie besser als die bisher bekannten? Bei schweren Krankheiten wie Krebs oder AIDS kann hier die "Wartezeit" bis zum Tod u. U. Antwort geben: bei besseren Therapien sollten die Wartezeiten länger sein. Bei psychischen Erkrankungen mit zyklischer Komponente kann die Zeitdauer bis zur nächsten Behandlung betrachtet werden; je größer diese Dauern, desto besser ist die Therapie.
4. In der Allgemeinen Psychologie können Reaktionszeiten als Wartezeiten interpretiert werden. Führen unterschiedliche Experimentalbedingungen zu unterschiedlichen Reaktionszeiten, so läßt dies u. U. Rückschlüsse auf die Reizverarbeitungsmechanismen zu.
5. In der Sozialpsychologie wird u.a. die Interaktion von Personen in Gruppen untersucht. Hier können die Zeiten, die zwischen den Interaktionen zwischen verschiedenen Gruppenmitgliedern verstreichen, Aufschluß über die interne Struktur und den Einfluß externer Variablen liefern.

Die Liste der Fragestellungen und Beispiele läßt sich verlängern, mag aber genügen. Man sieht aber, dass sich sofort eine Fragestellung anschließt, nämlich wie sich denn Wartezeiten miteinander vergleichen lassen. Dazu müssen die Verteilungen von Wartezeiten

charakterisiert werden. Weiter muß gefragt werden, wie sich externe Einflüsse auf die Verteilung von Wartezeiten auswirken. Im folgenden Abschnitt wird zuerst eine allgemeine Charakterisierung von Wartezeitverteilungen vorgestellt, die auf dem Begriff der *Hazardfunktion* (auch Hazardrate oder Conditional Failure Rate genannt) beruht. Die Hazardrate läßt bestimmte Deutungen hinsichtlich der zugrundeliegenden Prozesse zu und erlaubt überdies, auf relativ einfache Weise die Wirkung externer Variablen zu modellieren. Es schließt sich die Darstellung einiger nichtparametrischer und parametrischer Verfahren an.

## 1.2 Die Hazardfunktion und Verteilungen von Wartezeiten

### 1.2.1 Die Definition der Hazardfunktion

Da Wartezeiten zufällige Aspekte haben, wird die Diskussion der Wartezeiten oder ihrer Parameter (Mittelwert, Varianz, etc) in irgendeiner Weise an die Verteilung dieser Zeiten gebunden sein. Die Standardvermutung, dass zufällige Veränderliche normalverteilt sind, kann allerdings bestenfalls als Annäherung gelten: da Wartezeiten nie negativ sein können, scheidet die Normalverteilung als Verteilung für Wartezeiten grundsätzlich aus, obwohl diese Verteilung, wie bereits angedeutet, gelegentlich als Annäherung (Approximation) gute Dienste leisten mag. Es zeigt sich allerdings, dass die Vermutung, die jeweils betrachteten Ereignisse würden rein zufällig geschehen, überhaupt nicht auf die Normalverteilung führt, sondern auf eine Verteilung ganz anderen Typs. Da "reine Zufälligkeit" in Untersuchungen verschiedener Art oft in der Formulierung von Nullhypothesen auftritt, muß man diese Verteilung spezifizieren. Es zeigt sich nun, dass die Verteilung von Wartezeiten grundsätzlich durch eine bestimmte Funktion, die *Hazardfunktion*, definiert werden können; dementsprechend spielen Hazardfunktionen in Untersuchungen, in denen Wartezeiten oder Reaktionszeiten erhoben werden, eine zentrale Rolle. Die Hazardfunktion soll nun erklärt werden.

Man betrachtet ein Ereignis  $A$ , das zufällig in der Zeit eintrete. Die Wartezeit bis zum Eintreffen von  $A$  werde durch eine zufällige Veränderliche  $\tau$  repräsentiert. Die Verteilungsfunktion von  $\tau$  sei durch

$$F(t) = P(\tau \leq t) \quad (1.1)$$

gegeben, und die zugehörige Dichtefunktion sei

$$\frac{dF(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = F'(t) = f(t), \quad \Delta t > 0. \quad (1.2)$$

Angenommen, man wartet auf das Ereignis  $A$ , das aber bis zum Zeitpunkt  $t$  noch nicht eingetreten sei. Es wird also  $\tau > t$  sein. Man kann sich fragen, mit welcher Wahrscheinlichkeit  $A$  nun in dem Intervall  $[t, t + \Delta t)$  eintritt. Allgemein kann man nach der bedingten Wahrscheinlichkeit, dass das Ereignis im Intervall  $[t, t + \Delta t)$  eintritt unter der Bedingung, dass es bis zum Zeitpunkt  $t$  noch nicht eingetreten ist, fragen. Dies ist die Wahrscheinlichkeit

$$P(\tau \in [t, t + \Delta t) | \tau > t) = \frac{P(\tau \in [t, t + \Delta t) \cap \tau > t)}{P(\tau > t)} = \frac{P(\tau \in [t, t + \Delta t))}{1 - F(t)} \quad (1.3)$$

fragen; die rechte Seite dieser Gleichung ergibt sich, weil  $P(\tau \in [t, t + \Delta t) \cap \tau > t) = P(\tau \in [t, t + \Delta t) \cap \tau > t)$  ist, wie man sich leicht klar macht, und weil natürlich

$P(\tau > t) = 1 - F(t)$  ist. Gelegentlich wird

$$S(t) \stackrel{\text{def}}{=} 1 - F(t) \quad (1.4)$$

als "Survivor-Funktion" definiert; der Ausdruck ergibt sich sofort aus der Bedeutung von  $S(t)$ . Statt (1.3) kann man dann auch

$$P(\tau \in [t, t + \Delta t) | \tau > t) = \frac{P(\tau \in [t, t + \Delta t))}{S(t)} \quad (1.5)$$

schreiben. Die intuitive Bedeutung dieser bedingten Wahrscheinlichkeit macht man sich leicht klar: hängt  $P(\tau \in [t, t + \Delta t) | \tau > t)$  nicht von  $t$  ab, so gibt es keinen "unterliegenden" Prozess, der das Eintreten von  $A$  mit zunehmendem  $t$  wahrscheinlicher oder weniger wahrscheinlich macht. Insofern ist dann das Eintreten von  $A$  "rein zufällig". Umgekehrt wird man aus der möglichen Beobachtung, dass die bedingte Wahrscheinlichkeit mit  $t$  größer oder kleiner wird, schließen wollen, dass es eben einen unterliegenden Prozess gibt, der eine derartige Entwicklung bedingt. Diesen Schluß kann man aber nur mit Vorsicht ziehen; darauf wird weiter unten noch eingegangen.

Bevor nun auf die Bedeutung dieser bedingten Wahrscheinlichkeit näher eingegangen wird, soll der Fall  $\Delta t \rightarrow 0$  betrachtet und seine Implikation für die Definition der Verteilungsfunktion  $F$  hergeleitet werden. Es ist ja

$$\frac{P(\tau \in [t, t + \Delta t))}{1 - F(t)} = \frac{F(t + \Delta t) - F(t)}{1 - F(t)}$$

Weiter ist

$$\lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{dF(t)}{dt} = f(t),$$

so dass man

$$\lim_{\Delta t \rightarrow 0} (F(t + \Delta t) - F(t)) = f(t)dt$$

schreiben kann;  $dt$  steht hier für das "beliebig kleine"  $\Delta t$ . Dementsprechend folgt

$$\lim_{\Delta t \rightarrow 0} \frac{P(\tau \in [t, t + \Delta t))}{1 - F(t)} = \frac{f(t)dt}{1 - F(t)} \quad (1.6)$$

**Definition 1.1** Die Funktion

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{F'(t)}{1 - F(t)}, \quad F'(t) = \frac{dF(t)}{dt} \quad (1.7)$$

heißt Hazardfunktion.

Es sei noch einmal betont, dass  $\lambda(t)$  keine Wahrscheinlichkeit ist, vielmehr ist  $\lambda(t)dt$  die Wahrscheinlichkeit für das Ereignis  $\tau \in [t, t + dt)$  unter der Bedingung, dass  $\tau > t$  ist. Die Beziehung der Hazardfunktion zur Verteilungsfunktion  $F$  wird im folgenden Satz hergestellt

**Satz 1.1** Es sei  $\tau$  eine zufällige Wartezeit mit der Verteilungsfunktion  $F$  und der Hazardfunktion  $\lambda$ . Dann gilt

$$F(t) = 1 - \exp\left(-\int_0^t \lambda(\tau)d\tau\right). \quad (1.8)$$

**Beweis:** dass die in (1.8) auftretende Funktion  $\lambda$  tatsächlich die in Definition 1.1 eingeführte Hazardfunktion ist, ergibt sich sofort aus (1.8); es ist ja

$$\frac{dF(t)}{dt} = f(t) = -\exp\left(-\int_0^t \lambda(\tau)d\tau\right) \cdot (-\lambda(t)) = (1 - F(t))\lambda(t),$$

woraus sofort  $\lambda(t) = f(t)/(1 - F(t))$  folgt; also ist  $\lambda(t)$  die Hazardfunktion.

Nun gelte umgekehrt (1.7), es ist zu zeigen, dass daraus (1.8) folgt. Es sei

$$y(t) \stackrel{\text{def}}{=} -\log(1 - F(t)), \quad f(t) = F'(t)$$

Dann ist

$$\frac{dy(t)}{dt} = \frac{f(t)}{1 - F(t)} = \lambda(t),$$

so dass

$$\int^t \frac{dy(\tau)}{d\tau} = -\int^t \lambda(\tau)d\tau = y(t) = -\log(1 - F(t)).$$

Hieraus folgt wiederum

$$\exp\left(-\int^t \lambda(\tau)d\tau\right) = 1 - F(t),$$

und durch Auflösen nach  $F(t)$  erhält man (1.8). □

## 1.2.2 Die Exponentialverteilung und der Begriff der "reinen Zufälligkeit"

**Definition der Exponentialverteilung:** Die Exponentialverteilung wird zunächst formal eingeführt; eine Motivation für die Definition der Exponentialverteilung ergibt sich dann aus dem Begriff der "reinen Zufälligkeit", wie weiter unten erläutert werden wird.

Es werde der Spezialfall  $\lambda = \text{const}$  betrachtet, d.h.  $\lambda$  soll nicht von der Zeit abhängen. Dann kann das Eintreten des Ereignisses  $A$  als "rein zufällig" betrachtet werden. Die Verteilungsfunktion der Wartezeit ist nach (1.8) durch

$$F(t) = 1 - e^{-\lambda t} \tag{1.9}$$

gegeben; dies ist die *Exponentialverteilung*. Die Dichtefunktion der Exponentialverteilung ist durch

$$f(t) = \lambda e^{-\lambda t}, \tag{1.10}$$

und der Erwartungswert und die Varianz von  $\tau$  ist durch

$$E(\tau) = \frac{1}{\lambda}, \quad \text{Var}(\tau) = \frac{1}{\lambda^2} \tag{1.11}$$

gegeben.

**Anmerkung:** Die Exponentialverteilung ist eine stetige Verteilung; sie entspricht der geometrischen Verteilung im diskreten Fall: es wird eine Reihe von voneinander unabhängigen Versuchen (Bernoulli-Versuche) gemacht, bei denen ein Ereignis  $E$  mit der Wahrscheinlichkeit  $p$  eintritt oder mit der Wahrscheinlichkeit  $1 - p$  nicht eintritt.  $X$  sei

eine zufällige Veränderliche, die die Anzahl der Versuche bis zum ersten Versuch, bei dem  $E$  eintritt, repräsentiert. Dann ist

$$P(X = k) = p(1 - p)^k; \quad (1.12)$$

Die *Gesamtzahl* der Versuche ist hier gleich  $k + 1$ . Man kann durch gewisse Grenzwertbetrachtungen von (1.12) zur Definition der Exponentialverteilung gelangen, aber dieser Ansatz soll hier nicht weiter verfolgt werden.  $\square$

**Die Hazardfunktion der Exponentialverteilung:** Die Hazardfunktion wurde allgemein in (1.7), Seite 6, definiert. Für die Exponentialverteilung erhält man

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{\lambda \exp(-\lambda t)}{1 - (1 - \exp(-\lambda t))} = \lambda, \quad (1.13)$$

d.h.  $\lambda(t) = \lambda$  ist eine Konstante. Dies heißt, dass die bedingte Wahrscheinlichkeit, dass ein Ereignis im Intervall  $[t, t + dt)$  eintritt unter der Bedingung, dass es bis zum Zeitpunkt  $t$  noch nicht eingetreten ist, ist von  $t$  unabhängig. Die bedingte Wahrscheinlichkeit  $P(\tau \in [t, t + dt) | \tau > t)$  wird also nicht größer mit wachsendem  $t$ , und sie wird nicht kleiner. Bei einem Auto wird die Wahrscheinlichkeit, dass eine Reparatur nötig wird unter der Bedingung, dass keine Reparatur bis zum Zeitpunkt  $t$  nötig war, aufgrund der Verschleißprozesse größer. Gelänge es, verschleißfreie Autos zu bauen, so wäre  $\lambda(t) = \lambda$  konstant, aber nicht notwendig gleich Null, denn es ist denkbar, dass die Einzelteile einem "rein zufälligen" Zerfallsprozess ausgesetzt sind, der vom Gebrauch des Autos unabhängig ist. In diesem Sinne definiert (1.13) den "reinen Zufall".

Man kann die Hazardfunktion natürlich auch für den Fall einer diskreten Verteilung betrachten. Das diskrete Äquivalent der Exponentialverteilung ist die geometrische Verteilung: diese definiert die Wahrscheinlichkeit, dass die Anzahl der Versuche, die gemacht werden, bevor zum ersten Mal eine bestimmte Beobachtung gemacht wird, gleich  $n$  ist unter der Bedingung, dass die "Erfolgs"wahrscheinlichkeit bei jedem Versuch gleich  $p$  ist:

$$P(X = n) = p(1 - p)^{n-1}, \quad P(X \leq n) = 1 - (1 - p)^n \quad (1.14)$$

Die Hazardfunktion ist dann durch

$$\lambda(n) = \frac{p(1 - p)^{n-1}}{(1 - p)^n} = \frac{p}{1 - p} \quad (1.15)$$

gegeben. Offenbar ist wieder  $\lambda(n) = \lambda$  für alle  $n$ .

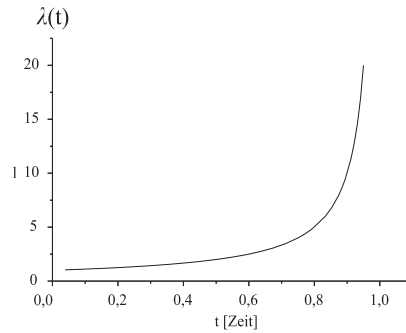
Der Begriff des "rein Zufälligen" wird oft mit der Gleichverteilung in Verbindung gebracht, – man denkt an den Würfel, bei dem angenommen wird, dass bei einem Wurf alle Seiten die gleiche Wahrscheinlichkeit haben. Nach (1.13) wird aber die Exponentialverteilung mit dem "rein Zufälligen" in Verbindung gebracht, und damit mit dem Begriff der bedingten Wahrscheinlichkeit. Das folgende Beispiel illustriert den Fall der Gleichverteilung.

**Beispiel 1.1** Das Ereignis  $A$  trete zufällig zwischen  $15^{00}$  und  $16^{00}$  ein, und  $\tau$  sei in dieser Zeit gleichverteilt, d.h. die Verteilungsfunktion sei durch

$$F(t) = \begin{cases} t/T, & t \in (0, T) \\ 0, & t \notin (0, T) \end{cases} \quad (1.16)$$



Abbildung 1.1: Hazardfunktion der Gleichverteilung



gegeben; dabei ist 0 für  $15^{00}$ ,  $T$  für  $16^{00}$  unter Berücksichtigung irgendwelcher Maßeinheiten (Sekunden, Minuten) gesetzt. Ebenso wird  $t$  in diesen Maßeinheiten gemessen; die Maßeinheiten müssen nicht explizit angegeben werden, weil sie sich im Quotienten  $t/T$  herauskürzen. Die Dichtefunktion ist dann

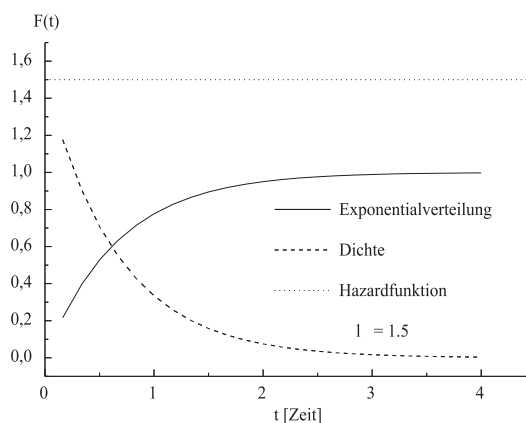
$$f(t) = 1/T, \quad t \in (0, T). \quad (1.17)$$

Für die Hazardfunktion findet man

$$\lambda(t) = \frac{1}{T(1 - t/T)} = \frac{1}{T - t}, \quad t \in (0, T). \quad (1.18)$$

Zu Beginn des Intervalles, um  $15^{00}$ , ist  $\lambda(0) = 1/T$ , und steigt dann mit  $t \rightarrow T$  gegen unendlich: ist  $A$  bis zum Zeitpunkt  $t$  nicht eingetreten, so wird natürlich das Auftreten von  $A$  immer wahrscheinlicher, je größer der Wert von  $t$  ist. Der Verlauf der Hazardfunktion wird in Abb. 1.1 gezeigt (für  $T = 1$ ).  $\square$

Abbildung 1.2: Exponentialverteilung, Dichte und Hazardfunktion



### Zum Begriff der reinen Zufälligkeit

Wie in den Beispielen deutlich wird, kann man zwei Verteilungen - die Exponentialverteilung und die Gleichverteilung - mit dem Begriff der "reinen Zufälligkeit" in Verbindung bringen, - offenbar ist der Begriff der reinen Zufälligkeit weniger eindeutig, als es auf den ersten Blick scheint. Die Gleichverteilung charakterisiert den Begriff gut, wenn ein endliches Intervall als Zeitraum, in dem das in Frage stehende Ereignis  $A$  eintreten kann, angenommen wird, und wenn man vor Beginn dieses Intervalls Aussagen über das Eintreten von  $A$  machen will. Ist man bereits in diesem Intervall und ist  $A$  bis zum Zeitpunkt  $t \in (0, T)$  noch nicht eingetreten, so ist die Wahrscheinlichkeit, dass es in einem Teilintervall des verbleibenden Zeitraums eintritt, wesentlich größer als die Wahrscheinlichkeit für dieses gleiche Intervall zu Beginn der Beobachtungszeit. Bei der Exponentialverteilung wird das Intervall  $[0, \infty)$  als möglicher Zeitraum betrachtet. Die Wahrscheinlichkeit, dass  $A$  in einem Intervall der Länge  $\Delta t$  zu Beginn des Beobachtungszeitraums eintritt ist höher als die Wahrscheinlichkeit, in einem Intervall der gleichen Länge  $\Delta t$ , dass zu einem späteren Zeitpunkt beginnt:

$$\begin{aligned} P(\tau \in [t, t + \Delta t)) &= F(t + \Delta t) - F(t) \\ &= 1 - e^{-\lambda(t+\Delta t)} - (1 - e^{-\lambda t}) \\ &= e^{-\lambda t} - e^{-\lambda(t+\Delta t)} \\ &= e^{-\lambda t}(1 - e^{-\lambda\Delta t}) \rightarrow 0, \quad t \rightarrow \infty \end{aligned}$$

denn  $1 - \exp(-\lambda\Delta t)$  ist ja eine Konstante. Bei der Gleichverteilung dagegen ist

$$P(\tau \in [t, t + \Delta t)) = \frac{t + \Delta t}{T} - \frac{t}{T} = \frac{\Delta t}{T} = \text{const.}$$

wobei die Nebenbedingung, dass  $(t, t + \Delta t)$  in  $(0, T)$  liegen muß, zu berücksichtigen ist.

**Zufällige Zeitpunkte:** Es ist aber nicht so, dass Gleich- und Exponentialverteilung auf völlig verschiedene Ansätze zurückgehen. Der Zusammenhang wird deutlicher, wenn man die Verteilung von *zufälligen Zeitpunkten* betrachtet. Man kann z.B. sagen, dass ein Ereignis zufällig in  $(0, T)$  auftritt, wenn

$$P(\tau \leq t) = \frac{t}{T} \tag{1.19}$$

gilt, wenn also die Zeitpunkte für das Auftreten des Ereignisses  $A$  gleichverteilt sind. Dann folgt für  $0 \leq t_1 < t_2 \leq T$

$$P(t_1 \leq \tau \leq t_2) = \frac{t_2 - t_1}{T}.$$

Man betrachte nun das Auftreten von  $n$  zufälligen, *voneinander unabhängigen* Ereignissen, und bestimme die Wahrscheinlichkeit, dass  $k$  davon im Intervall  $(t_1, t_2) \subset (0, T)$  auftreten. dass ein Ereignis in diesem Intervall auftritt, hat die Wahrscheinlichkeit

$$p = \frac{t_2 - t_1}{T} = \frac{\Delta t}{T}, \quad \Delta t = t_2 - t_1.$$

dass  $k$  Ereignisse auftreten, hat dann gemäß der Binomialverteilung die Wahrscheinlichkeit

$$P(K = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Es mögen nun die Bedingungen (i)  $n$  "groß",  $p = \Delta t/T$  "klein" und (ii)  $k \sim n\Delta t/T$  gelten; eine analoge Forderung ist (i)  $n \rightarrow \infty$  und  $p \rightarrow 0$  derart, dass (ii)  $np \rightarrow \lambda < \infty$  strebt. Dann strebt die Binomialverteilung gegen die Poisson-Approximation

$$P(K = k \in (t, t + \Delta t)) \rightarrow e^{-\lambda\Delta t} \frac{(\Delta t\lambda)^k}{k!} \quad (1.20)$$

Man sieht, dass die rechte Seite nicht von  $t$  abhängt, also gilt (1.20) für irgendein ein Intervall aus  $(0, T)$ , so dass man insbesondere  $t = 0$  annehmen kann. Nun sei  $k = 0$ , d.h. man betrachte den Fall, dass kein Ereignis während der Dauer  $\Delta t$  eintritt. Dies ist gleichbedeutend damit, dass man mindestens  $\Delta t$  Zeiteinheiten warten muß, bevor ein Ereignis  $A$  eintritt. Die Wahrscheinlichkeit dafür ist nach (1.20)

$$P(K = 0) = P(\tau > \Delta t) = e^{-\lambda\Delta t}. \quad (1.21)$$

Also ist die Verteilungsfunktion  $F(\Delta t) = P(\tau \leq \Delta t)$  der Wartezeit bis zum ersten Eintreten eines Ereignisses durch

$$F(\Delta t) = 1 - e^{-\lambda\Delta t}. \quad (1.22)$$

Dies ist aber die Exponentialverteilung.

Der Wert von  $\lambda$  kann "groß" oder "klein" sein; je größer der Wert von  $\lambda$  ist, desto schneller geht  $F(\Delta t)$  gegen 1. Ist  $\lambda$  "groß", so wird  $p$  "langsamer" gegen 0 gehen als  $T$  gegen unendlich strebt, so dass die Wahrscheinlichkeit, dass das in Frage stehende Ereignis in einem gegebenen Intervall eintritt, "groß" ist. Man kann sagen, dass der Prozess "intensiver" verläuft, weshalb  $\lambda$  auch "Intensitätsparameter" heißt.

**Beispiel: Alpha-Zerfall** Für hinreichend große Anzahl von Kernen ist die Anzahl von Zerfällen pro Zeit durch

$$\frac{dN(t)}{dt} = -\lambda N(t) \quad (1.23)$$

gegeben, woraus

$$N(t) = N_0 e^{-\lambda t} \quad (1.24)$$

folgt (Fließbach (1995), p. 164)<sup>1</sup>; zum Zeitpunkt  $t$  sind von  $N_0$  Teilchen noch  $N(t)$  übrig. Demnach ist

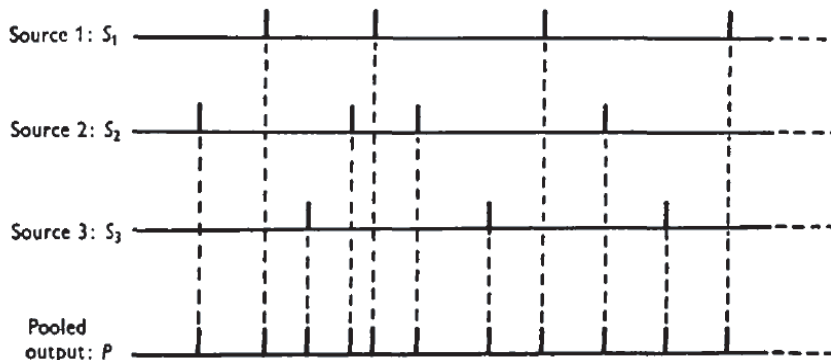
$$\frac{N_0 - N(t)}{N_0} = 1 - \frac{N(t)}{N_0} = 1 - e^{-\lambda t} \quad (1.25)$$

der Anteil der bis zur Zeit  $t$  zerfallenen Teilchen; dies ist gerade die Verteilungsfunktion der Exponentialverteilung und entspricht der Wahrscheinlichkeit, dass ein Teilchen bis zur Zeit  $t$  zerfallen ist. Die Wartezeit bis zum Zerfall eines Teilchens ist also exponentialverteilt, und nach dem Vorgegangenen ist dieser Befund mit der Interpretation, dass der Zerfall ein rein zufälliger Prozess ist, verträglich, wobei "rein zufällig" in diesem Fall mit "akausal" gleichgesetzt werden kann.

