Angewandte Stochastik

Dr. C.J. Luchsinger

3 Grundbegriffe von Markov-Ketten

Literatur Kapitel 3

* Grimmett & Stirzaker: Kapitel 6.1
* Krengel: § 15

Begrifflich:

1. A. A. Markov war ein russischer Mathematiker (1856-1922). Sein Name wird manchmal auf Deutsch auch Markoff oder Markof geschrieben.
2. Man muss immer klar angeben, wie die Zeit (diskret oder stetig) und der Zustandsraum (was für Werte kann X_n annehmen?) ist (diskret oder stetig). Wir verwenden die Bezeichnung "Kette", falls die Zeit diskret ist und "Prozess", falls die Zeit stetig ist. Im weiteren werden wir in der ganzen Vorlesung bis auf einen kurzen Abstecher in die Brownsche Bewegung nur diskrete Zustandsräume betrachten.

Die Markov-Eigenschaft kann einfach in Worte gefasst werden: Betrachten wir eine Folge von Zufallsgrössen $X_1, X_2, \ldots, X_n$. Wir wollen jetzt Aussagen machen über die Verteilung von X_{n+1}. Ein solcher Prozess hat die Markov-Eigenschaft, wenn für die Verteilung von X_{n+1} die Kenntnis von $X_1, X_2, \ldots, X_n$ gleichbedeutend ist mit der Kenntnis von X_n. Die weiter zurückliegenden Werte $X_1, X_2, \ldots, X_{n-1}$ spielen also keine Rolle für X_{n+1}, *wenn wir X_n kennen.*

Wir werden in 3.1 zur Motivation ein paar Anwendungsbeispiele kennenlernen. Dabei kann es vorkommen, dass ein paar Begriffe noch nicht bekannt sind (z.B. der Poisson-Prozess). Diese werden später ausführlich behandelt. Die meisten Beispiele werden im Verlauf der Vorlesung noch ausführlich besprochen.

3.1 Motivation

Beste Motivation:
Geht auf www.math-jobs.com
und schaut die Berufs-Anforderungen an!

* **Theoretisches Beispiel:** der Random Walk RW im $\mathbb{Z}^n$, $n = 1, 2, \geq 3$
* **BWL:** Als OR-Spezialist bei McDonald's (Warteschlangentheorie)
* **BWL & Computerwissenschaften:** Stochastische Netzwerke
* **Biologie & Computerwissenschaften:** Epidemienausbreitung (u.a. auch Viren)
* **Biologie & Sozialwissenschaften:** Wachstum einer Population
* **Militärwissenschaften & Biologie:** Sich gegenseitig bekämpfende Populationen
* **Versicherungswissenschaften:** Modellierung von Erdbebenschäden
* **Makroökonomie & Oekonometrie:** AR(1)-Modelle mit wechselnden Regimen zur Modellierung der Konjunktur

3.1.1 Theoretisches Beispiel: der Random Walk RW im $\mathbb{Z}^n$, $n = 1, 2, \geq 3$

Problembeschreibung: $n = 1$ zum Einstieg, diskrete Zeit: Sei X_i, $1 \leq i \leq k$, eine iid-Folge von Be(0.5)-Zufallsgrössen mit $P[X_i = -1] = P[X_i = 1] = 0.5$ (man spricht auch hier von Bernoulli-Zufallsgrössen, obschon die Werte nicht 0 und 1 sind). Definieren wir

$$S_k := \sum_{i=1}^{k} X_i.$$

Jetzt betrachten wir die Folge der $(S_k)_{k \geq 1}$ und setzen zusätzlich $S_0 = 0$. Da die X_i unabhängig voneinander sind, gilt wegen

$$S_k = S_{k-1} + X_k,$$

dass die Verteilung von S_k nur von S_{k-1} (und X_k) abhängt, nicht aber von weiter zurückliegenden Werten von S (wenn S_{k-1} gegeben). Die Folge der S_k hat die sogenannte Markov-Eigenschaft.

Fragestellungen:

* Kommt S_k wieder nach 0 zurück? Mit welcher Wahrscheinlichkeit?

* Wie oft?

* Wie lange dauert es im Durchschnitt, bis S_k wieder nach 0 zurückkommt?

* Wie sieht die Situation im $\mathbb{Z}^2$ aus (mit Wahrscheinlichkeiten 1/4 je Richtung)?

* Wie sieht die Situation im $\mathbb{Z}^3$ und höheren Dimensionen aus (mit Wahrscheinlichkeiten 1/(2n) je Richtung)?

Die Antworten auf diese Fragen folgen in Kapitel 4 und Kapitel 5.

