Kapitel 5: Kopplungen

Benedikt Schmidt

01. November 2021

1 Einleitung

Dieses Handout basiert auf Kapitel 5 (Kopplungen) in Markow Chains and Mixing Times [1]. Das Hauptziel besteht darin, eine effektive Methode zur Verfügung zu stellen, die es erlaubt, obere Schranken für die Totalvariationsmetrik anzugeben. Das ist nützlich, um Schranken für die Distanz zur Stationarität von Markow Ketten zu erhalten. Zuerst werden wir ein Einführungsbeispiel anschauen, das den Zweck illustriert. Dann werden wichtige Konzepte definiert, mit deren Hilfe wir beweisen können, dass es eine Schranke für die Totalvariationsmetrik gibt. Zum Schluss wird das Beispiel der zufälligen Irrfahrt (random walk) auf dem Kreis diskutiert und gezeigt, wie die Totalvariationsmetrik nach oben beschränkt werden kann. Wichtige Begriffe sind die folgenden:

- Kopplung von Wahrscheinlichkeitsverteilungen
- Totalvariationsmetrik
- Kopplung von Markow Ketten
- Markow'sche Kopplung

1.1 Eine einfache zufällige Irrfahrt

Als einführendes Beispiel betrachten wir eine einfache stochastische Irrfahrt auf dem Segment $\{0, 1, ..., n\}$. Das ist eine Markow-Kette. Sie bewegt sich entweder nach oben oder nach unten bei jeder Bewegung mit derselben Wahrscheinlichkeit. Die Irrfahrt kann das Intervall nicht verlassen.

Seien nun $0 \leq x \leq y \leq n$. Seien $(X_0, X_1, ...)$ und $(Y_0, Y_1, ...)$ Markow Ketten mit Übergangsmatrix P. Die Wahrscheinlichkeit, in t Schritten von x nach n zu gelangen ist gegeben durch

$$P_x(X_t = n) = P^t(x, n) \tag{1}$$

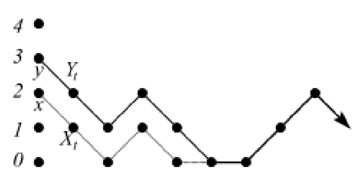

Abbildung 1: Gekoppelte zufällige Irrfahrten auf $\{0, 1, ..., n\}$. Die Irrfahrten trennen sich nicht mehr nach Zusammentreffen.

Entsprechend bezeichnen wir die Wahrscheinlichkeit, in t Schritten von y nach n zu gelangen mit

$$P_y(Y_t = n) = P^t(y, n) \tag{2}$$

Es ist intuitiv klar dass $P^t(x, n) \leq P^t(y, n)$. Ein Beweis für diese Intuition kann durch Kopplung der beiden Verteilungen $P^t(x, \cdot)$ und $P^t(y, \cdot)$ erbracht werden. Hierfür sei $\Delta_1, \Delta_2, ...$ eine Folge von i.i.d. Zufallsvariabeln mit Mittelwert null so dass die Δ_i nur Werte in $\{-1, 1\}$ annehmen. Weiter seien (X_t) und (Y_t) zufällige Irrfahrten, die auf demselben Wahrscheinlichkeitsraum definiert sind mit Startpunkten x respektive y. Die Bewegungen in beiden Ketten folgt dem Wert von Δ_t: für $\Delta_t = 1$ bewegt sich die Irrfahrt der Kette entlang nach oben. Für $\Delta_t =$ -1, bewegt sich die Irrfahrt der Kette entlang nach unten. Da beide Ketten jeweils eine Bewegung in dieselbe Richtung machen (nach oben oder nach unten), können sie nur an den Enden zusammentreffen (0 or n). Aus demselben Grund gilt, dass falls $x \leq y$, dann $X_t \leq Y_t$ für alle t. Falls X den oberen Endpunkt erreicht, also $X_t = n$, dann muss Y schon dort sein, also $Y_t = n$. Daher gilt

$$P^t(x, n) = P_x(X_t = n) \leq P_y(Y_t = n) = P^t(y, n) \tag{3}$$

2 Definitionen

Mit diesem Einführungsbeispiel im Hinterkopf wollen wir nun die Idee der Kopplung formalisieren.

