Vorstellung der Kopplung bei Markovketten

Leonie Weinhold

13. Mai 2011

1 Einleitung

In der Wahrscheinlichkeitstheorie und ihren Anwendungen ist die Coupling-Methode ein wichtiges Instrument, um z.B. Aussagen über Zufallsvariablen oder stochastische Prozesse treffen zu können. Insbesondere kann diese Methode bei Markov-Prozessen sehr geschickt eingesetzt werden. Im Folgenden wird die Coupling-Methode anhand von verschiedenen, einfachen Beispielen vorgestellt, damit dann im letzten Beispiel eine klassische Kopplung von Markov-Ketten gezeigt werden kann. Das erste Beispiel, dass wir anschauen nennt man das "Gambler's ruin problem"(Ruin des Spielers).

2 Beispiele

2.1 Gambler's ruin problem

In diesem einführenden Beispiel geht es um einen Spieler, der das Startkapital von m Geldeinheiten besitzt und damit in ein Spielkasino geht um zu spielen. Dort spielt er solange voneinander unabhängige Spiele bis er entweder insgesamt n Geldeinheiten ($n > m$) besitzt, oder Bankrott geht, also 0 Geldeinheiten besitzt. Die Gewinnwahrscheinlichkeit für ein einzelnes Spiel beträgt θ und die Wahrscheinlichkeit, dass er ein Spiel verliert beträgt 1-θ.
Ist $P(\theta)$ die Wahrscheinlichkeit, dass der Spieler das Kapital n erreicht ($\theta \in [0, 1]$, beliebig), d.h. $P(\theta) = P\{\text{Kapital } n \text{ wird erreicht}\}$, so wird nun mittels Coupling bewiesen, dass die Wahrscheinlichkeit $P(\theta)$ mit wachsendem θ steigt.

Beweis: Dafür seien $U_1, U_2, \ldots$ unabhängige, identisch verteilte Zufallsvariablen mit $U_i \sim U(0, 1)$. Definiere für alle θ die Zufallsvariablen:

$$\begin{aligned} Y_\theta(i) &= \begin{cases} +1 & U_i \leq \theta \\ -1 & \text{sonst} \end{cases} \qquad \forall i = 1, 2, \ldots \qquad \text{und} \\ X_\theta(\nu) &= m + \sum_{i=1}^{\nu} Y_\theta(i) \qquad \forall \nu = 1, 2, \ldots \end{aligned}$$

Der Prozess $X_\theta = \{X_\theta(\nu); \nu = 1, 2, \ldots\}$ beschreibt offensichtlich das Kapital des Spielers. Wählen wir jetzt ein θ', so dass gilt: $\theta < \theta'$ dann gilt wegen der Konstruktion von $Y_\theta(i)$:
$Y_\theta(i) \leq Y_{\theta'}(i)$ für alle $i = 1, 2, \ldots$ und $X_\theta(\nu) \leq X_{\theta'}(\nu)$ für alle $\nu = 1, 2, \ldots$.
Erreicht die Zufallsvariable $X_\theta(\nu)$ das Kapital n, so gilt das Gleiche also auch für $X_{\theta'}(\nu)$. Benennt man jetzt um, so dass $\widehat{X} = X_\theta$ und $\widehat{X}' = X_{\theta'}$, dann wurde soeben mit der Kopplung $(\widehat{X}, \widehat{X}')$ gezeigt, dass $\mathrm{P}(\theta) \leq \mathrm{P}(\theta')$. Das bedeutet insbesondere, dass diese Wahrscheinlichkeit nicht vom definierten zwei-dimensionalen Wahrscheinlichkeitsraum abhängt.

2.2 Stochastische Ordnung

Für das zweite Beispiel benötigt man zwei neue Begriffe: zum einen die Stochastische Ordnung, mit der man Zufallsvariablen vergleichen kann und zum anderen die Definition der verallgemeinerten Inversen für Verteilungsfunktionen.