3.1.2 BWL: Als OR-Spezialist bei McDonald's (Warteschlangentheorie)

Problembeschreibung: Kunden kommen in's Geschäft mit einer Rate von 100 pro Stunde. Wir modellieren diese Ankunft mit einem Poissonprozess (stetige Zeit). Diese Ankunft ist insofern zufällig, als dass diese 100 pro Stunde ein langjähriger Durchschnitt sind. Es können sehr wohl einmal nur 50 kommen und ein andermal 200. Sie wollen als Manager nicht zu viele Personen anstellen (falls nur 50 kommen) und auch den Peak von 200 abfangen (keine langen Wartezeiten (=unzufriedene Kunden, Leute kommen gar nicht rein)). Die Anzahl Wartende X_t zum Zeitpunkt t bei einem Single-Server-Queue genügt (dies wird später in der Vorlesung gezeigt) einem Markovprozess: Wir müssen nur die Ankunftsrate, die Leistungsrate des Single-Servers und die momentane Anzahl Wartenden kennen, um bereits bestens über die Zukunft informiert zu sein. Weiter zurück liegende Werte nützen uns nichts (diese Eigenschaft "erbt" X_t vom Ankunftsprozess (Poisson) und von der Exponentialverteilung, mit der man den Server modelliert).

Fragestellungen:

* Gibt es einen Gleichgewichtszustand der Anzahl Wartenden?
* Wächst die Anzahl Wartende bis ∞ (im Modell)? Nein, wenn Leistungsrate grösser als Ankunftsrate.
* Abschreckungseffekt einbauen bei $X_t > 10$
* Gibt es so etwas wie eine erwartete Wartezeit?

3.1.3 BWL & Computerwissenschaften: Stochastische Netzwerke

Problembeschreibung (Verallgemeinerung von 3.1.2):

1. Situation ohne Zufluss und ohne Abfluss (stetige Zeit): Knoten 1 bis 3; Leistungsfähigkeit pro Knoten ϕ_1, ϕ_2, ϕ_3 (z.B. 10 Kunden pro Stunde); 5 Personen mit Bedürfnissen unterwegs: $\mu_{ij}, 1 \leq i \leq 5, 1 \leq j \leq 3$; μ_{42}=Bedürfnis von Person 4 bei Knoten 2. p_{mn}=Wahrscheinlichkeit, dass Person nach Knoten m zu Knoten n wechselt; egal was weiter zurückliegende Vergangenheit (Markov-Eigenschaft!).

2. Zufluss und Abfluss (Shopping-Mall; Computer-Jobs): mutatis mutandis gleiche Situation wie oben, aber λ_i = Zuflussrate zu Knoten i von Aussen (z.B. Poissonprozess); p_i = Wahrscheinlichkeit, dass man bei Knoten i Austritt. Bedürfnisse der Kunden iid (z.B. exponentialverteilt).

Fragestellungen:

* Gibt es eine "Gleichgewichtsverteilung?
* Gibt es immer grössere Staus vor einzelnen Knoten? Unter welchen Bedingungen kommt das vor?
* Erwartete Wartezeit?
* "Einschwingen" des Prozesses!
* Wieviel Leistungsfähigkeit soll man jedem Knoten geben?

3.1.4 Biologie & Computerwissenschaften: Epidemienausbreitung (u.a. auch Viren)

Problembeschreibung: Feste Population von N Menschen; diskrete Zeit. Zum Zeitpunkt j seien (S_j, I_j, E_j) die Anzahl (Ansteckbare (Susceptibles), Infektiöse, Erholte), wobei damit für jeden Zeitpunkt gelten muss, dass $S_j + I_j + E_j = N$. Die Menschen verhalten sich im Modell unabhängig voneinander. Sei q die Wahrscheinlichkeit, dass ein gegebener Mensch einen Kontakt (genügend eng) mit einem anderen gegebenen Menschen zwischen zwei aufeinanderfolgenden Zeitpunkten hat.

$$S_{j+1} = S_j - (\text{die neu Infizierten}),$$

P[ein gegebener Mensch aus den Ansteckbaren nicht in diesem Intervall infiziert wird] $=$ $(1-q)^{I_j}$. Dann gilt

$$S_{j+1} \sim \text{Bin}(S_j, (1-q)^{I_j}).$$

Diese Verteilung enthält neben q nur (S_j, I_j) als Parameter und ist somit nur von der Gegenwart j abhängig (aber zweidimensional), falls diese bekannt ist. Weiter zurückliegende Werte haben keinen Wert für Vorhersagen. Wir haben also auch hier die Markov-Eigenschaft.