Definition 2.1 (Kopplung von Wahrscheinlichkeitsverteilungen)**.** *Eine Kopplung von zwei Wahrscheinlichkeitsverteilungen μ und ν ist ein Paar von Zufallsvariabeln (X, Y) definiert auf einem gemeinsamen Wahrscheinlichkeitsraum so dass μ der Randverteilung von X entspricht und ν derjenigen von Y.*

Also erfüllt eine Kopplung $P(X = x) = \mu(x)$ und $P(Y = y) = \nu(y)$. Um unsere Intuition zu schärfen, können wir ein anderes Beispiel machen: Seien μ und ν beide das sogenannte "Faire Münze-Mass". Dieses Mass gewichtet jeweils mit 1/2 die beiden Elemente von $\{0, 1\}$. Eine Möglichkeit, μ und ν zu koppeln, besteht darin, (X, Y) als Paar von unabhängigen Münzen zu definieren. Dann gilt $P(X = x, Y = y) = 1/4$ für alle $x, y \in \{0, 1\}$. Mittels Kopplung kann ein Vergleich von Verteilungen durch einen Vergleich von Zufallsvariabeln ersetzt werden. Auf ähnliche Weise können wir für zwei gegebene Prozesse deren gesamte Trajektorien koppeln:

Definition 2.2 (Kopplung von Markow Ketten)**.** *Seien (X_t) und (Y_t) Markow Ketten mit Übergangsmatrix P. Dann ist ein Prozess $(X_t, Y_t)_{t\geq 0}^{\infty}$ eine Kopplung von Markow Ketten mit Übergangsmatrix P.*

Kopplung von Markow Ketten ist nicht zu verwechseln mit Markow‘scher Kopplung.

Definition 2.3 (Markow‘sche Kopplung)**.** *Sei P eine Übergangsmatrix auf einem endlichen Zustandsraum $\mathcal{X}$. Eine Markow‘sche Kopplung von zwei P-Ketten ist eine Markow Kette $(X_t, Y_t)_{t\geq 0}$ mit Zustandsraum $\mathcal{X} \times \mathcal{X}$ welche für alle x, y, x', y'*

$$P(X_{t+1} = x' \mid X_t = x, Y_t = y) = P(x, x') \tag{4}$$

und

$$P(Y_{t+1} = y' \mid X_t = x, Y_t = y) = P(y, y') \tag{5}$$

erfüllt.

Gegeben eine Markow'sche Kopplung von Markow Ketten mit Übergangsmatrix P ist es immer möglich, die Kopplung so zu modifizieren, dass die beiden Ketten zusammen bleiben, nachdem sie sich ein einziges Mal getroffen hatten. Formalisiert bedeutet das

$$X_s = Y_s \implies X_t = Y_t \quad \forall\, t \geq s \tag{6}$$

Um das zu erreichen, müssen die Ketten laufen gelassen werden, bis sie zusammentreffen und ab dann zusammen laufen gelassen werden. Das führt auf den Begriff der Verschmelzungszeit:

Definition 2.4 (Verschmelzungszeit)**.** *Für eine Kopplung (X_t, Y_t) ist die Verschmelzungszeit τ_{kopp} definiert als*

$$\tau_{kopp} := \min\{t \geq 0\ : X_s = Y_s \ \forall\ s \geq t\} \tag{7}$$

3 Beschränkung der Totalvariationsmetrik

Für beliebige zwei Verteilungen μ und ν gibt es eine unabhängige Kopplung. Wenn jedoch μ und ν nicht identisch sind, ist es nicht immer möglich, dass X und Y jeweils denselben Wert annehmen. Wie nahe kann eine Kopplung an den Zustand gelangen, in dem X und Y identisch sind? Die Totalvariationsmetrik gibt die Antwort.

