Kapitel D

Markov–Spiele und Bellmans Optimalitätsprinzip

My first task in dynamic programming was to put it on a rigorous basis. I found that I was using the same technique over and over again to derive a functional equation. I decided to call it "The principle of optimality." Oliver Gross said one day, 'The principle is not rigorous.' I replied, 'Of course not. It's not even precise.' A good principle should guide the intuition.

Richard Bellman (1920–1984), *Eye of the Hurricane* (1984)

Inhalt dieses Kapitels D

1. Markov–Spiele: erste Beispiele
   - Erwartete Auszahlung in einem Markov–Belohnungsprozess
   - Irrfahrt, Gewinnerwartung und optimale Entscheidung
   - Google: Wo häuft sich die zufällige Irrfahrt im Internet?
2. Bellmans Optimalitätsprinzip
   - Banachs Fixpunktsatz und Blackwells Kriterium
   - Markov–Graphen, Erwartung und Optimalität
   - Bellmans Optimalitätsprinzip
3. Anwendung im maschinellen Lernen
   - Optimale Routenplanung eines Roboters
   - Gewinniteration vs Strategieiteration
   - Bestärkendes Lernen / *reinforcement learning*
4. Ausblick: die wunderbare Welt der Wahrscheinlichkeit
   - Zufällige Irrfahrten und Markov–Prozesse
   - Entropie verstehen und anwenden

Motivation und Überblick

When working in the field of analysis it is exceedingly helpful to have some underlying physical processes clearly in mind.

Richard Bellman (1920–1984), *Eye of the Hurricane* (1984)

In diesem Kapitel betrachten wir Markov–Spiele, in denen ein Akteur gegen den „blinden Zufall“ spielt und eine **optimale Strategie** sucht. Ich beginne simpel und möglichst **problemorientiert** [*problem driven*]: Wie können wir uns in sehr einfachen Bei-Spielen optimal verhalten? Daraus ergeben sich die richtigen Fragen, oft direkt von Studierenden, die wir dann durch Aufbau einer **passenden Theorie** zu lösen suchen [*method driven*]. Beide Arbeitsweisen ergänzen sich bestens.

Je considère comme complètement inutile la lecture de gros traités d'analyse pure: un trop grand nombre de méthodes passent en même temps devant les yeux. C'est dans les travaux d'application qu'on doit les étudier; c'est là qu'on juge leurs capacités et qu'on apprend la manière de les utiliser.

Joseph-Louis Lagrange (1736–1813)

Motivation und Überblick

In den vorigen Kapiteln haben wir die **Rekursion** / Rückwärtsinduktion kennen und nutzen und schätzen gelernt. In Anwendungen sind die zu lösenden Gleichungen oft nicht rekursiv aufgebaut, also nicht von klein nach groß „topologisch“ sortierbar, sondern selbstbezüglich / zyklisch. Dies sind allgemeine **Fixpunktgleichungen**. Ihre wunderschöne Theorie und praktische Anwendung sind überall von enormer Bedeutung.

📖 Nancy L. Stokey, Robert E. Lucas:
Recursive Methods in Economic Dynamics. Harvard Univ. Press 1989

Lars Ljungqvist, Thomas J. Sargent:
Recursive Macroeconomic Theory. The MIT Press (3rd ed.) 2012

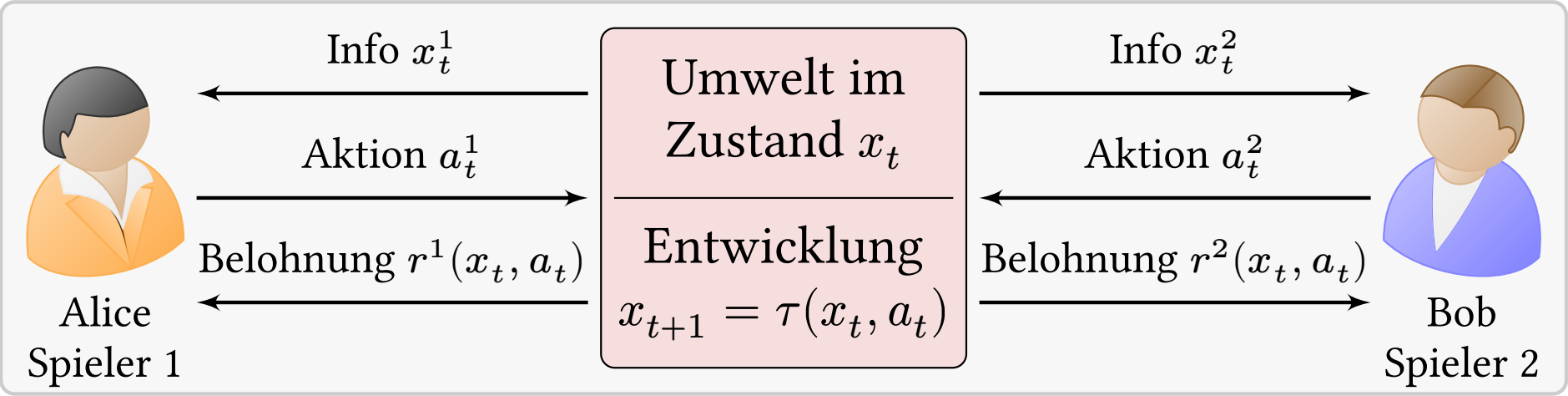
Die hier gesuchte Optimierung des strategischen Handelns führt aus algorithmischer Sicht zum **maschinellen Lernen** [*machine learning*].

📖 Stuart Russell, Peter Norvig: *Artificial Intelligence: A Modern Approach.* Addison Wesley (3rd ed.) 2016 (Kapitel 17: Making complex decisions)

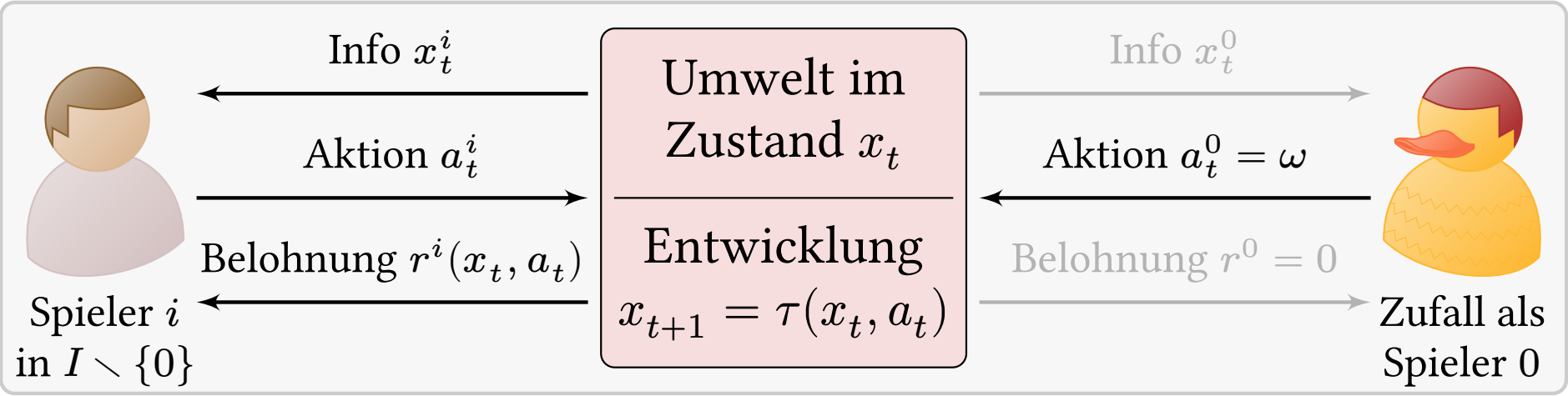
Richard S. Sutton, Andrew G. Barto: *Reinforcement Learning.*
The MIT Press (2nd ed.) 2018 (hier speziell §6.5: Q–learning).

Dynamische Spiele und strategische Interaktion

Dynamisches System mit Steuerung durch zwei Spieler: Markov–Spiel



Steuerung durch mehrere Spieler und Zufall: Markov–Spiel (MMDP)



Jeder Spieler will seinen individuellen Gesamtnutzen maximieren!

Dynamische Spiele und strategische Interaktion

In Kapitel C haben wir **neutrale kombinatorische Spiele** betrachtet: Spieler ziehen abwechselnd und haben dieselben Zugmöglichkeiten. Das vereinfacht die Analyse und enthüllt besonders klare Strukturen: Der Satz von Sprague–Grundy überführt jedes neutrale Spiel in ein äquivalentes Nim-Spiel und erlaubt (oft genug) eine effiziente Lösung.

Unser Ziel sind nun **allgemeine Spiele** für zwei oder mehr Personen, bei denen die Spieler:innen abwechselnd oder auch gleichzeitig ziehen. Bevor wir solche Spiele erklären und lösen, möchte ich noch etwas genauer einige spezielle und besonders einfache Fälle beleuchten.

In diesem Kapitel untersuchen wir Markov–Spiele, mit nur einem Spieler, aber Zufallszügen. Es handelt sich also um **Spiele gegen den Zufall**. Das Verhalten des Zufalls ist (in unseren Modellen) fest vorgegeben, daher handelt es sich um ein (stochastisches) **Optimierungsproblem**.

Der übliche Name hierfür ist **Markov–Entscheidungsprozess** [*Markov decision process*, MDP], doch auch das ist nichts anderes als ein Spiel. Hierzu erkläre ich erste schöne Beispiele und Lösungsmethoden.

Markov–Spiele und Entscheidungstheorie

Menschen stehen fortwährend vor Entscheidungen, kleinen wie großen. Dabei sind kurzfristige Vorteile gegen langfristige Ergebnisse abzuwägen. Für manche Probleme genügt es, den kurzfristigen Gewinn zu optimieren; dieses Verhalten heißt daher **gieriger Algorithmus** [*greedy algorithm*].

In vielen interessanten und relevanten Anwendungen führt jedoch eine kurzfristige Gewinnmaximierung langfristig zu schlechten Ergebnissen. Jetzige Entscheidungen beeinflussen spätere, daher können solche Fragen beliebig kompliziert werden und wir benötigen scharfe Denkwerkzeuge.

Im Laufe unserer Evolution hat sich dazu ein **leistungsfähiges Gehirn** als vorteilhaft erwiesen: Damit können wir planen, mögliche Handlungen im Geiste vollziehen sowie ihre Konsequenzen abschätzen und vergleichen. Wir können **fiktiv spielen** bevor wir **real handeln**. Denken hilft!

Markov–Entscheidungsprozesse behandeln solche Problemstellungen: Sie liefern einen gemeinsamen Rahmen zur Modellierung dieser Fragen sowie Werkzeuge zur optimalen Lösung sequentieller Entscheidungen. Die Mathematik hilft uns beim Planen und vorausschauenden Handeln.

Entscheidungen unter Unsicherheit sind ein allgegenwärtiges Problem. Markov–Entscheidungsprozesse sind darunter ein wichtiger Spezialfall. Wir gehen im Weiteren von folgenden starken Vereinfachungen aus, ihre klaren Vorteile und Einschränkungen sind uns wohl bewusst:

1. Wir betrachten nur einen Spieler und den „blinden Zufall“ als Gegner. (In vielen Anwendungen interagieren mehrere Spieler, dazu später. Wir lösen zunächst das Optimierungsproblem für einen Spieler, Nash–Gleichgewichte übertragen alles auf mehrere Spieler.)
2. Die Zeit ist diskret, also endlich $t \in \{0, \dots, N\}$ oder unendlich $t \in \mathbb{N}$. (Viele Anwendungen möchten wir mit kontinuierlicher Zeit $t \in [a, b]$ oder $t \in \mathbb{R}_{\geq 0}$ beschreiben, doch das zeitdiskrete Modell ist meist wesentlich einfacher und eine hinreichend gute Näherung.)
3. Die zukünftige Entwicklung ist zwar notorisch ungewiss, doch alle Wahrscheinlichkeiten sind uns exakt bekannt. (In den meisten realen Anwendungen müssen diese Wkten zuerst geschätzt oder mühsam gemessen werden und sind daher nahezu immer fehlerbehaftet.)

Erwartete Auszahlung u_s einer Strategie s

Satz D1A: erwartete Auszahlung

Sei $\Gamma = (X, A, \sigma, \tau)$ ein Markov–Graph mit terminaler Auszahlung $v : \partial X \to \mathbb{R}$, sofortiger Belohnung $r : A \times X \to \mathbb{R}$ und Diskont $\delta \in [0, 1]$.

Die Auszahlung entlang der Trajektorie $w = (x_0, a_0, x_1, a_1, x_2, \dots)$ ist

$$u(w) := \begin{cases} \sum_{t=0}^{n-1} \delta^t r(x_t, a_t, x_{t+1}) + \delta^n v(x_n) & \text{falls } x_n \in \partial X, \\ \sum_{t=0}^{\infty} \delta^t r(x_t, a_t, x_{t+1}) & \text{falls } w \in \Gamma_\infty. \end{cases}$$

Zur Vereinfachung sei hierzu $\delta \in [0, 1[$ sowie r und v beschränkt.

(1) Deterministisch: Zu jeder Strategie $s \in S(\Gamma)$ und jedem Startpunkt $x \in X$ ist die Trajektorie w rekursiv gegeben durch $x_0 = x$, $a_t = s(x_t)$, $x_{t+1} = \tau(x_t, a_t)$ für alle t. Die Auszahlung ist dann $u_s(x) = u(w)$.

(2) Stochastisch: Es existiert die erwartete Auszahlung

$$u_s : X \to \mathbb{R} : u_s(x) = \mathbf{E}_s^x[w \mapsto u(w)].$$

Diese ist beschränkt durch $|u_s|_X \leq |v|_{\partial X} + |r|_A/(1 - \delta) \in \mathbb{R}$.