2.2.1 Definition (Stochastisch geordnete Zufallsvariablen):

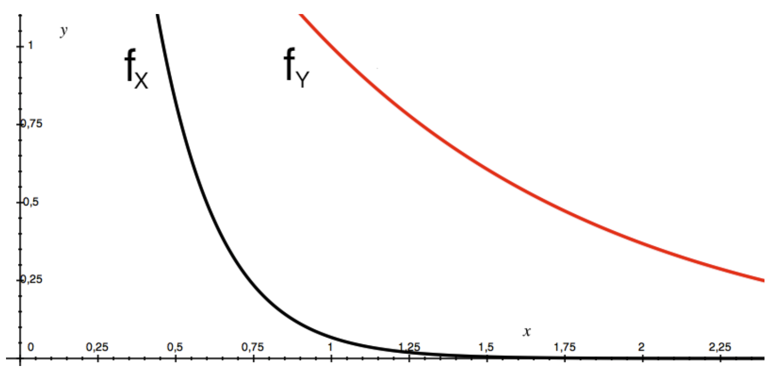
Seien X und Y zwei reelwertige Zufallsvariablen, dann heisst X *stochastisch kleiner* als Y wenn gilt:

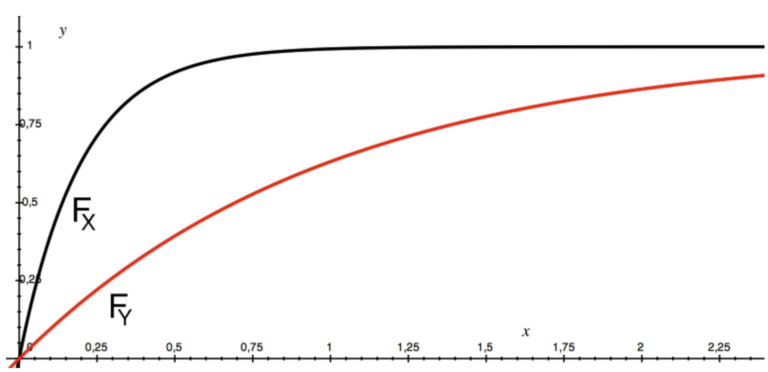
$$P(X \geq a) \leq P(Y \geq a) \qquad \forall a$$

Offensichtlich äquivalent dazu ist:

$$F_X(a) \geq F_Y(a) \qquad \forall a$$

Schreibweise: $X \overset{\mathcal{D}}{\leq} Y$, analog für $X \overset{\mathcal{D}}{\geq} Y$ und $X \overset{\mathcal{D}}{=} Y$
Das bedeutet, die Wahrscheinlichkeit, dass X ein bestimmtes Level erreicht, ist geringer (bzw. größer, gleich), als die Wahrscheinlichkeit, dass Y dieses Level erreicht.





2.2.2 Definition (Verallgemeinerte Inverse einer Verteilungsfunktion):

Für eine Verteilungsfunktion G definiert man die generalisierte Inverse G^{-1} durch

$$G^{-1}(u) = \inf\{x : G(x) \geq u\}, \qquad 0 < u < 1$$

Ist U gleichverteilt auf (0,1) so hat G^{-1}(U) die Verteilungsfunktion G.

Beweis: Seien $U_1, U_2, \ldots$ eine Folge von unabhängigen und gleichverteilten Zufallsvariablen und $G : \mathbb{R} \to [0,1]$ eine beliebige Verteilungsfunktion. Sei ausserdem $X := G^{-1}(U_i)$, $x \in \mathbb{R}$ und $u \in (0,1)$ dann gilt:

$$u \leq G(x) \qquad \leftrightarrow \qquad G^{-1}(u) \leq x$$

$$P(X \leq x) = P(G^{-1}(U_i) \leq x) = P(U_i \leq G(x)) = G(x)$$

2.2.3 Kopplung von stochastisch geordneten Zufallsvariablen

Unter Verwendung der Coupling-Methode werden wir jetzt folgendes beweisen: Sind X und X' Zufallsvariablen, so dass $X \overset{\mathcal{D}}{\leq} X'$, so existiert eine Kopplung $(\widehat{X}, \widehat{X}')$ von X und X', so dass $\widehat{X} \leq \widehat{X}'$.

