Technische Universität Berlin

Thema: Totalvariationabstand, Kopplung und Mischzeiten

Ausarbeitung zum Seminar Moderne Anwendung von Markovketten

Name:	Nina Kamswich
Matrikelnummer:	332677
Fachbereich:	Mathematik
	Prof. Dr. König
	Dr. Fackeldey

Inhaltsverzeichnis

1 Totalvariationsabstand

Dieser Ausarbeitung geht das Paper "Markov Chains and Mixing Times" von David A. Levin, Yuval Peres und Elizabeth L. Wilmer voraus, in dem unter anderem das Langzeitverhalten endlicher Markovketten betrachtet wird. Dabei ist auch die Konvergenzgeschwindigkeit von Familien von Markovketten interessant. Doch um von einer Konvergenz reden zu können, benötigen wir eine Metrik, um die Distanz zwischen Verteilungen zu messen. Mit dieser Metrik, dem Totalvariationsabstand, werden wir uns in diesem Abschnitt beschäftigen .

Definition.
Seien μ und ν Wahrscheinlichkeitsmaße auf einem endlichen Zustandsraum Ω, dann heißt

$$||\mu - \nu||_{\mathrm{TV}} = \max_{\mathrm{A} \subset \Omega} |\mu(\mathrm{A}) - \nu(\mathrm{A})|$$

Totalvariationsabstand zwischen μ und ν.

Da man bei dieser Definition das Maximum über alle Teilmengen von Ω bildet, ist sie sehr unhandlich in der Anwendung, da die Potenzmenge im Allgemeinen sehr groß ist. Deswegen werden wir nun drei weitere äquivalente Darstellungen des Totalvariationsabstands kennenlernen.
Als erste äquivalente Darstellung erhalten wir eine einfache Summe über den Zustandsraum:

Satz 1.

$$||\mu - \nu||_{\mathrm{TV}} = \frac{1}{2} \sum_{\mathrm{x} \in \Omega} |\mu(\mathrm{x}) - \nu(\mathrm{x})|$$

Beweis. Wir betrachten die geschickt gewählte Menge $\mathrm{B} := \{\mathrm{x} : \mu(\mathrm{x}) \geq \nu(\mathrm{x})\}$ und sei $\mathrm{A} \subset \Omega$ beliebig. Es ist leicht zu sehen, dass $\mathrm{A} = (\mathrm{A} \cap \mathrm{B}) \cup (\mathrm{A} \cap \mathrm{B}^c)$ und analog $\mathrm{B} = (\mathrm{A} \cap \mathrm{B}) \cup (\mathrm{A}^c \cap \mathrm{B})$ gelten. Damit folgt:
$\mu(\mathrm{A}) = \mu(\mathrm{A} \cap \mathrm{B}) + \mu(\mathrm{A} \cap \mathrm{B}^c)$ und $\nu(\mathrm{A}) = \nu(\mathrm{A} \cap \mathrm{B}) + \nu(\mathrm{A} \cap \mathrm{B}^c)$.
Damit erhalten wir
$\mu(\mathrm{A}) - \nu(\mathrm{A}) = \mu(\mathrm{A} \cap \mathrm{B}) - \nu(\mathrm{A} \cap \mathrm{B}) + \underbrace{\mu(\mathrm{A} \cap \mathrm{B}^c) - \nu(\mathrm{A} \cap \mathrm{B}^c)}_{\leq 0, \text{ da } \mu(\mathrm{x}) < \nu(\mathrm{x}) \ \forall \ \mathrm{x} \in \mathrm{B}^c} \leq \mu(\mathrm{A} \cap \mathrm{B}) - \nu(\mathrm{A} \cap \mathrm{B})$
und
$\mu(\mathrm{B}) - \nu(\mathrm{B}) = \mu(\mathrm{A} \cap \mathrm{B}) - \nu(\mathrm{A} \cap \mathrm{B}) + \underbrace{\mu(\mathrm{A}^c \cap \mathrm{B}) - \nu(\mathrm{A}^c \cap \mathrm{B})}_{\geq 0, \text{ da } \mu(\mathrm{x}) \geq \nu(\mathrm{x}) \ \forall \ \mathrm{x} \in \mathrm{B}} \geq \mu(\mathrm{A} \cap \mathrm{B}) - \nu(\mathrm{A} \cap \mathrm{B})$
Insgesamt folgt nun

$$\mu(\mathrm{A}) - \nu(\mathrm{A}) \leq \mu(\mathrm{A} \cap \mathrm{B}) - \nu(\mathrm{A} \cap \mathrm{B}) \leq \mu(\mathrm{B}) - \nu(\mathrm{B}) \tag{1}$$