Erwartete Auszahlung u_s einer Strategie s

Aufgabe: Führen Sie diese Konstruktion detailliert aus, als Übung und zur Wiederholung der Grundbegriffe der Wahrscheinlichkeitstheorie.

Lösung: (1) Gegeben sei $s \in S(\Gamma)$, also $s : X^\circ \to A$ mit $\sigma \circ s = \mathrm{id}_{X^\circ}$.
Zu jedem Startpunkt $x_0 \in X$ können wir s integrieren zur Trajektorie $w = (x_0, a_0, x_1, a_1, x_2, \dots)$ mit $a_t = s(x_t)$ und $x_{t+1} = \tau(x_t, a_t)$ für alle t.
Ist $w = (x_0, a_0, x_1, a_1, \dots, x_n) \in \Gamma_n$ endlich, mit $x_n \in \partial X$, so haben wir

$$u(w) := \textstyle\sum_{t=0}^{n-1} \delta^t r(x_t, a_t, x_{t+1}) + \delta^n v(x_n).$$

Statt $r : A \times X \to \mathbb{R}$ genügt hier $\bar{r} : A \to \mathbb{R}$ mit $\bar{r}(x, a) = r(x, a, \tau(x, a))$.
Ist die Trajektorie $w = (x_0, a_0, x_1, a_1, \dots) \in \Gamma_\infty$ unendlich, so haben wir

$$u(w) := \textstyle\sum_{t=0}^{\infty} \delta^t r(x_t, a_t, x_{t+1}).$$

Hier müssen wir die Konvergenz der Reihe sicherstellen. Am einfachsten gelingt dies, indem wir $\delta \in [0, 1[$ voraussetzen und $\bar{r}$ beschränken, also $|\bar{r}|_A \le M < \infty$; dann konvergiert die Reihe absolut.

Sind $\bar{r}$ und v beschränkt, so auch u_s dank $|u_s|_X \le |v|_{\partial X} + |\bar{r}|_A/(1-\delta)$.

Erwartete Auszahlung u_s einer Strategie s

(2) Im stochastischen Falle führt die Aktion $a \in A_x$ nicht sicher, sondern mit gewisser Wkt $p(x, a, y)$ zum Zustand y und zur Belohnung $r(x, a, y)$.

Entlang jeder un/endlichen Trajektorie w berechnen wir die Auszahlung

$$u(w) := \begin{cases} \sum_{t=0}^{n-1} \delta^t r(x_t, a_t, x_{t+1}) + \delta^n v(x_n) & \text{falls } x_n \in \partial X, \\ \sum_{t=0}^{\infty} \delta^t r(x_t, a_t, x_{t+1}) & \text{falls } w \in \Gamma_\infty. \end{cases}$$

Bei Start im Zustand $x \in X$ betrachten wir nun die erwartete Auszahlung

$$u_s(x) := \mathbf{E}_s^x [w \mapsto u(w)].$$

Ausführlich: Bei Start im Zustand $x \in X$ mit Strategie $s \in S(\Gamma)$ erhalten wir einen Markov–Prozess auf dem WRaum $(\Gamma_*, \mathscr{A}, \mathbf{P}_s^x)$ der Trajektorien, in unserem Modell möglicherweise mit terminalen Zuständen $x_n \in \partial X$.

Die sofortige Belohnung r und die terminale Auszahlung v erklären die Auszahlung $w \mapsto u(w)$ für jede Trajektorie $w \in \Gamma_*$. So erweitern wir den Markov–Prozess zu einen Markov–Belohnungsprozess, kurz MRP.

Erwartete Auszahlung u_s einer Strategie s

Sei $\Gamma_*(x_0)$ die Menge aller Trajektorien mit Start im Zustand $x_0 \in X$. Zu $w = (x_0, a_0, \dots, x_n)$ enthält $w \circ \Gamma_*(x_n)$ alle Trajektorien mit Anfang w. Bei Start in x mit Strategie s hat diese Menge die Wahrscheinlichkeit

$$\mathbf{P}_s^x(w \circ \Gamma_*(x_n)) = p(x_0, a_0, x_1) \cdot p(x_1, a_1, x_2) \cdots p(x_{n-1}, a_{n-1}, x_n)$$

falls $x_0 = x$ gilt sowie $a_t = s(x_t)$ für alle $t < n$, sonst $\mathbf{P}_s^x(w \circ \Gamma_*(x_n)) = 0$. Die Mengen $w \circ \Gamma_*(x_n)$ erzeugen die σ–Algebra $\mathscr{A}$ auf Γ_*. Darauf existiert genau ein WMaß $\mathbf{P}_s^x : \mathscr{A} \to [0,1]$, das die obige Pfadproduktformel erfüllt.
📖 Heinz Bauer: WTheorie, §35 Konstruktion stochastischer Prozesse.

So erhalten wir den WRaum $(\Gamma_*, \mathscr{A}, \mathbf{P}_s^x)$. Zu jeder messbaren Funktion $u : (\Gamma_*, \mathscr{A}) \to ([0,\infty], \mathscr{B}) : w \mapsto u(w)$ erhalten wir den Erwartungswert

$$\mathbf{E}_s^x[u : w \mapsto u(w)] \in [0,\infty].$$

Eine messbare Funktion $u : (\Gamma_*, \mathscr{A}) \to ([0,\infty], \mathscr{B})$ ist integrierbar, wenn $\mathbf{E}_s^x[u^\pm] < \infty$ gilt. In diesem Falle definieren wir $\mathbf{E}_s^x[u] = \mathbf{E}_s^x[u^+] - \mathbf{E}_s^x[u^-]$. Das gilt insbesondere, falls u messbar und zudem beschränkt ist.

Rekursive Berechnung der erwarteten Auszahlung

Satz D1B: rekursive Berechnung der erwarteten Auszahlung

Gegeben sei ein Markov–Graph $\Gamma = (X, A, \sigma, \tau)$ mit der Transition $\tau : A \to [X] : \tau(x, a) = \sum_{y \in X} p(x, a, y) \cdot y$, die terminale Auszahlung $v : \partial X \to \mathbb{R}$ und die sofortige Belohnung $r : A \times X \to \mathbb{R}$ mit $\delta \in [0, 1]$.

Zur Vereinfachung seien weiterhin r und v beschränkt sowie $\delta \in [0, 1[$.

(1) Für jede Strategie $s \in S(\Gamma)$ erfüllt die erwartete Auszahlung $u := u_s : X \to \mathbb{R}$ die rekursive **Erwartungsgleichung**:

$$u(x) = \begin{cases} v(x) & \text{für terminale Zustände } x \in \partial X, \text{ sonst} \\ \sum_{y \in X} p(x, s(x), y) \left[r(x, s(x), y) + \delta u(y) \right]. \end{cases}$$

(2) Diese Gleichung kann mehrere Lösungen $u : X \to \mathbb{R}$ haben. Die erwartete Auszahlung u_s ist die einzige beschränkte Lösung (D2K).

Aufgabe: Führen Sie diese Rechnung detailliert aus, als Übung und zur Wiederholung der Grundbegriffe der Wahrscheinlichkeitstheorie.

Rekursive Berechnung der erwarteten Auszahlung

Zu Satz D1B betrachten wir zunächst Summen von begrenzter Länge $\leq n$:

$$u^n(w) := \begin{cases} \sum_{t=0}^{n-1} \delta^t r(x_t, a_t, x_{t+1}) & \text{falls } x_t \in X^\circ \text{ für alle } t \leq n, \\ \sum_{t=0}^{\ell-1} \delta^t r(x_t, a_t, x_{t+1}) + \delta^\ell v(x_\ell) & \text{falls } x_\ell \in \partial X \text{ für ein } \ell \leq n. \end{cases}$$

Diese Auszahlung $u^n : \Gamma_* \to \mathbb{R}$ können wir rekursiv berechnen:

$$u^0(w) = \begin{cases} v(x_0) & \text{falls } x_0 \in \partial X, \\ 0 & \text{falls } x_0 \in X^\circ. \end{cases}$$

$$u^{n+1}(w) = \begin{cases} v(x_0) & \text{falls } x_0 \in \partial X, \\ r(x_0, a_0, x_1) + \delta u^n(w') & \text{falls } w = (x_0 \xrightarrow{a_0} x_1) \circ w'. \end{cases}$$

Nach Konstruktion konvergiert u^n gegen u, gleichmäßig auf Γ_*, denn

$$|u^n - u|_{\Gamma_*} \leq \delta^n \big[\, |v|_{\partial X} + |r|_{A \times X}/(1-\delta) \,\big] \searrow 0.$$

Sind alle Trajektorien / Spielverläufe einheitlich in der Länge $\leq n$ beschränkt, so gilt $u^n = u$. Allgemein nutzen wir $u^n \to u$ für $n \to \infty$.

Rekursive Berechnung der erwarteten Auszahlung

Entsprechend haben wir zu $u^n : \Gamma_* \to \mathbb{R}$ die erwartete Auszahlung

$$u_s^n \,:\, X \to \mathbb{R} \,:\, u_s^n(x) = \mathbf{E}_s^x[w \mapsto u^n(w)].$$

Der durch n begrenzte Zeithorizont hat den technischen Vorteil, dass wir die erwartete Auszahlung leicht rekursiv berechnen können:

$$u_s^0(x) \quad := \begin{cases} v(x_n) & \text{falls } x \in \partial X, \\ 0 & \text{falls } x \in X^\circ. \end{cases}$$

$$u_s^{n+1}(x) := \begin{cases} v(x) & \text{für terminale Zustände } x \in \partial X, \text{ sonst} \\ \sum_{y \in X} p(x, s(x), y)\,[r(x, s(x), y) + \delta u_s^n(y)]. \end{cases}$$

Nach Konstruktion konvergiert u_s^n gegen u_s, gleichmäßig auf X, denn

$$|u_s^n - u_s|_X \;\le\; \delta^n\big[\,|v|_{\partial X} + |r|_{A\times X}/(1-\delta)\,\big] \;\searrow\; 0.$$

In unserer Rekursionsgleichung können wir auf beiden Seiten den Limes $n \to \infty$ bilden. So erhalten wir die rekursive Gleichung von Satz D1B.

Rekursive Berechnung der erwarteten Auszahlung

$$0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \longrightarrow 5 \longrightarrow 6 \longrightarrow 7 \longrightarrow 8 \longrightarrow 9 \quad \cdots$$

Beispiel: Wir betrachten den Graphen $\Gamma = (X, A, \sigma, \tau)$ mit $X = \mathbb{N}$ und $A = \{\,(x, x+1) \mid x \in \mathbb{N}\}$ mit $\sigma = \mathrm{pr}_1$ und $\tau = \mathrm{pr}_2$ sowie $r = 0$. Die einzige Strategie in $S(\Gamma)$ ist demnach $s : X^\circ \to A : x \mapsto (x, x+1)$.

Genau dann löst eine Funktion $u : X \to \mathbb{R}$ die Erwartungsgleichung D1B, wenn $u(x) = c \cdot \delta^{-x}$ für eine Konstante $c \in \mathbb{R}$ und alle $x \in X$ gilt. Die einzige beschränkte Lösung ist also tatsächlich $u_s = 0$.

⚠ Die Erwartung u_s erfüllt Gleichung D1B. Zur Bestimmung von u ist diese Gleichung also *notwendig*, aber im Allgemeinen nicht *hinreichend*.

☺ Die Konstruktion D1A des WRaums $(\Gamma_*, \mathscr{A}, \mathbf{P}_s^x)$ definiert die erwartete Auszahlung $u_s : X \to \mathbb{R} : u_s(x) := \mathbf{E}_s^x[w \mapsto u(w)]$. Zur Gleichung D1B gilt demnach die *Existenz* einer Lösung, aber nicht immer die *Eindeutigkeit*.

☺ Für $\delta < 1$ und r, v beschränkt ist u_s die einzige beschränkte Lösung. Dies zeigen wir in Abschnitt D2 mit dem Fixpunktsatz von Banach. Zuerst will ich das Vorgehen mit eindrücklichen Beispielen illustrieren.

Zufällige Irrfahrt und harmonische Gewinnerwartung

0	1	2	3	4	5	6	7	8
7€								23€

Aufgabe: Ihre Spielfigur startet auf einem Spielfeld $x \in X^\circ = \{1, \dots, 7\}$. In jedem Zug rückt sie auf ein Nachbarfeld, gleichverteilt, unabhängig. Das Spiel endet am Rand $\partial X = \{0, 8\}$ mit der gezeigten Auszahlung.

(0) Welche Gewinnerwartung $u(x)$ hat jedes Feld $x \in X = \{0, \dots, 8\}$?
(1) Variante: Jeder Zug kostet $c(x) = -1$€. Ab wo würden Sie spielen?

Bitte schätzen Sie zunächst! Wie treffsicher ist Ihre intuitive Erwartung? Formulieren Sie dann nützliche Gleichungen und Lösungsmethoden…

Lösung: Für jedes innere Feld $x \in X^\circ = \{1, \dots, 7\}$ gilt

$$u(x) = \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1) + c(x).$$

Damit können Sie mögliche Lösungen prüfen. Auch berechnen? Wie?

7	9	11	13	15	17	19	21	23

7	2	-1	-2	-1	2	7	14	23

Zufällige Irrfahrt und harmonische Gewinnerwartung

Hier ist (0) ein extrem einfaches Spiel, schon (1) dürfte Sie überraschen: Ungeschult haben wir herzlich wenig Erfahrung mit zufälligen Irrfahrten. Erfahrungsgemäß fällt uns Menschen das rekursive Denken recht schwer, doch gerade dies ist sehr oft wesentlich für rationale Entscheidungen!

? Bevor wir die Lösung diskutieren, schätzen Sie bitte Ihre Erwartung: Ist Ihre Intuition recht präzise und treffsicher, oder allzu vage und irrig? Diese quantitativen Schätzfragen sind auch ein aufschlussreicher Test der vielzitierten Schwarmintelligenz und mahnen eindringlich zur Vorsicht: Betrügerische Geschäftspraktiken beruhen darauf, dass das Gegenüber die Situation schlecht einschätzen kann und Fehlentscheidungen trifft.