Beweis: Sind F und F' die Verteilungsfunktionen von X und X' so gilt $F \geq F'$. Ist $U \sim U(0,1)$, so ist $F^{-1}(U) \leq F'^{-1}(U)$ nach Definition der verallgemeinerten Inversen. Benennen wir nun so um, dass $\widehat{X} = F^{-1}(U)$ und $\widehat{X}' = F'^{-1}(U)$ so haben wir mit $(\widehat{X}, \widehat{X}')$ die gewünschte Kopplung gefunden.

2.3 Geburts- und Sterbeprozesse

Bevor wir uns diesem Beispiel zuwenden, benötigen wir nachfolgende Definitionen:

2.3.1 Definition (Die $\mathcal{O}$-Notation):

Gegeben seien die reelwertigen Funktionen f, g und $x, c \in \mathbb{R}$. Dann ist

$$g \in \mathcal{O}(f), \text{ wenn } 0 \quad \leq \quad \lim_{n\to\infty} \left(\frac{g(x)}{f(x)} \right) = c$$

d.h. g wächst nicht wesentlich schneller als f

$$g \in o(f), \text{ wenn } 0 \quad \leq \quad \lim_{x\to\infty} \left(\frac{g(x)}{f(x)} \right) = 0$$

d.h. g wächst langsamer als f

Kurzes Beispiel: Sei $g(x) = 2x^3 - 2x^2 - 7x - 3$, dann ist $g \in \mathcal{O}(x^3)$, da $\lim\limits_{x\to\infty} \frac{g(x)}{x^3} = 2$ und $g \in o(x^4)$, da $\lim\limits_{x\to\infty} \frac{g(x)}{x^4} = 0$ wie leicht nachzurechnen ist.

2.3.2 Geburts- und Sterbeprozesse

In diesem Beispiel werden zwei Geburts- und Sterbeprozesse miteinander verglichen.
Sei $X = \{X(t); t \geq 0\}$, und $X' = \{X'(t); t \geq 0\}$ zwei Geburts- und Sterbeprozesse. Sei dazu λ_i die Geburtsrate und μ_i die Sterberate von X und

$$P(X(t+dt) - X(t) = +1 | X(t) = i) = \lambda_i dt + o(dt)$$

$$P(X(t+dt) - X(t) = -1 | X(t) = i) = \mu_i dt + o(dt)$$

Ausserdem sei $X(0) = m$. Analog hat X' Geburtsrate λ'_i, die Sterberate μ'_i und den Startwert $X'(0) = m'$. Ist nun $\lambda_i \leq \lambda'_i$ für alle $i \geq 0$ und $\mu_i \geq \mu'_i$ für alle $i \geq 1$, so zeigen wir $X(t)$ ist stochastisch kleiner als $X'(t)$ für alle t (voraussgesetzt $m \leq m'$).

Beweis: Im ersten Schritt definiert man sich einen bivariaten Prozess $(\widehat{X}, \widehat{X}')$ mit den Startwerten (m, m') wie folgt:

von	nach	Rate
(i,j)	$(i+1,j)$	λ_i
	$(i,j+1)$	λ'_i
	$(i-1,j)$	μ_i
	$(i,j-1)$	μ'_i
(i,i)	$(i+1,i+1)$	λ_i
	$(i,i+1)$	$\lambda'_i - \lambda_i$
	$(i-1,i-1)$	μ'_i
	$(i-1,i)$	$\mu_i - \mu'_i$