Fragestellungen:

* Wie sollen wir den Genesungsprozess $I \to E$ modellieren?
* Gibt es für $j \to \infty$ noch Ansteckbare?
* Gibt es nichttriviale Gleichgewichte? Trivial ist zum Beispiel $(N, 0, 0)$.

3.1.5 Biologie & Sozialwissenschaften: Wachstum einer Population

Problembeschreibung: Wir betrachten das Wachstum einer Population in diskreter Zeit (Zeit z.B. in Generationen gemessen). Z_n sei die Anzahl Individuen zum Zeitpunkt n. Sei X_{nj} die Anzahl Nachkommen von Individuum j zur Zeit n. Jedes Individuum trägt zu jedem Zeitpunkt zum Wachstum der Population bei. Wir nehmen an, dass $X_{nj}, n \geq 0, j \geq 1$ iid (kaum realistisch). Sei z_0 die Grösse der Population zur Zeit 0 (fest). Dann gilt:

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_{nj} + Z_n.$$

Damit hängt Z_{n+1} nur von Z_n, nicht aber von weiter zurückliegenden Zeitpunkten ab (gegeben Z_n) (Markov-Eigenschaft!).

Fragestellungen:

* Überlappung der Generationen in stetiger Zeit (ist das ein Problem?)
* Wie schnell wächst Z_n?
* Einbau eines parallelen Sterbeprozesses, Immigration?
* Wann gibt es einen Gleichgewichtszustand?

3.1.6 Militärwissenschaften & Biologie: Sich gegenseitig bekämpfende Populationen

Frage an's Publikum: "Infanterie: Kampf bis zur letzten Patrone". Gleiche Qualität und Motivation der Armeen A und B. A hat 10'000 Mann Infanterie, B hat 5'000 Mann Infanterie. Wie wird etwa der Endstand sein?

Problembeschreibung: Gegeben sind zwei Populationen zur Zeit t (stetige Zeit); $x(t) \geq 0$ und $y(t) \geq 0$ ist die Menge oder Anzahl der einzelnen Populationen (z.B. Bakterien oder Infanterie). Diese Populationen bekämpfen sich nun derart, dass folgendes System von deterministischen Differentialgleichungen erfüllt ist (solange $x(t) \geq 0$ und $y(t) \geq 0$):

$$\dot{x} = -\beta y$$

$$\dot{y} = -\alpha x.$$

Offenbar ist α der Parameter, welcher angibt, wie stark die x-Population pro x-Einheit der y-Population zusetzt und β gibt an, wie stark die y-Population pro y-Einheit der x-Population schadet. Dies ist noch ein deterministisches Modell, welches ein stochastisches Analogon hat. Die Markov-Eigenschaft ist offenbar insofern gegeben, als dass nur der jetzige Zeitpunkt mit den jetzigen Zuständen über die Zukunft entscheidet und die Geschichte keine weiteren Hinweise für die Entwicklung in der Zukunft gibt.

Fragestellungen:

* Wer gewinnt bei gegebenen Anfangswerten; z.B. $x(0) = 5, \alpha = 1, y(0) = 3, \beta = 2$?

* Welche Kombination von $x(0), \alpha, y(0), \beta$ entscheidet über Sieg und Niederlage?

* Wieviel hat die Siegerpopulation am Schluss noch übrig?

Schlussfolgerungen aus diesem Beispiel:

1. Bei kleinen Populationen wird der Zufall wichtiger werden und ein stochastisches Modell ist dann adäquater.

2. Die Antwort auf die zweite Frage hat für die Militärwissenschaften grosse Konsequenzen, welche bereits Clausewitz (1780-1831, "Vom Krieg") bekannt waren: Konzentration der Kräfte - klotzen nicht kleckern (Guderian) - Guerillataktik. Der Name dieses Prinzips ist das sogenannte N^2-Gesetz von Lanchester (1868-1946). Wir werden in Kapitel 18 darauf zurückkommen.