? Bilden die oben angegebenen Zahlen eine Lösung? sogar die einzige? Mehrdeutigkeiten müssen wir erkennen und nötigenfalls auch beheben. Es ist schön und gut, die eigene Intuition zu nutzen und zu entwickeln. Leider hilft es wenig, irgendeine Antwort ohne Begründung zu orakeln. Wir wollen tragfähige Argumente, begründet und nachvollziehbar! Auch das ist ein Qualitätsmerkmal rationalen Handelns.

Zufällige Irrfahrt und harmonische Gewinnerwartung

0	1	2	3	4	5	6	7	8
7	9	11	13	15	17	19	21	23

Lösung: (0) Für $x \in X = \{0, \dots, 8\}$ suchen wir $u_x = u(x)$. Wir haben:

$$
\begin{aligned}
u_0 &= 7 \\
-\tfrac{1}{2}u_0 + u_1 - \tfrac{1}{2}u_2 &= 0 \\
-\tfrac{1}{2}u_1 + u_2 - \tfrac{1}{2}u_3 &= 0 \\
-\tfrac{1}{2}u_2 + u_3 - \tfrac{1}{2}u_4 &= 0 \\
-\tfrac{1}{2}u_3 + u_4 - \tfrac{1}{2}u_5 &= 0 \\
-\tfrac{1}{2}u_4 + u_5 - \tfrac{1}{2}u_6 &= 0 \\
-\tfrac{1}{2}u_5 + u_6 - \tfrac{1}{2}u_7 &= 0 \\
-\tfrac{1}{2}u_6 + u_7 - \tfrac{1}{2}u_8 &= 0 \\
u_8 &= 23
\end{aligned}
$$

Bandmatrix, dünn besetzt

$-\Delta u = 0$

☺ Lineare Gleichungssysteme können Sie lösen: Gauß gelingt immer!
Geschickte Substitution: Für $d_x := u_x - u_{x-1}$ gilt $d_1 = d_2 = \cdots = d_8 = d$ und $8d = 23 - 7 = 16$, also $d = 2$ und somit $u_x = 7 + d \cdot x$ für alle $x \in X$.

Zufällige Irrfahrt und harmonische Gewinnerwartung

0	1	2	3	4	5	6	7	8
7	2	-1	-2	-1	2	7	14	23

Lösung: (1) Für $x \in X = \{0, \dots, 8\}$ suchen wir $u_x = u(x)$. Wir haben:

$$
\begin{aligned}
u_0 &= 7 \\
-\tfrac{1}{2}u_0 + u_1 - \tfrac{1}{2}u_2 &= c_1 \\
-\tfrac{1}{2}u_1 + u_2 - \tfrac{1}{2}u_3 &= c_2 \\
-\tfrac{1}{2}u_2 + u_3 - \tfrac{1}{2}u_4 &= c_3 \\
-\tfrac{1}{2}u_3 + u_4 - \tfrac{1}{2}u_5 &= c_4 \\
-\tfrac{1}{2}u_4 + u_5 - \tfrac{1}{2}u_6 &= c_5 \\
-\tfrac{1}{2}u_5 + u_6 - \tfrac{1}{2}u_7 &= c_6 \\
-\tfrac{1}{2}u_6 + u_7 - \tfrac{1}{2}u_8 &= c_7 \\
u_8 &= 23
\end{aligned}
$$

Bandmatrix, dünn besetzt

$-\Delta u = 0$

☺ Allgemeine Faustregel: Ausrechnen ist mühsam. Prüfen ist leicht!
Wir vermuten, dass die Lösung eindeutig ist. Gilt das? Satz D1c hilft!
Negative Gewinnerwartung bedeutet: Ab hier besser nicht spielen!

Zufällige Irrfahrt und optimale Entscheidung

☺ Ist das Angebot lukrativ bei Start in $x = 4$? Nein! Bei Start in $x = 5$? Ja! Mathematik hilft Ihnen, die richtige Entscheidung zu treffen. Dieses Gleichungssystem ist **linear** und daher leicht zu lösen. Das System ist klein genug, um noch von Hand gelöst zu werden.

Die Matrix ist zudem dünn besetzt und hat hohe Symmetrie. Das nutzen wir gerne, hier etwa durch wiederholte Halbierung. Für (0) gelingt eine geschickte Substitution, auch ganz allgemein: Die Lösung ist die eindeutige Gerade durch die beiden Randwerte.

Die Gleichung $u_x = \frac{1}{2}u_{x-1} + \frac{1}{2}u_{x+1}$ nennen wir **Mittelwerteigenschaft**. Eine solche Funktion $u : X \to \mathbb{R}$ heißt **harmonisch**, als diskrete Analogie zur Laplace–Gleichung $\Delta u = 0$ mit $\Delta = \partial_1^2 + \cdots + \partial_n^2$ auf einem Gebiet $\Omega \subseteq \mathbb{R}^n$, die in der Analysis und der Physik eine wichtige Rolle spielt.

Die Differenz $-\frac{1}{2}u_{x-1} + u_x - \frac{1}{2}u_{x+1}$ entspricht der (negativen) **zweiten Ableitung** an der Stelle x. Ist sie gleich Null, so handelt es sich um eine Gerade. Ist sie konstant ungleich 0, so handelt es sich um eine Parabel. Die Zugkosten c_x im Punkt x entsprechen der Krümmung im Punkt x!

Zufällige Irrfahrt und optimale Entscheidung

0	1	2	3	4	5	6	7	8
7.00	2.67	0.33	0.00	0.60	3.20	7.80	14.40	23.00

Aufgabe: (2) Zusätzlich zu (1) dürfen Sie nun jederzeit aufgeben.

Lösung: Wir suchen $u : X \to \mathbb{R}$. Am Rand gilt $u(0) = 7$ und $u(8) = 23$.
In jedem aktiven Zustand $x \in \{1, \dots, 7\}$ sind zwei Aktionen möglich:

$$\begin{aligned} &\text{Weiterspielen:} && u(x) = c(x) + \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1) \\ &\text{Aufgeben:} && u(x) = 0 \end{aligned}$$

Ein rationaler Spieler wählt jeweils den besten Zug, also das Maximum:

$$\boxed{u(x) = \max\{\ 0,\ c(x) + \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1)\ \}}$$

⚠ Dieses Gleichungssystem ist **nicht-linear** und daher etwas kniffliger.
Wie immer gilt auch hier: Ausrechnen ist mühsam. Prüfen ist leicht!
Existiert immer eine Lösung u? Gibt es mehrere oder genau eine?
Wie können Sie die / eine berechnen? Zudem möglichst effizient?

☺ Hier hilft Ihnen ein Python-Skript oder eine Tabellenkalkulation!

Zufällige Irrfahrt und optimale Entscheidung

Eine numerische Näherung gelingt einfach und effizient durch Iteration:

$$u_{t+1}(x) = \max\{\ 0,\ c(x) + \tfrac{1}{2}u_t(x-1) + \tfrac{1}{2}u_t(x+1)\ \}$$

t	$u_t(0)$	$u_t(1)$	$u_t(2)$	$u_t(3)$	$u_t(4)$	$u_t(5)$	$u_t(6)$	$u_t(7)$	$u_t(8)$
0	7.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	23.00
1	7.00	2.50	0.00	0.00	0.00	0.00	0.00	10.50	23.00
2	7.00	2.50	0.25	0.00	0.00	0.00	4.25	10.50	23.00
3	7.00	2.63	0.25	0.00	0.00	1.13	4.25	12.63	23.00
4	7.00	2.63	0.31	0.00	0.00	1.13	5.88	12.63	23.00
5	7.00	2.66	0.31	0.00	0.00	1.94	5.88	13.44	23.00
6	7.00	2.66	0.33	0.00	0.00	1.94	6.69	13.44	23.00
7	7.00	2.66	0.33	0.00	0.00	2.34	6.69	13.84	23.00
8	7.00	2.66	0.33	0.00	0.17	2.34	7.09	13.84	23.00
9	7.00	2.67	0.33	0.00	0.17	2.63	7.09	14.05	23.00
...									
39	7.00	2.67	0.33	0.00	0.60	3.20	7.80	14.40	23.00
40	7.00	2.67	0.33	0.00	0.60	3.20	7.80	14.40	23.00

Zufällige Irrfahrt und optimale Entscheidung

☺ Vergleichen Sie die hier berechneten Werte mit Ihren intuitiven Schätzungen bei der ersten, informellen Annäherung an diese Frage. Unser kleines Markov–Spiel ist eine sehr einfache, aber durchaus realistische Illustration für einen Markov–Entscheidungsprozess.

☺ Die Lösung u_0 im linearen Fall, ohne Entscheidung, ist leicht zu berechnen durch das lineare Gleichungssystem, also als Fixpunkt: Wir nutzen auf $E = \{\, u : X \to \mathbb{R} \mid u(0) = 7,\ u(8) = 23 \,\}$ den linearen Operator $\Phi_0 : E \to E : u \mapsto \bar{u}$ mit $\bar{u}(x) = c(x) + \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$.

⚠ Die Lösung mit Entscheidungsmöglichkeit ist nicht $\max\{0, u_0\}$! Diese naive Fehlannahme führt tatsächlich zu Fehlentscheidungen.

☺ Wir erhalten sie vielmehr als Fixpunkt von $\Phi(u) = \max\{0, \Phi_0(u)\}$. Diesen nicht-linearen Operator können wir zur Iteration nutzen, wie oben gezeigt. Ist die beobachtete Konvergenz nicht wunderbar?

☺ In der Übung zeigen Sie allgemein, dass $\Phi : E \to E$ kontraktiv ist, so dass Sie Banachs Fixpunktsatz anwenden können. Alles wird gut. Damit können Sie u berechnen und Ihre optimale Strategie ablesen!

Vorsicht bei unendlichen Spielfeldern

0	1	2	3	4	5	6	7	8	9	10	
100€											⋯

Aufgabe: Das Spielbrett ist nun nach rechts unbegrenzt. Wie viel würden Sie als Teilnehmer zahlen / als Anbieter verlangen? Neutral gefragt:

(0) Was ist die Gewinnerwartung $u(x)$ für jedes Startfeld $x \in \mathbb{N}$?
(1) Jeder Zug kostet, $c = -1$€, und Sie müssen zu Ende spielen.
(2) Jeder Zug kostet, $c = -1$€, und Sie dürfen jederzeit aufgeben.

Bitte schätzen Sie zunächst! Wie treffsicher ist Ihre intuitive Erwartung? Formulieren Sie dann nützliche Gleichungen und Lösungsmethoden: Existiert eine Lösung $u : X \to \mathbb{R}$? Gibt es mehrere oder genau eine? Wie können Sie die / eine berechnen? Zudem möglichst effizient?

100	100	100	100	100	100	100	100	100	100	100	⋯
100	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	⋯
100	81	64	49	36	25	16	9	4	1	0	⋯

Vorsicht bei unendlichen Spielfeldern

Lösung: (0) Die Mittelwerteigenschaft $u(x) = \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$ hat als mögliche Lösungen $u_m(x) = 100 + mx$ mit $m \in \mathbb{R} \cup \{\pm\infty\}$.
⚠ Allein der eine Punkt $u(0) = 100$ legt die Gerade noch nicht fest!
Die simple Bilanzgleichung allein genügt hier also nicht zur Lösung.

Wir müssen etwas tiefer graben und das zu Grunde liegende **Modell** genauer präzisieren und auswerten: die Irrfahrt auf $\mathbb{Z}$ [*random walk*].
Der Übergang zu einem feineren Modell ähnelt der **Mikrofundierung**, um globale Bilanzgleichungen zu erklären oder notfalls zu ergänzen

☺ Wir vollenden die Rechnung, indem wir noch **weitere Bedingungen** einführen und nutzen: Der Erwartungswert $u(x) \in \mathbb{R}$ existiert, erfüllt die Bilanzgleichung $u(x) = \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$ und ist **beschränkt** durch $u(x) \in [0, 100]$ für alle $x \in \mathbb{N}$. Dann bleibt nur $u(x) = 100$ für alle $x \in \mathbb{N}$.

⚠ Wir sehen hier im Miniaturbeispiel den entscheidenden Unterschied zwischen einem endlichem und einem unendlichem Spielgraphen, zwischen einem kompakten und einem nicht-kompakten Gebiet.
Im Allgemeinen ist nur der endliche / kompakte Fall gutartig.

Vorsicht bei unendlichen Spielfeldern

(1) Die Lösungen sind $u_m(x) = 100 + mx + x^2$ mit $m \in \mathbb{R} \cup \{\pm\infty\}$.
Jede erfüllt $u_m(x) = \frac{1}{2}u_m(x-1) + \frac{1}{2}u_m(x+1) - 1$ für alle $x \in \mathbb{N}_{>0}$.

⚠ Die Lösung $u_{-\infty}$ ist auch auf $X = \{0, 1, 2, \dots, n\}$ mit $n \geq 3$ möglich.
Die Frage ist also: Warum ist dies auf $\mathbb{N}$ die einzig richtige Antwort?

☺ Wir vollenden die Rechnung, indem wir **weitere Bedingungen** einführen und nutzen: Nur $u_{-\infty}$ erfüllt die Beschränkung $u \leq 100$.

Jedes mathematische Phänomen lässt sich finanziell ausnutzen (in gut gemeinten Anwendungen) bzw. ausbeuten (in betrügerischer Absicht). Letzteres gilt besonders für wenig bekannte Regelmäßigkeiten (Sätze), noch besser eignen sich Unverständnis und weit verbreitete Fehler. Ausgenutzt wird hierbei nicht direkt der mathematische Sachverhalt, sondern vor allem die ungleich verteilte Kenntnis darüber.