Definition 3.1 (Totalvariationsmetrik)**.** *Für zwei Wahrscheinlichkeitsverteilungen μ und ν auf $\mathcal{X}$ ist die Totalvariationsmetrik definiert als*

$$||\mu - \nu||_{TV} = \max_{A \subset \mathcal{X}} |\mu(A) - \nu(A)| \tag{8}$$

Diese Definition ist explizit stochastisch: Die Distanz zwischen μ und ν entspricht der maximalen Differenz zwischen den Wahrscheinlichkeiten, die den verschiedenen Ereignissen durch die beiden Verteilungen zugeordnet werden. Für späteren Gebrauch sei folgende wichtige Gleichung festgehalten (der entsprechende Beweis kann nachvollzogen werden in [1, Prop. 4.7]). Seien μ und ν zwei Wahrscheinlichkeitsverteilungen auf $\mathcal{X}$. Dann gilt

$$||\mu - \nu||_{TV} = \inf\{\mathbf{P}(X \neq Y) : (X, Y) \text{ ist eine Kopplung von } \mu \text{ und } \nu\} \tag{9}$$

Nun können wir zum zentralen Resultat voranschreiten, wonach die Totalvariationsmetrik beschränkt werden kann. Wir möchten die Notation wie folgt vereinfachen: für die beiden Startpunkte x und y von zwei Markow Ketten X und Y schreiben wir $\mathbf{P}_{x,y}$ für die Wahrscheinlichkeit auf demjenigen Raum, auf dem (X_t) und (Y_t) definiert sind .

Theorem 3.1. *Sei (X_t, Y_t) eine Kopplung, die (6) erfüllt, so dass $X_0 = x$ und $Y_0 = y$. Sei τ_{kopp} die entsprechende Verschmelzungszeit. Dann gilt*

$$||P^t(x, \cdot) - P^t(y, \cdot)||_{TV} \leq \boldsymbol{P}_{x,y}(\tau_{kopp} > t) \tag{10}$$

Beweis. Es gilt dass $P^t(x, z) = \mathbf{P}_{x,y}(X_t = z)$ und $P^t(y, z) = \mathbf{P}_{x,y}(Y_t = z)$. Somit ist (X_t, Y_t) eine Kopplung von $P^t(x, \cdot)$ mit $P^t(y, \cdot)$. Nun wenden wir Gleichung (9) an und erhalten

$$||P^t(x, \cdot) - P^t(y, \cdot)||_{TV} \leq \mathbf{P}_{x,y}(X_t \neq Y_t) \tag{11}$$

Mit Gleichung (6) sehen wir, dass

$$\mathbf{P}_{x,y}(X_t \neq Y_t) = \mathbf{P}_{x,y}(\tau_{kopp} > t), \tag{12}$$

was die Behauptung impliziert. $\square$

Abbildung 2: Zufällige Irrfahrt auf $\mathbb{Z}_{10}$ (links) und auf $\mathbb{Z}_9$ (rechts)

Nun möchten wir dieses Resultat in Beziehung zu unserem Einführungsbeispiel setzen. Damals wollten wir eine Schranke für die Distanz zwischen der Startposition x_0 und einem der beiden Grenzzustände (entweder 1 oder n) finden. Diese Schranke entspricht formell der Schranke für die maximale Distanz zwischen $P^t(x_0, \cdot)$ und der stationären Verteilung π:

$$d(t) := \max_{x \in \mathcal{X}} ||P^t(x, \cdot) - \pi||_{TV} \tag{13}$$

Es ist hilfreich, die sogenannte *Mischzeit* zu definieren:

$$t_{mix}(\epsilon) := \min \{t : d(t) \leq \epsilon\} \tag{14}$$

Die Mischzeit beschreibt also die Zeit, die eine Markow Kette benötigt, um den Weg bis zur Stationarität zu beschreiten. Dabei ist es nützlich, einen Parameter einzuführen, der dafür sorgt, dass die Mischzeit klein ist. Deshalb setzten wir (arbitär)

$$t_{mix} := t_{mix}(1/4). \tag{15}$$

Nun können wir von Theorem 3.1 das folgende Korollar ableiten:

Korollar 3.1.1. *Angenommen für jedes Zustandspaar $x, y \in X$ gäbe es eine Kopplung (X_t, Y_t) mit $X_0 = x$ und $Y_0 = y$. Für jede solche Kopplung, sei τ_{kopp} die entsprechende Verschmelzungszeit, definiert wie in 2.4. Dann gilt:*

$$d(t) \leq \max_{x,y \in \mathcal{X}} \boldsymbol{P}_{x,y}(\tau_{kopp} > t) \tag{16}$$

und deshalb $t_{mix} \leq 4 \max_{x,y} \boldsymbol{E}_{x,y}[\tau_{kopp}]$.