Allerdings ist diese Gleichsetzung logisch nicht zwingend. Zur Begründung der Aussage, dass der Schluß nicht zwingend ist, muß nur ein Beispiel für einen Prozess angegeben werden, bei dem exponentialverteilte Wartezeiten beobachtet werden, die aber *nicht* auf Akausalität zurückzuführen sind. Cox und Smith (1953) betrachteten den Fall, dass es eine Reihe von "Quellen" oder Mechanismen gibt, die in regelmäßigen Abständen ein "Ereignis" erzeugen; die Quelle  $Q_j$  erzeugt eine Folge  $S_j$ . Man betrachtet nun die Folge  $S$  von

<sup>1</sup>Fließbach, T.: Quantenmechanik. Lehrbuch zur Theoretischen Physik III. Heidelberg 1995

Abbildung 1.3: Überlagerte deterministische Folgen (Cox & Smith (1953))



Ereignissen, die sich aus der Überlagerung der Folgen  $S_j$  ergibt (pooling of events, vergl. Abbildung 1.3). Man betrachtet etwa eine Menge von Neuronen, bei der jedes Neuron in regelmäßigen Abständen "feuert", also ein Aktionspotential erzeugt. Regelmäßig soll heißen, dass das  $i$ -te Neuron zu den Zeiten  $\Delta t_i, 2\Delta t_i, 3\Delta t_i, \dots$  feuert, wobei der Wert  $\Delta t_i$  für das  $i$ -te Neuron charakteristisch ist.  $S_i$  sei die Folge der Zeiten  $\Delta t_i, 2\Delta t_i, 3\Delta t_i, \dots$ . Das Intervall  $\Delta t_i$  ist die "Periode" für das  $i$ -te Neuron. Die  $\Delta t_1, \Delta t_2, \Delta t_3, \dots$  seien alle verschieden und es gelte  $\sum_{i=1}^N n_i \Delta t_i = 0$ , wobei  $\Delta t_i > 0$  für alle  $i$  und die  $n_i$  natürliche Zahlen sind, die nicht alle gleich Null sind, und die Quotienten  $\Delta t_i / \Delta t_j, i \neq j$ , seien alle irrational<sup>2</sup>. Ausgehend von einem zahlentheoretischen Theorem<sup>3</sup> von Weyl (1916) zeigen Cox et al. dann, dass die Verteilung der Zeiten zwischen benachbarten Ereignissen der Folge  $S$  gegen eine Exponentialverteilung strebt und dass benachbarte Zwischenzeiten stochastisch unabhängig sind; unter bestimmten Randbedingungen, auf die hier nicht weiter eingegangen werden muß, ist dann die Folge  $S$  als Poisson-Prozess zu beschreiben. Dies bedeutet, dass "reine" Zufälligkeit im eingangs erläuterten Sinne zur Gleichung (1.13) führt, dass umgekehrt von der im Prinzip beobachtbaren Tatsache, dass Zwischenzeiten exponentialverteilt sind, nicht auf die "reine" Zufälligkeit geschlossen werden kann, denn die Folgen  $S_i$  und damit auch die Folge  $S$  sind ja nicht zufällig, – man beachte, dass die  $\Delta t_i$  nicht zufällig gewählt werden müssen, die einzige Forderung ist ja, dass  $\Delta t_i / \Delta t_j$  für  $i \neq j$  irrational sein muß. Betrachtet man also eine Folge von Klicks, wie sie von einem Geigerzähler aufgrund zerfallender Teilchen erzeugt werden, so findet man zwar, dass die Zwischenzeiten der Klicks der Verteilung (1.25) genügen, aber man kann nicht zwingend folgern, dass der zugrundeliegende Zerfallsprozess "rein" zufällig im Sinne von akausal ist.

Das Ziel dieser Betrachtungen ist nicht, nachzuweisen, dass es den "reinen Zufall" nicht gibt, sondern dass man nicht notwendig auf Akausalität schließen kann, auch wenn manche Daten dies nahelegen. Die Idee des "reinen Zufalls" kann oft auch durch die Annahme des "averaging over uncertainty" (Passon (2004)) ersetzt werden. Es ist dann praktisch unmöglich, alle Aspekte zu kennen, die einen Prozess ausmachen, aber es muß

<sup>2</sup>Eine reelle Zahl  $x$  heißt *irrational*, wenn sie nicht in der Form  $p/q$  dargestellt werden kann, wobei  $p$  und  $q$  natürliche Zahlen sind. Das Wort 'irrational' hat hier nicht die Bedeutung von 'unvernünftig', sondern bezieht sich auf die ursprüngliche Bedeutung von 'ratio' als Verhältnis bzw. Quotient. 'Irrational' heißt dann eben nur, dass die Zahl nicht als Verhältnis darstellbar ist.

<sup>3</sup>Es sei  $\{\Theta\}$  der Dezimalteil von  $\Theta$ , also für  $\Theta = \pi = 3.14159\dots$  ist  $\{\Theta\} = .14159\dots$ , und es sei  $I_l$  irgendein Intervall der Länge  $l$  in  $(0, 1)$ , und  $p_m(I_l)$  sei der Anteil von  $\{\Theta\}, \{2\Theta\}, \dots, \{m\Theta\}$ , der in das Intervall  $I_l$  falle. Dann gilt  $\lim_{m \rightarrow \infty} p_m(I_l) = l$ .

nicht Akausalität postuliert werden.

### 1.2.3 Die Weibull-Verteilung

Die Exponentialverteilung ergab sich in (1.22) als Wartezeit bis zum ersten Ereignis in einem Poisson-Prozess, der als rein zufällige Folge von Ereignissen betrachtet werden kann, – falls die Ereignisse mit konstanter Wahrscheinlichkeit und unabhängig voneinander eintreten. Es ist aber möglich, dass Ereignisse immer wahrscheinlicher werden, je länger man wartet, oder dass sie immer unwahrscheinlicher werden, je länger man wartet. Die Weibull-Verteilung liefert eine Möglichkeit, derartige Fälle zu modellieren.

Der schwedische Ingenieur Ernst Hjalmar Waloddi Weibull (1887 – 1979) führte Betrachtungen zur Wahrscheinlichkeit des Brechens bei sprödem Material durch und wurde so auf die nach ihm benannte Verteilung geführt<sup>4</sup>, die in der Zuverlässigkeits- (Reliabilitäts-)analyse von Materialien oder Bauteilen eine zentrale Rolle spielt. Die Weibull-Verteilung wird in verschiedenen Texten auf verschiedene Weise eingeführt. Hier wird zunächst die in Johnson et al. (1994)<sup>5</sup> gegebene Definition gegeben, aus der sich die übrigen Formen ergeben.

**Definition 1.2** Die zufällige Veränderliche  $X$  ist Weibull-verteilt, wenn die Verteilungsfunktion von  $X$  durch

$$P(X \leq x) = \exp \left[ - \left( \frac{x_0 - x}{a} \right)^b \right], \quad x \leq x_0, \quad a > 0, \quad b > 0 \quad (1.26)$$

gegeben ist.

**Anmerkung:**  $X$  ist auf dem Intervall  $(-\infty, x_0]$  definiert; man wird  $X$  also nicht zur Repräsentation von Zeiten heranziehen. Um die Weibull-Verteilung auf Wartezeiten anwenden zu können, ist eine Transformation von  $X$  nötig (s. unten). Der Parameter  $a$  ist ein Skalierungsparameter und definiert *nicht* die Streuung von  $X$  (vergl. (1.29) unten). □

**Dichtefunktion:** Die Dichtefunktion ergibt sich sofort durch Differentiation von  $P(X \leq x)$  nach  $x$ :

$$f(x) = \frac{b}{a} \left( \frac{x_0 - x}{a} \right)^{b-1} \exp \left[ - \left( \frac{x_0 - x}{a} \right)^b \right]. \quad (1.27)$$

Für den Erwartungswert und die Varianz einer Weibull-verteiltern zufälligen Veränderlichen erhält man die Ausdrücke

$$E(X) = x_0 - a\Gamma(1 + 1/b) \quad (1.28)$$

$$Var(X) = a^2\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b) \quad (1.29)$$

Hier ist  $\Gamma(\cdot)$  die Gamma-Funktion

$$\Gamma(z) = \int_0^\infty y^{z-1} e^{-y} dy. \quad (1.30)$$

<sup>4</sup>Weibull, H.E.W.: A Statistical Distribution Function of Wide Applicability. *ASME Journal Of Applied Mechanics Paper*, 1951. Der Artikel gilt als "The hallmark paper of Weibull analysis".

<sup>5</sup>Johnson, N.L., Kotz, S., Balakrishnan, N. : Continuous univariate distributions, Vol. I, Wiley & Sons, New York, 1994

In Anwendungen wird häufig Gebrauch von der *Weibull-Funktion* gemacht. Diese Funktion definiert die Wahrscheinlichkeit eines Ereignisses in Abhängigkeit von einem physikalischen Parameter. Dosiswirkungsfunktionen und psychometrische Funktionen sind ein typisches Beispiel hierfür. Die Beziehung zur Definition (1.26) ist oft nicht klar, so dass kurz darauf eingegangen werden soll.

**Beispiel 1.2 (Dosiswirkungs- und psychometrische Funktionen)** Die Stimmung einer Person werde durch eine zufällige Veränderliche  $X$  repräsentiert; für den Fall  $X > S$  betrachtet sich die Person als "ausgeglichen", für  $X \leq S$  als "verspannt" bzw. nicht ausgeglichen.  $S$  ist ein bestimmter kritischer Wert, und es wird angenommen, dass  $X$  Weibull-verteilt ist, d.h. es existiert ein Wert  $x_0$ , den  $X$  nicht überschreiten kann,  $-x_0$  repräsentiert eine Intensität von Glücksgefühl, die nicht überschritten werden kann<sup>6</sup>. Da die betrachtete Person die Werte häufig unterhalb von  $S$  liegen, nimmt sie ein Medikament  $M$ , das als Aufheller wirkt. Gesucht ist nun die Wahrscheinlichkeit, dass bei einer Dosis  $m$  von  $M$  das Ereignis  $X > S$  eintritt.

Es sei  $g(m)$  die Wirkung von  $M$  bei einer Dosis  $m$ , und  $g$  sei eine wachsende Funktion von  $m$ . Die Stimmung werde durch den Ansatz  $U = X + g(m)$  repräsentiert;  $U$  ist eine zufällige Veränderliche. Das Glücksgefühl setzt ein, wenn  $U > S$  ist. Dann ist jedenfalls

$$P(U \leq S) = P(X + g(m) \leq S) = P(X \leq S - g(m)).$$

Da  $X$  als Weibull-verteilt vorausgesetzt wird, erhält man

$$P(U \leq S) = P(X \leq S - g(m)) = \exp \left[ - [x_0 - (S - g(m))]^b \right], \quad (1.31)$$

wobei  $a = 1$  gesetzt wurde (dies ist äquivalent der Aussage, dass  $a$  bereits in  $x_0$  und  $S - g(m)$  absorbiert wurde). Setzt man

$$S^* = x_0 - S \geq 0,$$

so findet man

$$\psi(m) = P(U > S) = 1 - \exp \left[ - [S^* + g(m)]^b \right], \quad (1.32)$$

Dies ist die Weibull-Funktion, wenn man  $S^* = 0$  (also  $x_0 = S$ ) annimmt und  $g(m) = km$  postuliert, d.h. man fordert, dass die Wirkung von  $M$  proportional zu  $m$  ist. Mit  $c = k^{1/b}$  erhält man dann die Dosiswirkungsfunktion (als Weibull-Funktion)

$$\psi(m) = 1 - \exp \left[ -cm^b \right]. \quad (1.33)$$

Dieser Ausdruck  $\psi$  findet man auch in der Psychophysik als psychometrische Funktion;  $U$  repräsentiert dann die Aktivität des entdeckenden neuronalen Systems,  $m$  ist die Stimulusintensität, und  $\psi(m)$  ist die Wahrscheinlichkeit des Entdeckens. Die häufig zu findende Bemerkung, die Weibull-Funktion definiere ein Hochschwellenmodell – für  $m = 0$  ist ja die Wahrscheinlichkeit des Entdeckens gleich Null, da  $\psi(0) = 0$ , ist aber nicht korrekt, denn es ist ja nur die stillschweigende Annahme  $S^* = 0$  gemacht worden. Eine analoge Aussage gilt für die Wirkung des Medikaments  $M$ . Die Annahme  $S^* = 0$ , d.h.  $S = x_0$ , bedeutet aber, dass normalerweise – also ohne Medikamentennahme bzw. ohne eine Stimulierung – die Stimmung nie besser als  $x_0 = S$  sein kann, die Person wäre dann in natürlicher Weise "unglücklich". Ebenso könnten die in der Psychophysik

<sup>6</sup>Ob eine solche Annahme sinnvoll ist, soll hier nicht weiter diskutiert werden.

zu beobachtenden wahren – also nicht auf Raten zurückführbaren – falschen Alarme nie auftreten. Der allgemeine Fall

$$\psi(m) = 1 - \exp \left[ - (S^* + km)^b \right]. \quad (1.34)$$

mit  $S^* \neq 0$  ist also plausibler (wobei das Postulat  $g(m) = km$  noch zu diskutieren wäre). Die Wahrscheinlichkeit eines "spontanen" Glücksgefühls oder eines wahren falschen Alarms ist dann

$$\psi(0) = 1 - \exp \left[ - (x_0 - S)^b \right]. \quad (1.35)$$

Anhand dieses Ansatzes kann die Modifizierbarkeit von  $S$  durch pay-off-Bedingungen getestet werden.  $\square$

Die in (1.26) gegebene Definition der Weibull-Verteilung ist weniger häufig zu finden, erleichtert aber die Diskussion von Dosiswirkungs- und psychometrischen Funktionen. Um nun Wartezeiten ("failure times") modellieren zu können, kann man eine Transformation der zufälligen Veränderlichen  $X$  vornehmen. Dazu sei

$$Y = 2x_0 - X, \quad (1.36)$$

und  $X$  sei wie in (1.26) verteilt. Dann findet man

$$P(Y \leq y) = P(2x_0 - X \leq y) = P(2x_0 - y \leq X) = 1 - P(X \leq 2x_0 - y),$$

d.h.

$$P(Y \leq y) = 1 - \exp \left[ - (x_0 - (2x_0 - y))^b \right],$$

also

$$P(Y \leq y) = 1 - \exp \left[ - (y - x_0)^b \right], \quad y \geq x_0, \quad (1.37)$$

wobei wieder  $a = 1$  gesetzt wurde. Der Erwartungswert ist nun – für beliebiges  $a > 0$ ,

$$E(Y) = x_0 + a\Gamma(1 + 1/b). \quad (1.38)$$

Die Varianz von  $Y$  wird wieder durch (1.29) angegeben, denn die Transformation  $Y = 2x_0 - X$  entspricht einer Transformation  $Y = \alpha X + \beta$  mit  $\alpha = -1$  und  $\beta = 2x_0$ , die die Varianz  $Var(Y) = \alpha^2 Var(X)$  impliziert, und  $\alpha^2 = (-1)^2 = 1$ .

**Beispiel 1.3 Dosiswirkung, zweite Art:** Auch die durch (1.37) definierte Verteilungsfunktion kann zur Modellierung von Dosiswirkungen benutzt werden. Die zufällige Veränderliche  $X$  repräsentiere z. B. Verspannung, und die Verteilungsfunktion sei durch (1.37) gegeben. Bis zum Grad  $S$  von Verspannung ist die Person kopfschmerzfrei, für  $X > S$  hat sie Spannungskopfschmerz. Das Medikament  $M$  soll helfen,  $g(m)$  sei wieder die Wirkung, wenn die Dosis gleich  $m$  ist, und  $g$  sei wieder eine wachsende Funktion von  $m$ . Es sei nun  $U = X - g(m)$ .  $g(m)$  wird von  $X$  subtrahiert, da  $M$  ja die Spannung reduzieren soll. Die Person ist frei von Kopfschmerzen, wenn  $U \leq S$ . Man hat

$$P(U \leq S) = P(X - g(m) \leq S) = P(X \leq S + g(m)),$$

d.h.

$$P(U \leq S|m) = \psi(m) = 1 - \exp \left[ - (S + g(m) - x_0)^b \right]. \quad (1.39)$$

Setzt man wieder  $S^* = S - x_0 \geq 0$ , so erhält man die Dosiswirkungskurve

$$\psi(m) = 1 - \exp \left[ - (S^* + g(m))^b \right] \quad (1.40)$$

und mit  $g(m) = km$  und  $S^* = 0$  erhält man wieder die übliche Weibull-Funktion; (1.40) und (1.34) sind formal gleich.

Die Anwendung von (1.37) auf die Definition von psychometrischen Funktionen ist allerdings ein wenig künstlich. Denn die Definition der korrespondierenden zufälligen Veränderlichen  $X$  ist jetzt so, dass der Stimulus entdeckt wird, wenn  $U = X - g(m) \leq S$  ist, wenn  $g$  den Effekt des Stimulus auf das entdeckende System repräsentiert. Da die Form (1.40) resultiert, muß  $g$  eine wachsende Funktion der Stimulusintensität  $m$  sein.  $X$  bzw.  $U$  kann nicht mehr als Aktivierung verstanden werden. Der Ansatz, psychometrische Funktionen über die Verteilung (1.37) zu definieren, funktioniert also nur, wenn  $X$  bzw.  $U$  eine Inhibierung des entdeckenden Systems bedeutet, die durch den Stimulus reduziert wird.  $\square$

Man sieht also, dass beide Formen der Weibull-Verteilung, (1.27) und (1.37), nützlich sind; die jeweiligen zufälligen Veränderlichen sind komplementär zueinander definiert.

**Weibull-verteilte Wartezeiten** Wartezeiten nehmen i.A. nur positive Werte an. Postuliert man, dass Wartezeiten Weibull-verteilt sind, so wird man auf die Form (1.37) rekurren. Es sei  $\tau$  die zufällige Veränderliche, die die Wartezeit repräsentiert. Im allgemeinen Fall – also für den Fall  $a \neq 1$  – erhält man dann die Verteilung

$$P(\tau \leq t) = 1 - \exp \left[ - \left( \frac{t - t_0}{a} \right)^b \right], \quad t \geq t_0 \geq 0, \quad (1.41)$$

wobei (nur aus kosmetischen Gründen!)  $t_0$  statt  $x_0$  geschrieben wurde. Der Fall  $t_0 > 0$  bedeutet, dass die Wartezeit den Betrag  $t_0$  nie unterschreiten kann. Die Dichtefunktion ist

$$f(t) = \frac{1}{a} \left( \frac{t - t_0}{a} \right)^{b-1} \exp \left[ - \left( \frac{t - t_0}{a} \right)^b \right], \quad (1.42)$$

und die Hazardfunktion ist

$$\phi(t) = \frac{f(t)}{1 - F(t)} = \frac{b}{a} \left( \frac{t - t_0}{a} \right)^{b-1}. \quad (1.43)$$

Für  $t_0 = 0$  erhält man üblicherweise betrachtete Form

$$P(\tau \leq t) = 1 - \exp \left[ - \left( \frac{t}{a} \right)^b \right], \quad t \geq 0, \quad (1.44)$$

bzw.

$$P(\tau \leq t) = 1 - \exp [-ct^b], \quad c = (1/a)^{1/b} > 0, \quad t \geq 0. \quad (1.45)$$

Für  $b = 1$  erhält man als Spezialfall die Exponentialverteilung. Die Dichtefunktion hat hier die Form

$$f(t) = cbt^{b-1} \exp(-ct^b). \quad (1.46)$$

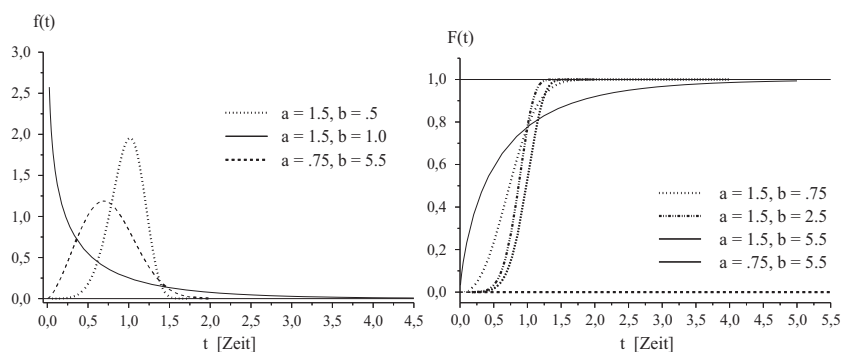
Der Erwartungswert ist

$$E(\tau) = \frac{1}{c} \Gamma(1 + 1/b) \quad (1.47)$$

$$Var(\tau) = \frac{1}{c^2} [\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)]. \quad (1.48)$$



Abbildung 1.4: Weibulldichten und -verteilungen für verschiedene Parameterwerte

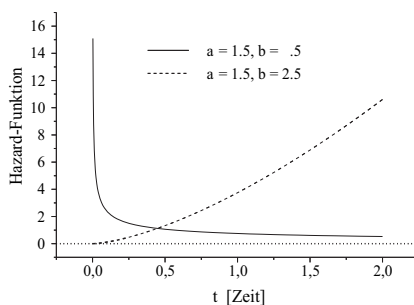


Sowohl der Erwartungswert wie die Varianz hängen von  $\alpha$  und  $\beta$  ab, so dass eine Veränderung des Erwartungswertes auch eine Veränderung der Varianz nach sich zieht.

Für die Hazardfunktion findet man

$$\phi(t) = \frac{f(t)}{1 - F(t)} = \frac{(1 - F(t))c b t^{b-1}}{1 - F(t)} = c b t^{b-1}. \quad (1.49)$$

Abbildung 1.5: Hazardfunktionen für Weibullverteilungen



**Die Weibull-Verteilung als Extremwertverteilung:** Es seien  $t_1, t_2, \dots, t_n$  Wartezeiten und es sei

$$\tau = \min(t_1, t_2, \dots, t_n). \quad (1.50)$$

Gesucht ist die Verteilung von  $\tau$ . Es läßt sich zeigen, dass unter bestimmten Bedingungen und  $n \rightarrow \infty$   $P(\tau \leq t)$  durch die Weibull-Verteilung gegeben ist, d.h. die Weibull-Verteilung ist eine von drei möglichen Extremwertverteilungen (vergl. Galambos (1978)<sup>7</sup>).

<sup>7</sup>Galambos, J.: The asymptotic theory of extreme order statistics. Wiley & Sons, New York 1978

### 1.2.4 Die Gompertz- bzw. Makeham-Verteilung

Gompertz<sup>8</sup> (1825) beschäftigte sich mit der Frage, warum Menschen sterben. Er nahm an, dass es eine bestimmte Kraft braucht, um den fortwährend auf sie einwirkenden destruktiven Einflüssen zu widerstehen. Zu Beginn seines Lebens wird der Mensch mit einer Gesamtkraft ausgestattet, die er nun im Laufe seines Lebens verbraucht. Zu Beginn eines bestimmten Zeitintervalls hat der Mensch noch einen gewissen Betrag an Kraft zur Todesvermeidung zur Verfügung, von dem ein Teil während des folgenden Zeitintervalles aufgebraucht wird. Er betrachtet dann den Grenzfall von beliebig ("unendlich") kleinen, aufeinanderfolgenden Zeitintervallen, und in jedem von ihnen verliert der Mensch einen jeweils gleich großen Anteil der noch verbleibenden Kraft. Aus dieser Annahme leitete Gompertz die Hazardfunktion

$$\lambda(t) = Bc^t, \quad t \geq 0, \quad B > 0, \quad c \geq 1 \quad (1.51)$$

ab, aus der sich die Verteilungsfunktion für die Wartezeit bis zum Ableben herleiten läßt. Die Hazardfunktion wird oft in etwas anderer Form angegeben:

$$\lambda(t) = \lambda_0 \exp(\gamma_0 t), \quad t \geq 0. \quad (1.52)$$

Tatsächlich sind (1.51) und (1.52) äquivalent. Denn es ist

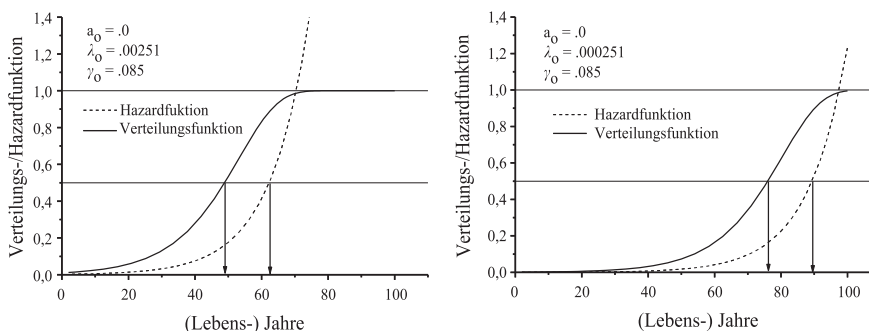
$$Bc^t = B \exp(t \log c),$$

und mit  $\lambda_0 = B$  und  $\gamma_0 = \log c$  hat man (1.52). Die Verteilungsfunktion ist dann

$$F(t) = 1 - \exp\left(-\alpha_0 t - \frac{\lambda_0}{\gamma_0} \exp(\gamma_0 t) - 1\right), \quad \lambda_0 > 0, \quad -\infty < \gamma_0 < \infty. \quad (1.53)$$

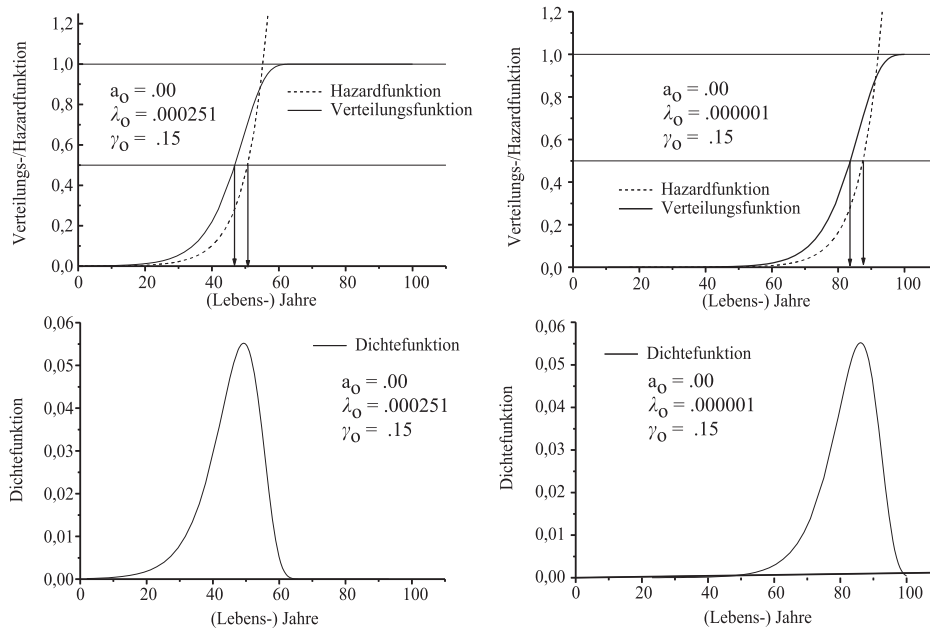
Man bemerke, dass hier negative Werte für  $\gamma_0$  zugelassen werden; die Gompertz'sche Forderung  $c \geq 1$  impliziert  $\gamma_0 = \log c \geq 0$ . Man kann aber den Fall  $c > 0$  ebenfalls zulassen. Für  $0 < c < 1$  ist dann  $\gamma_0 < 0$ . Abb. 1.6 zeigt die Gompertz-Verteilung für zwei Para-

Abbildung 1.6: Sterbewahrscheinlichkeit gemäß der Gompertz-Verteilung, I



etersätze: sie unterscheiden sich nur hinsichtlich des Parameters  $\lambda_0$ . Für den größeren Wert ( $\lambda_0 = .000251$ ) ist die Lebenserwartung deutlich höher. Mit der Wahrscheinlichkeit 1/2 wird man mindestens 76 Jahre alt, und die Wahrscheinlichkeit, mit 89.5 Jahren zu sterben unter der Bedingung, dass man überhaupt 89.5 Jahre alt wird, ist immer noch

Abbildung 1.7: Sterbewahrscheinlichkeit gemäß der Gompertz-Verteilung, II



erst  $1/2$ . Für den kleineren Wert  $\lambda_0 = .00251$  wird man mit der Wahrscheinlichkeit  $1/2$  höchstens 49 Jahre alt, und mit der Wahrscheinlichkeit von  $1/2$  stirbt man mit 62.5 Jahren, vorausgesetzt, man erreicht dieses Alter.