⚠ Ein Spiel oder Geschäft wie in (1) ist ein **Knebelvertrag** und zielt darauf, eine Vertragspartei möglichst langfristig zu binden und finanziell auszubeuten, meist wie hier durch Kündigungsfristen und -bedingungen. In klaren Fällen sind solche Verträge sittenwidrig und somit ungültig.

Vorsicht bei unendlichen Spielfeldern

(2) Ein rationaler Spieler wählt jeweils den besten Zug:

$$\boxed{u(x) = \max\{\ 0,\ \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1) - 1\ \}}$$

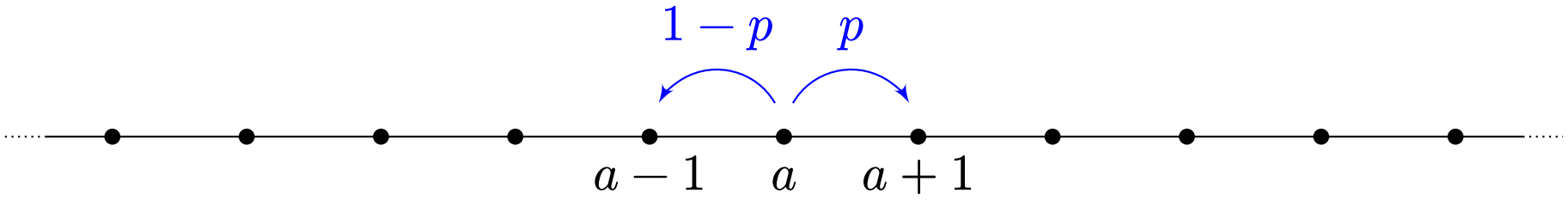
Als erstes führen wir dieses Problem zurück auf $X = \{0, \dots, n\}$:
Wir wählen den rechten Rand n hinreichend groß und setzen $u(n) = 0$.
Aufgrund der Beschränkung $u(x) \in [0, 100]$ für $x \geq 0$ und der Zugkosten $c = -1$ finden wir $u(x) \in [0, 99]$ für $x \geq 1$, sodann $u(x) \in [0, 98]$ für $x \geq 2$ und so weiter, bis schließlich $u(x) = 0$ für $x \geq 100$. Also genügt $n = 100$.

Das lineare Gleichungssystem ergibt unsere obigen Quadratzahlen!
Wenn wir erst einmal $u(10) = 0$ wissen, dann ist dies leicht zu sehen:
Von der Position $x \in \{0, \dots, 10\}$ bis zum Rand $\{0, 10\}$ ist die lineare Erwartung $100 - 10x$ und zudem die erwartete Reisezeit $x(10 - x)$.
Die Gewinnerwartung ist demnach $u(x) = 100 - 20x + x^2 = (10 - x)^2$.

☺ Wie immer gilt auch hier: Ausrechnen ist mühsam. Prüfen ist leicht!

☺ Im Gegensatz zum Knebelvertrag (1) scheint mir dieses Spiel (2) überschaubar und sowohl moralisch wie juristisch akzeptabel.

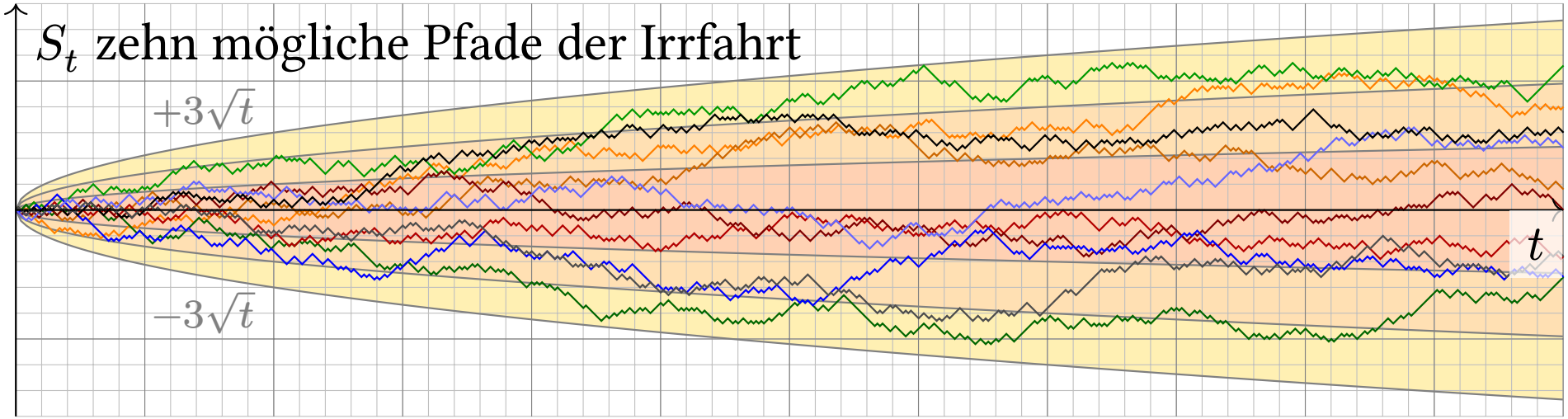
Zufällige Irrfahrt / *random walk*



Zufällige Irrfahrt in $X = \mathbb{Z}$: Zur Zeit $t = 0$ starten Sie im Punkt $S_0 = a$.

Im Schritt von S_t nach S_{t+1} gehen Sie mit Wahrscheinlichkeit $p \in [0,1]$ nach rechts und entsprechend mit Wahrscheinlichkeit $(1-p)$ nach links.

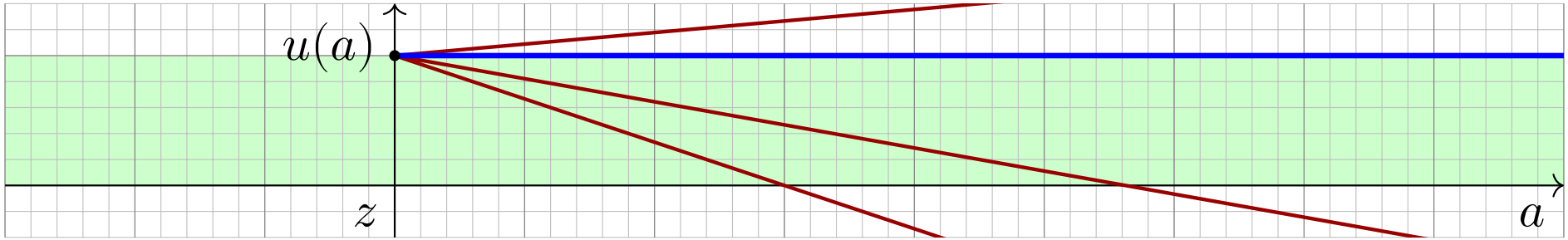
Das heißt, $S_t : \Omega \to \mathbb{Z}$ ist gegeben durch $S_t = a + X_1 + \cdots + X_t$ mit unabhängigen Zuwächsen, $\mathbf{P}_a(X_t{=}{+1}) = p$ und $\mathbf{P}_a(X_t{=}{-1}) = 1-p$.



Zufällige Irrfahrt / *random walk*

Aufgabe: (0) Bestimmen Sie zu S_t die Verteilung, Erwartung, Streuung.
Wir untersuchen speziell den symmetrischen Fall $p = 1/2$, ohne Drift.
Sie beginnen im Startpunkt $a \in \mathbb{Z}$ und fixieren einen Zielpunkt $z \in \mathbb{Z}$.
(1) Wie groß ist die Wkt $u(a) \in [0,1]$, das Ziel zu erreichen? (2) Reisezeit?

Lösung: (1) Wir erhalten eine Binomialverteilung, affin transformiert:
$\mathbf{P}_a(S_t = a + 2k - t) = \binom{t}{k} p^k (1-p)^{t-k} = \binom{t}{k} 2^{-t}$ für $t, k \in \mathbb{N}$ und $p = \frac{1}{2}$.
Somit gilt $\mathbf{E}(S_t) = a$ und $\mathbf{V}(S_t) = t$, also $\sigma(S_t) = \sqrt{t}$ und $S_t \approx N(a,t)$.



(1) Offensichtlich gilt $u(z) = 1$, denn hier ist der Start auch das Ziel.
Für $a > z$ gilt die Mittelwerteigenschaft $u(a) = \frac{1}{2}u(a+1) + \frac{1}{2}u(a-1)$.
Somit ist $u : \mathbb{Z}_{\geq z} \to [0,1]$ eine Gerade, $u(a) = 1 + m(a-z)$. (Warum?)
Zudem ist u beschränkt, $0 \leq u \leq 1$, daher folgt $m = 0$. Ebenso auf $\mathbb{Z}_{\leq z}$.

Zufällige Irrfahrt / *random walk*

(2) Sei $T_z \in \mathbb{N}$ die Zeit des ersten Besuchs im Zielpunkt z und $\mathbf{E}_a(T_z)$ die erwartete Reisezeit vom Startpunkt a zum Zielpunkt z. Dies ist invariant unter Verschiebungen, also $\mathbf{E}_{a+k}(T_{z+k}) = \mathbf{E}_a(T_z)$.

Zunächst gilt $\mathbf{E}_a(T_a) = 0$. Für $a \neq z$ zeigen wir nun $\mathbf{E}_a(T_z) = \infty$:

$$w := \mathbf{E}_0(T_1) \;= 1 + \tfrac{1}{2}\big[\underbrace{\mathbf{E}_1(T_1)}_{=0} + \mathbf{E}_{-1}(T_1)\big]$$

$$\mathbf{E}_{-1}(T_1) = \underbrace{\mathbf{E}_{-1}(T_0)}_{=w} + \underbrace{\mathbf{E}_0(T_1)}_{=w} \;= 2w$$

Hieraus folgt $w = 1 + w$. Das ist für $w \in \mathbb{R}$ unmöglich.
Es bleibt nur $\mathbf{E}_0(T_1) = w = \infty$, also $\mathbf{E}_0(T_z) = \infty$ für alle $z \neq 0$.

☺ Bei einer symmetrischen Irrfahrt ($p = 1/2$, ohne Drift) erreichen wir jeden Punkt mit Wkt 100%, aber die erwartete Reisezeit ist unendlich!

Die Rechnung ist einfach, dank unserer geschickten Formalisierung.
Die Interpretation hingegen muss man erst einmal verarbeiten.
Die naive Anschauung kann einen hier leicht narren.

Zufällige Irrfahrt / *random walk*

Die **Irrfahrt auf $\mathbb{Z}$** ist ein einfaches, aber wichtiges Modell. Typische Anwendung: Kontostand bei zufälligen Gewinnen und Verlusten, etwa als Modell für **Aktienkurse** (Louis Bachelier, 1900).

Ähnlich entsteht die **Brownsche Bewegung** durch Wärmebewegung. Der schottische Botaniker Robert Brown (1773–1858) entdeckte 1827 unter dem Mikroskop das unregelmäßige Zittern von Pollen in Wasser. Anfangs hielt er Pollen für belebt, doch er fand dasselbe bei Staubteilchen.

Albert Einstein (1879–1955) erklärte dies 1905 durch die ungeordnete **Wärmebewegung** der Wassermoleküle, die aus allen Richtungen in großer Zahl gegen die Pollen stoßen. Quantitativ konnte er so die Größe von Atomen bestimmen und die Anzahl pro Mol, die **Avogadro–Zahl**.

Experimentell bestätigte dies Jean Perrin (1870–1942): *Mouvement Brownien et Réalité Moléculaire* (1909). Für diesen Nachweis der Teilchennatur der Materie erhielt er 1926 den Physik-Nobelpreis.

Die Gerade finden wir durch vollständige Induktion: Aus $u(z) = 1$ und $u(z+1) = 1 + m$ folgt $u(a+1) = 2u(a) - u(a-1) = [2 + 2m(a-z)] - [1 + m(a-1-z)] = 1 + m(a+1-z)$.

🙂 Bei einer symmetrischen Irrfahrt ($p = 1/2$, ohne Drift) erreichen wir jeden Punkt mit Wkt 1! George Pólya (1887–1985) zeigte 1921: Jeden Punkt in $\mathbb{Z}$ besuchen wir mit Wkt 1 unendlich oft. Dies gilt ebenso in Dimension 2 bei Irrfahrt auf dem ebenen Gitter $\mathbb{Z}^2$. Erstaunlicherweise gilt es nicht mehr in Dimension $n \geq 3$ bei Irrfahrt auf dem Gitter $\mathbb{Z}^n$. Anschaulich: Ein betrunkener Mensch findet sicher irgendwann nach Hause, ein betrunkener Fisch oder Vogel hingegen nicht!

📖 Ausführung bei Feller, *Introduction to Probability*, vol. 1 (1968), §XIV.7: Das Sprichwort „Alle Wege führen nach Rom.“ stimmt zumindest zweidimensional. Dreidimensional ist die Rückkehrwahrscheinlichkeit nur etwa 34%, siehe `en.wikipedia.org/wiki/Random_walk`.

Erwartete Reisezeit bis zum Rand

Aufgabe: Wie lange ist die erwartete Reisezeit $u(x)$ bis zum Rand?
Erste Beispielrechnungen für $n = 2, \dots, 10$ ergeben folgende Daten:

0	1	0								
0	2	2	0							
0	3	4	3	0						
0	4	6	6	4	0					
0	5	8	9	8	5	0				
0	6	10	12	12	10	6	0			
0	7	12	15	16	15	12	7	0		
0	8	14	18	20	20	18	14	8	0	
0	9	16	21	24	25	24	21	16	9	0

Übung: Auf $X = \{0, \dots, n\}$ gilt $u(x) = x(n - x)$. Dann $n \to \infty$.