Der Prozess kann also, wenn er einmal auf der Diagonalen gelandet ist nicht mehr unter die Diagonale laufen. Man sieht, dass die erste Koordinate $\widehat{X}$ verteilt ist wie X und die zweite Koordinate ist verteilt wie X'. Wir haben jetzt also Versionen $\widehat{X}$ und $\widehat{X}'$ der Zufallsvariablen X und X' gefunden für die gilt: $\widehat{X}(t) \leq \widehat{X}'(t)$ für alle t. Insbesondere gilt: $\widehat{X}(t) \overset{\mathcal{D}}{\leq} \widehat{X}'(t)$ für alle t, da letzteres nichts mit dem gewählten Wahrscheinlichkeitsraum zu tun hat.

2.4 Konvergenz von Markov-Ketten

2.4.1 Einführung einiger Begriffe

Es sei $X = (X(0), X(1), \ldots)$ eine Markov-Kette. Folgende Begriffe sind wichtig:

- *periodisch/aperiodisch*: Ein Zustand i einer Markov-Kette hat die Periode k, wenn gilt: $k = \min(k \in \mathbb{N}$: *Zustand i ist von allen Zuständen aus in einem Vielfachen von k-Schritten erreichbar*). Ein Zustand i heisst aperiodisch wenn $k = 1$ ist.
- *rekurrent/transient*: Ein Zustand i heisst rekurrent, wenn die Wahrscheinlichkeit, dass dieser Zustand unendlich oft eintritt gleich 1 ist. Andernfalls heisst der Zustand transient.
- *irreduzibel*: Eine Markov-Kette heisst irreduzibel, wenn sie sich nicht in Teilketten zerlegen lassen kann. Anschaulich bedeutet das, jeder Zustand ist von jedem Zustand aus erreichbar.
- *stationäre Verteilung*: Eine Verteilung π heisst stationär, falls für alle j gilt $\pi_j = \sum_{i \in S} \pi_i p_{ij}$. Die Wahrscheinlichkeit einen bestimmten Zustand zu erreichen, hängt also nicht von der Zeit ab.

2.4.2 Kopplung von Markov-Ketten

Im letzten Beispiel über die Coupling-Methode beweisen wir, dass eine irreduzible, aperiodische Markov-Kette in einem endlichen m-dimensionalen Raum Stationarität erreicht, unabhängig von der Anfangsverteilung.
Sei $X = (X_0, X_1, \ldots)$ eine irreduzible, aperiodische Markovkette, P sei die dazugehörige Übergangsmatrix und λ die Anfangsverteilung. Ausserdem sei $\pi = (\pi_1, \pi_2, \ldots)$ die eindeutige, positive stationäre Verteilung, so dass $\pi = \pi P$. Wir werden jetzt beweisen, dass $P(X_\nu = y) \to \pi_y$ für $\nu \to \infty$, für alle festen $y = 1, \ldots, m$.

Beweis: Dieser Beweis ist ein klassischer Coupling-Beweis.
Sei nun $X' = (X'_0, X'_1, \ldots)$ unabhängig von X, mit der Übergangsmatrix P und Anfangsverteilung π. Somit ist X' stationär. Für ein festes y definiere $\widehat{X} = (\widehat{X}_0, \widehat{X}_1, \ldots)$ durch

$$\widehat{X}(\nu) = \begin{cases} X_\nu & \text{für } T > \nu \\ X'_\nu & \text{für } T \leq \nu \end{cases}$$

wobei $T = \min\{\nu \geq 0 : X_\nu = X'_\nu\}$, also ist T der Zeitpunkt an dem sich X und X' das erstemal treffen und es gilt $T < \infty$, da X irreduzibel und aperiodisch.
Zuerst wird nun bewiesen, dass $P(X_\nu = y | T \leq \nu) = P(\widehat{X}_\nu = y | T \leq \nu)$. Das ist aber gleichbedeutend mit $P(X_\nu = y | T \leq \nu) = P(X'_\nu = y | T \leq \nu)$.
Da aber gilt:

$$P(X_\nu = y | T \leq \nu) = \frac{P(X_\nu = y, T \leq \nu)}{P(T \leq \nu)}$$

und

$$P(X'_\nu = y | T \le \nu) = \frac{P(X'_\nu = y, T \le \nu)}{P(T \le \nu)}$$

genügt es $P(X_\nu = y, T \le \nu) = P(X'_\nu = y, T \le \nu)$ zu zeigen.