Anlog gilt:

$$\nu(\mathrm{A}) - \mu(\mathrm{A}) \leq \nu(\mathrm{B}^c) - \mu(\mathrm{B}^c)$$

$$\begin{aligned} \Rightarrow -(\mu(\mathrm{A}) - \nu(\mathrm{A})) &= \nu(\mathrm{A}) - \mu(\mathrm{A}) \\ &\leq \nu(\mathrm{B}^c) - \mu(\mathrm{B}^c) \\ &= 1 - \nu(\mathrm{B}) - 1 + \mu(\mathrm{B}) \\ &\overset{(*)}{=} \mu(\mathrm{B}) - \nu(\mathrm{B}) \end{aligned} \tag{2}$$

$$(1), (2) \Rightarrow |\mu(\mathrm{A}) - \nu(\mathrm{A})| \leq \mu(\mathrm{B}) - \nu(\mathrm{B})$$

$\Rightarrow \mu(\mathrm{B}) - \nu(\mathrm{B})$ ist eine obere Schranke für $||\mu - \nu||_{\mathrm{TV}}$.
Wenn wir die Ungleichung nun für A=B betrachten, wird diese obere Schranke mit Gleichheit erfüllt und es folgt:

$$\begin{aligned} ||\mu - \nu||_{\mathrm{TV}} = \mu(\mathrm{B}) - \nu(\mathrm{B}) &\overset{\text{mit}(*)}{=} \frac{1}{2}\left(\mu(\mathrm{B}) - \nu(\mathrm{B}) + \nu(\mathrm{B}^c) - \mu(\mathrm{B}^c)\right) \\ &= \frac{1}{2}\Big(\sum_{\mathrm{x}\in\mathrm{B}}(\mu(\mathrm{x}) - \nu(\mathrm{x})) + \sum_{\mathrm{x}\in\mathrm{B}^c}(-(\mu(\mathrm{x}) - \nu(\mathrm{x}))\Big) \\ &= \frac{1}{2}\sum_{\mathrm{x}\in\Omega}|\mu(\mathrm{x}) - \nu(\mathrm{x})| \end{aligned} \tag{3}$$

□

Bemerkung 1.
Aus der letzten Gleichung im Beweis und der Definition von B folgt, dass es genügt die Summe über B für den Totalvariationsabstand zu bilden:

$$||\mu - \nu||_{\mathrm{TV}} = \sum_{\mathrm{x}:\mu(\mathrm{x})>\nu(\mathrm{x})} \mu(\mathrm{x}) - \nu(\mathrm{x}) = \sum_{\mathrm{x}:\mu(\mathrm{x})<\nu(\mathrm{x})} \nu(\mathrm{x}) - \mu(\mathrm{x})$$

Mit dieser Darstellung können wir nun zeigen, dass der Totalvariationsabstand in der Tat eine Metrik ist:

Satz 2.
Sei P die Menge der Wahrscheinlichkeitsmaße über Ω. Dann ist d: P$\times$ P $\to \mathbb{R}$, $\mathrm{d}(\mu,\nu) = ||\mu - \nu||_{\mathrm{TV}}$ eine Metrik.

Beweis. Die meisten Metrikeigenschaften ergeben sich dabei direkt aus den Eigenschaften des Betrags:

1) $||\mu - \nu||_{\mathrm{TV}} \geq 0$ klar, da der Betrag einer Differenz immer größer 0 ist und somit auch das Maximum dieses Betrags größer 0 ist.

2) Symmetrie klar, da sich das Ergebnis der Differenz mit veränderter Summationsreihenfolge durch die Eigenschaften des Betrags nicht ändert.