Die Gompertzverteilung bildet die empirisch gefundene Hazardfunktion für das Überleben insofern nicht korrekt ab, als die erhöhten (relativ zu einem Lebensalter von einigen Monaten) Sterbewahrscheinlichkeiten kurz nach der Geburt nicht korrekt wiedergegeben werden. Für Alterungsprozesse bei älteren Menschen liefert die Verteilung allerdings oft eine sehr gute Beschreibung<sup>9</sup>.

Die Verteilung wird gelegentlich auch *Makeham-Verteilung* genannt. Man kann den von Gompertz beschriebenen Kampf gegen das Ableben auch als einen beliebigen, ressourcenverbrauchenden Teilprozess betrachten und annehmen, dass diese Prozesse unabhängig voneinander sind. Es sei noch angemerkt, dass die Gompertz-Verteilung offenbar ein Spezialfall der *doppelten Exponentialverteilung*, also einer weiteren Extremwertverteilung, ist.

Ein Spezialfall ergibt sich für  $c = 1$ . Dann ist  $\log c = \gamma_0 = 0$  und die Hazardfunktion wird auf den Fall  $\lambda(t) = \lambda_0$  eine Konstante reduziert; die Gompertz-Verteilung wird damit zu einer Exponentialverteilung.

<sup>8</sup>Gompertz, B. (1825) On the nature of the function expressive of the law of human mortality. Philosophical Transactions of the Royal Society of London, Series A, 115, 513-580

<sup>9</sup>**Centenarian Divorce.** The French group of demographers and gerontologists studying Madam Calment was also studying the life style and habits of centenarians. A centenarian couple in their study decided to seek a divorce. When asked why they had waited so long to make this decision they responded, "We waited out of consideration for our children. Thus we postponed the decision until they all had died." (story told by K Ritchie over dinner at Gerontological Meeting in New Orleans, Nov., 1993).

## 1.3 Zur Abschätzung der Hazardfunktion

### 1.3.1 Zensierungen

Häufig wird eine Gruppe von Personen über einen Zeitraum hinweg beobachtet: es wird notiert, wann ein bestimmtes, im gegebenen Zusammenhang interessierendes Ereignis eintritt. Woran man wirklich interessiert ist, ist die Zeit von einem bestimmten Zeitpunkt  $t_0$  an bis zum Eintreten des Ereignisses;  $t_0$  ist etwa das Ende einer Therapie (Operation), und das Ereignis ist z.B. der Tod der Person.

Nun hat man oft eine Anzahl von Personen in der Stichprobe, bei denen der Zeitpunkt  $t_0$  nicht bekannt ist. Diese Personen heißen *linkszensiert*. Ein Beispiel hierfür sind HIV-Infizierte, bei denen der Zeitpunkt  $t_0$  der Infektion nicht bekannt ist, so dass nicht klar ist, wie lange sie den Virus bereits in sich tragen.

Auf der anderen Seite kann man Personen in der Stichprobe haben, bei denen zu dem Zeitpunkt, zu dem die Studie beendet wird, das in Frage stehende Ereignis noch nicht eingetreten ist, oder bei denen man nicht feststellen kann, ob es eingetreten ist, etwa weil sie während der Studie unbekannt verzogen sind. Diese Personen heißen *rechtszensiert*.

Man könnte nun so vorgehen, dass man alle zensierten Personen, gleich ob sie links- oder rechtszensiert sind, aus der Stichprobe entfernt. Es zeigt sich aber, dass bei einem solchen Vorgehen Information verschwendet wird. Bei der Abschätzung der Hazardfunktion und damit der Survivorfunktion müssen dann die zensierten Fälle in geeigneter Weise berücksichtigt werden. In empirischen Untersuchungen ist die Verteilungsfunktion  $F$  der Wartezeiten oft nicht bekannt. Man versucht also zunächst, eine empirische Beschreibung der Hazardfunktion zu finden. Nun ist  $\lambda(t)$  eine kontinuierliche Funktion, die empirische Bestimmung von Wartezeiten setzt aber voraus, dass die Beobachtungszeiten in Intervalle kategorisiert werden. Dementsprechend kann die Charakterisierung der Hazardfunktion nur hinsichtlich dieser Intervalle geschehen. Die folgenden Betrachtungen dienen der Vorbereitung dieser Charakterisierung.

### 1.3.2 Die Survivor-Funktion für diskrete Intervalle

**Satz 1.2** *Das Zeitintervall  $(0, t)$  werde in nicht notwendig gleichgroße Teilintervalle  $(t_i, t_{i+1})$  zerlegt, d.h.  $0 = t_0 < t_1 < \dots < t_n = t$ . Dann gilt*

$$S(t) = \prod_{i=0}^{n-1} \frac{S(t_{i+1})}{S(t_i)} = \prod_{i=0}^{n-1} P(\tau > t_{i+1} | \tau > t_i). \quad (1.54)$$

**Beweis:** Es gilt ja allgemein

$$F(t) = 1 - \exp\left(-\int_0^t \lambda(\tau) d\tau\right).$$

Das Zeitintervall  $(0, t)$  kann nun in nicht notwendig gleichgroße Teilintervalle  $[t_i, t_{i+1})$  aufgeteilt werden. Es ist

$$\int_0^t \lambda(\tau) d\tau = \sum_i \int_{t_i}^{t_{i+1}} \lambda(\tau) d\tau,$$

so dass für diese Aufteilung  $0 < t_1 < t_2 < \dots < t_n = t$

$$F(t) = 1 - \exp\left(-\sum_i \int_{t_i}^{t_{i+1}} \lambda(\tau) d\tau\right) \quad (1.55)$$

gilt. Für die Survivor-Funktion

$$S(t) = 1 - F(t) = \exp\left(-\int_0^t \lambda(\tau) d\tau\right)$$

folgt also

$$S(t) = \exp\left(-\sum_i \int_{t_i}^{t_{i+1}} \lambda(\tau) d\tau\right) = \prod_{i=1}^{n-1} \exp\left(-\int_{t_i}^{t_{i+1}} \lambda(\tau) d\tau\right) \quad (1.56)$$

Nun ist

$$S(t_i) = \exp\left(-\int_0^{t_i} \lambda(\tau) d\tau\right), \quad S(t_{i+1}) = \exp\left(-\int_0^{t_{i+1}} \lambda(\tau) d\tau\right),$$

so dass

$$\int_{t_i}^{t_{i+1}} \lambda(\tau) d\tau = \log S(t_{i+1}) - \log S(t_i),$$

und (1.56) impliziert dann

$$S(t) = \prod_i \frac{S(t_{i+1})}{S(t_i)} \quad (1.57)$$

Es ist nun

$$P(\tau > t_{i+1} | \tau > t_i) = \frac{P(\tau > t_{i+1} \cap \tau > t_i)}{P(\tau > t_i)} = \frac{P(\tau > t_{i+1})}{P(\tau > t_i)} = \frac{S(t_{i+1})}{S(t_i)},$$

und damit folgt (1.54).  $\square$

### 1.3.3 Die Sterbetafelmethode

Die Gleichung (1.54) erlaubt eine Abschätzung der Survivor-Funktion anhand von Schätzungen  $\hat{S}/t_i$  auf der diskretisierten Zeitachse. Damit kann dann auch die Hazardfunktion  $\lambda_i$  für das  $i$ -te Intervall geschätzt werden: es ist einerseits

$$\lambda_i = P(\tau \in [t_i, t_{i+1}) | \tau > t_i) \quad (1.58)$$

und andererseits

$$P(\tau > t_{i+1} | \tau > t_i) = 1 - \lambda_i. \quad (1.59)$$

Mithin impliziert (1.54)

$$S(t) = \prod_{i=0}^{n-1} (1 - \lambda_i) \quad (1.60)$$

Die Hazardfunktion  $\lambda(t)$  lässt sich abschätzen, wenn man (i) Schätzungen  $\hat{\lambda}_i$  bekommt und (ii) eine Schätzung für die Dichte  $f(t) = dF(t)/dt$  gewinnen kann. Dazu werden die folgenden Größen benötigt:

1.  $N$  sei die Gesamtzahl der Individuen, die in der Studie beobachtet werden,
2.  $d_i$  sei die Anzahl der Individuen, für die im  $i$ -ten Intervall das in Frage stehende Ereignis eintritt,
3.  $w_i$  sei die Anzahl der Zensierungen im  $i$ -ten Intervall,  $R_i$  sei die Anzahl der Individuen, für die zu Beginn des  $i$ -ten Intervalles das in Frage stehende Ereignis noch nicht eingetreten ist und die auch nicht zensiert sind;  $R_i$  heit die *Risikomenge*.

Man findet insbesondere

$$R_1 = N \quad (1.61)$$

$$R_i = R_{i-1} - d_{i-1} - w_{i-1}, \quad i = 2, \dots, n \quad (1.62)$$

$$\hat{\lambda}_i = d_i/R_i, \text{ falls keine Zensierungen vorliegen} \quad (1.63)$$

$$\hat{\lambda}_i = d_i/(R_i - w_i/2), \text{ falls } w_i > 0 \quad (1.64)$$

Aus der Definition der Dichtefunktion, d.h. aus

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}$$

ergibt sich die Schätzung

$$\hat{f}(t_i) = \frac{\hat{S}(t_i) - \hat{S}(t_{i+1})}{\Delta t_i}, \quad (1.65)$$

wobei  $\Delta t_i = t_{i+1} - t_i$  ist. Also hat man mit

$$\hat{\lambda}_i = \frac{\hat{f}(t_i)}{\hat{S}(t_i)} \quad (1.66)$$

eine Schätzung der Hazardfunktion für den Zeitpunkt  $t_i$ .

### 1.3.4 Der Kaplan-Meier-Schätzer der Survivor-Funktion

Bei der Sterbetafelmethode durften die Intervallgrößen  $\Delta_i$  verschieden sein. Kaplan und Meier (1958) leiteten eine Schätzung für die Survivor-Funktion  $S(t)$  her, bei der  $\Delta_i = \text{const}$  ist. Demnach gilt

$$\hat{S}(t) = \begin{cases} 1, & t \leq t_1 \\ \prod_{i: t_i < t} (1 - d_i/R_i), & t > t_1 \end{cases} \quad (1.67)$$

Da  $\hat{S}(t)$  eine Schätzung ist, hat sie eine Varianz, die durch

$$\text{Var}(\hat{S}(t)) = \hat{S}^2(t) \sum_{t_i \leq t} \frac{d_i}{R_i(R_i - d_i)}. \quad (1.68)$$

Darüber hinaus läßt sich eine mittlere Wartezeit (Verweildauer, Lebenszeit) bestimmen:

$$\hat{\mu} = \sum_{i=1}^n \hat{S}(t_{i-1})/(t_i - t_{i-1}) \quad (1.69)$$

### 1.3.5 Der log-Rang-Test

#### Der log-Rang-Test für zwei Gruppen

Gegeben seien die Kaplan-Meier-(KM)-Kurven für zwei verschiedene Gruppen, etwa eine, die mit einer Therapie behandelt wurden, während die andere mit einem Placebo traktiert wurde. Die Frage ist, ob sich die beiden Kurven überzufällig voneinander unterscheiden oder ob sie statistisch äquivalent sind.

Es sei  $R_{kj}$  die Risikomenge in der  $k$ -ten Gruppe ( $k = 1, 2$ ) zum  $j$ -ten Zeitpunkt, und  $d_{kj}$  sei die Anzahl der Fälle (Personen), bei denen das Ereignis im  $j$ -ten Intervall eintritt. Dann können die für das  $j$ -te Intervall *erwarteten Häufigkeiten*

$$e_{kj} = \frac{R_{kj}}{R_{1j} + R_{2j}}(d_{1j} + d_{2j}) \quad (1.70)$$

berechnet werden. Weiter ist

$$s^2 \stackrel{\text{def}}{=} \sum_j \frac{d_{1j}d_{2j}(R_{1j} - R_{2j})(d_{1j} + d_{2j} - d_{1j} - d_{2j})}{(R_{1j} + R_{2j})^2(d_{1j} + d_{2j} - 1)} \quad (1.71)$$

Dann ist

$$LR = \frac{\sum_{j=1}^n (d_{kj} - e_{kj})}{s^2} \quad (1.72)$$

$LR$  ist unter der Nullhypothese, dass kein Unterschied zwischen den Gruppen besteht,  $\chi^2$ -verteilt mit  $df = 1$ .

Eine Approximation für den  $LR$ -Test ergibt sich durch

$$X^2 = \sum_{k,j} \frac{(d_{kj} - e_{kj})^2}{e_{kj}} \quad (1.73)$$

#### Der log-Rang-Test für mehrere Gruppen

Eine Approximation für diesen Fall ergibt sich gemäß (1.73), aber  $df = K - 1$ ,  $K$  die Anzahl der Gruppen.

## 1.4 Der Effekt exogener Variablen

Man möchte die Wirkung exogener (unabhängiger) Variablen auf die Verteilung der Wartezeiten erfassen. Ein einfaches Beispiel ist die Untersuchung der Effektivität von Therapien: ist die Therapie  $T_1$  besser als die Therapie  $T_2$ , so sollte die "Überlebenszeit" bei der Therapie  $T_1$  größer sein. Hat man bestimmte Prädiktorvariablen  $X_1, \dots, X_k$  so ist man daran interessiert, in welcher Weise die einzelnen Variablen auf die Überlebenszeit einwirken. Da die Verteilung der Wartezeit durch die Hazardfunktion definiert wird, muß man untersuchen, wie die Prädiktorvariablen auf die Hazardfunktion einwirken können. Es werden einige Standardansätze vorgestellt.

### 1.4.1 Das Exponentialmodell

Es wird davon ausgegangen, dass die Verteilungsfunktion  $F$  durch die Exponentialfunktion definiert ist, also

$$F(t) = 1 - e^{-\lambda t}, \quad f(t) = \lambda e^{-\lambda t}$$

gilt. Je größer der Wert von  $\lambda$  desto kürzer sind die Wartezeiten, und umgekehrt ( $E(t) = 1/\lambda$ ,  $Var(t) = 1/\lambda^2$ ). Soll die Überlebenszeit erhöht werden, so muß  $\lambda$  verringert werden. Man kann annehmen, dass

$$\lambda = \lambda(X_1, \dots, X_k)$$

ist. Man kann auch annehmen, dass die  $X_j$ ,  $j = 1, \dots, k$  auf den Erwartungswert wirken. Für die Exponentialverteilung bedeutet dies, dass

$$E(t) = \frac{1}{\lambda} = g(X_1, \dots, X_k) \quad (1.74)$$

ist. Dabei ist  $g$  eine noch näher zu spezifizierende Funktion.

Ein erster Ansatz besteht darin, anzunehmen, dass die  $X_j$  wie bei einer multiplen Regression entweder auf  $\lambda$  oder auf  $E(t)$  wirken. Dann existiert für jede der Prädiktorvariablen  $X_j$  ein Regressionsgewicht  $b_j$ , und

$$g(X_1, \dots, X_k) = g(b_0 + b_1 X_1 + \dots + b_k X_k).$$

Schreibt man die  $X_j$  als Vektor an und ebenso die Regressionsgewichte  $b_s$ ,  $s = 0, 1, \dots, k$ , d.h.

$$X = (1, X_1, X_2, \dots, X_k)', \quad b = (b_0, b_1, \dots, b_k)'$$

so ist

$$b_0 + b_1 X_1 + \dots + b_k X_k = \vec{x}' \vec{b}$$

und man kann (1.74) in der kompakten Form

$$E(t) = g(\vec{x}' \vec{b}) \quad (1.75)$$

schreiben.

Es werde nun

$$g(\vec{x}' \vec{b}) = \vec{x}' \vec{b}$$

gesetzt. Da wegen  $E(t) = 1/\lambda > 0$  sein muß, kann man die Koeffizienten  $b$  nur unter Restriktionen schätzen; der gleiche Sachverhalt gilt, wenn der Ansatz  $\lambda = g(\vec{x}' \vec{b})$  gemacht wird. Also muß für  $g$  eine Funktion gewählt werden, die unabhängig vom Vorzeichen von  $\vec{x}' \vec{b}$  nur positive Werte annehmen kann. Eine mögliche Wahl für  $g$  ist nun,

$$g(\vec{x}' \vec{b}) \stackrel{\text{def}}{=} \exp(\vec{x}' \vec{b}) \quad (1.76)$$

zu setzen. Setzt man  $X = (X_1, \dots, X_k)'$  und  $b = (b_1, \dots, b_k)'$ , so kann man

$$g(\vec{x}' \vec{b}) = \exp(b_0 + \vec{x}' \vec{b}) = \beta_0 \exp(\vec{x}' \vec{b}), \quad \beta_0 = e^{b_0} \quad (1.77)$$

schreiben. Für die Hazardrate erhält man dann

$$\lambda = 1/g(\vec{x}' \vec{b}) = \exp(-(b_0 + \vec{x}' \vec{b})) \quad (1.78)$$

### 1.4.2 Das Weibull-Modell

Die Weibull-Verteilung war durch

$$F(t) = 1 - \exp(-\alpha t^\beta), \quad f(t) = (1 - F(t)) \alpha \beta t^{\beta-1},$$



gegeben, und die Hazard-Funktion hat die Form

$$\phi(t) = \alpha\beta t^{\beta-1};$$

$\phi$  ist also eine Funktion der Zeit.

Die Frage ist, in welcher Weise die unabhängigen Variablen auf  $\phi$  einwirken. Sowohl  $\alpha$  wie auch  $\beta$  wirken auf den Erwartungswert wie auf die Varianz ein. Trägt man allerdings  $F(t)$  nicht gegen  $t$ , sondern gegen  $\log t$  auf, so sieht man, dass  $\log \alpha$  die Position auf der  $\log t$ -Achse bestimmt; verändert man  $\alpha$ , so wird die Funktion auf der  $\log t$ -Achse parallel verschoben. Insofern ist  $\alpha$  ein Positionsparameter, während  $\beta$  die Steilheit der Funktion und damit die Variabilität der  $\log t$ -Werte bestimmt. Nimmt man dementsprechend an, dass der Vektor  $\vec{x}$  der unabhängigen Variablen die Position der Verteilungsfunktion und damit auch der Dichte bestimmt (und damit die Wahrscheinlichkeiten der Zeiten), so kann man annehmen, dass  $\vec{x}$  auf  $\alpha$  wirkt: nach (1.47) gilt

$$E(\tau) = \frac{1}{\alpha^{1/\beta}} \Gamma\left(\frac{1+\beta}{\beta}\right).$$

Je größer  $\alpha$  desto kleiner der Erwartungswert, und umgekehrt. Man kann also den Ansatz

$$\frac{1}{\alpha} = \exp(b_0 + b_1 x_1 + \dots + b_n x_n) \quad (1.79)$$

machen. Je nach den Werten der Parameter  $b_0, b_1, \dots, b_n$  wird jetzt der Erwartungswert der Wartezeit größer oder kleiner; allerdings verändert sich auch die Varianz, die ja nach (1.47) durch

$$\text{Var}(\tau) = \frac{1}{\alpha^{2/\beta}} \left[ \Gamma\left(\frac{\beta+2}{\beta}\right) - \Gamma^2\left(\frac{\beta+1}{\beta}\right) \right]$$

gegeben ist und über  $\alpha$  ebenfalls von  $b_0 + b_1 x_1 + \dots + b_n x_n$  abhängt. Für  $\phi$  erhält man insbesondere

$$\phi(t|\vec{x}) = \alpha\beta t^{\beta-1} = \exp(-(b_0 + b_1 x_1 + \dots + b_n x_n))\beta t^{\beta-1}. \quad (1.80)$$

Man kann also die empirische Schätzung von  $\phi$  mit der durch (1.80) "vorhergesagten", von  $\vec{x}$  abhängigen Form vergleichen.

$\phi(t|\vec{x})$  hat eine bestimmte Form: die Hazard-Funktion besteht aus dem Produkt zweier Faktoren, von denen der eine nur von  $\vec{x}$  abhängt, und der andere von  $t$ . Man kann diese Beobachtung verallgemeinern und alle Modelle betrachten, die diese Eigenschaft haben. Man wird auf diese Weise zu dem allgemeinen Modell von Cox geführt.

### 1.4.3 Das Proportional-Hazard-Modell von Cox

Gesucht ist ein allgemeines Modell, das es erlaubt, die Wirkung von bestimmten unabhängigen Variablen auf die Wartezeiten (oder Verweildauern) zu überprüfen. Ein solches Modell besteht in Annahmen über die Hazardfunktion. Die folgende Definition einer Hazardfunktion liefert ein mögliches Modell:

**Definition 1.3** *Läßt sich die Hazardfunktion in der Form*

$$\phi(t|x) = \lambda_0(t)g(\vec{x}; \vec{b}) \quad (1.81)$$

*schreiben, so repräsentiert sie ein proportionales Hazardmodell;  $\phi$  heißt proportionale Hazardfunktion.  $g$  ist eine Link-Funktion.*

Der Ausdruck proportionales Hazardmodell ergibt sich aus der angenommenen Form der Hazardfunktion:  $\phi(t|\vec{x})$  ist proportional zu  $\lambda_0(t)$ , mit  $g(\vec{x}; \vec{b})$  als Proportionalitätskonstante, die allerdings von den Vektoren  $\vec{x} = (x_1, \dots, x_n)'$  und  $\vec{b} = (b_0, b_1, \dots, b_n)'$  abhängt. Das Wesentliche an der Definition 1.3 ist die Annahme, dass sich die Hazardfunktion als Produkt zweier Funktionen anschreiben lässt, wovon eine nur von  $t$ , und die andere nur von  $\vec{x}$  abhängt.

Es gelte insbesondere  $g(\vec{x}) = \exp(\vec{x}'\vec{b})$ ; dann gilt

$$\phi(t|\vec{x}) = \lambda_0(t) \exp(\vec{x}'\vec{b}). \quad (1.82)$$

Diese Form von  $g$  erlaubt es, dass  $\vec{x}'\vec{b}$  auch negativ sein kann;  $g(x) = \exp(\vec{x}'\vec{b})$  ist stets größer, mindestens gleich Null, so dass die Forderung  $\lambda(t|x) \geq 0$  stets erfüllt ist, wenn nur  $\lambda_0(t) \geq 0$  für alle  $t$  ist. (1.82) definiert das Cox'sche proportionale Hazardmodell, wobei  $\lambda_0(t)$  irgendeine beliebige, den Bedingungen einer Hazardfunktion entsprechende Funktion ist. Gelten für zwei Personen (allgemein: Objekte) die Hazardfunktionen  $\lambda(t|\vec{x}_1)$  und  $\lambda(t|\vec{x}_2)$ , d.h. unterscheiden sich die Funktionen nur hinsichtlich der Prädiktorvariablen, so folgt

$$\frac{\lambda(t|\vec{x}_1)}{\lambda(t|\vec{x}_2)} = \exp((\vec{x}_1 - \vec{x}_2)'b), \quad (1.83)$$

d.h. der Quotient der Hazardfunktionen hängt nur von der Differenz  $(\vec{x}_1 - \vec{x}_2)$ , nicht aber von  $t$  ab.

Die Survivorfunktion ist durch

$$\begin{aligned} S(t|\vec{x}) &= \exp\left(-\int_0^t \lambda(\tau|\vec{x}) d\tau\right) \\ &= \exp\left(\int_0^t \lambda_0(\tau) g(\vec{x}; \vec{b}) d\tau\right) \\ &= \exp\left(g(\vec{x}; \vec{b}) \int_0^t \lambda_0(\tau) d\tau\right). \end{aligned} \quad (1.84)$$

gegeben. Setzt man hier  $S_0(t) = \exp\left(\int_0^t \lambda_0(\tau) d\tau\right)$  und  $g(x) = \vec{x}'\vec{b}$ , so ergibt sich

$$S(t|\vec{x}) = S_0(t)^{\exp(\vec{x}'\vec{b})}. \quad (1.85)$$

wobei  $S_0(t)$  die durch  $\lambda_0(\tau)$  definierte, von  $\vec{x}$  unabhängige Survivorfunktion ist.