Erwartete Reisezeit bis zum Rand

Wir betrachten das Spielbrett $X = \{0, \dots, n\}$ mit Rand $\partial X = \{0, n\}$
und lösen $u(x) = 1 + \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$ mit $u(0) = u(n) = 0$.
Wir vermuten, dass die Lösung eindeutig ist. Gilt das? Satz D1c!

☺ Im endlichen Fall genügt unsere einfache Bilanzgleichung.

Wie immer ist es hilfreich, zunächst kleine Beispiele zu betrachten.
Die kleinen Fälle für $n = 2, \dots, 10$ lösen Sie leicht per Hand,
als lineares Gleichungssystem oder Probieren & Prüfen.

☺ Steht das Ergebnis erst einmal da, so ist es leicht zu prüfen!

Für jedes n beschreiben die Werte $x \mapsto u(x)$ eine Parabel:
Dies sehen Sie am leichtesten, wenn sie die Differenzen betrachten
(diskrete erste Ableitung) und diese als lineare Funktionen erkennen.

☺ Damit haben Sie die Formel gefunden, die Sie nun beweisen wollen.

Damit lösen Sie die obigen Spiele auf dem Spielbrett $X = \{0, \dots, n\} \subset \mathbb{Z}$
durch (0) lineare Gewinnerwartung plus (1) quadratische Reisezeit.

☺ Wir bestaunen das Zusammenspiel von Rekursion und Induktion!

Erwartete Reisezeit bis zum Rand

Beweis: Auf $X = \{0, \dots, n\}$ betrachten wir die Funktion

$$\boxed{u : X \to \mathbb{R} : x \mapsto x(n-x).}$$

Wir finden:

$$u(x-1) = xn - n - x^2 + 2x - 1$$
$$u(x+1) = xn + n - x^2 - 2x - 1$$

Also erfüllt u die geforderte Differenzengleichung:

$$\boxed{u(x) = 1 + \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1)}$$

Dies ist also *eine* Lösung, und dank Satz D1c zudem die einzige.

Beispiele: Für $r, s \in \mathbb{N}$ und $n = r + s$ gelten folgende schöne Formeln:

- Die erwartete Reisezeit von x bis $\{x - r, x + s\}$ beträgt rs.
- Diese Formel besteht weiter sogar für $r = \infty$ oder $s = \infty$.
- Die erwartete Reisezeit von x bis $\{x - r, x + r\}$ beträgt r^2.

Erwartete Reisezeit bis zum Rand

Diese Rechnung bietet uns einen zweiten und unabhängigen Beweis für die erwartete Reisezeit $u(x)$ in $X = \mathbb{N}$ bis zum Rand $\partial X = \{0\}$: Wir haben $u(x) \geq x(n-x)$ für jedes $n \geq x$, also

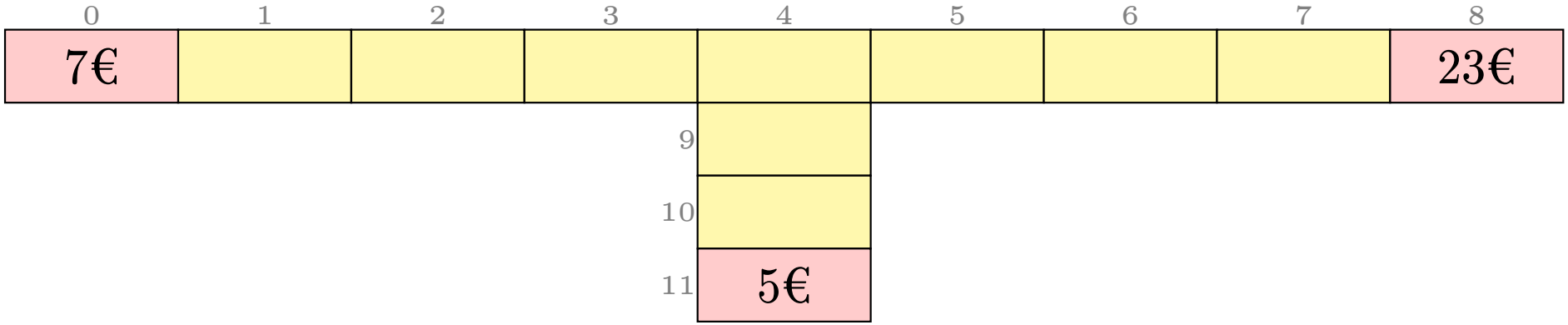
$$u(x) = \begin{cases} 0 & \text{falls } x = 0, \\ \infty & \text{falls } x > 0. \end{cases}$$

Es ist hilfreich, neben Funktionen $u : X \to \mathbb{R}$ auch $u : X \to \mathbb{R} \cup \{+\infty\}$ oder $u : X \to \mathbb{R} \cup \{-\infty\}$ zuzulassen. Diese treten natürlich auf, wie hier. Mittelwertgleichungen wir $u(x) = 1 + \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$ bleiben sinnvoll, solange niemals $+\infty$ und $-\infty$ addiert werden müssen.

☺ Die Erweiterung zu $\mathbb{R} \cup \{+\infty\}$ bzw. $\mathbb{R} \cup \{-\infty\}$ nützt für monotone Grenzwerte wie $u_0 \leq u_1 \leq u_2 \leq \ldots \nearrow u$ oder $u_0 \geq u_1 \geq u_2 \geq \ldots \searrow u$, und sei es nur als technisches Hilfsmittel für Zwischenrechnungen.

☺ Bilanzgleichungen funktionieren wunderbar auf *endlichen* Graphen. Auf unendlichen Graphen benötigen wir zusätzliche Bedingungen. Manchmal können wir durch endliche Teilgraphen ausschöpfen.

Zufällige Irrfahrt und optimale Entscheidung



Aufgabe: Selbes Spiel wie zuvor, aber auf einem neuem Spielbrett.
Wie viel würden Sie als Teilnehmer zahlen / als Anbieter verlangen?

(0) Was ist die Gewinnerwartung $u(x)$ für jedes Startfeld $x \in X$?
(1) Jeder Zug kostet, $c = -1$€, und Sie müssen zu Ende spielen.
(2) Jeder Zug kostet, $c = -1$€, und Sie dürfen jederzeit aufgeben.

Bitte schätzen Sie zunächst! Wie treffsicher ist Ihre intuitive Erwartung?
Formulieren Sie dann nützliche Gleichungen und Lösungsmethoden!
Die oben entwickelten Werkzeuge wirken auch in diesem Beispiel.

L Wir haben diese Aufgabe für Schüler:innen und Studieninteressierte sorgsam ausgearbeitet, siehe `eiserm.de/lehre/stukus/Gewinnerwartung.pdf`

Zufällige Irrfahrt und optimale Entscheidung

Lösung: (0) Gewinnerwartung ohne Zugkosten:

7	8	9	10	11	14	17	20	23
				9				
				7				
				5				

☺ Dies lösen Sie linear, mit neun Gleichungen für neun Unbekannte. Wir setzen hierzu $E = \{\, u : X \to \mathbb{R} \mid u(0) = 7,\ u(8) = 23,\ u(11) = 5 \,\}$ und $\Phi_0 : E \to E : u \mapsto \bar{u}$ mit $\bar{u}(x) = \sum_y p(x,y)u(y)$ und Übergangswkt $p(x,y) \in [0,1]$ von Feld x auf das Nachbarfeld y, wobei $\sum_y p(x,y) = 1$.

Die obigen Werte sind der eindeutige Fixpunkt von Φ_0, also die Lösung der Gleichung $\Phi_0(u) = u$. Dieses lineare Gleichungssystem können Sie exakt lösen, mit den Mitteln der Linearen Algebra, oder auch iterativ annähern durch Banachs Fixpunktsatz, mit den Mitteln der Analysis.

Eine Tabellenkalkulation löst dies für Sie iterativ. Wenn Sie handrechnen, zerlegen Sie das Problem geschickt in drei bereits gelöste Teilprobleme.

Zufällige Irrfahrt und optimale Entscheidung

Lösung: (1) Interessant ist die erwartete Reisezeit bis zum Rand:

0.00	6.30	10.60	12.90	13.20	12.90	10.60	6.30	0.00
				10.80				
				6.40				
				0.00				

Daraus erhalten wir folgende Gewinnerwartung mit Zugkosten:

7.00	1.70	−1.60	−2.90	−2.20	1.10	6.40	13.70	23.00
				−1.80				
				0.60				
				5.00				

☺ Beides ist leicht zu prüfen. Finden gelingt dank linearer Gleichungen.
Wir setzen hierzu $E = \{\, u : X \to \mathbb{R} \mid u(0) = 7,\ u(8) = 23,\ u(11) = 5 \,\}$ und $\Phi_0 : E \to E : u \mapsto \bar{u}$ mit den Daten $\bar{u}(x) = c(x) + \sum_y p(x,y)u(y)$.
Wir lösen die lineare Fixpunktgleichung $\Phi_0(u) = u$ wie im Fall (0).

Zufällige Irrfahrt und optimale Entscheidung

Lösung: (2) Gewinnerwartung mit Zugkosten und Abbruchmöglichkeit:

7.00	2.67	0.33	0.00	0.00	2.20	6.40	13.70	23.00
				0.00				
				1.50				
				5.00				

☺ Die Lösung u_0 im linearen Fall, ohne Entscheidung, ist leicht.
Wir lösen hierzu die lineare Fixpunktgleichung $\Phi_0(u_0) = u_0$ wie in (1).

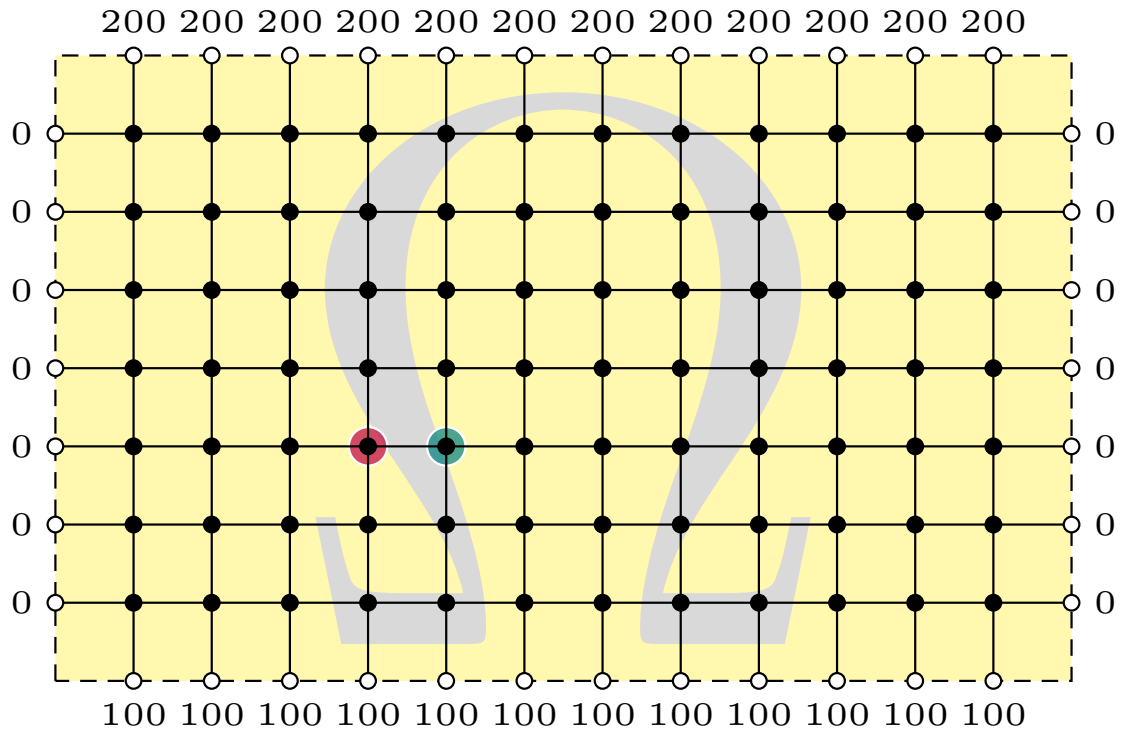
⚠ Die Lösung mit Entscheidungsmöglichkeit ist nicht $\max\{0, u_0\}$!
Diese naive Fehlannahme führt tatsächlich zu Fehlentscheidungen.

☺ Wir erhalten sie vielmehr als Fixpunkt von $\Phi(u) = \max\{0, \Phi_0(u)\}$.
Diesen Operator können wir zur Iteration nutzen, wie oben gezeigt.

Es ist bemerkenswert, dass die iterative Methode so oft funktioniert!
Hierzu beweist man möglichst allgemein, dass $\Phi : E \to E$ kontraktiv ist,
so dass Sie Banachs Fixpunktsatz anwenden können. Analysis sei Dank!
Damit können Sie getrost u berechnen und die optimale Strategie ablesen.

Irrfahrten und Potentiale: das Dirichlet–Problem

Zufällige Irrfahrt auf einem Spielfeld $\Omega \subseteq \mathbb{Z}^2$:



Sie ziehen mit Wkt $\frac{1}{4}$ nach links / rechts / oben / unten.