3) Sei $||\mu - \nu||_{\mathrm{TV}} = 0$:
Sei $|\mu(\mathrm{A}) - \nu(\mathrm{A})| = 0$ für alle $\mathrm{A} \subset \Omega$. Daraus folgt sofort, dass $\mu = \nu$, da die Gleichung für alle Teilmengen von Ω erfüllt ist.
Andererseits sei nun $\mu = \nu$ und offensichtlich gilt damit $||\mu - \nu||_{\mathrm{TV}} = 0$.

4) Dreiecksungleichung:

$$\begin{aligned}
||\mu - \nu||_{\mathrm{TV}} &= \frac{1}{2} \sum_{\mathrm{x} \in \Omega} |\mu(\mathrm{x}) - \nu(\mathrm{x})| \\
&= \frac{1}{2} \sum_{\mathrm{x} \in \Omega} |\mu(\mathrm{x}) - \xi(\mathrm{x}) + \xi(\mathrm{x}) - \nu(\mathrm{x})| \\
&\leq \frac{1}{2} \sum_{\mathrm{x} \in \Omega} (|\mu(\mathrm{x}) - \xi(\mathrm{x})| + |\xi(\mathrm{x}) - \nu(\mathrm{x})|) \\
&= ||\mu - \xi||_{\mathrm{TV}} + ||\xi - \nu||_{\mathrm{TV}}
\end{aligned} \tag{4}$$

□

Satz 3.

$$||\mu - \nu||_{\mathrm{TV}} = \frac{1}{2} \sup \left\{ \sum_{\mathrm{x} \in \Omega} \mathrm{f}(\mathrm{x})(\mu(\mathrm{x}) - \nu(\mathrm{x})) | \mathrm{f} : \Omega \to \mathbb{R} : \max_{\mathrm{x} \in \Omega} |\mathrm{f}(\mathrm{x})| \leq 1 \right\}$$

Beweis. Wir führen den Beweis, in dem wir die beiden Beziehungen $\geq$ und $\leq$ zeigen:
$\geq$: Sei $\mathrm{f} : \Omega \to \mathbb{R}$ beliebig mit $\max_{\mathrm{x} \in \Omega} |\mathrm{f}(\mathrm{x})| \leq 1$. Dann gilt:

$$\begin{aligned}
\frac{1}{2} \sum_{\mathrm{x} \in \Omega} (\mathrm{f}(\mathrm{x})\mu(\mathrm{x}) - \mathrm{f}(\mathrm{x})\nu(\mathrm{x})) &\leq \frac{1}{2} | \sum_{\mathrm{x} \in \Omega} (\mathrm{f}(\mathrm{x})\mu(\mathrm{x}) - \mathrm{f}(\mathrm{x})\nu(\mathrm{x}))| \\
&\leq \frac{1}{2} \sum_{\mathrm{x} \in \Omega} |(\mathrm{f}(\mathrm{x})(\mu(\mathrm{x}) - \nu(\mathrm{x})))| \\
&\leq \frac{1}{2} \sum_{\mathrm{x} \in \Omega} |\mu(\mathrm{x}) - \nu(\mathrm{x})| \\
&= ||\mu - \nu||_{\mathrm{TV}}
\end{aligned} \tag{5}$$

Da diese Ungleichungskette für alle f, die die Bedingung $\max_{\mathrm{x} \in \Omega} |\mathrm{f}(\mathrm{x})| \leq 1$ erfüllen, gilt, gilt sie auch für das Supremum über alle f.
$\leq$: Wir konstruieren nun ein f^*, welches (5) mit Gleichheit erfüllt. Wenn wir ein solches f finden, sind wir fertig.
Sei also

$$\mathrm{f}^*(\mathrm{x}) = \begin{cases} 1, & \mu(\mathrm{x}) \geq \nu(\mathrm{x}) \\ -1, & \mu(\mathrm{x}) < \nu(\mathrm{x}) \end{cases}$$

Dann gilt:

$$\frac{1}{2}\sum_{x\in\Omega}(f^*(x)(\mu(x)-\nu(x))) = \frac{1}{2}(\underbrace{\sum_{x:\mu(x)\geq\nu(x)}(\mu(x)-\nu(x))}_{=||\mu-\nu||_{TV}} + \underbrace{\sum_{x:\mu(x)<\nu(x)}(\nu(x)-\mu(x))}_{=||\mu-\nu||_{TV}})$$
$$= ||\mu-\nu||_{TV}$$

□

2 Kopplung von Wahrscheinlichkeitsmaßen

In diesem Abschnitt lernen wir das Prinzip der Kopplung zwischen Wahrscheinlichkeitsmaßen kennen, um die letzte Darstellung des Totalvariationsabstands zu erhalten.