Die Interpretation der Regressionsgewichte  $\vec{b} = (b_0, b_1, \dots, b_n)'$  ergibt sich aus (1.82). Denn es ist ja

$$\vec{x}'\vec{b} = b_0 + b_1 x_1 + \dots + b_n x_n, \quad (1.86)$$

so dass

$$\lambda(t|\vec{x}) = e^{b_0} e^{b_1 x_1} e^{b_2 x_2} \dots e^{b_n x_n}. \quad (1.87)$$

Es werde nun  $x_i$ ,  $1 \leq i \leq n$  um eine Einheit  $\Delta x_i$  verändert, so dass  $x_i + \Delta x_i$  in die Gleichung für  $\lambda(t|\vec{x})$  eingeht. Man kann dann den Quotienten

$$\delta_{\Delta x_i} = \frac{e^{b_0} e^{b_1 x_1} e^{b_2 x_2} \dots e^{b_i(x_i + \Delta x_i)} \dots e^{b_n x_n}}{e^{b_0} e^{b_1 x_1} e^{b_2 x_2} \dots e^{b_n x_n}} = e^{b_i \Delta x_i} \quad (1.88)$$

betrachten.  $\delta_{\Delta x_i}$  ist der Quotient zweier Hazard-Funktionen: über dem Bruchstrich steht  $\lambda(t|\vec{x})$ , wenn statt  $x_i$  der Wert  $x_i + \Delta x_i$  eingeht, und unter dem Bruchstrich steht  $\lambda(t|\vec{x})$

wie in (1.87) definiert. Der Quotient ist offenbar unabhängig von  $t$ , d.h. er gilt für alle  $t$ , vorausgesetzt, das proportionale Hazard-Modell gilt. Die rechte Seite von (1.88) ist  $e^{b_i \Delta x_i}$ ; dieser Wert gibt an, wie sich die Hazard-Funktion verändert, wenn man  $x_i$  um die Einheit  $\Delta x_i$  verändert. Multipliziert mit 100 ergibt  $\delta_{\Delta x_i}$ , also  $\delta_{\Delta x_i} 100$ , die prozentuale Veränderung der Hazard-Funktion hat. Die Größe dieser Veränderung wird durch den Wert von  $b_i$  bestimmt; ist  $b_i = 0$ , so ergibt sich keine Veränderung in Abhängigkeit von  $x_i$  bzw.  $\Delta x_i$ . So sei  $x_1 = 1$ , wenn eine Person männlich ist, und  $x_1 = 0$ , wenn sie weiblich ist, und  $x_i$  sei das Alter. Betrachtet werde die Wartezeit, bis eine Person zum ersten Male arbeitslos wird, oder eine bestimmte Krankheit bekommt.  $b_1$  liefert dann, über den Faktor  $e^{b_1}$ , die prozentuale Veränderung der Hazard-Rate für das Geschlecht, und  $b_i$  liefert über den Faktor  $e^{b_i \Delta x_i}$  die Veränderung der Hazard-Rate mit dem Alter. Findet man z.B.  $e^{b_i \Delta x_i} \times 100 = 1.5\%$ , so erhöht sich die Wahrscheinlichkeit, die in Frage stehende Krankheit zu bekommen (oder arbeitslos zu werden), pro Jahr um 1.5 %, - vorausgesetzt, man hat sie nicht schon (oder ist schon arbeitslos).

#### 1.4.4 Competing-Risk-Modelle

Bis jetzt ist stillschweigend davon ausgegangen worden, dass es nur ein Ereignis gibt, dass das Ende der Wartezeit definiert. Es ist aber durchaus möglich, dass das Ende der Wartezeit durch verschiedene Ereignisse definiert wird. So kann der Tod nicht nur durch eine bestimmte Krankheit, die gerade therapiert worden ist, herbeigeführt werden, sondern er kann andere Ursachen haben, z.B. durch einen Unfall, oder durch Mord, oder durch eine andere, unvorhergesehene Krankheit (Ebola-Virus). Analysiert man die Dauer der Arbeitslosigkeit, so können für die Beendigung der der Arbeitslosigkeit mehrere Gründe in Frage kommen:

- Arbeit ist durch Vermittlung zustande gekommen,
- Arbeit ist selbst gefunden worden,
- Erkrankung, die zu Arbeitsunfähigkeit führt,
- vorzeitiger Übergang in die Rente.

Die verschiedenen Ereignisse konkurrieren deshalb gewissermaßen miteinander, um die Wartezeit zu beenden, weshalb auch der Name *Competing Risk* für diese Situation gewählt wird.

Es sei  $\{z_1, \dots, z_m\}$  die Menge der möglichen Ereignisse, die die Wartezeit beenden. Weiter sei  $Y$  eine zufällige Veränderliche mit  $Y = z_j$ , wenn das Ereignis  $z_j$  eingetreten ist, und  $\vec{x}$  sei ein Vektor mit Prädiktorvariablen; jede Person ist durch einen solchen Vektor definiert. Dann wird die zu  $z_j$  korrespondierende Hazarrate

$$\phi_j(t|\vec{x}) = \lim_{\Delta t \rightarrow 0} \frac{P(\tau \in [t, t + \Delta t), Y = z_j | \tau > t, \vec{x})}{\Delta t} \quad (1.89)$$

eingeführt. Die Gesamthazardrate ergibt sich dann gemäß

$$\phi(t|\vec{x}) = \sum_j \phi_j(t|\vec{x}) \quad (1.90)$$

und für die Survivorfunktion findet man

$$S(t|\vec{x}) = \exp\left(-\int_0^t \phi(\tau|\vec{x}) d\tau\right). \quad (1.91)$$

### 1.4.5 Der Mehr-Episoden-Fall

Die Ereignisse, die die Wartezeit beenden, sind implizit als *absorbierend* angenommen worden, d.h. der beobachtete Prozess startet nach Eintreten eines (oder des) Ereignisses nicht wieder neu. In vielen Fällen ist oder sind die Ereignisse aber nicht absorbierend:

1. ein therapierter Drogenabhängiger kann rückfällig werden,
2. ein Arbeitsloser, der einen Job gefunden hat, kann wieder arbeitslos werden,
3. nach einer Scheidung kann ein Mensch noch einmal heiraten, etc.

Es seien  $t_0 < t_1 < \dots < t_k < \dots$  die Wartezeiten zu den Ereignissen, die eine Episode beenden. Die Länge der  $k$ -ten Episode ist durch

$$v_k = t_k - t_{k-1} \quad (1.92)$$

gegeben. Weiter werde für jede Person für die  $k$ -te Episode ein Vektor  $\vec{x}_k$  von Prädiktorvariablen gemessen, die im Prinzip auch selbst zeitabhängig sein dürfen.

Die Hazardfunktion ist über die bedingte Wahrscheinlichkeit

$$P(t_i \leq \tau \leq t_{i+1} | \tau > t_i) = \frac{P(t_i \leq \tau \leq t_{i+1})}{S(t_i)}$$

eingeführt worden. Es ist aber

$$P(t_i \leq \tau \leq t_{i+1}) = S(t_i) - S(t_{i+1}),$$

und mithin ist

$$P(t_i \leq \tau \leq t_{i+1} | \tau > t_i) = \frac{S(t_i) - S(t_{i+1})}{S(t_i)}$$

Für die erste Episode hat man die Hazardfunktion

$$\phi_1(t | \vec{x}_1) = \lim_{\Delta t \rightarrow 0} \frac{P(\tau_1 \in [t, t + \Delta t] | \tau_1 > t, x_1)}{\Delta t} \quad (1.93)$$

mit der zugehörigen Survivor-Funktion

$$S_1(t | \vec{x}_1) = \exp \left( - \int_0^t \phi_1(\tau | \vec{x}_1) d\tau \right). \quad (1.94)$$

Zum Zeitpunkt  $t_2$  beginnt die zweite Episode. Für diese Episode ist die Hazardfunktion durch

$$\phi_2(t | \vec{x}_2), H_1) = \lim_{\Delta t \rightarrow 0} \frac{P(\tau_1 \in [t_1, t_1 + \Delta t] | \tau_2 > t_1, \vec{x}_2)}{\Delta t}, \quad (1.95)$$

gegeben, wobei  $H_1 = (t, \vec{x}_1)$  die Vorgeschichte des Prozesses bis  $t_1$  repräsentiert; die Survivor-Funktion wird dann analog zu (1.94) definiert. Allgemein erhält man für die  $k$ -te Episode die Hazardfunktion

$$\phi_k(t | \vec{x}_k) = \lim_{\Delta t \rightarrow 0} \frac{P(\tau_k \in [t_{k-1}, t_{k-1} + \Delta t] | \tau_k > t_{k-1}, \vec{x}_k, H_k)}{\Delta t}, \quad (1.96)$$

mit  $H_k = (t_1, \vec{x}_1, \dots, t_{k-1}, \vec{x}_{k-1})$ .

Die Regressionsansätze können alle auf den Mehr-Episoden-Fall übertragen werden. Für den Fall, dass die Hazardfunktion durch die Weibull-Verteilung definiert ist, erhält man für  $\phi_k$

$$\phi(t|\vec{x}_k) = \delta_k \phi_{0k} (\phi_{0k} t)^{\delta_k - 1} \exp(\vec{x}'_k \vec{\beta}_k). \quad (1.97)$$

Geht man vom allgemeinen Proportional-Hazard-Modell aus, so hat man

$$\phi_k(t|\vec{x}_k) = \phi_{0k}(t) \exp(\vec{x}'_k \vec{\beta}_k), \quad (1.98)$$

wobei  $\phi_{0k}(t)$  eine nicht näher spezifizierte Baseline-Hazardfunktion ist.

## 1.5 Markov-Ketten

### 1.5.1 Grundlegende Definitionen

Viele in der Psychologie betrachteten Prozesse können als Folge von Ereignissen

$$S_0, S_1, \dots, S_n, \dots$$

angesehen werden. So kann es von Interesse sein, in einer Fabrik die Folge der Unfälle zu analysieren, oder in einer psychiatrischen Klinik die Folge der als "schizophren" eingestuft Patienten, oder die Abfolge bestimmter Klassen von Interaktionen in einer sozialen Gruppe, die Folge der Gewalttaten gegen Ausländer, etc. Der Übergang von einem Ereignis zu einem anderen ist dabei zufällig in dem Sinne, als nicht mit Sicherheit gesagt werden kann, welches Ereignis auf ein bereits eingetretenes folgt. Dementsprechend sind die Übergangswahrscheinlichkeiten  $p(S_k|S_{k-1}, S_{k-2}, \dots, S_0)$  von Interesse; sie reflektieren in einem noch zu spezifizierenden Sinne die Struktur des betrachteten Prozesses.

**Definition 1.4** Zu einem Zeitpunkt  $t_k$  sei der Prozess im Zustand  $S_j$ , und im Zeitpunkt  $t_{k-1}$  sei der Prozess im Zustand  $S_i$ . Weiter gelte

$$P(S_k|S_{k-1}, S_{k-2}, \dots, S_0) = P(S_j|S_i), \quad (1.99)$$

so dass der Übergang nach  $S_j$  nur vom Zustand zur Zeit  $t_{k-1}$  abhängt. Dann heißt die Folge der Zustände  $S_0, S_1, \dots, S_k, \dots$  Markovkette.

Prozesse mit der Eigenschaft (1.99) nehmen deswegen eine zentrale Rolle bei dem Anwendungen der Theorie der stochastischen Prozesse ein, weil die Analyse solcher Prozesse wesentlich einfacher ist als die von Prozessen, die (1.99) nicht erfüllen. In vielen Fällen ist (1.99) auch nicht nur eine Vereinfachung, sondern durchaus plausibel.

Es sei  $p_{ij} = p(S_i|S_j)$ . Die  $p_{ij}$  können zu einer Matrix zusammengefaßt werden:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & \dots & \vdots \\ p_{j1} & p_{j2} & \dots & p_{jn} \\ \vdots & \vdots & \dots & \vdots \end{pmatrix} \quad (1.100)$$

$P$  heißt auch *Matrix der Übergangswahrscheinlichkeiten*. Die Übergangswahrscheinlichkeiten  $p_{ij}$  können dabei eine Funktion der Zeit sein, so dass  $p_{ij} = p_{ij}(t)$ . Gilt  $p_{ij} = \text{constant}$  für alle  $i, j$ , so heißt die Markov-Kette *homogen*. Die Anzahl der Zustände kann, muß aber nicht endlich sein.

**Beispiel 1.4** (Random Walk) Gegeben sei eine Folge von Zuständen

$$\{\dots - S_3, -S_2, -S_1, S_0, S_1, S_2, S_3, \dots\}.$$

Es gelte  $P(S_{i-1}|S_i) = 1 - p$ ,  $p(S_{i+1}|S_i) = p$ , und  $p$  sei unabhängig von  $i$ , mithin ist die Folge eine Markov-Kette. Offenbar ist  $p(S_j|S_i) = 0$  für  $j \neq i-1$  oder  $j \neq i+1$ .  $S_i$  kann mit dem Wert  $x_i$  einer zufälligen Veränderlichen, die ein Merkmal repräsentiert, gleichgesetzt werden. Demnach heißt die Folge *Random Walk*, da sie unter anderem den zufälligen Weg eines Teilchens beschreibt, das zu bestimmten Zeitpunkten mit der Wahrscheinlichkeit  $p$  nach rechts und mit der Wahrscheinlichkeit  $1 - p$  nach links gestoßen wird. Wird nicht festgelegt, wann der Random Walk endet, so gibt es unendlich viele Zustände. Für eine endliche Anzahl von möglichen Zuständen ergibt sich insbesondere

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \dots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{pmatrix} \quad (1.101)$$

□

**Beispiel 1.5** (Entscheidungsverhalten: das VTE-Modell (von Vicarious Trial and Error)). Eine Person muß sich zwischen zwei Alternativen  $A_1$  und  $A_2$  entscheiden. Um in der gegebenen Situation zu einer Entscheidung zu gelangen, muß sie verschiedene Aspekte der Situation abwägen. Die Betrachtung eines bestimmten Aspekts erlaubt ihr den Übergang zu  $A_1$ , falls der Aspekt zu einer bestimmten Menge  $M_1$  gehört, gleichwohl ist der Übergang dann nicht sicher, falls der Aspekt nicht genügend Information über die Alternative liefert, sucht sie einen neuen Aspekt. Wenn sie einen neuen Aspekt sucht, sei sie im Zustand  $S_0$ . Hat sie einen Aspekt der Menge  $M_1$  gewählt, so ist sie in Zustand  $S_1$ . Komplementär zu  $M_1$  gibt es eine Menge  $M_2$ ; hat sie einen Aspekt aus der Menge  $M_2$  gewählt, so ist die Person im Zustand  $S_2$ . Für einen geeigneten Aspekt aus  $M_2$  folgt eine Entscheidung für  $A_2$ , sonst kehrt sie zu  $S_0$  zurück. Die Übergangswahrscheinlichkeiten seien wie folgt festgelegt:

	$S_0$	$S_1$	$A_1$	$A_2$
$S_0$	$p_{01}$	$p_{02}$	0	0
$S_1$	$p_{10}$	0	0	$p_{12}$
$A_1$	0	0	1	0
$A_2$	0	0	0	1

d.h.

$$P = \begin{pmatrix} p_{01} & p_{02} & 0 & 0 \\ p_{10} & 0 & 0 & p_{12} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Es muß natürlich  $p_{01} + p_{02} = 1$ ,  $p_{10} + p_{12} = 1$  gelten. Da  $p(A_1|A_1) = p(A_2|A_2) = 1$ , bleibt die Person bei ihrer Entscheidung, hat sie sie erst einmal getroffen. Die Zustände  $A_1$  und  $A_2$  heißen demnach *absorbierend*. Die Übergangswahrscheinlichkeiten  $p_{ij}$  hängen nicht von der Geschichte des Prozesses ab, also ist der Prozess eine Markov-Kette. Die spezielle Definition der Zustände und der Übergangswahrscheinlichkeiten definieren die Struktur des Prozesses. Dieses Entscheidungsmodell ist das VTE-Modell von Bowers (1959, 1962). □

Die Matrix  $P$  der Übergangswahrscheinlichkeiten bestimmt allerdings den Prozess noch nicht vollständig. Um den Prozess vollständig zu charakterisieren müssen noch die Wahrscheinlichkeiten, mit denen der Prozess in einem der Zustände  $S_1, S_2, \dots$  ist, angegeben werden. Diese werden im *Startvektor*  $p_0 = (p_0(S_1), p_0(S_2), \dots, p_0(S_n))'$  angegeben. Der Einfachheit halber werde  $p_j^0$  für  $p_0(S_j)$  geschrieben, so dass  $p_0 = (p_1^0, p_2^0, \dots, p_n^0)'$ . Sind nun  $P$  und der Startvektor  $p_0$  gegeben, können die Wahrscheinlichkeiten dafür, dass der Prozess nach  $k$  Schritten im Zustand  $S_i, i = 1, 2, \dots$  ist, ausgerechnet werden. Für endliches  $n$  gilt dann etwa für den Zeitpunkt  $t_1$ , wenn der Prozess in  $t_0$  startet, nach dem Satz der totalen Wahrscheinlichkeit:

$$\begin{aligned} p_1(S_i) &= p(S_i|S_1)p_1^0 + \dots + p(S_i|S_n)p_n^0 = \\ &= p_{i1}p_1^0 + \dots + p_{in}p_n^0. \end{aligned} \quad (1.102)$$

Man sieht, dass nach den Regeln der Matrixmultiplikation der Vektor  $p_1 = (p_1^1, p_2^1, \dots, p_n^1)'$  kompakt durch

$$p_1' = p_0' P \quad (1.103)$$

angegeben werden kann. Die Wahrscheinlichkeit, dass der Prozess im zweiten Schritt im Zustand  $S_i$  ist, ergibt sich analog zu (1.102) und (1.103) als

$$p_2' = p_1' P. \quad (1.104)$$

Für  $p_1$  kann man den in (1.103) angegebenen Ausdruck einsetzen und erhält sofort  $p_2 = p_0 P \cdot P = p_0 P^2$ . So kann man fortfahren und erhält dann für den Vektor der Zustandswahrscheinlichkeiten nach dem  $k$ -ten Schritt

$$p_k = p_0' P^k. \quad (1.105)$$

Die Zustände einer Markov-Kette können wie folgt klassifiziert werden.

**Definition 1.5** *Ein Zustand heißt transient, wenn der Prozess diesen Zustand wieder verlassen, aber dann nie wieder erreichen kann. Ein Zustand heißt ergodisch, wenn er nicht transient ist. Ist ein Zustand einziger Zustand einer Menge ergodischer Zustände, so heißt er absorbierend*

Es sei  $C_k$  eine Menge von ergodischen Zuständen derart, dass ein Übergang von einem Zustand dieser Menge in einen anderen Zustand dieser Menge möglich ist. Dann kann der Prozess die Menge dieser Zustände nie verlassen. Daraus kann man folgern, dass die Menge der Zustände einer Markov-Kette zerlegt werden kann in eine Menge  $T$  von transienten Zuständen und in Teilmengen  $C_1, C_2, \dots$  jeweils ergodischer Zustände.

**Definition 1.6** *Eine Markov-Kette, deren Zustandsmenge nicht weiter zerlegt werden kann, heißt irreduzibel.*

Eine Markov-Kette, deren Zustandsmenge aus verschiedenen ergodischen Mengen  $C_1, C_2, \dots$  besteht, besteht also eigentlich aus verschiedenen Markov-Ketten, zwischen denen kein Übergang möglich ist; die Kette geht immer in den gleichen Zustand über, hat sie ihn erstmalig erreicht. Ist  $S_j$  ein solcher Zustand, so gilt also  $p_{jj} = 1$ . Dann folgt, dass eine endliche Markov-Kette nie nur transiente Zustände haben kann, denn nach Durchlaufen aller dieser Zustände kann sie in keinen Zustand mehr zurück; dieser Zustand wäre dann aber ein absorbierender und mithin ein ergodischer Zustand.

Es sei eine ergodische Markov-Kette mit der Matrix  $P$  von Übergangswahrscheinlichkeiten gegeben. Es sei  $p_n$  der Vektor der Zustandswahrscheinlichkeiten nach dem  $n$ -ten Schritt. Es ist  $p_n' = p_0' P^n$ ,  $p_{n-1}' = p_{n-2}' P$ , d.h.  $p_n' = p_{n-1}' P$  etc., so dass schließlich  $p_n' = p_0' P^n$ ,  $p_0$  der Startvektor.

**Definition 1.7** Gilt  $P^n > 0$  für irgendein  $n \in \mathbb{N}$ , wie auch immer der Prozess startet, so heißt die Markov-Kette regulär. Ein ergodischer Prozess heißt zyklisch, wenn die Zustände in definiter Ordnung durchlaufen werden.

Es sei  $p_0$  wieder ein Startvektor für eine reguläre Markov-Kette, und  $p'_n = p'_0 P^n$ . Es ergibt sich die Frage, ob  $p_n$  für  $n \rightarrow \infty$  gegen einen bestimmten Vektor  $\pi$  konvergiert, für den dann  $\pi' = \pi' P$  gilt (denn dann ist ja  $p_{n-1}$  ebenfalls gleich  $\pi$ ). Die Bedingung  $\pi' P = \pi' P$  besagt aber, dass  $\pi$  ein Eigenvektor von  $P'$  ist mit dem zugehörigen Eigenwert  $\lambda = 1$ , denn es ist ja  $(\pi' P)' = P' \pi$ , d.h. es soll  $P' \pi = \pi$  gelten. Eine notwendige Bedingung für die Existenz von  $\pi$  ist also die Existenz eines Eigenwerts  $\lambda = 1$ .

**Beispiel 1.6** In einem einfachen Lernexperiment soll die Versuchsperson (Vp) lernen, bestimmte Stimuli entweder in die Kategorie A oder in die Kategorie  $\bar{A}$  zu klassifizieren; die Aufgabe der Vp besteht also darin, die Regel herauszufinden, nach der die Stimuli klassifiziert werden sollen. Sie bekommt einen Stimulus dargeboten, dann klassifiziert sie, und danach wird ihr vom Versuchsleiter (Vl) gesagt, welche Kategorie die korrekte war (unabhängig davon, ob die Antwort richtig oder falsch war). Es wird angenommen, dass der Lernprozeß als Markov-Kette dargestellt werden kann; die Matrix der Übergangswahrscheinlichkeiten und der Startvektor seien durch

$$P = \begin{bmatrix} 1 & 0 \\ c & 1-c \end{bmatrix}, \quad p_0 = \begin{bmatrix} q \\ 1-q \end{bmatrix}$$

gegeben. Es werden also zwei mögliche Zustände angenommen: hat die Vp die Regel gefunden, so ist sie im Zustand  $S_1$ , hat sie sie noch nicht gefunden, so ist sie im Zustand  $S_2$ . Am Anfang bildet sie eine mögliche Regel, die mit der Wahrscheinlichkeit  $q$  korrekt ist. Ist sie in  $S_1$  angelangt, so kann sie den Zustand nicht wieder verlassen, d.h. sie kann, zumindest im Rahmen des Lernexperiments, die Regel nicht wieder vergessen. Der Zustand  $S_1$  ist demnach ein Beispiel für einen *absorbierenden Zustand*. Mit der Wahrscheinlichkeit  $c$  geht sie von  $S_2$  nach  $S_1$  über, und es wird angenommen, dass  $c$  konstant über die Zeit ist. Man muß sich klarmachen, dass die Vp eine falsche Hypothese über die Regel haben kann und trotzdem eine korrekte Antwort geben kann. In diesem Fall wird sie ihre Hypothese nicht aufgeben, sie bleibt also in  $S_2$ . Daraus folgt, dass sie nur nach einer *falschen* Antwort in den Zustand  $S_1$  übergehen kann. Nach (1.105) kann man nun ausrechnen, mit welcher Wahrscheinlichkeit die Vp nach  $k$  Lerndurchgängen (Darbietung eines Stimulus, Antwort und "Reinforcement" des Vl) im Zustand  $S_i$ ,  $i = 1, 2$  ist. Nach dem ersten Durchgang gilt

$$p'_1 = (q, 1-q) \begin{bmatrix} 1 & 0 \\ c & 1-c \end{bmatrix} = (q + (1-q)c, (1-q)(1-c)).$$

Für den zweiten Durchgang erhält man den Vektor der Zustandswahrscheinlichkeiten

$$\begin{aligned} p'_2 &= (q, 1-q) \begin{bmatrix} 1 & 0 \\ c & 1-c \end{bmatrix}^2 \\ &= (q, 1-q) \begin{bmatrix} 1 & 0 \\ c+c(1-c) & (1-c)^2 \end{bmatrix} \\ &= (q + c(2-c) - 2qc + qc^2, (1-q)(1-c)^2), \end{aligned}$$

und für  $p_3$

$$p'_3 = (q, 1-q) \begin{bmatrix} 1 & 0 \\ c(2-c) + c(1-c)^2 & (1-c)^3 \end{bmatrix}, \quad \text{etc.}$$



Man findet, dass die Wahrscheinlichkeit, *nicht* in  $S_1$  zu sein, nach dem  $k$ -ten Schritt durch  $(1 - c)^k$  gegeben ist, so dass die Wahrscheinlichkeit, in  $S_1$  zu sein, gleich  $1 - (1 - c)^k$  ist; für  $k \rightarrow \infty$  folgt also  $p(S_1) \rightarrow 1$  und damit  $p(S_2) \rightarrow 0$ ; also ist der Vektor  $\pi$  mit  $P\pi = \pi$  durch  $\pi = (1, 0)'$  gegeben. Die Kette ist ergodisch.  $\square$

Eine weitere Einführung in diese Modelle findet man in Atkinson, Bower und Crothers (1965)<sup>10</sup> bzw. in den Kapiteln 9 und 10 von Luce, Bush, Galanter (1963)<sup>11</sup>.

**Beispiel 1.7** (2-Personen Interaktion, Atkinson et al (1965), p. 313) Die Interaktion zweier Personen (einer *Dyade*) läßt sich in bestimmten Situationen dadurch charakterisieren, dass die Handlungen der Personen jeweils in eine von zwei Klassen fallen: eine Handlung ist entweder kooperativ oder kompetitiv (nicht kooperativ). Wenn die Person  $A$  kooperiert,  $B$  aber nicht, so erleidet  $A$  einen "Schaden" (worin auch immer der besteht) und  $B$  erzielt etwa einen Gewinn, und umgekehrt erleidet  $B$  einen Schaden, wenn  $B$  kooperiert,  $A$  aber nicht; dann hat  $A$  einen Gewinn. Kooperieren beide, so haben sie den höchsten Gewinn bzw. den geringsten Verlust, kooperieren beide nicht, so ist ihr Verlust maximal. Die einfachste Lösung für beide besteht darin, stets zu kooperieren. Erzielt man aber den höchsten Gewinn, wenn man selber *nicht* kooperiert, aber der andere kooperiert, so kann es reizvoll sein, eben nicht zu kooperieren. So muß man stets aufs neue entscheiden, ob man kooperieren will oder nicht. Denn nicht zu kooperieren hat den Vorteil, dem anderen keinen Gewinn zukommen zu lassen für den Fall, dass er nicht kooperiert.

Die Dyade kann also in drei verschiedenen Zuständen sein:  $S_1$  — keiner der beiden kooperiert ( $\bar{A}\bar{B}$ ),  $S_2$  — beide kooperieren ( $AB$ ),  $S_3$  — einer der beiden kooperiert nicht ( $\bar{A}B \cup A\bar{B}$ ). Das Verhalten der Dyade werde nun durch die folgenden Annahmen beschrieben:

- A1: Macht eine der beiden Personen  $A$  oder  $B$  einen Gewinn, so behält sie das Verhalten, das zu diesem Gewinn führte, mit Wahrscheinlichkeit 1 bei.
- A2: Macht eine der beiden Personen  $A$  oder  $B$  keinen Gewinn, macht sie also einen Verlust, so ändert sie das Verhalten, dass zu diesem Verlust führte, mit der Wahrscheinlichkeit  $\theta$ .