4 Zufällige Irrfahrt auf dem diskreten Kreis

Wir wollen eine obere Schranke für t_{mix} in der zufälligen Irrfahrt auf dem Kreis finden. Dies ist gleichbedeutend dazu, eine obere Schranke für die Totalvariationsmetrik zwischen $P^t(x, \cdot)$ und der stationären Verteilung π zu finden. Der dieser Irrfahrt zugrundeliegende Graph ist $\mathbb{Z}_n$ mit den Eckpunkten $\{1, 2, ..., n\}$. Eine Kante zwischen j und k besteht jeweils dann, wenn $j \equiv k \pm 1$. Die Irrfahrt verbleibt in ihrer aktuellen Position mit Wahrscheinlichkeit $1/2$, bewegt sich im Uhrzeigersinn mit Wahrscheinlichkeit $p/2$, und im Gegenuhrzeigersinn mit Wahrscheinlichkeit $q/2$, wobei $p + q = 1$.

Wir wollen zeigen, dass $t_{mix} \leq n^2$. Zu diesem Zweck konstruieren wir eine Kopplung (X_t, Y_t) von zwei Ketten, welche zufällig auf $\mathbb{Z}_n$ umher irren. Ein Teilchen startet bei x, das andere bei

y. Die Bewegung eines Paares gestaltet sich so, dass sich jeweils ein Teilchen nach dem anderen bewegt. Dadurch kann verhindert werden, dass die beiden Teilchen in einem einzigen, simultanen Sprung ihren Platz wechseln. Um zu bestimmen, welches der beiden Teilchen springt, wird zu jeder Zeiteinheit eine faire Münze geworfen. Jeder Münzwurf ist unabhängig von den vorausgegangenen Münzwürfen. Dies geschieht so lange, bis sich die beiden Teilchen an derselben Stelle befinden. Das durch den Münzwurf bestimmte Teilchen befolgt eine Inkrementierung im Uhrzeigersinn mit Wahrscheinlichkeit p und eine Inkrementierung im Gegenuhrzeigersinn mit Wahrscheinlichkeit q. Sobald sich die beiden Teilchen an derselben Stelle befinden, machen sie nur noch identische Bewegungen. Sei nun D_t die Distanz zwischen X_t zu Y_t im Uhrzeigersinn . Der Prozess (D_t) ist dann auch eine Irrfahrt auf $\{0, 1, 2, ..., n\}$. Zudem sei vorausgesetzt, dass (D_t) stoppt an den Positionen 0 und n. Setze $\tau = \min\{t \geq 0 : D_t \in \{0, n\}\}$. Dann folgt (siehe [1, Prop. 2.1], S.21) $\mathbf{E}_{x,y}(\tau) = k(n - k)$. Hierbei bezeichnet k die Distanz im Uhrzeigersinn zwischen x und y. Per Definition von τ gilt $\tau = \tau_{kopp}$. Mit Korollar 3.1.1 folgt

$$d(t) \leq \max_{x,y \in \mathbb{Z}_n} \mathbf{P}_{x,y}(\tau > t) \leq \frac{\max_{x,y} \mathbf{E}_{x,y}(\tau)}{t} \leq \frac{n^2}{4t} \tag{17}$$

Die rechte Seite wird zu 1/4 falls $t = n^2$. Mit Gleichung (14) folgt dann $t_{mix} \leq n^2$.

Bibliographie

[1] David Asher Levin, Yuval Peres und Elizabeth L. Wilmer. *Markov chains and mixing times*. Second edition. Providence, Rhode Island: American Mathematical Society, 2017. 447 S. ISBN: 978-1-4704-2962-1.