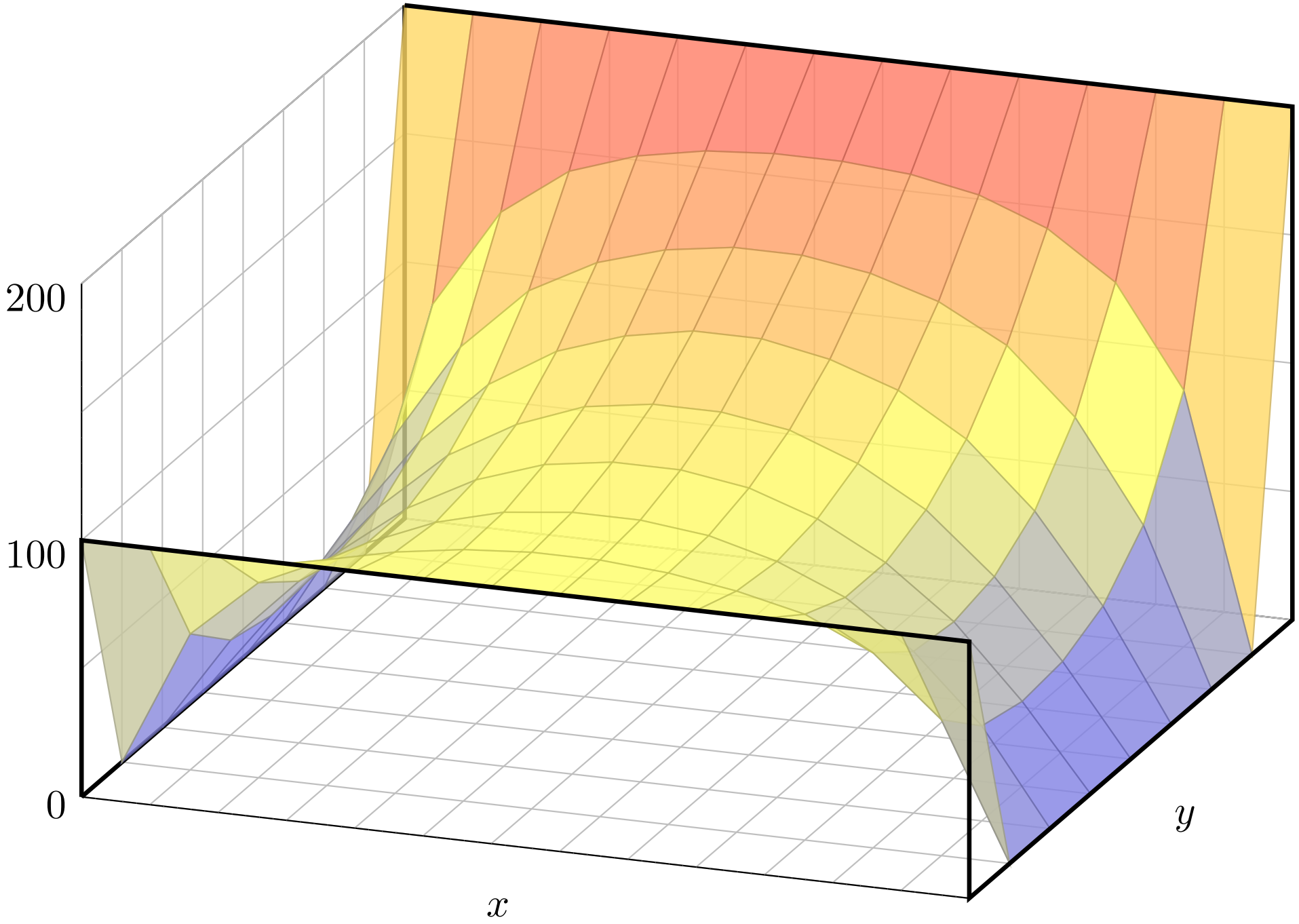
Das Spiel endet am Rand mit dem gezeigten Gewinn.

Für welche Startpunkte würden Sie 100 zahlen?

Aufgabe: (1) Wie groß ist die Gewinnerwartung $u(x, y)$ auf jedem Feld? Existiert eine Lösung $u : \Omega \to \mathbb{R}$? Ist die gesuchte Lösung eindeutig? Wie berechnet man sie? möglichst effizient? näherungsweise?

☺ Kontext und Anwendung ändern sich, die Rechnung bleibt dieselbe! (2) Hooke: Netz aus Massen und Federn. (3) Kirchhoff: Spannung einer elektrischen Schaltung. (4) Fourier: diskrete Wärmeleitung / Diffusion.

Irrfahrten und Potentiale: das Dirichlet–Problem



Irrfahrten und Potentiale: das Dirichlet–Problem

Lösung: (1) Sei $u(x,y)$ die Gewinnerwartung auf dem Feld $(x,y) \in \Omega$.
In jedem Randpunkt $(x,y) \in \partial\Omega$ ist der Gewinn $u(x,y)$ fest vorgegeben.
In jedem inneren Punkt $(x,y) \in \Omega^\circ$ gilt die **Mittelwerteigenschaft**:

$$u(x,y) = \tfrac{1}{4}u(x{-}1,y) + \tfrac{1}{4}u(x{+}1,y) + \tfrac{1}{4}u(x,y{-}1) + \tfrac{1}{4}u(x,y{+}1)$$

Eine solche diskrete Funktion $u : \mathbb{Z}^2 \supset \Omega \to \mathbb{R}$ nennen wir **harmonisch**.
Übung: Lösen Sie diese Gleichungen, etwa mit einer Tabellenkalkulation!

	200	200	200	200	200	200	200	200	200	200	200	200	
000	100	139	158	167	172	174	174	172	167	157	139	100	000
000	061	100	125	139	147	151	151	147	139	125	100	061	000
000	043	077	102	118	127	132	132	127	118	102	077	043	000
000	035	065	088	103	113	117	117	113	103	088	065	035	000
000	033	061	081	095	104	108	108	104	095	081	061	033	000
000	037	063	081	092	099	102	102	099	092	081	063	037	000
000	053	076	088	094	098	100	100	098	094	088	076	053	000
	100	100	100	100	100	100	100	100	100	100	100	100	

Irrfahrten und Potentiale: das Dirichlet–Problem

Allgemein sind harmonische Funktionen ein wunderschönes Thema in Analysis, Numerik, WTheorie, Physik, ..., hier der diskrete Spezialfall: Wir betrachten eine endliche Teilmenge $\Omega \subset \mathbb{Z}^2$. Innere Punkte $z \in \Omega^\circ$ sind solche, deren vier Nachbarn in Ω liegen. Der Rand ist $\partial\Omega = \Omega \smallsetminus \Omega^\circ$.

Dirichlet–Problem: In jedem Randpunkt $z \in \partial\Omega$ ist der Wert $u(z)$ festgelegt durch die vorgegebene Randfunktion $v = u|_{\partial\Omega} : \partial\Omega \to \mathbb{R}$. Gesucht sind dazu alle harmonischen Funktion $u : \Omega \to \mathbb{R}$ mit $u|_{\partial\Omega} = v$. Existiert eine Lösung? Ist sie eindeutig? Wie können wir sie berechnen?

☺ Die Aufgabe führt uns direkt zu einem **linearen Gleichungssystem**: Wir haben $n = 7 \times 12 = 84$ Unbekannte und ebensoviele Gleichungen. Das sieht vernünftig aus, bedeutet aber noch nicht, dass es genau eine Lösung gibt. Hierzu müssen wir genauer hinschauen und begründen!

Diese Anwendung ist faszinierend, sie fördert sowohl die physikalische Anschauung als auch die mathematisch-methodische Vorgehensweise. Hier gilt das wunderbare Minimum-Maximum-Prinzip, siehe Satz D1c. Daraus folgt die Eindeutigkeit und damit sogar die Existenz einer Lösung!

Harmonische Funktionen: Minimum-Maximum-Prinzip

Satz D1c: Minimum-Maximum-Prinzip, Eindeutigkeit und Existenz

(1) Jede harmonische Funktion $u : \Omega \to \mathbb{R}$ nimmt ihr Minimum und ihr Maximum am Rand $\partial\Omega$ an: $\min_\Omega u = \min_{\partial\Omega} u$ und $\max_\Omega u = \max_{\partial\Omega} u$.

(2) Für je zwei harmonische Funktionen $u, \tilde{u} : \Omega \to \mathbb{R}$ folgt:
Monotonie: Aus $u \geq \tilde{u}$ auf dem Rand $\partial\Omega$ folgt $u \geq \tilde{u}$ auf ganz Ω.
Eindeutigkeit: Aus $u = \tilde{u}$ auf dem Rand $\partial\Omega$ folgt $u = \tilde{u}$ auf ganz Ω.

(3) Existenz: Zu jeder beliebig vorgegebenen Randfunktion $v : \partial\Omega \to \mathbb{R}$ existiert genau eine harmonische Funktion $u : \Omega \to \mathbb{R}$ mit $u|_{\partial\Omega} = v$.

Skizze: (1) Sei $z \in \Omega$ eine Minimalstelle. Im Fall $z \in \partial\Omega$ sind wir fertig. Für $z \in \Omega^\circ$ gilt die Mittelwerteigenschaft, also sind alle Nachbarn von z ebenfalls Minimalstellen. Es gibt einen Weg von z zu einem Randpunkt $z' \in \partial\Omega$. Also ist auch z' eine Minimalstelle. (Ebenso für das Maximum.)

(2) Auch $v = u - \tilde{u}$ ist harmonisch. Aus $v \geq 0$ auf $\partial\Omega$ folgt $v \geq 0$ auf Ω.
(3) Wir haben so viele Unbekannte wie Gleichungen. Für jede rechte Seite existiert nach (2) höchstens eine Lösung, also genau eine. QED

Harmonische Funktionen: Eindeutigkeit und Existenz

Aufgabe: Führen Sie diese Beweisskizze sorgfältig aus!

Harmonische Funktionen sind wichtig in der Mathematik, der Physik und den Ingenieurwissenschaften. Die Analysis erklärt Ihnen hierzu das kontinuierliche Modell und zugehörige partielle Differentialgleichungen. Unsere kleine Illustration ist Teil eines größeren Gesamtkunstwerks.

Wir betrachten hier ganz bescheiden ein endlich-diskretes Modell und erhalten ein beliebig großes, doch einfaches lineares Gleichungssystem. Wir haben geeignete Werkzeuge! Dies schult, wie gesagt, wunderbar unsere physikalische Anschauung und unsere mathematische Methodik.

Wir sehen zudem, dass unsere Koeffizientenmatrix hier dünn besetzt ist; das ist typisch für Anwendungen, in denen jeder Punkt nur mit wenigen Nachbarn interagiert. Für so strukturierte Probleme bietet die Numerik spezialisierte und besonders effiziente Näherungsverfahren.

Wenn Ihnen partout nichts Besseres einfällt: Gauß geht immer…
und für Matrizen moderater Größe auch ausreichend schnell.

Harmonische Funktionen: Eindeutigkeit und Existenz

Lösung: Sei $\Omega \subset \mathbb{Z}^n$ endlich. Wir zerlegen Ω in Inneres Ω° und Rand $\partial\Omega$: Jeder innere Punkt $z \in \Omega^\circ$ hat alle seine $2n$ Nachbarn $z \pm e_k$ in Ω; jeder Randpunkt $z \in \partial\Omega$ hat mindestens einen Nachbarn außerhalb von Ω.

Zu $u : \Omega \to \mathbb{R}$ definieren wir den **Mittelwert** $\bar{u} : \Omega^\circ \to \mathbb{R}$ durch

$$\boxed{\bar{u}(z) := \frac{1}{2n} \sum_{k=1}^{n} u(z + e_k) + u(z - e_k).}$$

Wir nennen u **harmonisch**, wenn $u(z) = \bar{u}(z)$ für alle $z \in \Omega^\circ$ gilt.

(1a) Sei $u : \Omega \to \mathbb{R}$ harmonisch. Da die Menge Ω endlich ist, existiert eine Minimalstelle $z_0 \in \Omega$. Es gilt also $u(z_0) \leq u(z)$ für alle $z \in \Omega$. Liegt z_0 im Inneren, so ist jeder Nachbar $z_1 = z_0 \pm e_k$ ebenfalls eine Minimalstelle: Wir haben $u(z_1) \geq u(z_0)$. Gäbe es einen Nachbarn z_1 mit $u(z_1) > u(z_0)$, so wäre $\bar{u}(z_0) > u(z_0)$, Widerspruch! Also folgt $u(z_1) = u(z_0)$.

(1b) Es gibt einen Weg $z_0, \dots, z_\ell \in \Omega$ zu einem Randpunkt $z_\ell \in \partial\Omega$, wobei $z_i \in \Omega^\circ$ gilt für alle $0 \leq i < \ell$ sowie $z_{i+1} - z_i \in \{\pm e_1, \dots, \pm e_n\}$. Nach (1a) gilt $u(z_0) = u(z_1) = \dots = u(z_\ell)$, also ist auch z_ℓ eine Minimalstelle von u.

Harmonische Funktionen: Eindeutigkeit und Existenz

(2) Die Differenz $v = u - \tilde{u} : \Omega \to \mathbb{R}$ ist harmonisch.

(2a) Gilt $u \geq \tilde{u}$ auf dem Rand $\partial\Omega$, so dort auch $v \geq 0$.
Nach (1) folgt $v \geq 0$ auf ganz Ω, also dort auch $u \geq \tilde{u}$.

(2b) Hier gilt $v = 0$ auf $\partial\Omega$, nach (1) also $v = 0$ auf ganz Ω.

(3) Wir definieren $L : \mathbb{R}^\Omega \to \mathbb{R}^\Omega : u \mapsto \hat{u}$ durch $\hat{u}(z) = u(z)$ für $z \in \partial\Omega$ und $\hat{u}(z) = u(z) - \bar{u}(z)$ für $z \in \Omega^\circ$; das misst den Fehler zur Harmonizität.