Durch diese Annahmen wird die Folge der Zustände, in der sich die Dyade befindet, als Markov-Kette charakterisiert. Denn das folgende Verhalten hängt immer nur vom Ergebnis des zuletzt gewählten Verhaltens ab, nicht aber von den davor gewählten Verhaltensweisen. Bezeichnet man mit  $S_{i,n}$  den Sachverhalt, dass die Dyade im  $n$ -ten "Durchgang", d.h. beim  $n$ -ten Austausch von Verhaltensweisen, im Zustand  $S_i$  ist, so folgen aus den Annahmen A1 und A2 die folgenden Übergangswahrscheinlichkeiten:  $P(S_{1,n+1} | S_{1,n}) = (1 - \theta)^2$ , denn  $S_1$  heißt, dass beide nicht kooperiert haben, und beide ändern - stochastisch unabhängig voneinander — ihren Verhaltenstyp von kompetitiv zu kooperativ bzw. bleiben beim kompetitiven Verhaltenstyp, und dies geschieht mit der Wahrscheinlichkeit  $(1 - \theta)(1 - \theta) = (1 - \theta)^2$ . Weiter ist  $p(S_{2,n+1} | S_{1,n}) = \theta^2$ , - den mit Wahrscheinlichkeit  $\theta$  gehen beide zum kooperativen Verhaltenstyp über. Schließlich folgt noch  $p(S_{3,n+1} | S_{1,n}) = 2\theta(1 - \theta)$ , denn es gilt entweder  $\bar{A}\bar{B} \rightarrow \bar{A}B$ ; beide machen einen Verlust, der erste bleibt mit der Wahrscheinlichkeit  $\theta$  bei seinem Verhaltenstypus, und der andere ändert in von kompetitiv zu kooperativ, und da sie stochastisch unabhängig

<sup>10</sup>Atkinson, R.C., Bower, G.H., Crothers, E.J.: An Introduction to Mathematical Learning Theory. New York 1965

<sup>11</sup>Luce, R.D., Bush, R.R., Galanter, E. Handbook of Mathematical Psychology, Vol. II, New York 1963

voneinander handeln, hat dies die Wahrscheinlichkeit  $\theta(1 - \theta)$ . Die andere Möglichkeit ist besteht im Übergang  $\bar{A}\bar{B} \rightarrow A\bar{B}$ , der ebenfalls mit der Wahrscheinlichkeit  $(1 - \theta)\theta$  erfolgt, und zusammen ergibt sich die Wahrscheinlichkeit  $2\theta(1 - \theta)$ .

Die Übergangswahrscheinlichkeiten  $p(S_{j,n+1} | S_{i,n})$  ergeben sich durch analoge Betrachtungen, so dass man die Matrix

	$S_1$	$S_2$	$S_3$
$S_1$	$(1 - \theta)^2$	$\theta^2$	$2\theta(1 - \theta)$
$S_2$	$\theta^2/4$	$1 - \theta + \theta^2/4$	$\theta - \theta^2/2$
$S_3$	$\theta(1 - \theta)$	$\theta(1 - \theta)$	$1 - \theta + 2\theta^2$

bzw.

$$P = \begin{pmatrix} (1 - \theta)^2 & \theta^2 & 2\theta(1 - \theta) \\ \theta^2/4 & 1 - \theta + \theta^2/4 & \theta - \theta^2/2 \\ \theta(1 - \theta) & \theta(1 - \theta) & 1 - \theta + 2\theta^2 \end{pmatrix}$$

erhält.

Das Modell definiert eine ergodische Markov-Kette und stellt sicherlich eine drastische Vereinfachung dar, denkt man an die Komplexität menschlichen Handelns. Um so mehr überrascht es, zu sehen, wie gut experimentelle Daten dem Modell entsprechen.  $\square$

Markov-Ketten haben die mathematisch sehr angenehme Eigenschaft der Gedächtnislosigkeit: die Wahrscheinlichkeit, in einen bestimmten Zustand zu gehen, hängt nur von dem Zustand ab, in dem der Prozess gerade ist. Deswegen lassen sich Markov-Ketten (verhältnismäßig leicht) analysieren. Es ist also nicht verwunderlich, dass sie als Modelle für psychologische Prozesse herangezogen werden. Der klassische Artikel von Miller (1952)<sup>12</sup> liefert eine Reihe wichtiger Einzelheiten über Markov-Ketten, die bei der Modellierung von Nutzen sind. Unter anderem wird gezeigt, wie die zunächst oft psychologisch unplausible Annahme der Gedächtnislosigkeit durch geeignete Definition der Zustände gerechtfertigt werden kann, so dass die mathematischen Vorteile der Markov-Ketten erhalten bleiben. Weitere Details findet man in Kemeny und Snell (1967)<sup>13</sup>. Coleman (1964)<sup>14</sup> macht extensiven Gebrauch von Markov-Ketten bei der Modellierung sozialer Prozesse.

Die Zeit ist bisher nicht explizit in die Beschreibung der Markov-Ketten eingegangen. Natürlich ist es plausibel, anzunehmen, dass die Zustandsübergänge in einem bestimmten Zeitraster erfolgen. So kann man z.B. annehmen, dass die Übergänge stets nach Ablauf eines Zeitintervalls der Länge  $T$  erfolgen; die Werte der  $p_{ij}$  können dann von  $T$  abhängen. Eine zweite Möglichkeit besteht darin, die  $p_{ij}$  explizit als Funktionen der Zeit zu betrachten,  $p_{ij} = p_{ij}(t)$ . Spezielle Formen dieser Abhängigkeit, die für die Modellierung sozialer Prozesse von Belang sind, diskutiert Singer (19??)<sup>15</sup>. Andere Formen der Zeitabhängigkeit werden im folgenden Kapitel behandelt.

## 1.5.2 Verweildauern und Semi-Markov-Prozesse

Bei der Betrachtung von Markov-Ketten werden die Zeiten, die der Prozess in einem gegebenen Zustand verbringt - dies sind die *Verweildauern* - nicht explizit betrachtet.

<sup>12</sup>Miller, G.A.: (1952) Finite Markov-Processes in Psychology. Psychometrika, **17**, 149-167. Reprint in Luce, R.D., Bush, R.R., Galanter, E.: Readings in Mathematical Psychology, Vol. I, New York 1963

<sup>13</sup>Kemeny, J.G., Snell, J.L.: Finite MARKov Chains. Toronto, 1967

<sup>14</sup>Coleman, J. Introduction into mathematical sociology. New York 1964

<sup>15</sup>Singer, B., Spilerman, S.: (19??), The representation of Social Processes by Markov Models. American Journal of Sociology, **82**, 1 — 54

Gleichwohl kann es nützlich sein, diese Zeiten zu betrachten. Denn häufig ist man ja an einem Prozess *in der Zeit* interessiert: man beobachtet die Folge von Zuständen und die Zeiten (*Verweildauern*), die der Prozess in den einzelnen Zuständen bleibt. Man spricht dann auch von *Markov-Ketten in kontinuierlicher Zeit*.

**Beispiel 1.8** (Dyadische Interaktion) Man ist an der Interaktion zweier "Spieler" interessiert, die sich in einer Prisoner's Dilemma-Situation befinden. Die Interaktion bestehe aber nicht in einer diskreten Folge solcher Spiele, sondern sei ein kontinuierlicher Verlauf: zu jedem Zeitpunkt soll es möglich sein, sich

- kooperativ ( $A_1$ ), oder
- kompetitiv ( $A_2$ ), oder
- neutral ( $A_3$ )

zu verhalten; die Menge der möglichen Verhaltensweisen soll also in drei disjunkte Klassen aufteilbar sein. Man hat eine Matrix der Art (1.1), bei der die Zeilen den Partner I und

Tabelle 1.1: Übergänge bei Interaktionen

	$A_1$	$A_2$	$A_3$
$A_1$	$p_{11}$	$p_{12}$	$p_{13}$
$A_2$	$p_{21}$	$p_{22}$	$p_{23}$
$A_3$	$p_{31}$	$p_{32}$	$p_{33}$

die Spalten den Partner II repräsentieren; für jeden Übergang  $\{A_i \rightarrow A_j\}$  kann man die entsprechende Wartezeit beobachten.  $\square$

In Definition 1.1, p. 6, ist der Begriff der Hazardfunktion eingeführt worden. Auf den Fall von Markov-Ketten kann er wie folgt angewandt werden:

**Definition 1.8** *Es seien  $A_i$  und  $A_j$  zwei Zustände einer Markov-Kette, und  $\{A_i \rightarrow A_j\}$  bezeichne den Übergang von  $A_i$  nach  $A_j$ . Dann heißt*

$$\alpha_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P(\{A_i \rightarrow A_j\} \in (t, t + \Delta t] | A_i \text{ zur Zeit } t)}{\Delta t} \quad (1.106)$$

die Übergangsrate von  $A_i$  nach  $A_j$ . Weiter bezeichne  $\{A_i^\uparrow\}$ , dass der Prozess den Zustand  $A_i$  verläßt. Dann heißt

$$\lambda_i = \lim_{\Delta t \rightarrow 0} \frac{P(\{A_i^\uparrow\} \in (t, t + \Delta t] | A_i \text{ zur Zeit } t)}{\Delta t} \quad (1.107)$$

die Terminationsrate des Zustands  $A_i$ .

$\lambda_i$  ist offenbar die Hazardfunktion für die Wartezeit, die verstreicht, bis der Prozess den Zustand  $A_i$  verläßt. Ein Markov-Prozess ist aber gedächtnislos, d.h. die Wahrscheinlichkeit, dass ein Markov-Prozess zur Zeit  $t + \Delta t$  in einem Zustand  $A_i$  für irgend ein  $i$  ist,

ist unabhängig von der Geschichte des Prozesses bis zur Zeit  $t$ . Auf die Definition von  $\lambda_i$  bezogen heißt dies, dass

$$\lambda_i = \textit{konstant}$$

Daraus folgt nach Beispiel ?? dann, dass die Verteilung der Verweildauer, d.h. der Wartezeit, die verstreicht, bis der Prozess den Zustand  $A_i$  verläßt, durch die Exponentialverteilung

$$F_i(t) = 1 - e^{-\lambda_i t} \quad (1.108)$$

gegeben ist. Nun verläßt der Prozess den Zustand  $A_i$ , indem er in irgend einen anderen Zustand übergeht, also in  $A_1$  oder  $A_2$  oder  $A_3$  etc., und er kann nicht in zwei Zustände zugleich übergehen. Dementsprechend muß

$$\lambda_i = \sum_j \alpha_{ij} \quad (1.109)$$

gelten.

Aus den Gleichungen (1.106) und (1.107) ergibt sich ein Ausdruck für die Übergangswahrscheinlichkeit  $p_{ij} = P(A_i \rightarrow A_j)$ . Es ist ja

$$\frac{\alpha_{ij}(t)}{\lambda_i(t)} = \lim_{\Delta t \rightarrow 0} \frac{P(\{A_i \rightarrow A_j\} \in (t, t + \Delta t] | A_i \text{ zur Zeit } t)}{P(\{A_i^\uparrow\} \in (t, t + \Delta t] | A_i \text{ zur Zeit } t)}$$

Dieser Quotient ist eine bedingte Wahrscheinlichkeit, nämlich die Wahrscheinlichkeit, zum Zeitpunkt  $t$  von  $A_i$  nach  $A_j$  zu wechseln unter der Bedingung, dass der Zustand  $A_i$  überhaupt zum Zeitpunkt  $t$  verlassen wird. Damit ist dieser Quotient gleich der Übergangswahrscheinlichkeit

$$p_{ij}(t) = \frac{\alpha_{ij}(t)}{\lambda_i(t)} \quad (1.110)$$

Es ist möglich, dass die Wechsel der Zustände durch eine Markov-Kette beschrieben werden können, dass aber die Verweildauern nicht exponentialverteilt sind. Prozesse dieser Art heißen *Semi-Markov-Prozesse*. Ein Beispiel für einen solchen Prozess ist eine Folge von Zuständen, bei die Aufenthaltsdauern durch die Weibull-Verteilung gegeben ist.

### 1.5.3 Schätzungen für die Übergangswahrscheinlichkeiten

Die Übergangswahrscheinlichkeiten  $p_{ij}(t)$  sind im allgemeinen nicht bekannt und müssen deshalb aus den beobachteten Häufigkeiten der Übergänge geschätzt werden. Anderson und Goodman (1957)<sup>16</sup> haben hierfür Maximum-Likelihood-Schätzungen hergeleitet. Es seien  $t = 1, 2, \dots, T$  die Zeitpunkte, zu denen die Zustände beobachtet werden, und  $p_{ij}(t)$  sei die Wahrscheinlichkeit, dass der Prozess zur Zeit  $t$  in Zustand  $S_i$  (oder  $A_i$ ) ist unter der Bedingung, dass er in  $S_j$  (oder  $A_j$ ) war zur Zeit  $t - 1$ ,  $i, j = 1, \dots, m$ . Für  $p_{ij}(t) = p_{ij}$  für alle  $t$  ist der Prozess stationär. Weiter sei  $n_{ij}(t)$  die Anzahl der Individuen, die von  $S_i$  zur Zeit  $t - 1$  zu  $S_j$  zur Zeit  $t$  übergehen, und  $n_i(t)$  seien die Gesamtzahl der Individuen in Zustand  $S_i$  zur Zeit  $t$ . Die Maximum-Likelihood-Schätzung für  $p_{ij}(t)$  ist dann durch

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{n_i(t-1)} = n_{ij}(t) / \sum_{k=1}^m n_{ik}(t) \quad (1.111)$$

<sup>16</sup> Anderson, T.W., Goodman, L.A., Statistical Inference about Markov Chains. Ann. Math. Stat. 1957, 28, 89-110; Abdruck in Luce, R., Bush, R.R., Galanter, E. Readings in Mathematical Psychology, Vol. I, New York 1963

### 1.5.4 Dyadische Interaktion

In diesem Abschnitt soll ein Beispiel für die Anwendung von Markov-Ketten als Beschreibung und Analyse von dyadischen Interaktionen relativ ausführlich gegeben werden.

Gardner & Griffin (1989)<sup>17</sup> untersuchen die Dynamik der Variablen "Anschauen" in einer Zweierinteraktion. Ein Ehepaar (6 Jahre verheiratet, Ehefrau 25 Jahre alt, Ehemann 31 Jahre alt) diskutieren Fragen ihrer Beziehung, und es wird aufgezeichnet, wann die Frau ihren Mann anschaut, wieder wegschaut, etc, und ebenso wird das "Schau"-verhalten des Ehemannes aufgezeichnet. Es werden zunächst zwei Indikatorvariablen definiert,  $y_w(t)$  und  $y_m(t)$ , wobei der Index  $w$  für "wife" und der Index  $h$  für "husband" steht. Es soll

$$y_w(t) = \begin{cases} 1, & \text{Ehefrau schaut den Ehemann an} \\ 2 & \text{Ehefrau schaut den Ehemann nicht an} \end{cases} \quad (1.112)$$

Die Definition von  $y_m(t)$  ist analog. Das Verhalten der Frau wird dann durch eine Folge der Art 1, 1, 2, 2, 1, 2, 2, 1, 1, 1, etc beschrieben, und das des Mannes durch eine korrespondierende Folge. Die Variablen  $y_w(t)$  und  $y_h(t)$  sind zufällige Veränderliche, die das Verhalten des Paares zur Zeit  $t$  repräsentieren.

Die Zeiten, zu denen die Frau ihr Verhalten ändert, werden mit  $t_{w1}, t_{w2}, \dots$  bezeichnet, und analog dazu die Zeiten, zu denen der Mann sein Verhalten ändert, durch  $t_{h1}, t_{h2}, \dots$ . Die Dauer des  $k$ -ten Verhaltens ist dann

$$u_{wk} = t_{w(k+1)} - t_{wk}, \quad k = 1, 2, \dots \quad (1.113)$$

und es ist

$$t_{w(k+1)} = \sum_{i=0}^k u_{wi}. \quad (1.114)$$

Im Folgenden werden  $t_{wk}$  und  $u_{wk}$  - wie in der Statistik üblich - als Realisierungen von zufälligen Veränderlichen  $T_{wk}$  und  $U_{wk}$  aufgefaßt. Für die Zeiten des Mannes gelten analoge Definitionen.

Die Autoren nehmen an, dass die folgenden vier Faktoren für die Verteilungen der Zeiten  $T$  relevant sind:

- Die Dauer des Verhaltens,
- die Art des Verhaltens,
- das Individuum (Frau oder Mann), und
- das unmittelbar vorangegangene Verhalten des Partners in der Dyade.

Inbesondere werden die folgenden Annahmen gemacht:

1. Die Parameter der Verteilungen ändern sich nicht während der Interaktion.
2. Für kleine Zeitintervalle ist das Partnerverhalten, bedingt auf die vorangegangene Interaktion, unabhängig voneinander für kleine Zeitintervalle.
3. Die Zeiten zwischen den Verhaltensänderungen sind weibullverteilt.

<sup>17</sup>Gardner, W., Griffin, W.A. (1989) Methods for the analysis of parallel streams of continuously recorded social behaviours. Psychol. Bulletin, 105, 446-455

Weiter wird angenommen, dass sich

Der dyadische Prozess kann durch einen Vektor  $Y_d(t)$  beschrieben werden, der durch

$$Y_d(t) = (Y_w(t), Y_m(t))' \quad (1.115)$$

definiert ist. Zu jedem Zeitpunkt  $t$  ist  $Y_d$  durch eines der folgenden vier Paare definiert:

$$(1, 1), (1, 2), (2, 1), (2, 2)$$

Im ersten Fall ist also zum Zeitpunkt  $t$   $Y_w(t) = 1$  und  $Y_m(t) = 1$ , d.h. beide Partner schauen sich an. Im zweiten Fall ist  $Y_w(t) = 1$  und  $Y_m(t) = 2$ , d.h. die Frau schaut den Mann an, der Mann sieht aber nicht seine Frau an, etc. Man kann sagen, dass sich das Paar zum Zeitpunkt  $t$  in einen von vier möglichen Zuständen befindet, der durch  $Y_d(t)$  definiert ist. Von jedem Zustand ist ein Übergang zu einem der anderen Zustände möglich. So bedeutet der Übergang  $(1, 1) \rightarrow (2, 1)$  dass die Frau wegschaut, nachdem sich die Partner beide angesehen haben. Man kann also die Übergangsmatrix definieren.

Tabelle 1.2: Matrix der Zustandsübergänge

	(1, 1)	(1, 2)	(2, 1)	(2, 2)
(1, 1)	$n_{11}$	$n_{12}$	$n_{13}$	$n_{14}$
(1, 2)	$n_{21}$	$n_{22}$	$n_{23}$	$n_{24}$
(2, 1)	$n_{31}$	$n_{32}$	$n_{33}$	$n_{34}$
(2, 2)	$n_{41}$	$n_{42}$	$n_{43}$	$n_{44}$

Dem Übergang vom  $i$ -ten Zustand ( $i = 1$  wenn (1, 1) vorliegt,  $i = 2$ , wenn (1, 2) gegeben ist, etc) zum  $j$ -ten Zustand entsprechen nun die Übergangsraten oder Übergangsintensitäten  $\alpha_{ij}$ . Für den Übergang  $(1, 1) \rightarrow (2, 1)$  hat man insbesondere die Übergangsraten  $\alpha_{13}$ , d.h.

$$\alpha_{13}(t) = \alpha_{(1,1),(2,1)} = \lim_{\Delta t \downarrow 0} \frac{P(Y_d(t + \Delta t) = (2, 1) | Y_d(t) = (1, 1))}{\Delta t} \quad (1.116)$$

Je größer  $\alpha_{ij}(t)$ , desto größer ist das "Risiko", dass sich das Verhalten ändert. Die Übergangsintensitäten werden in einer Matrix zusammengefasst:

$$R(t) = \begin{pmatrix} 0 & \alpha_{(1,1)(1,2)} & \alpha_{(1,1)(2,1)} & \alpha_{(1,1)(2,2)} \\ \alpha_{(1,2)(1,1)} & 0 & \alpha_{(1,2)(2,1)} & \alpha_{(1,2)(2,2)} \\ \alpha_{(2,1)(1,1)} & \alpha_{(2,1)(1,2)} & 0 & \alpha_{(2,1)(2,2)} \\ \alpha_{(2,2)(1,1)} & \alpha_{(2,2)(1,2)} & \alpha_{(2,2)(2,1)} & 0 \end{pmatrix} \quad (1.117)$$

Man bemerke, dass die Diagonalzellen dieser Matrix alle gleich Null sind. Dies bedeutet, dass Übergänge zwischen identischen Zuständen, also etwa von (1, 1) nach (1, 1), "unmöglich" sind; die Matrix enthält nur die Intensitäten oder Raten für Übergänge zwischen verschiedenen Zuständen. Man kann sagen, dass der Begriff des Risikos sich nur auf die Veränderung von Zuständen bezieht; ein Übergang etwa der Art  $(1, 1) \rightarrow (1, 1)$  hat das Risiko Null.

Aus Annahme 2 folgt nun für hinreichend kleine Werte von  $\Delta t$

$$P(Y_d(t + \Delta t) | Y_d(t)) \approx P(Y_h(t + \Delta t) | Y_d(t)) P(Y_w(t + \Delta t) | Y_d(t)) \quad (1.118)$$

wobei  $\approx$  für "in guter Näherung" steht<sup>18</sup> Inhaltlich bedeutet die Gleichung, dass z.B. die Frau nicht auf das Verhalten ihres Mannes zum "jetzigen" Zeitpunkt  $t$  reagiert, sondern auf das Verhalten unmittelbar davor. Hieraus folgt wiederum, dass die Wahrscheinlichkeit, dass die Partner ihr Verhalten zum exakt gleichen Zeitpunkt ändern, gleich Null ist, was wiederum bedeutet, dass die Übergangintensität für Übergänge, bei denen beide Partner ihr Verhalten zum gleichen Zeitpunkt äbnern, gleich Null ist: der Übergang  $(1, 1) \rightarrow (2, 2)$  hat demnach ein Risiko gleich Null. Die Risikomatrix  $R(t)$  nimmt also die folgende Form an:

$$R(t) = \begin{pmatrix} 0 & \alpha_{(1,1)(1,2)} & \alpha_{(1,1)(2,1)} & 0 \\ \alpha_{(1,2)(1,1)} & 0 & 0 & \alpha_{(1,2)(2,2)} \\ \alpha_{(2,1)(1,1)} & 0 & 0 & \alpha_{(1,2)(2,2)} \\ 0 & \alpha_{(2,2)(1,2)} & \alpha_{(2,2)(2,1)} & 0 \end{pmatrix} \quad (1.119)$$

Inspiziert man diese Matrix, so sieht man, dass nur die Übergangintensitäten für die Frau einerseits und den Mann andererseits geschätzt werden müssen (diese Ersparnis hinsichtlich der Anzahl der zu schätzenden Parameter ist gewissermaßen der "tiefere Sinn" der Annahme 2; die Annahme stellt andererseits keine Einschränkung der Allgemeinheit dar, denn die Wahrscheinlichkeiten von Punkten einer kontinuierlichen Variablen, hier der Zeit  $t$ , sind stets gleich Null.

Da angenommen wurde, dass das Verhalten der Partner unabhängig voneinander ist, folgt u.a.

$$\alpha_{(1,1)(2,1)} = \lim_{\Delta t \downarrow 0} \frac{P(Y_w(t + \Delta t) = 2 | Y_d(t) = (1, 1))}{\Delta t} \times \frac{P(Y_h(t + \Delta t) = 1 | Y_d(t) = (1, 1))}{\Delta t}. \quad (1.120)$$

Mit  $\Delta t \rightarrow 0$  folgt, dass die Wahrscheinlichkeit, ein Partner sein Verhalten ändert, gegen 1 strebt, d.h.

$$\alpha_{(1,1)(2,1)} = \lim_{\Delta t \downarrow 0} \frac{P(Y_w(t + \Delta t) = 2 | Y_d(t) = (1, 1))}{\Delta t} \quad (1.121)$$

Dies ist das *instantane Risiko*, dass die Frau nach  $t$  Zeiteinheiten (Sekunden) wegsieht, wenn sich beide Partner angesehen haben.

Zur Vereinfachung werde die folgende Notation eingeführt:

$$\alpha_{(1,1)(2,1)} = \alpha_{w1}(t | Y_h(t) = 1) = \alpha_{w1}(t | 1) \quad (1.122)$$

und analog für die anderen Intensitäten. Für die Risikomatrix erhält man dann

$$R(t) = \begin{pmatrix} 0 & \alpha_{h1}(t|1) & \alpha_{w1}(t|1) & 0 \\ \alpha_{h2}(t|1) & 0 & 0 & \alpha_{w1}(t|2) \\ \alpha_{w2}(t|2) & 0 & 0 & \alpha_{h1}(t|2) \\ 0 & \alpha_{w2}(t|2) & \alpha_{h2}(t|2) & 0 \end{pmatrix} \quad (1.123)$$

Hier bezeichnet der erste Index die Frau ( $w$ ) oder den Mann ( $h$ ), der zweite Index das Verhalten, das gezeigt wird, und der dritte Index zeigt das Verhalten des Partners an.

<sup>18</sup>Die Gleichung lautet exakt

$$P(Y_d(t + \Delta t) | Y_d(t)) = P(Y_h(t + \Delta t) | Y_d(t)) P(Y_w(t + \Delta t) | Y_d(t)) + o(\Delta t),$$

wobei  $o(\Delta t)$  eine beliebige Funktion ist, die schneller als  $\Delta t$  gegen Null geht, d.h.  $o(\Delta t)/\Delta t \rightarrow 0$  für  $\Delta t \rightarrow 0$ .

Die Wahrscheinlichkeitsverteilungen der Zeiten (zwischen den Verhaltensänderungen) definieren die Risiken und umgekehrt. Nach Annahme 3 sind die Zwischenzeiten weibullverteilt. Die Hazardfunktion der Weibullverteilung ist nach (??) durch

$$\alpha_{w1}(t|2) = a_{w12}t^{-b_{w12}}, \quad a_{w12} > 0, b_{w12} > 0 \quad (1.124)$$

gegeben.

Ein großer Wert von  $\alpha$  impliziert ein höheres "Risiko" der Verhaltensänderung, und der Wert von  $\beta$  bestimmt die Gestalt der Übergangswahrscheinlichkeit als Funktion von  $t$ . Insbesondere gilt

- $0 < b < 1$ : Die Übergangintensität wird kleiner mit  $t$ ,
- $b = 1$ : Die Übergangintensität ist konstant (dies ist der Fall der Exponentialverteilung),
- $b > 1$ : Die Übergangintensität wird mit steigendem Wert von  $t$  größer, d.h. das gegenwärtige Verhalten wird mit steigender Wahrscheinlichkeit abgebrochen.