$$\begin{aligned}
P(X_\nu = y, T \le \nu) &= \sum_{m=1}^{\nu} \sum_x P(T = m, X_m = x, X_\nu = y) \\
&\overset{\text{totale Wkt.}}{=} \sum_{m=1}^{\nu} \sum_x P(T = m, X_m = x) P(X_\nu = y | X_m = x, T = m) \\
&= \sum_{m=1}^{\nu} \sum_x P(T = m, X'_m = x) P(X'_\nu = y | X'_m = x, T = m) \\
&= \sum_{m=1}^{\nu} \sum_x P(T = m, X'_m = x, X'_\nu = y) \\
&= P(X'_\nu = y, T \le \nu)
\end{aligned}$$

Damit gilt jetzt:

$$\begin{aligned}
P(X_\nu = y) &= P(X'_\nu = y, T \le \nu) + P(X_\nu = y, T > \nu) \\
&\le P(X'_\nu = y) + P(X_\nu = y, T > \nu)
\end{aligned}$$

analog gilt für $X' : P(X'_\nu = y) \le P(X_\nu = y) + P(X'_\nu = y, T > \nu)$.
Daraus folgt

$$\begin{aligned}
|P(X_\nu = y) - P(X'_\nu = y)| &= |P(X'_\nu = y, T \le \nu) + P(X_\nu = y, T > \nu) \\
&\quad -(P(X_\nu = y, T \le \nu) + P(X'_\nu = y, T > \nu))| \\
&\le P(X_\nu = y, T > \nu) + P(X'_\nu = y, T > \nu)
\end{aligned}$$

Summiert man nun über y auf so ergibt sich:

$$\begin{aligned}
\sum_y |P(X_\nu = y) - P(X'_\nu = y)| &\le \sum_y \big(P(X_\nu = y, T > \nu) + P(X'_\nu = y, T > \nu)\big) \\
&= 2P(T > \nu)
\end{aligned}$$

Sei nun $X_0 = x$ und X'_0 hat nach Vorraussetzung die stationäre Verteilung π, dann folgt das gewünschte Ergebnis mit der Kopplung $\widehat{X}' = X'$ und $\widehat{X}$ wie oben definiert.

$$\sum_y |p^\nu(x, y) - \pi_y| \le 2P(T > \nu) \to 0 \text{ für } \nu \to \infty$$

3 Kopplung

Die formale Definition des Begriffs Kopplung kommt zuletzt, da sie ohne Beispiele schwer verständlich ist.
Es seien im Folgenden $(\Omega, \mathcal{F}, P)$ und $(\Omega', \mathcal{F}', P')$ Wahrscheinlichkeitsräume über E. Hierbei kann $E \in \mathbb{N}$, $E \in \mathbb{R}$, usw.

Definition 1:
Unter der Kopplung von Zuvallsvariablen $X\colon \Omega \to E$ und $X'\colon \Omega' \to E$ versteht man einen Wahrscheinlichkeitsraum $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P})$ und die Zuvallsvariablen $(\widehat{X}, \widehat{X}')$: $\widehat{\Omega} \to E^2$, so dass

$$X \overset{\mathcal{D}}{=} \widehat{X} \qquad \text{und} \qquad X' \overset{\mathcal{D}}{=} \widehat{X}'$$

Definition 2:
Seien μ, ν Verteilungen auf einer endlichen Menge Ω. Eine Verteilung ω auf $\Omega \times \Omega$ heisst Kopplung, falls gilt:

$$\begin{aligned} \forall x \in \Omega, \qquad & \sum_{y \in \mathbb{R}} \omega(x, y) = \mu(x) \\ \forall y \in \Omega, \qquad & \sum_{x \in \mathbb{R}} \omega(x, y) = \nu(y) \end{aligned}$$

Mit anderen Worten: ω ist eine gemeinsame Verteilung, deren Randverteilungen μ und ν sind.