(3a) Die Abbildung $L : \mathbb{R}^\Omega \to \mathbb{R}^\Omega$ ist $\mathbb{R}$–linear, wird also dargestellt als $L(u) = Au$ durch eine quadratische (!) Matrix $A \in \mathbb{R}^{\Omega\times\Omega}$. Nun der Clou:

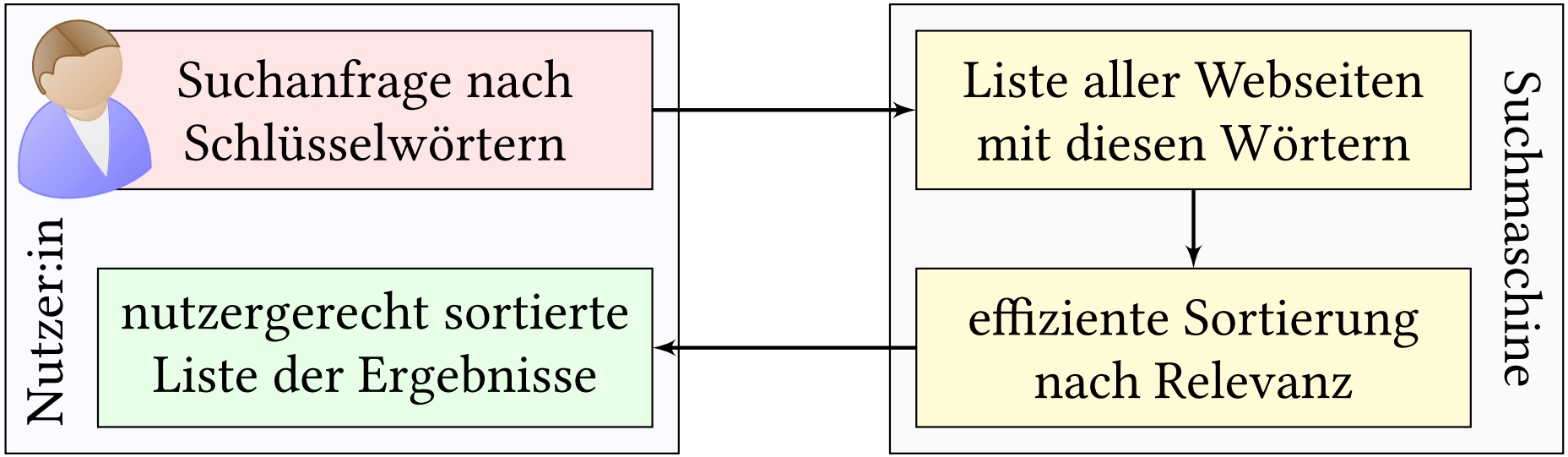
(3b) Gilt $L(u) = 0$, so ist u harmonisch mit $u|_{\partial\Omega} = 0$; dank (2) folgt $u = 0$.
Demnach ist L injektiv, und somit ist die Matrix A in $\mathbb{R}^{\Omega\times\Omega}$ invertierbar!

🙂 Das ist ein höchst eleganter Beweis und scharfsinniges Vorgehen:
Wir zeigen zunächst das allgemeine Minimum-Maximum Prinzip (1), daraus folgt die Eindeutigkeit (2) und damit schließlich die Existenz (3).

Die Argumente gelten für jeden Markov–Graphen, mit Übergangswkten auf den Kanten, und seinen diskreten Laplace–Operator $L : \mathbb{R}^\Omega \to \mathbb{R}^\Omega$.

Die Mathematik hinter Google

„Wo simmer denn dran? Aha, heute krieje mer de Suchmaschin.
Wat is en Suchmaschin? Da stelle mer uns janz dumm. …“



Mathematik: Wie misst man Relevanz von Informationen?
Artificial Intelligence (AI), Machine Learning (ML), …

Informatik: Wie verarbeitet man enorm große Datenmengen?
Big Data, Data Mining, Data Science, … „Data is the new oil.“

Finanzstrategie: Wie verdient man Geld mit einem Gratisprodukt?
„If you’re not paying for it, you’re not the customer, you are the product.“

Die Mathematik hinter Google

Als das World Wide Web Mitte der 1990er noch klein war, da genügte es, zu einer Suchanfrage einfach *alle* Treffer aufzulisten. Die Liste war noch kurz, jede:r Nutzer:in konnte sie leicht selbst überblicken. Das Internet blieb jedoch nicht lange so überschaubar.... Das Volumen explodierte!

Als Versuch einer Lösung ging 1998 die Suchmaschine Google in Betrieb und dominiert seither den Markt. Sie wird ständig weiterentwickelt. Die meisten Optimierungen hütet Google streng als Firmengeheimnis, doch das ursprüngliche Grundprinzip ist veröffentlicht und genial einfach:

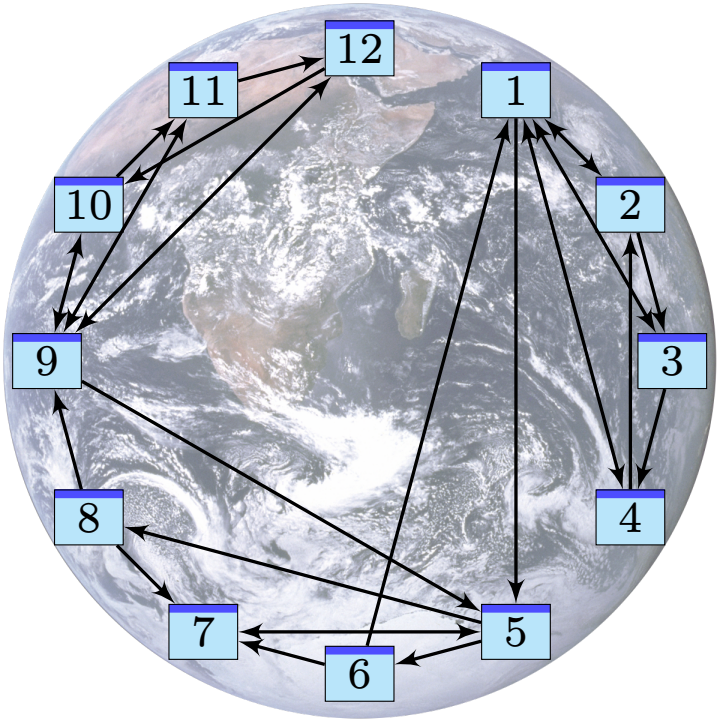
📖 Sergey Brin, Larry Page: *The anatomy of a large-scale hypertextual web search engine.* Stanford University 1998, online verfügbar unter `infolab.stanford.edu/pub/papers/google.pdf`

Bei vorherigen Suchmaschinen musste man endlose Trefferlisten durchforsten, bis man auf die ersten interessanten Ergebnisse stieß. Bei Google stehen sie auf wundersame Weise ganz oben. Wie ist das möglich? Die Antwort liegt (zu einem großen Teil) in folgender genial-einfachen Idee. Google misst die Popularität p_i (*PageRank*) jeder Seite i durch folgendes Gleichungssystem:

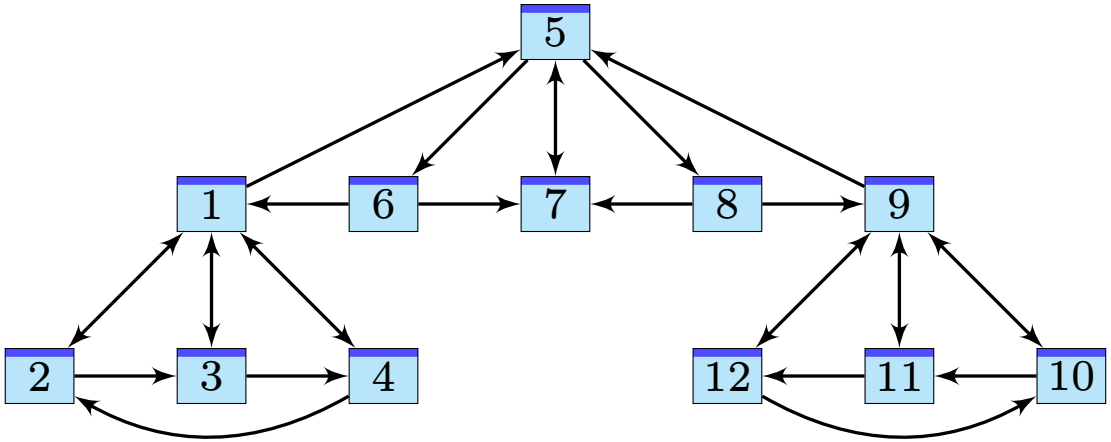
$$\text{PageRank} \qquad p_i = \frac{q}{N} + \sum_{j \to i} \frac{1-q}{\ell_j} p_j$$

Keine Angst, die Formel sieht nur auf den ersten Blick kompliziert aus. Ich werde sie anhand von Beispielen Schritt für Schritt erläutern. Wer sowas schon gesehen hat, weiß, dass es sich um eine besonders einfache Formel handelt, nämlich ein *lineares Gleichungssystem*, das keine Quadrate oder komplizierteres enthält. Schon die Formel von Pythagoras $a^2 + b^2 = c^2$ ist komplizierter.

Chaos und Struktur im Web



Miniaturbeispiel des Web als ein Graph aus Seiten $i = 1, \dots, N$ und Links $i \to j$.
Versuch einer hierarchischen Anordnung:



Eine Seite ist populär, wenn viele Seiten auf sie verweisen? Zu naiv! ☹
Eine Seite ist populär, wenn viele populäre Seiten auf sie verweisen. ☺

Ein zufälliger Surfer folgt von der aktuellen Seite irgendeinem der Links.

Aufgabe: Berechnen Sie die Aufenthaltswkten. Konvergieren sie gegen ein Gleichgewicht? Wie schnell? Immer dasselbe, d.h. ist es eindeutig?

☺ Im Rückblick ist die abstrakt-mathematische Idee genial einfach.
Wer diese Aufgabe bis 1998 professionell löste, ist heute Milliardär.

Chaos und Struktur im Web

Klassische Texte sind von einer Person geschrieben und linear: Ein Buch hat einen Anfang und ein Ende, typischerweise liest man es von vorne nach hinten in der Reihenfolge der Seiten. Meist gibt es zudem ein Inhaltsverzeichnis oder einen Index zum leichteren Nachschlagen. (Ja, liebe Kinder, unsere Vorfahren konnten Texte mit hunderttausend Buchstaben am Stück lesen, ohne Clicks und ohne Werbung. Man nannte das „Buch“ und speicherte es auf „Papier“. Damals!)

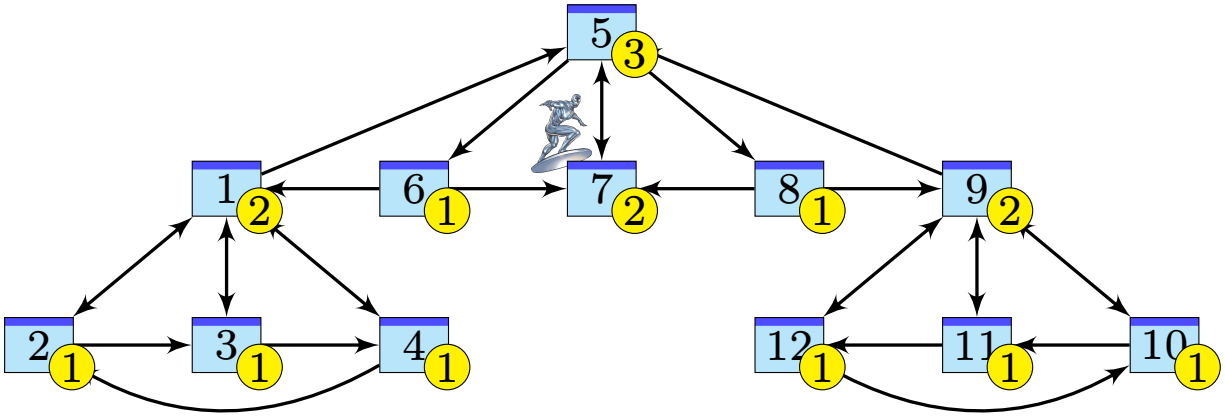
Webseiten dagegen bilden eine gänzlich andere Struktur. Niemand käme auf die Idee, das Internet von Anfang bis Ende anzuordnen, durchzulesen oder auszudrucken: Es hat keine lineare Struktur, keine erste und keine letzte Seite, es ist zudem viel zu groß, und das meiste ist uninteressant.

Die Webseiten verweisen gegenseitig aufeinander und bilden einen *Hypertext*. Zur Illustration betrachten wir ein Miniaturbeispiel bestehend aus 12 Webseiten. Unter den Seiten $1, 2, 3, 4$ wird 1 am häufigsten zitiert. Die Seite 1 scheint daher besonders relevant oder populär. Gleiches gilt für $9, 10, 11, 12$ mit 9 an der Spitze. Die Struktur von $5, 6, 7, 8$ ist ähnlich mit 7 an der Spitze. Aber die Seiten $1, 7, 9$, die wir schon als relevant erkannt haben, verweisen alle auf die Seite 5. Diese scheint daher populär / wichtig / zentral und für eine spätere Suche besonders relevant.

Diese Anordnung war Handarbeit. Lässt sie sich automatisieren? Nach welchem Algorithmus? Erster Versuch einer Bewertung: Eine Seite ist populär, wenn viele Seiten auf sie verweisen. Nachteil: Die naive Linkzählung ist leichte Beute für Manipulationen, z.B. Linkfarmen.

Zweiter Versuch: Eine Seite ist populär, wenn viele populäre Seiten auf sie verweisen. Das klingt zunächst zirkulär, lässt sich aber als lineare Gleichung auffassen und lösen. Ich erläutere dazu die besonders anschauliche Betrachtungsweise des zufälligen Surfers.