Man kann spezielle Modelle des Verhaltens testen, indem man Annahmen über die Elemente der Intensitätsmatrix  $R(t)$  macht und diese anhand von Daten überprüft. So kann man z.B. von der Annahme ausgehen, dass die Verteilung der Zeiten nur von der Identität des Individuums abhängt, nicht aber von der Art des eigenen Verhaltens oder der des Partners. Dann folgt

$$\alpha_{h1}(t|1) = \alpha_{h1}(t|2) = \alpha_{h2}(t|1) = \alpha_{h2}(t|2)$$

und

$$\alpha_{w1}(t|1) = \alpha_{w1}(t|2) = \alpha_{w2}(t|1) = \alpha_{w2}(t|2).$$

Eine grundlegende Frage ist, ob das Verhalten eines Individuums vom Verhalten des Partners abhängt oder nicht. Die Nullhypothese ist dann, dass das Verhalten nicht vom Partner abhängt. Dieser Hypothese entspricht das

#### Modell der wechselseitigen Unabhängigkeit:

$$\alpha_{w1}(t|1) = \alpha_{w1}(t|2) \quad (1.125)$$

$$\alpha_{w2}(t|1) = \alpha_{w2}(t|2) \quad (1.126)$$

$$\alpha_{h1}(t|1) = \alpha_{h2}(t|2) \quad (1.127)$$

$$\alpha_{h2}(t|1) = \alpha_{h2}(t|2) \quad (1.128)$$

Die Verteilung der Zeiten sei durch die Weibull-Verteilung gegeben. Es gilt

$$a_{w11} = a_{w12} \quad (1.129)$$

$$b_{w11} = b_{w12} \quad (1.130)$$

Es gibt insgesamt acht Parameter  $a$  und acht Parameter  $b$ , die in Vektoren  $P_a$  und  $P_b$  zusammengefaßt werden können:

$$P_a = (a_{w11}, a_{w12}, a_{w21}, a_{w22}, a_{h11}, a_{h12}, a_{h21}, a_{h22})' \quad (1.131)$$

$$P_b = (b_{w11}, b_{w12}, b_{w21}, b_{w22}, b_{h11}, b_{h12}, b_{h21}, b_{h22})' \quad (1.132)$$



Die Parameter in  $P_a$  können nun auf vier Parameter

$$\gamma = (\gamma_m, \gamma_p, \gamma_b, \gamma_{bp})'$$

zurückgeführt werden, die in  $P_b$  auf Parameter

$$\phi = (\phi_m, \phi_p, \phi_b, \phi_{bp})'$$

Über die Parameter  $\gamma$  können die Effekte, die in der Untersuchung erfaßt werden sollen, auf den Parameter  $a$  erfaßt werden, und über die Parameter  $\phi$  werden die Effekte auf den Parameter  $b$  erfaßt. Man geht dabei so vor, dass man geeignete Matrizen  $A_{ind}$  und  $B_{ind}$  definiert, über die die *Logarithmen* der  $a$  und  $b$  als Linearkombination der  $\gamma$  und  $\phi$  dargestellt werden können. Damit ergibt sich methodisch ein Zugang zu den Daten, der dem bei der Anwendung eines log-linearen Modells entspricht. Insbesondere sei

$$A_{ind} = B_{ind} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad (1.133)$$

Dann soll gelten

$$\log P_a = A_{ind}\gamma, \quad \log P_b = B_{ind}\phi. \quad (1.134)$$

So ist z.B.

$$\log a_{w11} = \gamma_m + \gamma_p + \gamma_b + \gamma_{bp},$$

und ebenso folgt

$$\log b_{w11} = \phi_m + \phi_p + \phi_b + \phi_{bp}$$

Hierin bezeichnen nun  $\gamma_m$  und  $\phi_m$  die Gesamtmittelwerte der  $\log a$ - und  $\log b$ -Parameter,  $\gamma_p$  und  $\phi_p$  repräsentieren einen Personeneffekt;  $\gamma_p > 0$  bedeutet, dass die Risikoniveaus für eine Verhaltensänderung (hinschauen, wenn man weggesehen hat, und wegsehen, wenn man hingesehen hat) für die Frau größer sind als für den Mann. Für  $\gamma_p < 0$  sind die Risiken für den Mann größer.  $\phi_p > 0$  bedeutet, dass die  $b$ -Werte für die Frau größer sind als für den Mann. Die Wartezeitverteilungen sind dann steiler für die Frau, - was noch nicht bedeutet, dass sie auch kürzer sind, eher ist die Variabilität der Zeiten geringer. Für  $\phi_p < 0$  sind die Verteilungen des Mannes steiler als die der Frau.

Die Parameter  $\gamma_b$  und  $\phi_b$  repräsentieren einen Verhaltenseffekt.  $\gamma_b > 0$  bedeutet, dass das Risikoniveau für das Hinschauen größer ist als das für das Wegschauen.  $\phi_b > 0$  bedeutet, dass die  $b$ -Werte für die Verteilungen für das Hinschauen größer sind als für das Wegschauen, das Schauen also eine eher kleinere Variabilität hat als das Wegschauen. Für negative Parameter gilt wieder die Umkehrung.

Die Parameter  $\gamma_{pb}$  und  $\phi_{pb}$  repräsentieren eine Person  $\times$  Verhalten-Interaktion: für  $\gamma_{pb} > 0$  hat die Frau im Vergleich zu ihrem Mann ein höheres Risiko, das Hinschauen abzubrechen, als das Wegschauen.  $\phi_{pb} > 0$  bedeutet, dass die  $b$ -Werte für die Verteilungen der Frau, wenn sie ihren Mann anschaut, größer sind als die für die Verteilungen, wenn sie wegschaut; für den Mann gilt dann die Umkehrung.

Die Schätzungen für die  $\gamma$  und  $\phi$  sind in der folgenden Tabelle 1.3 zusammengefaßt worden: Die Schätzungen für die  $a$ - und  $b$ -Werte für das Unabhängigkeitsmodell erge-

Tabelle 1.3: Parameterschätzungen für das Unabhängigkeitsmodell

Parameter	Schätzung	SD	$t$
$\gamma_m$	-1.399	.049	-28.30
$\gamma_p$	-.042	.049	-.84
$\gamma_b$	.085	.049	1.72
$\gamma_{pb}$	-.0224	.049	-4.53
$\phi_m$	.060	.030	2.00
$\phi_p$	-.090	.030	-3.00
$\phi_b$	-.057	.030	-1.90
$\phi_{pb}$	-.045	.030	-1.50

Tabelle 1.4: Schätzungen der  $a$ - und  $b$ -Parameter der Wartezeitverteilungen

Person	Verhalten	Verh. d. and.	$a$	$b$
Ehefrau	Ansehen	Ansehen	.206	.876
		nicht Anseh.	.206	.876
	nicht Anseh.	Ansehen	.272	1.075
		nicht Anseh.	.272	1.075
Ehemann	Ansehen	Ansehen	.350	1.148
		nicht Anseh.	.350	1.148
	nicht Anseh.	Ansehen	.189	1.175
		nicht Anseh.	.189	1.175

ben sich daraus wie in Tabelle 1.4 angegeben Zur Illustration werde der Wert von  $a_{w21}$  betrachtet. Dieser Parameter beschreibt das Risiko, d.h. die Neigung einer Verhaltensänderung, den Mann anzusehen, wenn sie ihn vorher nicht angesehen hat, und wenn er sie ansieht. Aus (1.134) folgt

$$a_{w21} = \frac{e^{\gamma_m} e^{\gamma_p}}{e^{\gamma_b} e^{\gamma_{pb}}} = \frac{e^{-1.399} e^{-.042}}{e^{.085} e^{-.244}} = .272$$

Der Tabelle 1.3 entnimmt man, dass  $|\gamma_p| < |\gamma_{pb}|$  und  $|\gamma_b| < |\gamma_{pb}|$ , - das Verhalten wird also in erster Linie durch die *Interaktion* zwischen den Partnern, nicht so sehr durch die Personen unabhängig von der Interaktion bestimmt.  $\gamma_{pb} < 0$  bedeutet, dass die Ehefrau ein höheres "Risiko" hat, das Wegschauen abzubrechen als das Anschauen abzubrechen; für ihren Mann gilt die Umkehrung.

Das Unabhängigkeitsmodell kann mit einem Modell, in dem bestimmte Abhängigkeiten postuliert werden, verglichen werden.

**Ein Abhängigkeitsmodell:** Für dieses Modell werden die folgenden  $\gamma$ - und  $\phi$ -Parameter spezifiziert:

$$\gamma = (\gamma_m, \gamma_p, \gamma_b, \gamma_o, \gamma_{bp}, \gamma_{po}, \gamma_{bo}, \gamma_{pbo})' \quad (1.135)$$

$$\phi = (\phi_m, \phi_p, \phi_b, \phi_o, \phi_{bp}, \phi_{po}, \phi_{bo}, \phi_{pbo})' \quad (1.136)$$

Die Design-Matrizen werden wie folgt spezifiziert:

$$A_{depend} = B_{depend} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{pmatrix} \quad (1.137)$$

Die Matrix  $R(t)$  hat insgesamt  $8 \times 8$  Übergangintensitäten. Beim Abhängigkeitsmodell werden 8  $\gamma$ - und 8  $\phi$ -Parameter postuliert; damit ist das Modell saturiert und die Parameter in  $R$  werden nicht durch ihre Abhängigkeit von den  $\gamma$  und  $\phi$  eingeschränkt. Die  $\gamma$  und  $\phi$ -Parameter entsprechen den Parametern eines  $2 \times 2 \times 2$ -log-linearen Modells, wobei der Partner jeweils als dritter Faktor eingeht.

Die folgenden Parameter sind noch nicht im Unabhängigkeitsmodell enthalten:

- $\gamma_o, \phi_o$ : *Effekt des jeweils anderen Partners* -  $\gamma_o > 0$  bedeutet, dass das "Risiko" einer Verhaltensänderung größer ist, wenn der andere Partner *schaute*, als wenn er *wegschaute*.  $\phi_o > 0$  bedeutet, dass der *b*-Parameter der Wartezeitverteilung größer ist, wenn der andere Partner schaut.
- $\gamma_{po}, \phi_{po}$ : *Interaktion zwischen Person und Partner*. Für  $\gamma_{po} > 0$  folgt, dass das Risiko einer Verhaltensänderung größer ist, wenn sie angesehen wird als wenn ihr Mann wegschaute. Für den Ehemann ist das Risiko einer Verhaltensänderung größer wenn seine Frau wegschaute.  $\phi_{po} > 0$  bedeutet, dass der *b*-Wert der Verteilung größer ist, wenn sie angesehen wird (steilere Verteilung, d.h. kleinere Varianz der Zeiten), und größer für den Ehemann, wenn er nicht angesehen wurde.

- $\gamma_{bo}, \phi_{bo}$ : *Interaktion Verhalten  $\times$  Partner*:  $\gamma_{op} > 0$  bedeutet, dass sich das Risiko einer Verhaltensänderung erhöht, wenn sich beide Partner ansehen.  $\phi_{op} > 0$  bedeutet, dass die  $b$ -Werte der Verteilungen erhöht sind, wenn sich die Partner ansehen.
- $\gamma_{pbo}, \phi_{pbo}$ : *Interaktion Person  $\times$  Verhalten  $\times$  Partner*: Die Risiken hängen von allem ab.

**Modellvergleich:** Das Unabhängigkeitsmodell kann als Spezialfall des Abhängigkeitsmodells betrachtet werden: alle Parameter des ersteren sind in letzterem enthalten. Der relative Fit der beiden Modell läßt sich dann über eine Likelihood-Ratio-Statistik abschätzen:

$$\chi^2 = -2(\log L_{ind} - \log L_{dep}), = 25.02 \quad df = 8 \quad (1.138)$$

wobei  $L$  die Likelihood des jeweiligen Modells ist. Es gibt insgesamt  $N = 311$  Verhaltensübergänge. Der  $\chi^2$ -Wert ist auf dem  $p = .01$ -Wert signifikant, d.h. das Abhängigkeitsmodell erklärt signifikant mehr als das Unabhängigkeitsmodell: die Partner handeln nicht tunabhängig voneinander. Die  $a$ - und  $b$ -Parameter für das Abhängigkeitsmodell werden in der Tabelle 1.5 zusammengefaßt: Die Parameter  $a$  und  $b$  der Wartezeitverteilungen

Tabelle 1.5:  $a$ - und  $b$ -Parameter für das Abhängigkeitsmodell

Person	Verhalten	Partner	$a$	$b$
Ehefrau	Ansehen	Ansehen	.157	.687
		nicht Anseh.	.247	.868
	nicht Anseh.	Ansehen	.318	1.283
		nicht Anseh.	.248	.961
Ehemann	Ansehen	Ansehen	.322	1.172
		nicht Anseh.	.411	1.130
	nicht Anseh.	Ansehen	.11914	1.252
		nicht Anseh.	.180	1.106

der Frau variieren mehr als Funktion des Verhaltens des Mannes als die entsprechenden Parameter des Mannes. Dies zeigt, dass das Verhalten der Frau durch das des Mannes beeinflusst wird, aber nicht umgekehrt.

Gardner und Griffin diskutieren noch ein weiteres Abhängigkeitsmodell, das die Asymmetrien zwischen dem Verhalten der Frau und dem des Mannes explizit in Rechnung stellt, dazu führen sie den Parametervektor

$$\phi = (\phi_m, \phi_p, \phi_b, \phi_{bp}, \phi_{wo}, \phi_{wbo})' \quad (1.139)$$

ein, zusammen mit der Design-Matrix

$$A_{asym} = B_{asym} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 \end{pmatrix} \quad (1.140)$$

Es werden also die Parameter für die Haupt- und Interaktionseffekte des Partnerverhaltens weggelassen, dafür werden zwei neue parameter eingeführt, die den Einfluß des Verhaltens des Ehemannes auf die Ehefrau spezifizieren. Berechnet man die zu (1.138) analoge Statistik, so erhält man  $\chi^2 = 5.44$  bei  $df = 4$  Freiheitsgraden; dieser Wert ist nicht signifikant. Das Asymmetriemodell ist aber sparsamer als das Abhängigkeitsmodell, so dass es letzterem vorzuziehen ist.

Faßt man die Ergebnisse zusammen, so zeigt sich, dass insgesamt die Ehefrau sensibler für nichtverbale "Cues" ist als der Ehemann. Formal äußert sich dies u.a. darin, dass die Parameter der Wartezeitverteilungen vom Verhalten des Partners abhängen.

# Kapitel 2

## Zeitreihenanalyse

### 2.1 Einführung

Gegeben sei eine Variable, die in der Zeit variiert und die durch eine Funktion  $Y$  der Zeit  $t$  repräsentiert werden kann; zur Zeit  $t$  sei als  $Y$  durch  $Y(t)$  gegeben. Die Variation von  $Y$  enthalte zufällige Komponenten, so dass  $Y(t)$  als zufällige Veränderliche aufgefaßt werden kann.  $Y$  als zufällige Funktion von  $t$  ist damit eine Realisierung (Trajektorie) eines stochastischen Prozesses  $Y_t$ .  $Y(t)$  werde zu bestimmten Zeitpunkten  $t_1, t_2, t_3 \dots$  gemessen. Die Folge  $\{Y(t_i)\}_{i=1,2,\dots}$ , d.h.  $Y(t_1), Y(t_2), Y(t_3) \dots$  ist dann eine *Zeitreihe*; die Zeitreihe kann dann als Realisierung des stochastischen Prozesses in den diskreten Zeitpunkten  $t_i, i = 1, 2, \dots$  betrachtet werden.

Ein stochastischer Prozess läßt sich durch gewisse statistische Größen beschreiben. So ist die Mittelwertfunktion durch

$$\mu(t) = E(Y(t)) = \int_{-\infty}^{\infty} yf(y; t)dy \quad (2.1)$$

definiert, wobei  $f$  die Dichte von  $Y$  zur Zeit  $t$  ist. Eine zweite Statistik ist die Autokovarianzfunktion

$$Kov(t_1, t_2) = E(Y(t_1) - \mu(t_1))(Y(t_2) - \mu(t_2)) \quad (2.2)$$

wobei

$$\sigma^2(t) = Kov(t_1, t_2), \quad t = t_1 = t_2 \quad (2.3)$$

die Varianz des Prozesses zur Zeit  $t$  ist. Man erhält dann die Autokorrelationsfunktion

$$R(t_1, t_2) = \frac{Kov(t_1, t_2)}{\sigma(t_1)\sigma(t_2)} \quad (2.4)$$

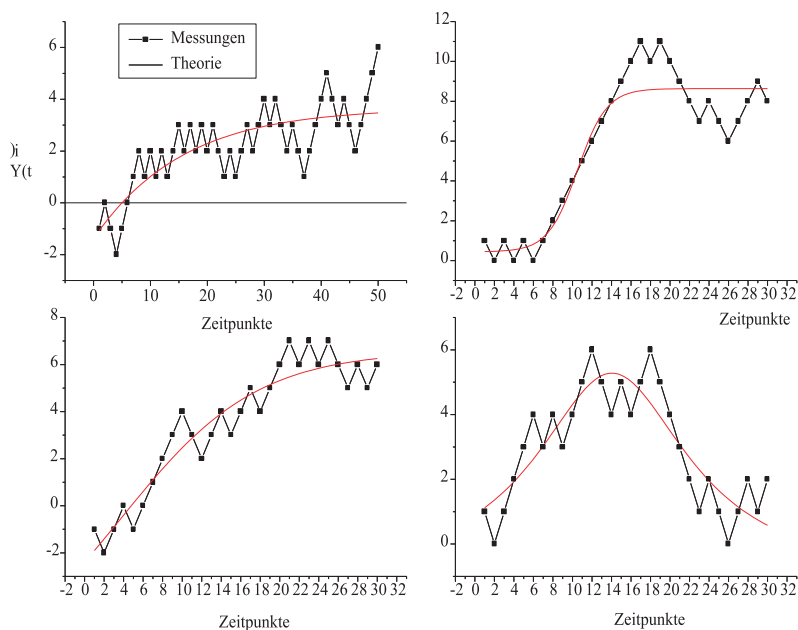
Die Begriffe der Mittelwertfunktion, der Kovarianz- und Varianzfunktion werden für die Zeitreihe  $\{Y(t_i)_{i=1,2,\dots}\}$  analog definiert. Statt von einer Mittelwertfunktion spricht man bei Zeitreihen auch von einem *Trend*. Spezielle Trends sind z.B. Wachstumskurven, oder saisonale Schwankungen bzw. saisonale Trends. Vielfach stehen insbesondere die Trends im Mittelpunkt des Interesses, etwa wenn im Rahmen einer Therapie das Gewicht einer Person reduziert werden soll. Bei zyklischen Erkrankungen werden angenähert periodische Mittelwertfunktionen oder Trends - also saisonale Schwankungen - von Interesse sein. Damit ist ein Hauptziel einer Zeitreihenanalyse, nämlich die Abschätzung eines

Trends, genannt. Die Möglichkeiten der Zeitreihenanalyse sind aber, zumindest im Prinzip, noch reichhaltiger: da  $Y$  als von der Zeit abhängige Größe betrachtet wird, ergibt sich die Möglichkeit, die der zeitlichen Variation von  $Y$  entsprechende *Dynamik* zu erfassen. Fragen nach der unterliegenden Dynamik werden insbesondere interessant, wenn mehr als nur eine Zeitreihe beobachtet wird.

Um zu einer solchen Abschätzung zu gelangen, muß die Zeitreihe in geeigneter Weise charakterisiert werden. Ist der stochastische Prozess stationär, so wird auch die Zeitreihe stationär sein. Dies bedeutet, dass der Trend durch eine Konstante beschrieben werden kann. Stationarität bedeutet natürlich noch mehr, nämlich dass die Kovarianzfunktion  $Kov(t, t + \Delta t)$  nur von  $\Delta t$ , nicht aber von  $t$  abhängt, so dass man  $Kov(t, t + \Delta t) = Kov(\Delta t)$  schreiben kann. Tatsächlich ist diese Bedingung neben dem Trend von Interesse: so könnte eine Therapie, in der adipöse Patienten lernen sollen, ihr Gewicht zumindest halten, besser noch zu reduzieren, sich so auswirken, dass  $\mu(t) = konstant$ , aber  $Kov(t, t + \Delta t)$  und  $\sigma(t)$  eine Funktion nicht nur von  $\Delta t$ , sondern auch von  $t$  ist. Das Gewicht der Therapierten fluktuiert dann in der Zeit, aber es kommt zu keiner systematischen Reduktion des Gewichts. Von einem Therapieerfolg wird man dann kaum sprechen können.

In Abb. 2.1 werden einige Beispiele gezeigt;  $Y(t)$  möge die Befindlichkeit eines Patienten bezeichnen. Die Interpretation der Daten könnte wie folgt lauten:

Abbildung 2.1: Entwicklung von Variablen in der Zeit



**Interpretation:** Die mit "Theorie" bezeichneten Kurven sind Mittelwertfunktionen, die aus den Daten auf Grund einer Theorie geschätzt wurden. In (a) scheint die Theorie die Daten gut zu repräsentieren, während in (b) die Theorie bis zur 15-ten Woche gut vorhersagt, dann aber den "Overshoot" und den darauf folgenden "Undershoot" nicht darstellen kann. In der Theorie der linearen Systeme sind der Over- und der Undershoot als *Einschwingvor-*

*gänge* bekannt; biologisch oder psychologisch lassen sie sich als Ausdruck der Interaktion zwischen Aktivierungs- und Inhibierungsprozessen deuten. Solche Prozesse könnten auch im Fall der Beobachtungsreihe (c) wirksam sein; vielleicht ist die Reihe nicht lang genug, um den zu erwartenden konstanten Wert der Mittelwertfunktion zu liefern. Möglicherweise kann die Therapie aber auch Oszillationen der Befindlichkeit erzeugen, denn wirklich konstant ist die Befindlichkeit ja bei kaum einem Menschen. In (d) schließlich ergibt sich nach einer mittleren Verbesserung des Zustandes wieder eine Verschlechterung nach dem fünfzehnten Zeitabschnitt; vielleicht sollte die behandelte Person die Therapie wechseln.

Die Interpretation mag direkt und vernünftig erscheinen; ob sie auch korrekt ist, wird in Abschnitt 2.2.3 diskutiert werden.

Man muß allgemein fragen, wie eine Zeitreihe repräsentiert werden kann. Dazu sind zunächst zwei Arten, sie als lineare Kombination bestimmter Werte darzustellen, bekannt: man kann versuchen, eine Zeitreihe entweder als *autoregressiven Prozess* (abgekürzt: AR-Prozess) oder als "gleitendes Mittel" (*moving average*, abgekürzt MA-Prozess, zu beschreiben. In einem nächsten Schritt kann man dann eine Repräsentation durch eine Kombination dieser beiden Typen versuchen; dies führt zu einer ARMA-Repräsentation. Bei einer Darstellung als ARMA-Prozess wird der Wert  $Y(t)$  immer noch als eine lineare Kombination vergangener Größen dargestellt. Erweist sich auch eine solche Darstellung als inadäquat, kann man zu einer *integrierten* Darstellung als ARIMA-Prozess übergehen. Die meisten in der Praxis vorkommenden Prozesse lassen sich als ARMA- bzw. ARIMA-Prozesse darstellen. Wie diese Repräsentationen definiert sind, wird im folgenden Abschnitt näher ausgeführt.

## 2.2 ARIMA-Prozesse

### 2.2.1 Autoregressive und Moving-Average Prozesse

#### Der autoregressive Prozess

**Definition 2.1** *Es sei  $\{Y(t_i)_{i=1,2,\dots}\}$  eine Zeitreihe und es seien  $\xi_i$  unabhängig und identisch verteilte (i.i.d.)<sup>1</sup> zufällige Veränderliche, mit  $E(\xi_i) = 0$  und  $Var(\xi) = \sigma^2$  für alle  $i$ . Gilt*

$$Y(t_i) = \phi_1 Y(t_{i-1}) + \phi_2 Y(t_{i-2}) + \dots + \phi_p Y(t_{i-p}) + \xi_i, \quad (2.5)$$

mit  $\phi_k \in \mathbb{R}$ ,  $k = 1, 2, \dots, p$ , so heißt die Zeitreihe ein autoregressiver Prozess der Ordnung  $p$  oder  $p$ -ter Ordnung; man schreibt  $AR(p)$  für diesen Prozess.

#### Anmerkungen:

1. Autoregressive Prozesse wurden zuerst von Yule (1927) bei der Diskussion von Sonnenflecken eingeführt.
2. Die Größen  $\phi_1, \phi_2, \dots, \phi_p$  sind die *Parameter* der Zeitreihe. Sie sind im allgemeinen unbekannt und müssen aus den Beobachtungen, d.h. aus den Messungen  $Y(t_i)$ , geschätzt werden.

---

<sup>1</sup>Die für "unabhängig und identisch verteilt" gebräuchliche Abkürzung i.i.d ergibt sich aus dem englischen "independent and identically distributed".



3. Für die  $\xi_i$  ist auch der Ausdruck *random shocks* gebräuchlich. Da die  $\xi_i$  unkorreliert sein sollen, wird die Folge  $\{\xi_i\}$  auch als "weißes Rauschen" bezeichnet<sup>2</sup>
4. Ist insbesondere  $p = 1$ , so ist der Prozess ein AR(1)-Prozess, d.h.  $Y(t_i)$  hängt nur von  $Y(t_{i-1})$  ab. Damit ist der AR(1)-Prozess ein Markov-Prozess.
5. Hat ein AR( $p$ )-Prozess größere Werte zu den vergangenen Zeitpunkten angenommen, so werden wahrscheinlich (abhängig von den Parametern  $\phi_i$ ) auch die zukünftigen Werte groß sein; eine analoge Aussage gilt für kleine Werte. Dies deutet darauf hin, dass ein AR( $p$ )-Prozess nicht notwendig stationär ist.

### Der Moving-Average Prozess

**Definition 2.2** *Es sei  $\{Y(t_i)_{i=1,2,\dots}\}$  eine Zeitreihe und es seien  $\xi_i$  unabhängig und identisch verteilte (i.i.d.) zufällige Veränderliche, mit  $E(\xi_i) = 0$  und  $Var(\xi) = \sigma^2$  für alle  $i$ . Es gelte*

$$Y(t_i) = \xi_i - \Theta_1 \xi_{i-1} - \dots - \Theta_q \xi_{i-q}, \quad (2.6)$$

mit  $\Theta_j \in \mathbb{R}$ ,  $j = 1, \dots, q$ , so heißt die Zeitreihe Moving Average Prozess der Ordnung  $q$  ( $q$ -ter Ordnung); man schreibt  $MA(q)$  für einen solchen Prozess.