2.1 Definition und Beispiele

Definition.
Seien μ und ν Wahrscheinlichkeitsmaße auf Ω und X,Y Zufallsvariablen auf demselben Wahrscheinlichkeitsraum mit $\mu(x) = \mathbb{P}(X = x)$ bzw. $\nu(x) = \mathbb{P}(Y = y)$ die Marginalverteilungen von X bzw. Y. Dann heißt das Tupel (X,Y) die Kopplung von μ und ν.

Um das Prinzip der Kopplung etwas besser zu verstehen, betrachten wir zunächst zwei unterschiedliche Kopplungen des fairen Münzwurfs:

Beispiel 1.
Seien μ, ν beide die Wahrscheinlichkeitsverteilungen des fairen Münzwurfs, das heißt $\mu = \nu$ mit $\mu(0) = \mu(1) = 0,5 = \nu(0) = \nu(1)$.
(i) Für die erste Kopplung beschreibe (X,Y) zwei unabhängige Münzen, sodass $\mathbb{P}(X = x, Y = y) = \frac{1}{4}$ für alle $x \in \{0, 1\}$.
(ii) Eine andere Möglichkeit μ und ν zu koppeln ist, dass X einen fairen Münzwurf beschreibt und wir Y mit Y=X definieren. Dann ergeben sich die Wahrscheinlichkeiten:
$\mathbb{P}(X = Y = 0) = \frac{1}{2}, \mathbb{P}(X = Y = 1) = \frac{1}{2}, \mathbb{P}(X \neq Y) = 0$

Allerdings kann man eine Kopplung nicht nur durch ein Tupel (X, Y) darstellen, sondern auch durch ein Wahrscheinlichkeitsmaß auf dem Raum $\Omega \times \Omega$.

Bemerkung 2.
(i) Sei q eine Wahrscheinlichkeitsverteilung auf dem Produktraum $\Omega \times \Omega$, welche $\sum_{y\in\Omega} q(x, y) = \mu(x)$ und $\sum_{x\in\Omega} q(x, y) = \nu(y)$ erfüllt. Dann existiert ein Tupel (X, Y) von Zufallsvariablen, das q als gemeinsame Verteilung besitzt. Und dieses Tupel ist eine Kopplung von μ und ν.
(ii) Sei (X, Y) eine Kopplung von μ und ν und sei $q(x, y) = \mathbb{P}(X = x, Y = y)$ die gemeinsame Verteilung von X und Y auf $\Omega \times \Omega$, dann gilt für q:

$\sum_{y \in \Omega} q(x,y) = \sum_{y \in \Omega} \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) = \mu(x)$
und
$\sum_{x \in \Omega} q(x,y) = \sum_{x \in \Omega} \mathbb{P}(X = x, Y = y) = \mathbb{P}(Y = y) = \nu(x)$

Durch die Definition der Kopplung erhalten wir eine weitere äquivalente Darstellung des Totalvariationsabstands:

Satz 4.

$$||\mu - \nu||_{TV} = \inf \{\mathbb{P}(X \neq Y) | (X,Y) \text{Kopplung von } \mu \text{ und } \nu\}$$

Beweis.
$\leq$: Sei $A \in \Omega$ beliebig und μ und ν die Marginalverteilungen von X bzw. Y.

$$\begin{aligned} |\mu(A) - \nu(A)| &= |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \\ &= |\mathbb{P}(X \in A, X \neq Y) + \mathbb{P}(X \in A, X = Y) - \mathbb{P}(Y \in A, X \neq Y) - \mathbb{P}(Y \in A, X = Y)| \\ &= |\mathbb{P}(X \in A, X \neq Y) - \mathbb{P}(Y \in A, X \neq Y)| \\ &\leq \max\{\mathbb{P}(X \in A, X \neq Y), \mathbb{P}(Y \in A, X \neq Y)\} \\ &\leq \mathbb{P}(X \neq Y) \end{aligned} \tag{6}$$