4 Abschluss: Ein Zahlentrick

Als Abschluss ein sehr anschauliches Beispiel für das Coupling folgender Zahlentrick, den Eugene B. Dynkin seinen Studenten präsentierte. Ein Student wurde zufällig ausgewählt, der dann willkürlich 100 Zahlen zwischen 0 und 9 an die Tafel schrieb. Ein anderer Student wählte eine der ersten zehn Zahlen aus, die er aber für sich behielt. War diese Zahl zum Beispiel 5 dann zählte der Student von dort aus 5 Zahlen weiter, notierte sich die an dieser Stelle stehende Zahl und fuhr dann so fort. Wenn der Student zu einer 0 kam zählte er 10 Zahlen vorwärts. Eine mögliche Folge wären also die unterstrichenen in der folgenden Reihe:

$$2\ 9\ \underline{5}\ 7\ 0\ 2\ 1\ \underline{0}\ 3\ 8\ 6\ 1\ 9\ 4\ 5\ 6\ 2\ \underline{3}\ 8\ 4\ \underline{2}\ 4\ \underline{7}\ 6\ 9\ 2\ 7\ 4\ \ldots$$

Der Trick besteht nun darin, dass Dynkin, ohne zu wissen welche Zahl der Student am Anfang ausgewählt hatte die letzte Zahl bestimmen konnte auf die der Student beim Zählen kam.
Um auf diese Zahl zu kommen, wählte sich Dynkin ebenfalls eine Zahl unter den ersten zehn aus und bildete auf die Gleiche Weise die Zahlenreihe. Durch dieses Vorgehen findet ab einem Zeitpunkt T die Kopplung statt, ab der dann die zwei Zahlenreihen zusammen laufen. Einer von Dynkins Studenten berechnete die Wahrscheinlichkeit, dass dieser Trick schief geht. Sie betrug 0,026.

5 Ausblick

Wir haben nun anhand verschiedener Beispiele die Coupling-Methode kennengelernt. Nachdem die einführenden Beispiele zum Kennenlernen der Coupling-Methode gedacht waren, zeigte das letzte Beispiel einen längeren und schwierigeren Beweis mittels der Coupling-Methode.
Weiterführend wäre es jetzt interessant zu sehen, wie die Coupling-Methode in der Epidemiologie angewandt werden kann und welche Erkenntnisse sich dadurch beweisen lassen.

Literatur

[1] H.Andersson, T.Britton,*Stochastic Epidemic Models and Their Statistical Analysis*.LNS,Springer,2000.

[2] Torgny Lindvall, *Lectures on the Coupling Method*,Dover Pubn Inc; 2002.

[3] Richard Durrett, *Probability: Theory and Examples*, Duxbury Press; 1995.

[4] Eric Vigoda , *Markov Chains, Coupling, Stationary Distribution*, http://www.cc.gatech.edu/~vigoda/MCMC_Course/MC-basics.pdf, 21.04.11;

[5] Sanjay K. Bose, *Stochastic Processes Markov Processes and Markov Chains Birth Death Processes*,http://www.iitg.ac.in/skbose/qbook/Slide_Set_2.PDF, 26.04.11

[6] Yevgeniy Kovchegov, *From Markov Chains to Gibbs Fields.*, http://www.yaroslavvb.com/papers/kovchegov-from.pdf, 20.04.11