#### Anmerkungen:

1. In der deutschsprachigen Literatur wird für den  $MA(q)$  auch der Ausdruck *Gleitmittelprozeß* gebraucht. Der Begriff des *gleitenden Mittels* besagt, dass bei einer Folge von Meßwerten über eine bestimmte Teilfolge benachbarter Werte gemittelt wird; die entstehenden Mittelwerte sind dann selbst eine Folge bzw. Zeitreihe. Der  $MA(q)$  ist eine solche Folge, wobei die vorangegangenen Werte (hier die Random Shocks) gewichtet gemittelt werden; die Gewichte sind die Parameter  $\Theta_1, \dots, \Theta_q$ . Natürlich kann auch der AR( $p$ )-Prozess als Folge gleitender Mittel aufgefaßt werden, und insofern ist die Bezeichnung Moving-Average Prozess oder Gleitmittelprozeß für den in (2.6) mißverständlich. Die Bezeichnung hat sich aber durchgesetzt.

Im Folgenden wird von Moving-Average Prozessen statt von Gleitmittelprozessen geredet werden, einmal wegen des Bezuges zur Bezeichnung  $MA(q)$ , zum anderen weil der Ausdruck "Gleitmittelprozeß" Assoziationen hervorrufen mag, die nicht zur Sache gehören.

2. dass die Gewichte  $\Theta_j$  mit negativem Vorzeichen eingehen hat lediglich formale Gründe, die die Diskussion von Prozessen erleichtern, die sowohl einen AR- wie auch einen MA-Anteil enthalten (s. weiter unten).
3. Bei einem  $MA(q)$ -Prozess werden gewissermaßen "von außen" kommende Einflüsse - die Random Shocks  $\xi_i$  - aufsummiert, um den Prozess  $\{Y(t_i)\}_{i=1,2,\dots}$  zu erzeugen. Der  $MA(q)$ -Prozess kann als Antwort eines durch die Parameter  $\Theta_1, \dots, \Theta_q$  definierten Systems betrachtet werden, das als Input die Folge  $\xi_i, \xi_{i-1}, \dots, \xi_{i-q}$  unabhängiger, gleichverteilter Random Shocks hat.

<sup>2</sup>In der Theorie stochastischer Prozesse heißt ein stationärer Prozess "weißes Rauschen", wenn die Autokorrelationsfunktion für alle  $t \neq 0$  gleich Null ist und nur für  $t = 0$  einen Wert ungleich Null annimmt. Es läßt sich zeigen, dass sich die Trajektorien eines solchen Prozesses als Überlagerung von Sinusschwingungen darstellen läßt, bei der alle Frequenzen mit gleicher Amplitude vorkommen. Eine solche Darstellung findet man auch für weißes Licht. Da der Prozess im akustischen Bereich als Rauschen wahrgenommen wird, ergibt sich der Ausdruck "weißes Rauschen".

## Statistiken

Die Mittelwertsfunktion bzw. der Trend der Zeitreihe  $\{Y(t_i)\}_{i=1,2,\dots}$  ergibt sich aus (2.5) gemäß

$$\mu(t_i) = E(Y(t_i)) = \phi_1 E(Y(t_{i-1})) + \dots + \phi_p E(Y(t_{i-p})), \quad (2.7)$$

da ja  $E(\xi_i) = 0$  für alle  $i$ . Dieser Ausdruck zeigt, wie sich  $\mu(t_i)$  aus den vergangenen Werten  $\mu(t_{i-1}), \mu(t_{i-2}), \dots, \mu(t_{i-p})$  zusammensetzt, nämlich als mit den Parametern  $\phi_1, \dots, \phi_p$  gewogene Summe dieser Erwartungswerte. Auf die Schätzung der  $\phi_i$  und  $\mu(t_i)$  wird weiter unten gesondert eingegangen.

Die Autokorrelationsfunktion des AR( $p$ )-Prozesses ergibt sich, indem man den in (2.5) definierten Ausdruck für  $Y(t_i)$  mit  $Y(t_{i+k})$  multipliziert und dann den Erwartungswert dieses Produkts bestimmt. Es ist

$$Y(t_i)Y(t_{i+k}) = \phi_1 Y(t_{i-1})Y(t_{i+k}) + \phi_2 Y(t_{i-2})Y(t_{i+k}) + \dots + \phi_p Y(t_{i-p})Y(t_{i+k}) + \xi_i Y(t_{i+k}).$$

Der Erwartungswert des Produkts  $Y(t_i)Y(t_{i+k})$  ist dann durch

$$\begin{aligned} E(Y(t_i)Y(t_{i+k})) &= \phi_1 E(Y(t_{i-1})Y(t_{i+k})) + \phi_2 E(Y(t_{i-2})Y(t_{i+k})) \\ &\quad + \dots + \phi_p E(Y(t_{i-p})Y(t_{i+k})) + E(\xi_i Y(t_{i+k})). \end{aligned} \quad (2.8)$$

gegeben. Hierin ist  $E(\xi_i Y(t_{i+k})) = 0$ , da ja  $\xi_i$  unabhängig von  $Y(t_{i+k})$  ist<sup>3</sup>. Setzt man nun

$$\begin{aligned} \rho_k &= E(Y(t_i)Y(t_{i+k})) \\ \rho_{k-1} &= E(Y(t_{i-1})Y(t_{i+k})) \\ &\vdots = \vdots \\ \rho_{k-p} &= E(Y(t_{i-p})Y(t_{i+k})) \end{aligned}$$

so ergibt sich

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (2.9)$$

Läßt man in dieser Gleichung den Index  $k$  die Werte 1, 2, 3  $\dots$  durchlaufen, so erhält man ein lineares Gleichungssystem, die

### Yule-Walker-Gleichungen:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_3 + \dots + \phi_p \rho_{p-2} \\ &\vdots = \vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \phi_3 \rho_{p-3} + \dots + \phi_p \end{aligned} \quad (2.10)$$

Setzt man für die Autokorrelationen  $\rho_1, \dots, \rho_p$  Schätzungen  $r_1, \dots, r_p$  ein, so erhält man ein lineares Gleichungssystem in den Unbekannten Parametern  $\phi_1, \dots, \phi_p$ ; die Lösung dieses Gleichungssystem in bezug auf die  $\phi_k, k = 1, 2, \dots$  liefert dann Schätzungen für die Parameter  $\phi_k$ .

**Beispiel 2.1** Gegeben sei der AR(1)-Prozess

$$Y(t_i) = \phi_1 Y(t_{i-1}) + \xi_i$$

<sup>3</sup>Ist aber auch  $Y(t_{i+k})$  unabhängig von  $\xi_i$ ?

Für die Autokorrelation ergibt sich

$$\rho_k = \phi_1 \rho_{k-1}, \quad k > 1. \quad (2.11)$$

Da  $\rho_0 = 1$  ergibt sich durch sukzessives Einsetzen, d.h.  $\rho_2 = \phi_1 \rho_1$ ,  $\rho_3 = \phi_1 \rho_2 = \phi_1^2 \rho_1$ , etc

$$\rho_k = \phi_1^k. \quad (2.12)$$

Dies heißt, dass für  $0 < \phi_1 < 1$  die Autokorrelation  $\rho_k$  exponentiell abfällt. Für  $-1 < \phi_1 < 0$  oszilliert  $\rho_k$  mit exponentiell kleiner werdender Amplitude, denn  $\phi_1^k < 0$  für ungerades  $k$ . Für  $|\phi_1| > 0$  wächst  $\rho_k$  exponentiell, wenn  $\phi_1 > 1$ , und oszilliert mit exponentiell wachsender Amplitude für  $\phi_1 < -1$ , d.h. mit größer werdendem zeitlichen Abstand der Meßwerte *wächst* die Autokorrelation des Prozesses. Dies ist physikalisch, biologisch oder psychologisch im allgemeinen nicht plausibel; man wird also eher  $|\phi_1| < 1$  erwarten.

Für die Varianz des Prozesses findet man

$$\text{Var}(Y(t_i)) = \frac{\sigma^2}{1 - \rho_1^2}. \quad (2.13)$$

□

**Beispiel 2.2** Gegeben sei ein AR(2)-Prozess

$$Y(t_i) = \phi_1 Y(t_{i-1}) + \phi_2 Y(t_{i-2}) + \xi_i$$

Die Autokorrelationen sind durch die Differenzgleichungen

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k > 0 \quad (2.14)$$

gegeben. Dabei ist  $\rho_0 = 1$ ,  $\rho_1 = \phi_1 / (1 - \phi_2)$ . Für die Varianz ergibt sich

$$\text{Var}(Y(t_i)) = \frac{\sigma^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2}, \quad \sigma^2 = \text{Var}(\xi) \quad (2.15)$$

Hat man Schätzungen für  $\rho_1$  und  $\rho_2$ , so ergeben die Yule-Walker-Gleichungen Schätzungen für die Parameter  $\phi_1$  und  $\phi_2$ :

$$\phi_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2} \quad (2.16)$$

$$\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (2.17)$$

□

### Die partielle Autokorrelationsfunktion (PCF)

Die Parameter  $\phi_k$  eines AR( $p$ )-Prozesses können als Regressionskoeffizienten einer multiplen, linearen Regression aufgefaßt werden. Die Ordnung  $p$  eines AR-Prozesses ist aber im allgemeinen nicht a priori bekannt. Im Prinzip könnte man so vorgehen, dass man den Wert von  $p$  so lange erhöht, bis die Werte von  $Y(t_{i-p})$  nicht mehr zur Vorhersage von  $Y(t_i)$  beitragen. Nun sind die Regressionsgewichte proportional zu den partiellen Korrelationskoeffizienten zwischen  $Y(t_i)$  und  $Y(t_{i-1})$ ; die restlichen Variablen  $Y(t_{i-2})$ ,  $Y(t_{i-3})$  etc sind auspartialisiert. Betrachtet man nun einen AR-Prozess der Ordnung  $k$  und bezeichnen die Parameter, d.h. die Regressionsgewichte, mit  $\phi_{kj}$ ,  $j = 1, \dots, k$ . Der erste Index  $k$  zeigt die Ordnung des Prozesses an, der zweite bezieht sich auf die Variable  $Y(t_{i-j})$ ; dies ist die Variable mit dem Lag  $j$ . Da es sich um einen AR( $k$ )-Prozess handelt, kann man  $\phi_{kj} \neq 0$  für  $j \leq k$  annehmen und  $\phi_{kj} = 0$  für  $j > 0$ . Dies bedeutet:

**Satz 2.1** Die partiellen Autokorrelationen für einen AR( $p$ )-Prozess sind gleich Null für  $k > p$ .

Die partiellen Autokorrelationen können bestimmt werden, indem man sukzessive AR-Prozesse der Ordnung  $p = 1, 2, \dots$  mittels der Methode der Kleinsten Quadrate schätzt und jeweils die Schätzungen  $\hat{\phi}_{11}, \hat{\phi}_{22}, \hat{\phi}_{33}, \dots$  betrachtet. Man kann sie über die Yule-Walker-Gleichungen schätzen, indem man die Stichprobenschätzungen  $r_j$  für die  $\rho_j$  einsetzt; dies liefert

$$r_j = \hat{\phi}_{k1}r_{j-1} + \hat{\phi}_{k2}r_{j-2} + \dots + \hat{\phi}_{k,k-1}r_{j-k+1} + \hat{\phi}_{kk}r_{j-k} \quad (2.18)$$

für  $j = 1, 2, \dots, k$ . Diese Gleichung kann man nach den  $\hat{\phi}_{kj}$  auflösen.

Die  $\hat{\phi}_{kj}$  sind natürlich zufällige Veränderliche, da sie ja Schätzungen aufgrund von Stichprobenwerten der Zeitreihe sind. Da die Entscheidung über die Ordnung eines AR-Prozesses von den  $\hat{\phi}_{kk}$  abhängt, ist die Varianz der Schätzung dieser Größe von Interesse. Man findet (Box und Jenkins (1976), p. 65)

$$\text{Var}(\hat{\phi}_{kk}) \approx 1/n, \quad k \geq p + 1 \quad (2.19)$$

Der Standardfehler ist damit

$$\sqrt{\text{Var}(\hat{\phi}_{kk})} \approx 1/\sqrt{n} \quad (2.20)$$

Eindeutige Entscheidungen über die Ordnung eines AR-Prozesses sind im allgemeinen schwierig.

### Die Autokorrelationen des MA( $q$ )

Ausgehend von der allgemeinen Formel

$$\rho_k = E(Y(t_i)Y(t_{i-k}))$$

findet man

$$\rho_k = \begin{cases} (-\Theta_k + \Theta_1\Theta_{k+1} + \dots + \Theta_{q-k}\Theta_q)/(1 + \Theta_1^2 + \dots + \Theta_q^2), & 1 \leq k \leq q \\ 0, & k > q \end{cases} \quad (2.21)$$

Für die Varianz erhält man

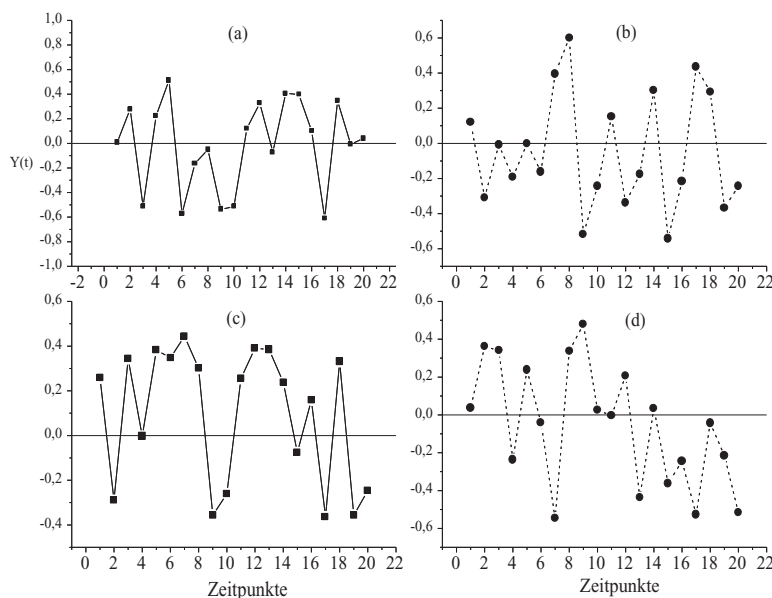
$$\rho_0 = (1 + \Theta_1^2 + \Theta_2^2 + \dots + \Theta_q^2)\sigma^2, \quad \sigma^2 = \text{Var}(\xi) \quad (2.22)$$

Für  $k > q$  sind die Autokorrelationen gleich Null. Die  $\Theta_j$  lassen sich aus den Gleichungen (2.21) schätzen, wenn man bereits Schätzungen  $r_k$  für die  $\rho_k$  hat. Unangenehmerweise sind die Gleichungen (2.21) nicht linear in den  $\Theta_j$ ; die Schätzungen  $\hat{\theta}_j$  haben eine geringe statistische Effizienz. In Abb. 2.2 werden Trajektorien von AR(2)- und MA(2)-Prozessen gezeigt. Die Parameter sind für alle Trajektorien gleich, nämlich  $T_1 = .15$  und  $T_2 = -.25$ .

### Die Beziehung zwischen AR( $p$ )- und MA( $q$ )-Prozessen

AR( $p$ )- und MA( $q$ )-Prozesse scheinen auf den ersten Blick völlig verschieden zu sein, wenn man davon absieht, dass sie analog zur multiplen Regression als lineare Kombination von "Prädiktoren"  $Y(t_{i-1}), \dots, Y(t_{i-p})$  bzw.  $\xi_i, \dots, \xi_{i-q}$  definiert sind. Tatsächlich gibt

Abbildung 2.2: Verschiedene Trajektorien von AR(2)- ((a) und (b)) und MA(2)-Prozessen ((c) und (d)) mit jeweils gleichen Parametern



es aber einen Zusammenhang zwischen ihnen, der für die Diskussion der Statistik solcher Prozesse von Bedeutung ist.

Zunächst soll der Begriff des "Operators" eingeführt werden. Ein Operator ist eine Regel, nach der von einem Objekt zu einem anderen übergegangen wird. Es sei  $B$  ein Operator, und es gelte

$$BY(t_i) = Y(t_{i-1}) \tag{2.23}$$

$B$  definiert dann die Anweisung, von dem "Objekt"  $Y(t_i)$  zu dem "Objekt"  $Y(t_{i-1})$  überzugehen. Der spezielle, in (2.23) eingeführte Operator erweist sich als sehr nützlich für die Diskussion des Zusammenhanges zwischen AR( $p$ )- und MA( $q$ )-Prozessen, weshalb er durch eine eigene Definition geehrt wird:

**Definition 2.3** *Der durch die Beziehung (2.23) eingeführte Operator heißt Backwardshift Operator. Der durch die Beziehung  $FY(t_i) = B^{-1}Y(t_i) = Y(t_{i+1})$  erklärte Operator  $F$  heißt Forwardshift Operator. Der durch  $\nabla Y(t_i) = Y(t_i) - Y(t_{i-1})$  definierte Operator  $\nabla$  heißt Backward Difference Operator.*

Insbesondere kann der Operator  $\nabla$  durch den Operator  $B$  ausgedrückt werden, denn es gilt

$$\nabla Y(t_i) = Y(t_i) - Y(t_{i-1}) = Y(t_i) - BY(t_i) = (1 - B)Y(t_i) \tag{2.24}$$

Der Operator  $B$  ist also die Anweisung, in der Zeitreihe einen Wert zurückzugehen, während  $F$  die Anweisung ist, einen Wert nach vorn zu gehen. Die Schreibweise  $F = B^{-1}$  zeigt, dass  $F$  gewissermaßen das inverse Vorgehen der durch  $B$  erklärten Regel darstellt. Es zeigt sich, dass diese Bezeichnung mehr als nur eine bei Zahlen übliche Schreibweise ist: sie erlaubt, mit Operationen wie mit Zahlen zu "rechnen". So bedeutet etwa  $B^p$ , dass der Operator  $B$   $p$ -mal nacheinander ausgeführt werden soll. Dies führt zu der folgenden

**Definition 2.4** *Der durch*

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.25)$$

*erklärte Operator heißt autoregressiver Operator.*

Die  $\phi_j$  sind die Parameter (Gewichte), die bei der Definition des AR( $p$ )-Prozesses eingeführt wurden. In der Tat kann der AR( $p$ )-Prozess durch den Operator  $\phi(B)$  definiert werden: es ist

$$\phi(B)Y(t_i) = Y(t_i) - \phi_1 Y(t_{i-1}) - \dots - \phi_p Y(t_{i-p}).$$

Der Vergleich mit (2.5) zeigt dann, dass also gerade die Beziehung

$$\phi(B)Y(t_i) = \xi_i \quad (2.26)$$

gilt. Für den MA( $q$ )-Prozess erhält man auf analoge Weise die Darstellung

$$\Theta(q) = 1 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_q B^q \quad (2.27)$$

und damit

$$Y(t_i) = \Theta(B)\xi_i \quad (2.28)$$

AR( $p$ )- und MA( $q$ )-Prozesse sind ineinander überführbar. So kann man im AR( $p$ )-Prozess für  $Y(t_i)$  den Ausdruck für  $Y(t_{i-1})$  den gemäß Definition 2.1 zu gebildeten Ausdruck

$$Y(t_{i-1}) = \phi_1 Y(t_{i-2}) + \dots + \phi_p Y(t_{i-p-1}) + \xi_{i-1}$$

anschreiben und ihn für  $Y(t_{i-1})$  in (2.5) einsetzen. Für  $Y(t_{i-2})$ ,  $Y(t_{i-3})$ , etc verfährt man ebenso. Man erhält damit einen Moving-Average-Prozess MA( $q$ ) mit  $q = \infty$ , d.h. einen MA-Prozess unendlicher Ordnung. Wendet man auf (2.26) den Operator  $\phi^{-1}$  an, so ergibt sich

$$\phi_{-1}(B)\phi(B)Y(t_i) = Y(t_i) = \phi^{-1}(B)\xi_i$$

wobei von der bei Zahlen bekannten Beziehung  $\phi^{-1}\phi = 1$  Gebrauch gemacht wurde.  $\phi^{-1}(B)$  definiert also einen MA-Prozess der Ordnung  $\infty$ . Ausgeschrieben erhält man

$$Y(t_i) = \xi + \sum_{j=1}^{\infty} \Gamma_j \xi_j. \quad (2.29)$$

**Definition 2.5** *Der durch (2.29) definierte Prozess heißt allgemeiner linearer Prozess.*

Der allgemeine lineare Prozess ist also ein autoregressiver Prozess der Ordnung  $p$ , und für  $\Gamma_j = 0$  für alle  $j > q$  entspricht er einem Moving-Average Prozess MA( $q$ ).

Unter Verwendung des Backshift-Operators läßt sich der allgemeine lineare Prozess nun in der Form

$$Y(t_i) = \left( 1 + \sum_{j=1}^{\infty} (\Gamma_j B^j) \right) \xi_i = \Gamma(B)\xi_i \quad (2.30)$$

schreiben. Unter bestimmten - noch näher zu kennzeichnenden - Bedingungen gilt nun gleichzeitig

$$Y(t_i) = \sum_{j=1}^{\infty} \pi_j Y(t_{i-j}) + \xi_i, \quad (2.31)$$

d.h. die Zeitreihe ist als AR-Prozess unendlicher Ordnung darstellbar. Über den Backshift-Operator läßt sich die Gleichung (2.31) in der Form

$$Y(t_i) \left( 1 - \sum_{j=1}^{\infty} \pi_j B^j \right) = \xi_i = \pi(B)Y(t_i) \quad (2.32)$$

schreiben. Multipliziert man beide Seiten mit  $\Gamma(B)$ , so erhält man  $\Gamma(B)\pi(B)Y(t_i) = \Gamma(B)\xi_i = Y(t_i)$ , d.h. aber

$$\Gamma(B)\pi(B) = 1. \quad (2.33)$$

Daraus folgt dann

$$\pi(B) = \Gamma^{-1}(B) \quad (2.34)$$

Diese Gleichung liefert die Gewichte  $\pi_j$ , wenn man die  $\Gamma_j$  kennt, und umgekehrt die Gewichte  $\Gamma_j$ , wenn man die Gewichte  $\pi_j$  kennt.

Bekanntlich gilt

$$S_n = 1 + a^2 + a^3 + \dots + a^n = \begin{cases} (1 + a^n)/(1 - a), & n < \infty \\ 1/(1 - a), & |a| < 1, \quad n \rightarrow \infty \end{cases} \quad (2.35)$$

Betrachtet man nun die Reihe  $1 + \Theta B + \theta^2 B^2 + \dots$ , so kann man  $B$  als "Dummy", d.h. als Pseudovariablen betrachten; für  $|\Theta| < 1$  erhält man aus (2.35) dann

$$1 + \Theta B + \theta^2 B^2 + \dots = \frac{1}{1 - \Theta B} \quad (2.36)$$

Rechnet man in dieser Weise mit dem Operator, so bleibt alles richtig.

**Beispiel 2.3** Die Zeitreihe  $\{Y(t_i)\}_{i=1,2,\dots}$  sei durch einen MA(1)-Prozess definiert, d.h. es soll

$$Y(t_i) = \Theta_1 \xi_{i-1} + \xi_i \quad (2.37)$$

gelten. Dann gilt

$$Y(t_i) = \xi_i - \Theta_1 B \xi_i = (1 - \Theta_1 B) \xi_i.$$

In bezug auf die Repräsentation (2.29) heißt dies, dass

$$\Gamma_i = \begin{cases} -\Theta_1, & i < 2 \\ 0, & i \geq 2 \end{cases}$$

Die Anwendung von (2.36) liefert dann

$$\frac{1}{1 - B\Theta} Y(t_i) = \xi_i = Y(t_i) + \Theta Y(t_{i-1}) + \Theta^2 Y(t_{i-2}) + \dots \quad (2.38)$$

und damit

$$Y(t_i) = \xi - \Theta Y(t_{i-1}) - \Theta^2 Y(t_{i-2}) - \dots \quad (2.39)$$

Damit ist die ursprünglich als MA(1)-Prozess definierte Zeitreihe als autoregressiver Prozess unendlicher Ordnung beschrieben. Der Vergleich von (2.31) mit dieser letzten Gleichung liefert dann

$$\pi_j = -\Theta^j,$$

d.h. die  $\pi_j$  sind alle aus einem Parameter  $\Theta$  vorhersagbar. Der Informationswert der  $\pi_j$  ist bereits in  $\pi_1 = -\Theta$  enthalten. Dies bedeutet, dass zwar die Darstellung der Zeitreihe als MA(1)-Prozess beliebig ist, - denn sie ist ja auch als AR( $\infty$ )-Prozess darstellbar, aber die MA(1)-Darstellung hat den Vorteil größerer Ökonomie.  $\square$

Betrachtet man die Erwartungswerte der durch (2.37) definierten Zeitreihe, d.h. des MA(1)-Prozesses, so sieht man, dass

$$E(Y(t_i)) = E(\Theta_1 \xi_{i-1} + \xi_i) = \Theta_1 E(\xi_i) + E(\xi_i) = 0,$$

denn  $E(\xi) = 0$  nach Voraussetzung. Weiter folgt

$$\text{Var}(Y(t_i)) = \text{Var}(\xi) + \Theta^2 \text{Var}(\xi) = (1 + \Theta^2)\sigma^2 = \text{konstant}$$

Für die Autokorrelation findet man

$$\rho_{i,i-k} = \frac{E((Y(t_i) - \mu(t_i))(Y(t_{i-k}) - \mu(t_{i-k})))}{(1 + \Theta^2)\sigma^2} = \frac{E(Y(t_i)Y(t_{i-k}))}{(1 + \Theta^2)\sigma^2}$$

denn  $\mu(t_i) = 0$  für alle  $i$ , damit auch für  $i - k$ .

Wählt man  $|\Theta| > 1$ , so werden in der AR-Darstellung des MA(1)-Prozesses die Gewichte  $\Theta^j$  immer größer, was bedeutet, dass der Einfluß von vergangenen Werten der Zeitreihe um so größer ist, je weiter sie zurückliegen. Physikalische, biologische und psychologische Gründe legen aber nahe, dass Zeitreihenwerte im allgemeinen einen um so geringeren Einfluß haben, je weiter sie zurückliegen. Deshalb wird man  $|\Theta| < 1$  fordern, was dazu führen kann, dass die Charakterisierung eines Prozesses als MA(1)-Prozesses nicht aufrecht zu erhalten ist. Für  $|\Theta| < 1$  hat man einen *invertierbaren Prozess*. Allgemein hat man die folgende

**Definition 2.6** Gilt  $\pi_j \rightarrow 0$  für  $j \rightarrow \infty$ , so heißt die Zeitreihe  $Y(t_i)_{i=1,2,\dots}$  invertierbar.