3.1.7 Versicherungswissenschaften: Modellierung von Erdbebenschäden

Problembeschreibung: Erdbebenzeitpunkte kann man nicht sicher vorhersagen. Aber aus der Vergangenheit kann man für Regionen die Häufigkeitsrate schätzen (Modell: Poissonprozess). Trotz der Schätzung des Häufigkeitsparameters aus der Vergangenheit geht man im Modell von einem Markovprozess (in stetiger Zeit) aus.
Man kann für jeden Schadenzeitpunkt unabhängig von den anderen Zeitpunkten die Schadengrösse ebenfalls zufällig modellieren. Es ergibt sich dann etwa folgendes Bild:

Fragestellungen:

* Sind Erdbebenzeitpunkte weltweit unabhängig voneinander? (Griechenland & Türkei)
* Wieviel Prämie soll verlangt werden?
* Soll man für die Prämienberechnung lediglich auf Durchschnittswerte achten (20 Milliarden Franken Schaden alle 10 Jahre also Jahresprämie 2 Milliarden Franken plus Administrationskosten) oder die grosse Variabilität einbeziehen?
* Stopp-Loss-Versicherungen

3.1.8 Makroökonomie & Oekonometrie: AR(1)-Modelle mit wechselnden Regimen zur Modellierung der Konjunktur

Problembeschreibung: Für makroökonomische Grössen wie Wechselkursrelationen oder Wachstumsraten des realen Bruttosozialprodukts oder Realzinsniveaus kann man (nach allfälligen Transformationen) sogenannte autoregressive (AR) Prozesse zu deren Beschreibung verwenden (diskrete Zeit; z.B. Quartalsdaten). Bei einem AR(1)-Prozess genügt X_t beispielsweise folgender Gleichung:

$$X_t = \alpha X_{t-1} + \epsilon_t,$$

wobei ϵ_t z.B. eine Folge von iid $\mathcal{N}(0,1)$-Zufallsgrössen sei.

Wenn man den Mittelwert μ ungleich Null einbauen will, kann man dies mit

$$X_t - \mu = \alpha(X_{t-1} - \mu) + \epsilon_t$$

berücksichtigen. Wenn man verschiedene Mittelwerte μ_1, μ_2 einbauen will (Regime 1 und 2; z.B. Rezession und Boom-Phase), kann man

$$X_t - \mu_1 = \alpha(X_{t-1} - \mu_1) + \epsilon_t \tag{3.1}$$

und

$$X_t - \mu_2 = \alpha(X_{t-1} - \mu_2) + \epsilon_t \tag{3.2}$$

simultan betrachten. Markov-Switching-Modelle werden jetzt folgendermassen konstruiert: Zum Zeitpunkt t sei der Prozess in Regime 1 und genügt damit Gleichung (3.1). Dann wird mit (3.1) X_{t+1} berechnet.

Jetzt wird mit folgenden Übergangswahrscheinlichkeiten entschieden, ob das Regime gewechselt wird oder nicht: mit Wahrscheinlichkeit p_{00} bleiben wir in Regime 1, mit Wahrscheinlichkeit p_{01} wechseln wir zu Regime 2 und berechnen dann X_{t+2} mit Hilfe von (3.2) (mit anderem μ!). Die Übergangswahrscheinlichkeiten von Regime 2 zu Regime 1 sind analog. Dem Beobachter (Notenbankchef) ist jeweils nicht bekannt, in welchem Regime er gerade ist; er kennt nur den Prozess X.

Fragestellungen:

* Wann gab es Regimewechsel (ex post)?
* Wann gibt es wieder einen Regimewechsel? Kann man solche Regimewechsel voraussagen (und damit Geld verdienen)?
* Wie soll man die Parameter μ schätzen?
* Sind diese Modelle besser als die grossen Modelle aus der Makroökonomie; was heisst hier besser?

3.2 Definition von Markov-Ketten und elementare Eigenschaften

3.2.1 Definition von Markov-Ketten

Definition 3.1 [Markov-Kette] *Eine Markov-Kette ist eine Folge $X_0, X_1, \ldots$ von Zufallsgrössen, die Werte in $\mathbb{Z}^+$ annehmen (allg. in einem abzählbaren Zustandsraum), mit der Eigenschaft, dass $\forall n \geq 0$ und $(j, i, i_{n-1}, i_{n-2}, \ldots, i_0) \in (\mathbb{Z}^+)^{n+2}$ gilt:*

$$P[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] = P[X_1 = j | X_0 = i] =: p_{ij}.$$

In dieser Definition ist auch gleich die sogenannte *Zeithomogenität* miteingebaut. Ohne diese muss es heissen:

$$P[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] =: p_{ij,n},$$

um anzugeben, dass man im Zeitpunkt n ist (und lediglich den bisherigen Verlauf vernachlässigt). Weitaus häufiger (in dieser Vorlesung immer!) ist jedoch der Fall, wo man sowohl die Markov-Eigenschaft wie auch die Zeithomogenität fordert.