Die letzte Ungleichung gilt, da $\{X \in A\} \cap \{X \neq Y\} \subseteq \{X \neq Y\}$ und $\{Y \in A\} \cap \{X \neq Y\} \subseteq \{X \neq Y\}$. Diese Ungleichungskette ist für alle $A \subseteq \Omega$ erfüllt:

$$\Rightarrow \max_{A \subset \Omega} |\mu(A) - \nu(A)| \leq \mathbb{P}(X \neq Y)$$

Dies ist wiederum für alle Kopplungen (X,Y) von μ, ν erfüllt:

$$\Rightarrow ||\mu - \nu||_{TV} = \inf \{\mathbb{P}(X \neq Y) | (X,Y) \text{Kopplung von } \mu \text{ und } \nu\}$$

$\geq$: Mit dem gleichen Argument, das wir im Beweis zu Satz 3 benutzt haben, genügt es für diese Richtung zu zeigen, dass wir eine Kopplung finden, für die gilt: $\mathbb{P}(X \neq Y) = ||\mu - \nu||_{TV}$.

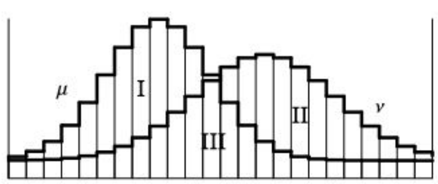

ABBILDUNG 1: Darstellung der Wahrscheinlichkeitsmaße μ und ν (siehe [1]).

Seien μ und ν nun zwei unterschiedliche Wahrscheinlichkeitsmaße über demselben endlichen Zustandsraum Ω. Unser Ziel ist es nun, die Zufallsvariablen

X und Y so zu wählen, dass sie möglichst oft gleich sind. Laut Bemerkung 2(i) genügt es eine gemeinsame Verteilung q von X und Y zu finden, so dass $\sum_{x\in\Omega} q(x, y) = \nu(y)$ und $\sum_{y\in\Omega} q(x, y) = \mu(x)$ gelten.
Wir definieren nun unsere gemeinsame Verteilung q für X und Y, wie folgt:

$$q(x, x) := \mu(x) \wedge \nu(x)$$
$$q(x, y) := \frac{(\mu(x) - \mu(x) \wedge \nu(x))(\nu(y) - \mu(y) \wedge \nu(y))}{||\mu - \nu||_{TV}}$$

Man kann sich die beiden Wahrscheinlichkeitsmaße μ und ν oBdA wie in Figur 1 visualisieren. Wir wählen X und Y nun abhängig vom Wurf einer unfairen Münze. Dabei geben wir dem Kopf-Ereignis die Wahrscheinlichkeit $p = \sum_{x\in\Omega} \mu(x) \wedge \nu(x)$, wobei $\mu(x) \wedge \nu(x) = \min\{\mu(x), \nu(x)\}$ und setzen dabei $X = Y$. Es gilt somit:

$$\begin{aligned}
\mathbb{P}(X = Y) = p &= \sum_{x\in\Omega} \mu(x) \wedge \nu(x) \\
&= \sum_{x\in\Omega,\mu(x)\leq\nu(x)} \mu(x) + \sum_{x\in\Omega,\mu(x)>\nu(x)} \nu(x) \\
&= \underbrace{\sum_{x\in\Omega,\mu(x)\leq\nu(x)} \mu(x) + \sum_{x\in\Omega,\mu(x)>\nu(x)} \mu(x)}_{=1} - \sum_{x\in\Omega,\mu(x)>\nu(x)} \mu(x) + \sum_{x\in\Omega,\mu(x)>\nu(x)} \nu(x) \\
&= 1 - \sum_{x\in\Omega,\mu(x)>\nu(x)} (\mu(x) - \nu(x)) \\
&\overset{\text{Bem.1}}{=} 1 - ||\mu - \nu||_{TV}
\end{aligned} \tag{7}$$