Es gilt nun der

**Satz 2.2** Es sei  $Y_t = Y(t_i)_{i=1,2,\dots}$  eine Zeitreihe, die als linearer Prozess

$$Y(t_i) = \sum_{j=1}^{\infty} \Gamma_j \xi_{i-j} + \xi_i$$

(vergl. (2.29)) darstellbar sei. Die Zeitreihe ist stationär, wenn  $\Gamma(B)$  konvergiert für  $|B| \leq 1$ .

**Beweis:** Box und Jenkins (1976), p. 49 □

## 2.2.2 ARMA-Prozesse

Eine allgemeine Klasse von Zeitreihen erhält man, wenn man AR- und MA-Prozesse kombiniert:

**Definition 2.7** Es sei  $\{Y(t_i)\}_{i=1,2,\dots}$  eine Zeitreihe und es seien  $\xi_i$  wieder i.i.d. zufällige Veränderliche mit  $E(\xi_i) = 0$ ,  $\text{Var}(\xi_i) = \sigma^2$  für alle  $i$ . Gilt

$$Y(t_i) = \phi_1 Y(t_{i-1}) + \dots + \phi_p Y(t_{i-p}) + \xi_i - \Theta_1 \xi_{i-1} - \Theta_2 \xi_{i-2} - \dots - \Theta_q \xi_{i-q}, \quad (2.40)$$

so heißt die Zeitreihe gemischter Autoregressiv-Moving Average Prozess; man schreibt auch kurz ARMA( $p, q$ )-Prozess.



Ausdrücke für die Autokorrelationen eines ARMA( $p, q$ )-Prozesses werden nach dem gleichen Schema wie in den vorangegangenen Fällen hergeleitet: man multipliziert den Ausdruck für  $Y(t_i)$  mit  $Y(t_{i-k})$  für  $k = 1, 2, \dots$  und bildet die Erwartungswerte. Es ergibt sich

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p} + \rho_{xy}(k) - \Theta_1 \rho_{xy}(k-1) - \dots - \Theta_q \rho_{xy}(k-q), \quad (2.41)$$

wobei  $\rho_{xy}(k) = EY(t_{i-k}\xi_i)$  ist. Für  $\rho_{xy}(k-1)$  etc. ist die Interpretation entsprechend.  $Y(t_{i-k})$  hängt nur von den Random Shocks  $\xi$  ab, die zur Zeit  $t-k$  aufgetreten sind, so dass

$$\begin{aligned} \rho_{xy}(k) &= 0, & k > 0 \\ \rho_{xy}(k) &\neq 0, & k \leq 0 \end{aligned}$$

Deshalb gilt

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, \quad k \geq q+1 \quad (2.42)$$

Für einen ARMA( $p, q$ )-Prozess gibt es also  $q$  Autokorrelationen  $\rho_q, \rho_{q-1}, \dots, \rho_1$ , die sowohl von den Parametern  $\Theta_1, \dots, \Theta_q$  wie von den Parametern  $\phi_1, \dots, \phi_p$  abhängen. Ist  $q-p < 0$ , so besteht die gesamte Autokorrelationsfunktion aus einer Mischung gedämpfter Exponentialterme und/oder gedämpfter Sinusterme. Gilt  $q-p \geq 0$ , so folgen die erste Werte  $\rho_0, \rho_1, \dots, \rho_{q-p}$  nicht diesem Muster.

Die partielle Autokorrelation bricht nicht nach endlich vielen Gliedern ab; von einem gewissen Lag an verhält sie sich wie die partielle Autokorrelationsfunktion eines MA( $q$ )-Prozesses.

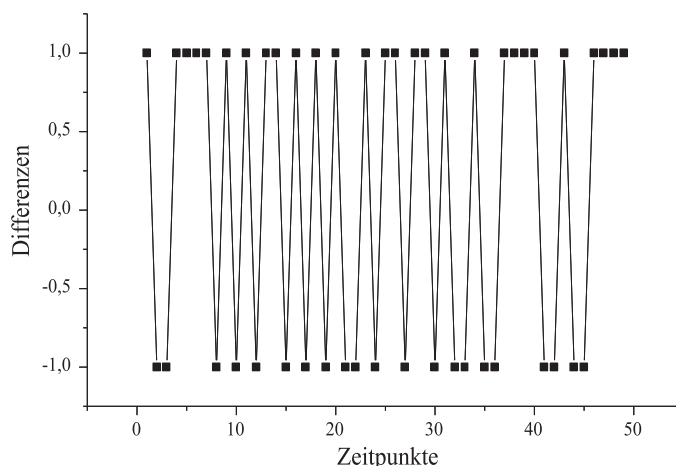
## 2.2.3 Integrierte Prozesse

### Der Random Walk

Ein ARMA-Prozess ist stationär, wenn die Wurzeln (d.h. die Werte der Dummy-Variablen  $B$ ) größer als 1 sind bzw. außerhalb des Einheitskreises liegen. Der Prozess ist "explosiv-nichtstationär", wenn die Wurzeln innerhalb des Einheitskreises liegen (Beispiel: Bakterienwachstum). Für den Fall, dass sie auf dem Einheitskreis liegen, spricht man von "homogener Nichtstationarität".

Nichtstationarität kann sich auf verschiedene Weise äußern: so kann die Mittelwertfunktion nicht konstant sein, und/oder die Varianz/Kovarianzfunktion kann eine Funktion der Zeit sein. Beispiele für Verläufe, die Realisierungen nichtstationärer Prozesse sind, sind in Fig. 2.1 abgebildet worden; eine Interpretation wurde gleich mitgeliefert. Man betrachte nun die Differenzen  $\Delta Y(t) := Y(t) - Y(t-1)$  der in Abb. 2.1, gezeigten Realisierungen des Prozesses; sie werden in Fig. 2.3 für die in Abb. 2.1, (a) gegebene Trajektorie gezeigt. Es fällt auf, dass die Differenzen entweder den Wert +1 oder -1 annehmen, und dies wahrscheinlich in zufälliger Folge. Die Mittelwertfunktion dieses Prozesses scheint konstant, insbesondere gleich 0 zu sein. Die Varianz des Prozesses ist offenbar konstant, und wegen der vermuteten Zufälligkeit des Auftretens der Werte +1 und -1 liegt es nahe, zu vermuten, dass auch die Kovarianzfunktion nicht von  $t$  abhängt.  $\Delta Y(t)$  ist also aller Wahrscheinlichkeit nach ein stationärer Prozess. Nehmen wir an, dass diese Vermutung korrekt ist; die Frage ist dann, was daraus über den ursprünglichen Prozess  $Y(t)$  gefolgert werden kann. Um dies zu diskutieren, wird zunächst die folgende Definition eingeführt:

Abbildung 2.3: Differenzen  $Y(t) - Y(t - 1)$



**Definition 2.8** Es sei  $\{Y(t_i)_{i=1,2,\dots}\}$  eine Zeitreihe für die die Aussage

$$Y(t_i) = Y(t_{i-1}) + \xi_i, \quad i = 1, 2, \dots \quad (2.43)$$

wobei  $\xi_i$  eine i.i.d. zufällige Veränderliche ist mit

$$p(\xi_i) = \begin{cases} p, & \xi_i = 1 \\ 1 - p, & \xi_i = -1 \end{cases} \quad (2.44)$$

Dann heißt der Prozess ein Random Walk mit der Drift  $p$ .

Nach Gleichung (2.43) ist der Random Walk ein autoregressiver Prozess mit  $\phi_1 = 1$ . Die  $\xi_i$  bilden wegen der Forderung, i.i.d zu sein, ein "weißes Rauschen" ab, wobei in Gleichung (2.44) allerdings noch festgelegt wird, dass  $\xi_i$  nur die Werte +1 oder -1 annehmen kann. Es sei aber angemerkt, dass der Random Walk nicht nur wegen  $\phi_1 = 1$  ein spezieller AR-Prozess ist: die  $\xi_i$  können beim Random Walk nur die Werte +1 oder -1 annehmen. Beim allgemeinen AR-Prozess ist die Einschränkung nicht gegeben; dort wird angenommen, dass die  $\xi_i$   $N(0, 1)$ -verteilte Variablen sind, d.h. dass die  $\xi_i$  normalverteilt mit dem Erwartungswert 0 und der Varianz 1 sind. Es folgt sofort

$$E(\xi_i) = 1 \cdot p + (-1) \cdot (1 - p) = 2p - 1,$$

d.h. die Mittelwertsfunktion des "Rauschens" hat den konstanten Wert  $2p - 1$ . Für den Erwartungswert von  $Y(t_i)$  erhält man

$$E(Y(t_i)) = E(Y(t_{i-1})) + E(\xi_i) = E(Y(t_{i-1})) + 2p - 1 \quad (2.45)$$

Nun ist

$$\begin{aligned} E(Y(t_2)) &= E(Y(t_1)) + 2p - 1 \\ E(Y(t_3)) &= E(Y(t_2)) + 2p - 1 = E(Y(t_1)) + 2(2p - 1) \\ E(Y(t_4)) &= E(Y(t_3)) + 2p - 1 = E(Y(t_1)) + 3(2p - 1) \\ &\text{etc} \end{aligned}$$

d.h.

$$E(Y(t_i)) = i(2p - 1), \quad E(Y(t_1)) = 2p - 1 \quad (2.46)$$

wobei angenommen wurde, dass  $Y(t_1) = \xi_1$  und damit  $EY(t_1) = 2p - 1$  ist. Insbesondere folgt

$$E(Y(t_i)) = 0 \text{ für alle } i, \text{ wenn } p = 1/2 \quad (2.47)$$

Für den Fall  $p = 1/2$  ist die Mittelwertfunktion des Prozesses also konstant. Für  $p \neq 1/2$  wächst oder fällt die Mittelwertfunktion linear mit  $i$ , je nachdem ob  $p < 1/2$  oder  $p > 1/2$  gilt.

Eine alternative Ableitung des Erwartungswertes ergibt sich aus der Tatsache, dass ja

$$Y(t_i) = \xi_i + \xi_{i-1} + \dots + \xi_1$$

und demnach

$$E(Y(t_i)) = E(\xi_i) + \dots + E(\xi_1) = i(2p - 1). \quad (2.48)$$

Für die Varianz ergibt sich sofort

$$Var(Y(t_i)) = \sum_{k=1}^i Var(\xi_k) = i \quad (2.49)$$

d.h. die Varianz des Prozesses wächst linear mit  $i$ !

Es werde noch die Autokorrelationsfunktion des Prozesses betrachtet. Da der Prozess autoregressiv ist, ist sie bereits allgemein bekannt; für den Spezialfall  $p = 1$  folgt nach (2.11) insbesondere

$$\rho_k = \phi_1 \rho_{k-1} = \phi_1^k = 1 \text{ für alle } k \quad (2.50)$$

denn es ist ja  $\phi_1 = 1$ .

Wegen des Anwachsens der Varianz des Prozesses mit  $i$  ist also der Random Walk in jedem Falle ein nichtstationärer Prozess.

Es werden nun die Differenzen  $\Delta y(t_i) = Y(t_i) - Y(t_{i-1})$  des Prozesses betrachtet. Es ist

$$\Delta Y(t_i) = Y(t_i) - Y(t_{i-1}) = Y(t_{i-1}) + \xi_i - Y(t_{i-1}) = \xi_i$$

d.h.

$$\Delta Y(t_i) = \begin{cases} +1, & \text{mit Wahrscheinl. } p \\ -1, & \text{mit Wahrscheinl. } 1 - p \end{cases} \quad (2.51)$$

Der Erwartungswert und die Varianz der Differenzen ist

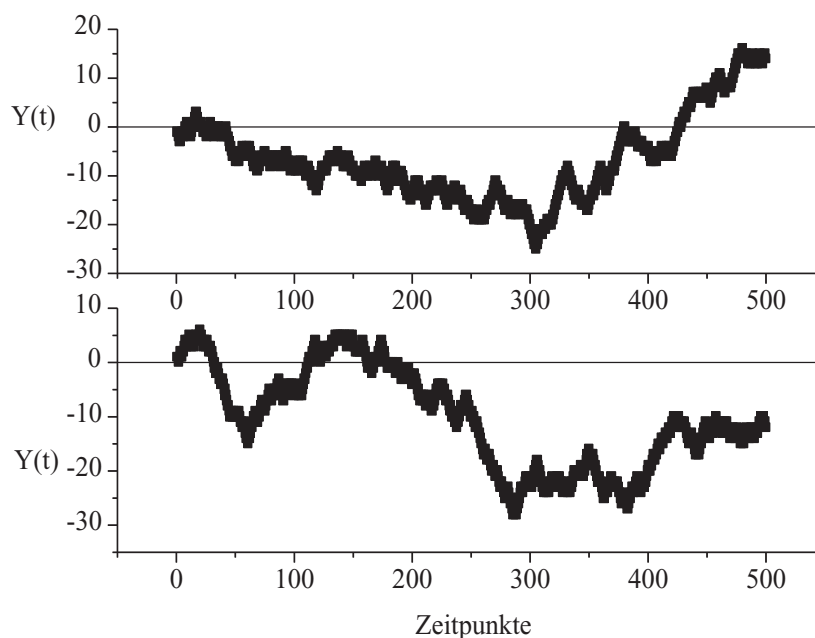
$$E(\Delta y(t_i)) = E(\xi_i) = 2p - 1, \quad Var(\Delta Y(t_i)) = 1 \quad (2.52)$$

Für die Kovarianz erhält man

$$Kov(\Delta(Y(t_i)), \Delta Y(t_{i-k})) = Kov(\xi_i, \xi_{i-k}) = 0, \quad (2.53)$$

denn die  $\xi_i$  sind ja unkorreliert. Es folgt, dass die Differenzen einen stationären Prozess bilden. Kehren wir noch einmal zu Abbildung 2.1 zurück. Die dort gezeigten Trajektorien zufälliger Prozesse sind allesamt Trajektorien eines Random Walks mit dem Parameter  $p = 1/2$ . Das Auge läßt sich täuschen: man sieht eine Struktur, die nicht vorhanden ist. Die Mittelwertfunktionen sind zwar in der Tat aus den Daten errechnet worden, gelten aber nur für die beobachteten Werte und sind *deswegen* nicht konstant. Die wahre Mittelwertfunktion ist konstant gleich Null! Offenbar führt eine atheoretische Betrachtung

Abbildung 2.4: Random Walks,  $p = 1/2$



der Daten zu einer groben Fehlinterpretation. Erst die Analyse der Differenzen, der Autokorrelationen etc. führt zu der Einsicht, dass hier keine durchschnittliche Entwicklung zu entnehmen ist, die irgendeine psychologische Bedeutung haben könnte. Psychologisch interessant kann natürlich die Aussage sein, dass es sich bei dem Prozess um einen AR-Prozess erster Ordnung mit dem Regressionsgesicht 1 handelt, aber diese Feststellung ist von anderer Natur als die in der Psychologie so häufig anzutreffende Betrachtung von Mittelwerten.

Die Eigenschaften eines Random Walks werden noch einmal in der Abb. 2.4 verdeutlicht. Die Trajektorien folgen der gleichen mathematischen Gesetzmäßigkeit wie die in Abb. 2.1, nur werden hier 500 Meßpunkte betrachtet. Die beiden Trajektorien zeigen, wie lange, d.h. wieviele Meßpunkte lang, eine Trajektorie unterhalb (oder oberhalb) des Erwartungswerts 0 bleiben kann. Es wird auch deutlich, dass die Abweichungen von 0 immer größer werden. Die Mittelwertfunktion hat zwar den Wert  $E(Y) = 0$ , kann aber, wenn man sie aus wenigen Daten schätzt ohne etwas über die Struktur des Prozesses zu kennen, einen völlig falschen Eindruck vermitteln, wie in Abb. 2.1 gezeigt wurde. Dies macht noch einmal deutlich, wie notwendig es ist, die *Struktur* des Prozesses zu analysieren, bevor man sich auf vorschnelle Deutungen der Daten einläßt.

Ein weiterer Punkt sollte angemerkt werden: es ist in den Abbildungen 2.1 und 2.4 nicht ganz klargestellt worden, auf welcher Zeitskala der Prozess läuft. In Abb. 2.1 ist unterstellt worden, dass es sich bei der Einteilung auf der  $t$ -Achse um Tage oder Wochen handelt. Ob in der Realität ein therapeutischer Prozess bei dieser Zeitskala durch einen Random Walk beschrieben werden kann, sei dahingestellt; Zweck der Betrachtung war ja nur, zu demonstrieren, dass die Mittelwertfunktion zu völlig falschen Schlußfolgerungen führen kann, wenn sie anhand einer Stichprobe von Meßwerten errechnet wird. Wichtiger

noch ist der Hinweis auf eine andere Eigenschaft. Die Trajektorien in Abb. 2.4 scheinen "dicker" zu sein als die in Abb. 2.1. Tatsache ist, dass die *Makrostruktur* in Abb. 2.4 die gleiche ist wie die *Mikrostruktur* in Abb. 2.1; letztere Abbildung zeigt ja einen Teil der Trajektorien von Abbn. 2.4. Betrachtet man einen solchen Prozess für eine gewisse Zeitspanne  $(0, T)$ , so ergibt sich für eine bestimmte Unterteilung in  $n$  Schritte das gleiche Bild (d.h. die gleiche stochastische Struktur) wie für das Intervall  $(0, T/2)$  mit  $2n$  Schritten. Unterscheiden sich die Schritte nur um ein beliebig kleines Zeitintervall  $dt$ , so ergibt sich eine beliebig zerknitterte Trajektorie.

### Der ARIMA-Prozess

Der Random Walk ist nichtstationär. Die Differenzen  $\Delta Y(t_i) = Y(t_i) - Y(t_{i-1})$  bilden aber einen stationären Prozess. Man kann nun einen beliebigen, aus AR- und MA-Komponenten zusammengesetzten Prozess betrachten, der nicht stationär ist. Man kann nun wieder die Differenzen  $\Delta Y(t_i)$  betrachten und nachsehen (wie, wird weiter unten behandelt), ob sie eine stationären Prozess bilden. Wenn nicht, kann man den  $\Delta Y$ -Prozess noch einmal "differenzieren", d.h. die Differenzen  $\Delta^2 Y(t_i) = \Delta Y(t_i) - \Delta Y(t_{i-1})$  bilden, und testen, ob dieser Prozess stationär ist. Wenn nicht, fährt man weiter fort und bildet den Prozess  $\Delta^3 Y(t_i)$ , etc. Bevor aber auf die Bedeutung dieser Differenzenbildung weiter eingegangen wird, soll noch eine formale Betrachtung eingeschoben werden, weil sie Hilfsmittel liefert, die die Diskussion vereinfacht.

Dazu werde der Einfachheit halber noch einmal der Random Walk betrachtet: es sei also  $y(t_i) = Y(t_{i-1}) + \xi$ . Dann ist also

$$Y(t_i) - Y(t_{i-1}) = \xi_i$$

Bei Verwendung des Backshift-Operators kann man aber

$$Y(t_{i-1}) = BY(t_i)$$

schreiben. Dann ist

$$\Delta Y(t_i) = Y(t_i) - Y(t_{i-1}) = Y(t_i) - BY(t_i) + \xi_i = (1 - B)Y(t_i) + \xi_i$$

$\Delta Y(t_i) = (1 - B)Y(t_i)$  ist stationär. Nun seien die  $Y(t_i)$  Realisierungen eines beliebigen Prozesses (d.h. einer beliebigen *Zeitreihe*). Angenommen, die Folge der  $\Delta Y(t_i)$  sei noch *nicht* stationär; dann kann man also die Folge der zweiten Differenzen

$$\Delta^2 Y(t_i) = \Delta Y(t_i) - \Delta Y(t_{i-1})$$

bilden und wiederum nachsehen, ob sie stationär ist. Auf der rechten Seite kann man aber nun

$$\Delta Y(t_i) = (1 - B)Y(t_i), \quad \Delta Y(t_{i-1}) = (1 - B)Y(t_{i-1})$$

schreiben; aus der Definition von  $1 - B$  folgt ja, dass sie nicht nur für den Random Walk gilt. Also läßt sich für  $\Delta^2 Y(t_i)$  auch

$$\Delta^2 Y(t_i) = (1 - B)Y(t_i) - (1 - B)Y(t_{i-1}) = (1 - B)(Y(t_i) - Y(t_{i-1})) = (1 - B)^2 Y(t_i)$$

schreiben. Damit hat man eine elegante Möglichkeit, die  $d$ -fache Differenzenbildung anzuschreiben: es gilt

$$\Delta^d Y(t_i) = (1 - B)^d Y(t_i), \tag{2.54}$$

worüber man sich durch Nachrechnen versichern kann.

Es werde nun angenommen, dass die Folge  $\Delta^d Y(t_i)$  für ein  $d \geq 1$  stationär ist. Man macht sich nun leicht klar, dass sich die ursprünglichen  $Y(t_i)$  aus den Differenzen durch Summation wieder herstellen lassen. So hat man beim Random Walk  $\Delta Y(t_i) = Y(t_i) - Y(t_{i-1}) = \xi_i$ , dann ist einerseits

$$\begin{aligned} Y(t_i) &= Y(t_{i-1}) + \xi = Y(t_{i-1}) + \Delta Y(t_i) \\ &= Y(t_{i-1}) + (1 - B)Y(t_i) \\ &= BY(t_i) + (1 - B)Y(t_i) \\ &= BY(t_i) + Y(t_i) - BY(t_i) = Y(t_i) \end{aligned}$$

Andererseits ist  $Y(t_{i-1}) = Y(t_{i-2}) + \xi_{i-1}$  und damit  $\xi_{i-2} = Y(t_{i-1}) - Y(t_{i-2}) = \Delta y(t_{i-1})$ ; analog sind die  $\Delta Y(t_{i-k})$  für  $k > 1$  definiert. Man rechnet dann leicht nach, dass

$$Y(t_i) = \sum_k \Delta Y(t_{i-k}) + \xi_i \quad (2.55)$$

gelten muß: die Werte  $Y(t_i)$  entstehen also durch Summation der Differenzen. Mußten, um zu einer stationären Zeitreihe zu gelangen, die  $d$ -fachen Differenzen  $\Delta^d Y(t_i)$  gebildet werden, so lassen sich die  $Y(t_i)$  aus den entsprechenden  $d$  Summen "zusammensetzen".

Nun kann man nach Gleichung (2.26) für einen autoregressiven Prozess auch

$$\phi(B)Y(t_i) = \xi_i$$

schreiben. Hat man einen ARMA( $p, q$ )-Prozess, so gilt

$$\phi_p(B)Y(t_i) = \Theta_q(B)\xi_i$$

Gilt weiter (2.54), so erhält man schließlich

$$\phi_p(B)Y(t_i) = \Theta_q(B)\xi_i = \Theta_q(1 - B)^d Y(t_i) \quad (2.56)$$

**Definition 2.9** Der durch (2.56) definierte Prozess heißt ARIMA( $p, d, q$ )-Prozess.

Das "I" in ARIMA steht für "integriert", und integriert steht für "summiert" im Sinne von (2.55). Der ARIMA( $p, d, q$ )-Prozess ist der allgemeinste Prozess, den man anhand von AR- und MA-Prozessen definieren kann, und in der Praxis lassen sich viele Prozesse, d.h. gegebene Zeitreihen, tatsächlich als ARIMA-Prozesse beschreiben.

Es gibt allerdings noch eine Komponente, die bei einem ARIMA-Prozess noch nicht berücksichtigt worden ist, nämlich die *saisonale Komponente*. Auf sie wird im folgenden Abschnitt kurz eingegangen.

## Periodische ARIMA-Modelle

Es werde zunächst ein Beispiel betrachtet (vergl. Tabelle 2.1). Angenommen, die  $Y_i = Y(t_i)$  repräsentieren Unfallhäufigkeiten (vergl. Fig. 4). Unfälle treten an verschiedenen Wochentagen verschieden häufig auf, man kann vermuten, dass es einen 7-Tage-Zyklus gibt. Die Autokorrelationsfunktion bzw. die partielle Autokorrelationsfunktion weisen dann für den "Lag" (der Wert von  $k$ )  $k$  hohe Werte auf. Es gibt also "nachbarschaftliche" Abhängigkeiten und zusätzlich Abhängigkeiten zwischen  $Y_7$  und  $Y_{14}$ ,  $Y_{14}$  und  $Y_{21}$  sowie zwischen  $Y_7$  und  $Y_{21}$ , etc. Man kann dann den AR(1)-Prozess (dies ist ein ARIMA(1, 0, 0)-Prozess)

$$(1 - \phi_1 B)Y(t_i) = \epsilon_i \text{ kein weißes Rauschen!} \quad (2.57)$$

Tabelle 2.1: Struktur einer Zeitreihe mit saisonaler Komponente

	Mo	Di	Mi	Do	Fr	Sa	So
1. Woche	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$	$Y_6$	$Y_7$
2. Woche	$Y_8$	$Y_9$	$Y_{10}$	$Y_{11}$	$Y_{12}$	$Y_{13}$	$Y_{14}$
3. Woche	$Y_{15}$	$Y_{16}$	$Y_{17}$	$Y_{18}$	$Y_{19}$	$Y_{20}$	$Y_{21}$

mit dem Prozess

$$(1 - \phi_7 B^7 - \phi_{14} B^{14})\epsilon_i = \xi_i \quad (2.58)$$

kombinieren. Die Kombination ist dabei nicht additiv, sondern *multiplikativ*:

$$(1 - \phi_1 B)(1 - \phi_7 B^7 - \phi_{14} B^{14})Y(t_i) = \xi_i \quad (2.59)$$

Multipliziert man diesen Ausdruck aus, so erhält man

$$(1 - \phi_7 B^7 - \phi_{14} B^{14} - \phi_1 B + \phi_1 \phi_7 B^8 + \phi_1 \phi_{14} B^{15})Y(t_i) = \xi_i. \quad (2.60)$$

Für  $Y(t_i)$  ergibt sich

$$Y(t_i) = \phi_1 Y(t_{i-1}) + \phi_7 Y(t_{i-7}) + \phi_{14} Y(t_{i-14}) + \phi_1 \phi_7 Y(t_{i-8}) + \phi_1 \phi_{14} Y(t_{i-15}) + \xi_i \quad (2.61)$$

und

$$Y(t_{21}) = \phi_1 Y(t_{20}) + \phi_7 Y(t_{14}) + \phi_{14} Y(t_7) - \phi_1 \phi_7 Y(t_{13}) - \phi_1 \phi_{14} Y(t_6) + \xi_{21} \quad (2.62)$$

Ein saisonaler ARIMA-Prozess wird durch die Schreibweise  $ARIMA(p, d, q)(P, D, Q)$  gekennzeichnet; die Parameter  $P$ ,  $D$  und  $Q$  beziehen sich dabei auf die saisonale Komponente.