Die Matrix $P = (p_{ij})_{i,j \in \mathbb{Z}^+}$ nennen wir *Übergangsmatrix*. Wir müssen noch angeben, wie wir starten:

$$\lambda_i := P[X_0 = i], i \in \mathbb{Z}^+.$$

Häufig werden wir deterministisch starten, das heisst, dass für ein i gelten wird $\lambda_i = 1$. Wir legen die λ_i's im Vektor λ ab. Es muss natürlich immer gelten: $\sum_{i \geq 0} \lambda_i = 1$. Jetzt können wir erste Rechnungen durchführen (mit Begründung jedes Schrittes):

$P[X_0 = i, X_1 = j] =$

$$P[X_0 = i, X_1 = j, X_2 = k] = \tag{3.3}$$

Durch Induktion zeigt man auch (freiwillige Übung):

$$P[X_0 = i_0, \ldots, X_n = i_n] = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}.$$

Wenn $X_0 = i_0$ nicht explizit angegeben ist, bedienen wir uns der FTW: Weil $\Omega = \cup_{i_0 \in \mathbb{Z}^+} \{X_0 = i_0\}$ (disjunkte Vereinigung), können wir folgern:

$$P[X_1 = i_1] = \sum_{i_0 \in \mathbb{Z}^+} P[X_1 = i_1 | X_0 = i_0] P[X_0 = i_0] = \sum_{i_0 \in \mathbb{Z}^+} p_{i_0 i_1} \lambda_{i_0} = \sum_{i_0 \in \mathbb{Z}^+} \lambda_{i_0} p_{i_0 i_1}$$

(falls ein $\lambda_{i_0} = 0$ ist, definieren wir Œ $P[X_1 = i_1 | X_0 = i_0] = 0$). Bei mehreren Schritten ergibt sich

$$P[X_2 = i_2] = \tag{3.4}$$

$= \sum_{i_0, i_1 \in \mathbb{Z}^+} \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2}.$

Wir bedienen uns manchmal der matriziellen Schreibweise: im ersten Fall erhalten wir eine Rechtsmultiplikation der Art

$$(\lambda^t P)_{i_1};$$

im zweiten Fall

$$(\lambda^t P^2)_{i_2}.$$

Dabei ist λ^t das transponierte λ. Mit mehreren Schritten geht man analog vor. Man beachte, dass in (3.3) der Pfad von 0 bis 2 vorgegeben ist, während in (3.4) z.B. X_1 jeden beliebigen Wert annehmen darf. Wir wollen uns kurz mathematisch in 3.2.2 absichern, dass wir nicht unbemerkt unerlaubte Operationen durchführen.

3.2.2 Übergangsmatrix P als Operator, Konsistenz der Wahrscheinlichkeit P

Bereits in 3.2.1 haben wir übrigens einen Notationsmissbrauch begangen, der sich bei den Markov-Ketten vollständig eingebürgert hat und nie zu Problemen führt: P als Übergangsmatrix und P als Wahrscheinlichkeit. Zu beiden müssen wir jetzt noch kurz mathematische Abklärungen treffen.

Übergangsmatrix P als Operator

1. Die p_{ij} als Einträge der Matrix P sind alle ≥ 0 und ≤ 1; zudem summieren die Zeilen jeweils auf 1:

$$\sum_k p_{ik} = \sum_k P[\{X_1 = k\}|\{X_0 = i\}] = P[\cup_k \{X_1 = k\}|\{X_0 = i\}] = P[\Omega|\{X_0 = i\}] = 1\,\forall i.$$

Hingegen gilt im Allgemeinen nicht $\sum_i p_{ik} = 1$.

2. Betrachten wir die Einträge der Matrix P^2:

$$(P^2)_{ij} = \sum_k p_{ik} p_{kj} < \infty,$$

da $\sum_k p_{ik} = 1$ und die Summe durch die p_{kj} (≤ 1) nur kleiner wird. Die Zeilen-Einträge von $P^n, n \geq 1$, summieren übrigens auch auf 1 (freiwillige Übung).

3. P als Operator (von rechts): Bis jetzt haben wir ja mit λ von rechts multipliziert und es gilt $\sum_k \lambda_k = 1$. Somit werden wir Ausdrücke der Art

$$xP$$

mit $x \in l_1 := \{y|y = (y_0, y_1, \ldots); \sum_{j\geq 0} |y_j| < \infty\}$ betrachten müssen. Offensichtlich liegt xP wieder in l_1: Definieren wir $\|x\|_1 := \sum_i |x_i|$, so gilt:

$$\|xP\|_1 = \sum_i |\sum_j x_j p_{ji}| \leq \sum_i \sum_j |x_j| p_{ji} = \sum_j |x_j| \sum_i p_{ji} = \|x\|_1.$$

P ist bei Rechtsmultiplikation also schwach kontraktiv: $\|xP\|_1 \leq \|x\|_1$.