Trifft also das Ereignis Kopf ein, wählen wir den Wert unserer Zufallsvariable X (bzw. Y, da die beiden hier gleich gewählt werden) gemäß des Wahrscheinlichkeitsmaßes:

$$\gamma_{\text{Kopf}}(x) = \frac{\mu(x) \wedge \nu(x)}{p}$$

Trifft das Gegenereignis ein, wählen wir X gemäß des Wahrscheinlichkeitsmaßes:

$$\gamma_{\text{Zahl}_X}(x) = \begin{cases} \frac{\mu(x)-\nu(x)}{||\mu-\nu||_{TV}}, & \text{falls } \mu(x) > \nu(x) \\ 0, & \text{sonst} \end{cases}$$

und Y unabhängig davon, gemäß

$$\gamma_{\text{Zahl}_Y}(x) = \begin{cases} \frac{\nu(x)-\mu(x)}{||\mu-\nu||_{TV}}, & \text{falls } \nu(x) > \mu(x) \\ 0, & \text{sonst} \end{cases}$$

Dabei ist durch Bemerkung 1 sicher gestellt, dass γ_{Zahl_X} und γ_{Zahl_Y} tatsächlich Wahrscheinlichkeitsmaße sind.

Um nun die oben genannten Eigenschaften der gemeinsamen Verteilung q zu beweisen, betrachten wir wieder die Menge $B = \{y : \mu(y) \geq \nu(y)\}$.
Sei $y \in B^c$, dann gilt:

$$\begin{aligned}
\sum_{x\in\Omega} q(x,y) &= q(y,y) + \sum_{x,x\neq y} q(x,y) \\
&\overset{y\in B^c}{=} \mu(y) + \frac{\nu(y)-\mu(y)}{||\mu-\nu||_{TV}} \Bigg[\underbrace{\sum_{x,x\neq y} \mu(x)}_{=1-\mu(y)} - \underbrace{\sum_{x,x\neq y} \mu(x) \wedge \nu(x)}_{\overset{(7)}{=} 1-||\mu-\nu||_{TV}-\mu(y)} \Bigg] \qquad (8) \\
&= \mu(y) + \frac{\nu(y)-\mu(y)}{||\mu-\nu||_{TV}} ||\mu-\nu||_{TV} \\
&= \mu(y) + (\nu(y)-\mu(y)) \\
&= \nu(y)
\end{aligned}$$

Und für $y \in B$ folgt:

$$\begin{aligned}
\sum_{x\in\Omega} q(x,y) &= q(y,y) + \sum_{x,x\neq y} q(x,y) \\
&\overset{y\in B}{=} \nu(y) + \frac{\nu(y)-\nu(y)}{||\mu-\nu||_{TV}} \Bigg[\sum_{x,x\neq y} \big(\mu(x) - \mu(x)\wedge\nu(x)\big) \Bigg] \qquad (9) \\
&= \nu(y)
\end{aligned}$$

$\sum_{y\in\Omega} q(x,y) = \mu(x)$ gilt analog.
Nun zeigen wir noch, dass unser q tatsächlich ein Wahrscheinlichkeitsmaß ist. Das $q(x,y) \geq 0$ für alle $x, y \in \Omega$ gilt, ist klar, da μ und ν Wahrscheinlichkeitsmaße und somit größer gleich 0 sind. Außerdem gilt, dass der Totalvariationsabstand in diesem Fall echt größer als 0 ist, da wir $\mu \neq \nu$ voraussetzen. Wie in (7) gezeigt, gilt

$$\sum_{x\in\Omega} q(x,x) = \sum_{x\in\Omega} \mu(x) \wedge \nu(x) = 1 - ||\mu-\nu||_{TV}$$

und

$$\sum_{(x,y),x\neq y} q(x,y) = \sum_{y\in B^c} \sum_{x\in B} q(x,y) \overset{\text{Bem.1}}{=} \sum_{y\in B^c} \nu(y) - \mu(y)\wedge\nu(y) = \sum_{y\in B^c} \nu(y)-\mu(y) \overset{\text{Bem.1}}{=} ||\mu-\nu||_{TV}$$

und damit abschließend:

$$\sum_{x\in\Omega} q(x,x) + \sum_{(x,y),x\neq y} q(x,y) = 1$$

(X,Y) ist also eine Kopplung von μ und ν, für die unsere Bedingung erfüllt ist:

$$\mathbb{P}(X \neq Y) \ = \ 1 - \mathbb{P}(X = Y) \ = \ 1 - p \ = \ ||\mu-\nu||_{TV}$$

□

3 Mischzeit

Im Folgenden machen wir uns an einem realen Beispiel (vgl. [2], S.45) klar, dass die Definition der Mischzeit einer Markovkette sinnvoll ist.