4. P als Operator (von links): Wir werden später auch von links multiplizieren. Wir werden hier mit $x \in l_\infty := \{y | y = (y_0, y_1, \ldots)^t; \sup_{j \geq 0} |y_j| < \infty\}$ arbeiten und

$$Px$$

untersuchen (die Absolutbeträge lassen wir weg, sowieso ≥ 0):

$$(Px)_j = \sum_k p_{jk} x_k \leq \sup_m x_m = \|x\|_\infty,$$

also liegt insbesondere (j war frei gewählt) Px wieder in l_∞. Zudem ist auch die Linksmultiplikation schwach kontraktiv: $\|Px\|_\infty \leq \|x\|_\infty$

<u>Konsistenz der Wahrscheinlichkeit P - nicht Prüfungsstoff</u>

Wir repetieren aus der Vlsg WT: Nehmen wir einmal an, wir haben mathematisch sauber ein $(\Omega, \mathcal{A}, P)$ und eine Folge von Zufallsgrössen $(X_n)_{n \geq 0}$ konstruiert. Dann muss wegen der Stetigkeit von P (vgl Vlsg WT) sicher gelten:

$$P[X_1 \leq t_1, \ldots, X_n \leq t_n] = \lim_{t \to \infty} P[X_1 \leq t_1, \ldots, X_n \leq t_n, X_{n+1} \leq t].$$

Falls wir zu einer gegebenen Folge von Verteilungsfunktionen $(F_n)_{n \geq 0}$ eine Folge von Zufallsgrössen $(X_n)_{n \geq 0}$ konstruieren wollen, müssen wir also sicher fordern, dass

$$\lim_{t \to \infty} F_{n+1}(t_1, \ldots, t_n, t) = F_n(t_1, \ldots, t_n).$$

In der Tat ist diese *Konsistenzbedingung* auch genügend. Dies ist ein fundamentales Resultat aus der Masstheorie und lautet (Beweis in A.N. Sirjaev: Wahrscheinlichkeit)

WT-Satz 2.27 [Satz von Kolmogorov über die Existenz stochastischer Prozesse] *Für alle n gelte, dass F_n eine Verteilungsfunktion auf $\mathbb{R}^n$ ist. Es gelte zudem die Konsistenzbedingung*

$$\lim_{t \to \infty} F_{n+1}(t_1, \ldots, t_n, t) = F_n(t_1, \ldots, t_n)$$

für alle n und $(t_1, \ldots, t_n)$. Dann gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine Folge von Zufallsgrössen $(X_n)_{n \geq 0}$ so, dass F_n für alle n die Verteilungsfunktion von $(X_1, \ldots, X_n)$ ist.

3.2.3 Ein paar Beispiele

In "3.1 Motivation" haben wir viele Beispiele kennengelernt, ohne die genaue Ausformulierung der Markoveigenschaft zu kennen. In Kenntnis der exakten Definition der Markoveigenschaft wollen wir jetzt weitere Beispiele anfügen.

Beispiel 1: Das einfachste, nichttriviale Beispiel

2 Zustände: 0 und 1. Übergangswahrscheinlichkeiten: $p_{01} = p, p_{10} = q$, wo $p, q \in (0, 1)$. Schematisch folgendermassen dargestellt:

Die Übergangsmatrix sieht folgendermassen aus:

Kurze Besprechung des Falls p oder $q \in \{0, 1\}$. Verteilung Zeit bis Absorption?

Beispiel 2: Single Server Warteschlange (mehrere Varianten)

Zustandsraum ist $\{0, 1, \ldots, N\}$; $p_{01} = 1$, $p_{i,i+1} = p$, wo $p \in (0, 1)$ und $i \leq N - 1$, $p_{N,N-1} = 1$ (zu lange Warteschlange, Leute kommen gar nicht erst wenn zu lange). X_n ist dann die Anzahl Wartender zur Zeit n; wir definieren die Zeiteinheit als Zeit bis zur nächsten Änderung (*entweder* Kunde ist bedient und geht weg *oder* neuer Kunde kommt). Schematisch stellt man dies folgendermassen dar, Matrix:

Alternativ kann man den Zustandsraum auch als $\{0, 1, \ldots, \}$ wählen und die Übergangswahrscheinlichkeiten mit $p_{01} = 1, p_{12} = 0.5$ und für $i \geq 2$

$$p_{i,i+1} = \frac{1}{i}$$

oder ähnlich modellieren. Damit fängt man auf, dass es kaum eine klare Schranke (N) gibt, ab der die potentiellen Kunden durch die zu lange Warteschlange abgeschreckt werden. Der Abschreckungseffekt kommt so graduell.