3.1 Das Beispiel Kartenmischen

Ein natürliches Beispiel für die Definition von Mischzeiten ist das Mischen eines Kartenspiels mit N Spielkarten. Wir fragen uns nun wie groß die Anzahl der Mischschritte t im Vergleich zu N gewählt werden muss, damit das Blatt am Ende gut gemischt ist.
Zuerst spezifizieren wir, was genau wir unter gut gemischt verstehen.
Vor und nach jedem Mischschritt ist die Reihenfolge der Karten durch eine Permutation der N Karten gegeben, ist also ein Element der Permutationsgruppe S_N. Da die Mischschritte zufällig sind, haben wir nach t Schritten eine Wahrscheinlichkeitsverteilung μ_t auf S_N gegeben. μ_0 ist also das Punktmaß auf der Identität von S_N. Gut gemischt ist unser Blatt also, wenn jede Permutation nach t Schritten (nahezu) gleich wahrscheinlich ist und damit μ_t (nahezu) die Gleichverteilung π auf S_N ist. Da es allerdings unwahrscheinlich ist, dass der Kartenstapel nach einer gegebenen und festen Anzahl von Schritten gut gemischt ist, betrachtet man den Abstand zwischen π und μ_t. Je näher die beiden Wahrscheinlichkeitsverteilungen beieinander liegen, desto besser gemischt ist das Kartenblatt. Diesen Abstand liefert dabei der Totalvariatioinsabstand.
Das betrachtete Mischverfahren ist in unserem Beispiel Top to Random: Dabei wird die oberste Karte des Stapels abgehoben und zufällig zwischen die restlichen Karten geschoben. Das die Karte dabei ganz oben bleibt, ist nicht ausgeschlossen. Die gewählten Positionen sind unabhängig voneinander. Daraus ergibt sich, dass die Mischschritte unabhängig, identisch verteilte Zufallsvariablen X_i sind und $Y_t := X_t \circ ... \circ X_1$ die Kartenreihenfolge nach t Schritten als Hintereinanderausführung von Elementen aus S_N beschreibt. Die Folge $Y_1, Y_2...$ ist damit eine Irrfahrt auf S_N und somit eine Markovkette mit Zustandsraum S_N und die stationäre Verteilung auf diesem Raum ist die Gleichverteilung.
Man kann zeigen, dass das Blatt schon nach $N \log N + 10N$ Mischschritten gut gemischt ist.

3.2 Definition Mischzeit

Sei $X = (X_0, X_1, \ldots)$ eine irreduzible Markovkette mit Übergangsmatrix P und stationärer Verteilung π. Im Folgenden betrachten wir den maximalen Abstand der Verteilung der Markovkette zu ihrer stationären Verteilung, als eine vom Zeitschritt t abhängige Funktion d:

$$d(t) := \max_{x \in \Omega} ||P_x^t - \pi||_{TV}$$

Wir abstrahieren nun das oben angesprochene Beispiel des Kartenmischens und definieren für eine irreduzible, aperiodische Markovkette mit endlichem Zustandsraum die Konvergenzgeschwindigkeit der Übergangswahrscheinlichkeiten gegen die invariante Verteilung:

Definition. (Mischzeit)

$$\mathrm{t}_{\mathrm{mix}}(\epsilon) := \min\left\{\mathrm{t} : \mathrm{d}(\mathrm{t}) \leq \epsilon\right\}, \epsilon > 0$$

Die Mischzeit einer Markovkette ist also die Zeit, die die Markovkette benötigt, damit die Distanz zwischen ihr und der stationären Verteilung klein ist.

Literaturverzeichnis

[1] David A. Levin, Yuval Peres, Elizabeth L. Wilmer, Markov Chains and Mixing Times, American Mathematical Society, 2006

[2] Michael Scheutzow, Stochastische Modelle Vorlesungsskript, Berlin, 2010