Beispiel 3: 2-Allel Modell aus der Genetik

Wir unterstellen in diesem einfachen Modell, dass es genau 2 Typen von Genen oder Individuen gibt: A und B. Angenommen $X_0 = i, i \in \{0, 1, \ldots, N\}$ ist die Anzahl Individuen vom Typ A zur Zeit 0. Es gibt konstant total N Individuen. Jedes Individuum stirbt gleichwahrscheinlich $(1/N)$. Wenn $X_0 = i$, dann wird nach Annahme ein zur Zeit 1 sterbendes Individuum mit Wahrscheinlichkeit i/N durch ein Individuum vom Typ A ersetzt. Folgende 4 Fälle (mit Wahrscheinlichkeiten) sind zu unterscheiden:

$$\begin{aligned}
&i \rightarrow\ i+1 \text{ (B stirbt, mit A ersetzt) mit Wahrscheinlichkeit } \frac{i(N-i)}{N^2} \\
&i \rightarrow\ i \text{ (A stirbt, mit A ersetzt) mit Wahrscheinlichkeit } \frac{i}{N}\frac{i}{N} \\
&i \rightarrow\ i \text{ (B stirbt, mit B ersetzt) mit Wahrscheinlichkeit } \frac{N-i}{N}\frac{N-i}{N} \\
&i \rightarrow\ i-1 \text{ (A stirbt, mit B ersetzt) mit Wahrscheinlichkeit } \frac{i(N-i)}{N^2}.
\end{aligned}$$

Es ergeben sich damit folgende Übergangswahrscheinlichkeiten:

$$p_{ij} = \begin{cases} \frac{i(N-i)}{N^2} & \text{falls } j = i \pm 1 \\ 1 - 2\frac{i(N-i)}{N^2} & \text{falls } j = i \\ 0 & \text{sonst.} \end{cases}$$

Man beachte: Wenn wir nur wissen wollen, ob der Prozess in 0 oder N endet, könnten wir gerade so gut eine symmetrische Irrfahrt auf $\{0, 1, \ldots, N\}$ untersuchen. Offenbar ist mit dieser Änderung dann auch hier die Situation derart, dass wenn $X_n = 1$ gilt, die Wahrscheinlichkeit "für einen Schritt nach links oder nach rechts" jeweils je 1/2 ist (vgl. die Situation in 3.1.6 (sich gegenseitig bekämpfende Populationen)).

3.2.4 Chapman-Kolmogoroff Gleichung und Ungleichung

Notation: $p_{ij}^{(n)} := P[X_n = j | X_0 = i]$. Mit der Klammer im Exponent von $p_{ij}^{(n)}$ geben wir also an, dass wir nicht einfach das Element p_{ij} der Übergangsmatrix P hoch n nehmen, sondern in n Schritten von i nach j kommen wollen. Es gilt folgende Formel:

Lemma 3.2 [Chapman-Kolmogoroff Gleichung und Ungleichung] *Es gilt*

$$p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)}.$$

Eine einfache Folgerung ist die entsprechende Ungleichung:

$$p_{ij}^{(m+n)} \geq p_{ik}^{(m)} p_{kj}^{(n)} \ \forall k.$$

Beweis

$$\begin{aligned}
p_{ij}^{(m+n)} &:= P[X_{m+n} = j | X_0 = i] \\
&\overset{\circ}{=} \sum_k P[X_{m+n} = j | X_m = k, X_0 = i] P[X_m = k | X_0 = i] \\
&= \sum_k p_{kj}^{(n)} p_{ik}^{(m)} \\
&= \sum_k p_{ik}^{(m)} p_{kj}^{(n)}.
\end{aligned}$$

Das zweite Gleichheitszeichen "$\overset{\circ}{=}$" folgt aus einer Verallgemeinerung von FTW, siehe WTS-Aufgabenblatt 3. Kurz: wir können in Anlehnung an die FTW auch mit $P[C] > 0$ benutzen:

$$P[A|C] = \sum_{i=1}^{\infty} P[A|C \cap B_i] P[B_i|C]. \qquad (bFTW)$$

Dabei steht das "b" in bFTW für bedingt.

□

3.2.5 Stoppzeit

In der Markoveigenschaft (wie wir sie definieren) spielt der konkrete Zeitpunkt nie eine Rolle (Zeithomogenität). Hingegen war die Zeit doch immerhin deterministisch. Wir wollen uns fragen, ob wir auch mit zufälligen Zeiten $N(\omega)$ beispielsweise

$$P[X_{N+1} = j | X_N = i] = p_{ij} \tag{3.5}$$

erhalten. Dazu definieren wir zuerst:

Definition 3.3 [Stoppzeit (Markov-Zeit)] *Eine Zufallsgrösse N mit Werten in $\{0, 1, 2, 3 \ldots\} \cup \{\infty\}$ heisst Stoppzeit (oder Markov-Zeit) für den Prozess $(X_0, X_1, X_2, \ldots)$, wenn für alle $n \geq 0$ das Ereignis $\{N = n\}$ sich als $\{(X_0, X_1, \ldots, X_n) \in A_n\}$ darstellen lässt, wobei $A_n \subseteq (\mathbb{Z}^+)^{n+1}$.*

Die Markov-Eigenschaft bedeutet, dass man keinen Vorgriff auf die Zukunft hat. Ob die Zufallsgrösse N den Wert n annimmt, muss mit ausschliesslicher Hilfe der Markov-Kette bis zur Zeit n entschieden werden können.

Ein Beispiel hierfür ist:

$$N := \inf\{m \geq 1 | X_m = 0\}.$$

Hier gilt nämlich (von LeserIn auszufüllen):

$\{N = n\} =$

Die Stoppzeit erfüllt in Spielen insbesondere eine gewisse Fairnessforderung. Man kann nicht sagen, ich höre jetzt zur Zeit n auf, falls ich in Zeit $n + 1$ verlieren werde. Beispiele zur Stoppzeit sind auf Übungsblatt 3 zu untersuchen.

3.2.6 Starke Markov-Eigenschaft

Wir sind hiermit gerüstet, um den Sinn von (3.5) zu untersuchen. Wir überlassen es aufmerksamen LeserInnen, folgende Gleichung zu beweisen (welche noch nicht Stoppzeiten benutzt, aber für den Beweis des Theorems benutzt wird): Es gilt für $A \subseteq (\mathbb{Z}^+)^n$:

$$P[X_{n+1} = j | X_n = i, (X_{n-1}, X_{n-2}, \ldots, X_0) \in A] = p_{ij}. \tag{3.6}$$

Theorem 3.4 [Starke Markov-Eigenschaft] *Sei N eine Stoppzeit, $B = \cup_{n \geq 0} B_n$, wobei $B_n \subseteq (\mathbb{Z}^+)^{n+1}$ und $P[N < \infty, X_N = i, (X_{N-1}, X_{N-2}, \ldots, X_0) \in B] > 0$. Dann gilt:*

$$P[X_{N+1} = j | N < \infty, X_N = i, (X_{N-1}, X_{N-2}, \ldots, X_0) \in B] = p_{ij}.$$

Die umständliche Konstruktion von B ist notwendig, weil wir ja a priori gar nicht wissen, wie gross N sein wird und damit wissen wir auch nicht, wieviele Schritte wir retour bis 0 gehen müssen.

Beweis Wir definieren $D := \{N < \infty, X_N = i, (X_{N-1}, X_{N-2}, \ldots, X_0) \in B_{N-1}\}$ und verwenden an der Stelle "$\doteq$" wieder bFTW.

$$\begin{aligned}
&P[X_{N+1} = j | N < \infty, X_N = i, (X_{N-1}, X_{N-2}, \ldots, X_0) \in B] \\
&= P[X_{N+1} = j | N < \infty, X_N = i, (X_{N-1}, X_{N-2}, \ldots, X_0) \in B_{N-1}] = P[X_{N+1} = j | D] \\
&\doteq \sum_{n \geq 0} P[X_{N+1} = j | N = n, D] P[N = n | D] \\
&= \sum_{n \geq 0} P[X_{N+1} = j | N = n, N < \infty, X_N = i, (X_{N-1}, \ldots, X_0) \in B_{N-1}] P[N = n | D] \\
&= \sum_{n \geq 0} P[X_{n+1} = j | X_n = i, (X_{n-1}, X_{n-2}, \ldots, X_0) \in B_{n-1}] P[N = n | D] \\
&= \sum_{n \geq 0} P[X_{n+1} = j | X_n = i] P[N = n | D] \\
&= \sum_{n \geq 0} p_{ij} P[N = n | D] \\
&= p_{ij}.
\end{aligned}$$

Dabei haben wir im dritt-letzten Gleichheitszeichen (3.6) und im letzten Gleichheitszeichen verwendet, dass bedingte Wahrscheinlichkeiten auch Wahrscheinlichkeiten sind (und auf 1 summieren).

□