Irrfahrt eines zufälligen Surfers

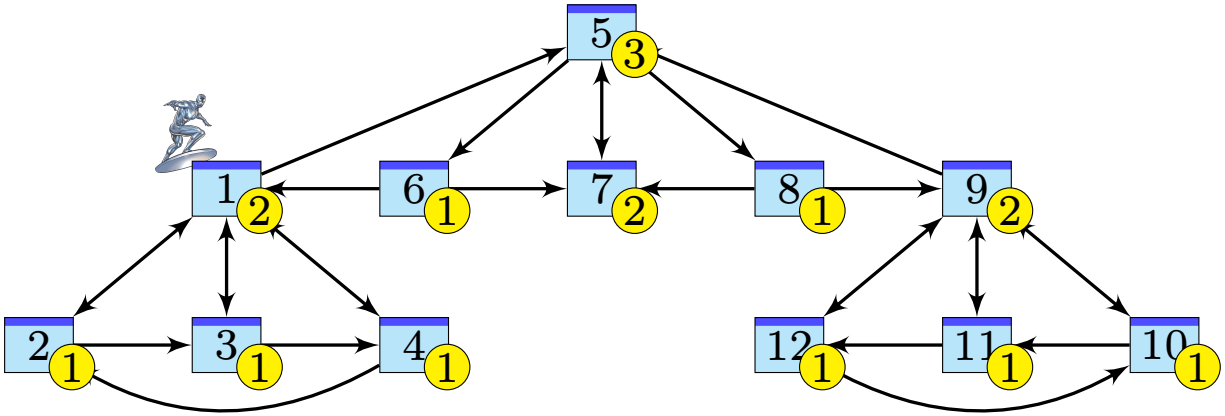


Googles Heuristik: Aufenthaltswkt $\sim$ Popularität $\sim$ Relevanz

Aufgabe: Berechnen Sie die Aufenthaltswkten bei Start auf Seite 7.

	1	2	3	4	5	6	7	8	9	10	11	12
$t = 0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t = 1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t = 2$	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
$t = 3$	.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000
$t = 4$	.000	.042	.042	.042	.417	.111	.111	.111	.000	.042	.042	.042
$t = 5$	.118	.021	.021	.021	.111	.139	.250	.139	.118	.021	.021	.021
...												
$t = 29$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t = 30$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Irrfahrt eines zufälligen Surfers



Googles Heuristik: Aufenthaltswkt ∼ Popularität ∼ Relevanz

Aufgabe: Berechnen Sie die Aufenthaltswkten bei Start auf Seite 1.

	1	2	3	4	5	6	7	8	9	10	11	12
$t = 0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t = 1$	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
$t = 2$	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
$t = 3$	.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000
$t = 4$	.234	.135	.135	.135	.151	.059	.059	.059	.000	.010	.010	.010
$t = 5$	.233	.126	.126	.126	.118	.050	.109	.050	.045	.005	.005	.005
...												
$t = 69$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t = 70$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Irrfahrt eines zufälligen Surfers

Jeder kleine Rechenschritt ist einfach, doch davon sind sehr viele nötig.
Das möchten Sie nicht selbst rechnen, es ist ideal für einen Computer!
Dazu ist es hilfreich, zunächst das Modell präzise zu formulieren:

Aufgabe: Sei $p_j(t)$ die Aufenthaltswkt auf Seite j zur Zeit t.
Von Seite j gibt es genau $\ell_j \geq 1$ weiterführende Links $j \to i$.
Bestimmen Sie daraus die Aufenthaltswkten zur Zeit $t+1$.

Lösung: Die Wkt, auf Seite i zu landen, ist die Summe aller Links $j \to i$:

Diffusion $\quad p_i(t+1) = \sum_{j \to i} \frac{1}{\ell_j} p_j(t) \quad$ kurz $\quad p(t+1) = L\,p(t)$

Gleichgewicht $\quad p_i = \sum_{j \to i} \frac{1}{\ell_j} p_j \quad$ kurz $\quad p = L\,p$

☺ Die Matrix $L \in \mathbb{R}^{N \times N}$ codiert die Linkstruktur des Internets:
Wie hier gezeigt, setzen wir $L_{ij} = 1/\ell_j$ falls $j \to i$ und $L_{ij} = 0$ sonst.
Sie mag groß sein, doch alles ist linear, also sehr einfach strukturiert.

Irrfahrt eines zufälligen Surfers

Dieses einfache Modell hat einige beachtliche Vorteile:

☺ Eine Seite ist populär, wenn viele populäre Seiten auf sie verweisen.
Das klingt zirkulär, doch als lineare Gleichung ist es einfach und klar.

☺ Die Matrix L ist sehr dünn besetzt; sie hat extrem viele Nullen.
Wir können sie effizient speichern und $p(t+1) = L\,p(t)$ berechnen.

☺ Wir beobachten Konvergenz, zumindest in unseren obigen Beispielen:
Die Wkten diffundieren für größte t gegen eine stationäre Verteilung!

Dank dieser Betrachtungsweise löst sich unser LGS sozusagen von allein!
Das ist gerade für große Matrizen viel effizienter als andere Verfahren.
(Der sonst so erfolgreiche Gauß–Algorithmus ist hier weniger ratsam.)

Leider hat dieses einfache Modell auch noch erhebliche Nachteile:

☹ Für manche Graphen konvergieren die Wkten nicht.
Wir wünschen uns Konvergenz, immer, zudem unabhängig vom Start!

☹ Schwarze Löcher: Was tun bei Seiten oder Cliquen ohne Ausgang?
Unser Surfer bleibt irgendwo gefangen, das scheint wenig plausibel!

Zufall und Notwendigkeit: Googles PageRank

Um diese Nachteile zu beheben, nutzt Google ein verfeinertes Modell:
Unser Surfer befindet sich auf Seite j und hat zwei Möglichkeiten.
Sprung: Mit Wkt $q = 0.15$ startet er neu (irgendwo, zufällig)
Fluss: Mit Wkt $(1 - q)$ folgt er zufällig einem Link $j \to i$.

Aufgabe: Wie lauten die Formeln im verfeinerten Modell? **Lösung:**

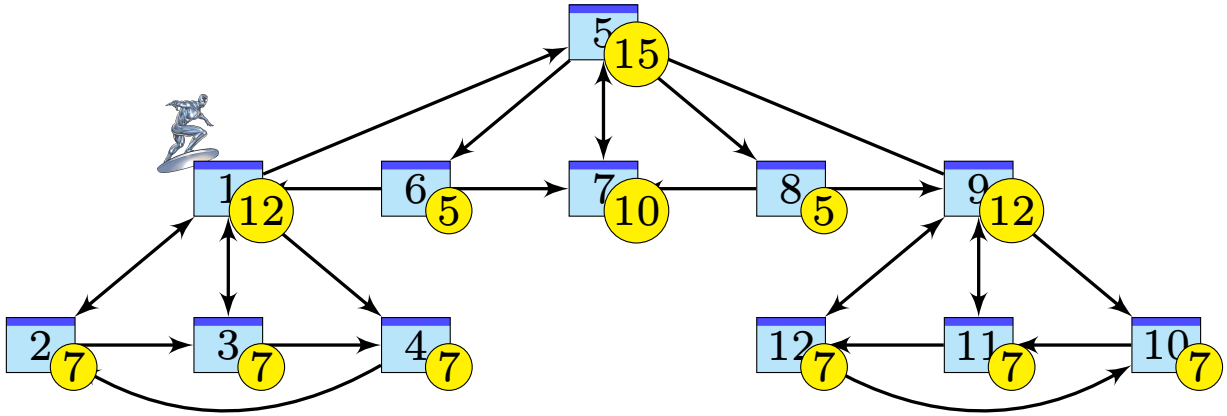
Diffusion
$$p_i(t+1) = \frac{q}{N} + \sum_{j \to i} \frac{1-q}{\ell_j} p_j(t)$$

Gleichgewicht
$$p_i = \frac{q}{N} + \sum_{j \to i} \frac{1-q}{\ell_j} p_j$$

☺ Voilà, schon haben wir es: Das ist Googles berühmter *PageRank*!

Dieses verfeinerte Modell mit Sprung / Teleportation / Neuanfang lässt sich ebenso leicht berechnen wie das einfache Modell. Zudem entkommt es schwarzen Löchern und garantiert Konvergenz, immer, sicher, sogar schnell. Der etwas wilkürliche Wert $q = 0.15$ entspricht dem typischen Verhalten, sechs bis sieben Links zu folgen, bevor man neu anfängt.

Zufall und Notwendigkeit: Googles PageRank



Googles Heuristik: Aufenthaltswkt $\sim$ Popularität $\sim$ Relevanz

Aufgabe: Wie verläuft die Diffusion bei Sprunghaftigkeit $q = 0.15$?

	1	2	3	4	5	6	7	8	9	10	11	12
$t = 0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t = 1$	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
$t = 2$	.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020
$t = 3$	.186	.124	.124	.124	.158	.021	.085	.021	.071	.028	.028	.028
$t = 4$	.180	.105	.105	.105	.140	.057	.075	.057	.057	.040	.040	.040
$t = 5$	.171	.095	.095	.095	.126	.052	.101	.052	.087	.042	.042	.042
...												
$t = 29$	.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066
$t = 30$	.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066

Zufall und Notwendigkeit: Googles PageRank

Dieses Modell hat solide theoretische Grundlagen! Bei positiver Sprunghaftigkeit $0 < q \leq 1$ garantiert **Banachs Fixpunktsatz**:

1. Es gibt genau ein Gleichgewicht $p : \{1, \dots, N\} \to \mathbb{R}$. Dieses erfüllt $p_1, \dots, p_N > 0$ und $p_1 + \cdots + p_N = 1$.
2. Für jede Anfangsverteilung $p(0)$ konvergiert die Diffusion gegen diese Gleichgewichtsverteilung, $p(t) \to p$ für $t \to \infty$.
3. Die Konvergenz ist mindestens so schnell wie die der geometrischen Folge $(1-q)^n \searrow 0$, in unserem Beispiel $q = 0.15$ also wie $0.85^n \searrow 0$.

Mit diesen Aufenthaltswkten p_i misst Google die Popularität / Relevanz! Der Satz sichert die mathematischen Grundlagen und beschert uns einen effizienten Algorithmus. Das ist der erste Schritt zur Implementierung. (Alternativ rechtfertigt der Satz nachträglich die Implementierung.)

Übung: Sobald Sie Banachs wunderbaren Fixpunktsatz kennen, können Sie diese schöne Anwendung selbst nachrechnen!

Zufall und Notwendigkeit: Googles PageRank

Dieses Modell bewährt sich in der Praxis!

1. Die berechnete Popularität p_i kommt der Nutzererwartung recht nahe: Die so berechnete Reihenfolge entspricht der „gefühlten Relevanz“.
2. Die Ergebnisse sind robust gegen Manipulationen wie Linkfarmen. Der PageRank ist nicht perfekt, aber gut genug als Ausgangspunkt.
3. *Search Engine Optimization* (SEO) ist ein großes Geschäft, alle Seiten haben über die Jahre aufgerüstet, Google behält bisher die Oberhand.

Hier arbeiten Theorie und Praxis wunderbar zusammen. Unsere obigen Beobachtungen zur schnellen Konvergenz und guten Eigenschaften sind nicht zufällig, sondern beruhen auf mathematischen Gesetzmäßigkeiten. Diese kann man beweisen und darf sich anschließend darauf verlassen. Das ist der Garant für jede stabil funktionierende Implementierung.

Mehr zur Geschichte siehe `en.wikipedia.org/wiki/History_of_Google`.
Nostalgia Nerd: *Before Google.* `youtu.be/812YkCrmksk`

Rückblick auf ein Vierteljahrhundert Google

Google wurde 1998 eingeführt und hat den Markt für Suchmaschinen über ein Vierteljahrhundert dominiert – eine Ewigkeit in Anbetracht schnelllebiger Trends. Seit Kurzem destillieren Modelle der generativen künstlichen Intelligenz (genAI) recht brauchbare Zusammenfassungen des gesammelten Internetkorpus. Wir werden sehen, wie die Nutzung sich weiter entwickelt. Im Rückblick zitiere ich Brin ud Page von 1998:

PageRank: bringing order to the Web. The citation (link) graph of the Web is an important resource that has largely gone unused in existing Web search engines. We have created maps containing as many as 518 million of these hyperlinks, a significant sample of the total. These maps allow rapid calculation of a Web page's "PageRank", an objective measure of its citation importance that corresponds well with people's subjective idea of importance. Because of this correspondence, PageRank is an excellent way to prioritize the results of Web key-word searches. For most popular subjects, a simple text matching search that is restricted to Web page titles performs admirably when PageRank prioritizes the results.

Rückblick auf ein Vierteljahrhundert Google

The most important measure of a search engine is the quality of its search results. […] our own experience with Google has shown it to produce better results than the major commercial search engines for most searches.

Diese Einschätzung hat sich anschließend als zutreffend erwiesen und war die Grundlage des Erfolges. Der PageRank misst die Aufenthaltswkt, und diese entspricht erstaunlich gut der „Popularität“ oder „Relevanz“:

Intuitive justification: PageRank can be thought of as a model of user behavior. We assume there is a “random surfer” who is given a Web page at random and keeps clicking on links, never hitting “back”, but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank. […]

Another intuitive justification is that a page can have a high PageRank if there are many pages that point to it, or if there are some pages that point to it and have a high PageRank. Intuitively, pages that are well cited from many places around the Web are worth looking at.

Rückblick auf ein Vierteljahrhundert Google

Auch zur Finanzierung durch Werbung schreiben Brin und Page schon 1998 erstaunlich weitsichtig, was sich später bewahrheiten sollte:

Currently, the predominant business model for commercial search engines is advertising. The goals of the advertising business model do not always correspond to providing quality search to users. For example, in our prototype search engine one of the top results for cellular phone is "The Effect of Cellular Phone Use Upon Driver Attention", a study which explains in great detail the distractions and risk associated with conversing on a cell phone while driving. This search result came up first because of its high importance as judged by the PageRank algorithm, an approximation of citation importance on the web. It is clear that a search engine which was taking money for showing cellular phone ads would have difficulty justifying the page that our system returned to its paying advertisers. For this type of reason and historical experience with other media, we expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers.

Rückblick auf ein Vierteljahrhundert Google

Für den Erfolg von Google war das Timing entscheidend. Zunächst die Entwicklung der Technologie und des Internets: Die Suchmaschine löste ein dringendes Problem, das erst kurz zuvor Mitte der 1990er entstand. Für den Start von Google war 1998 genau der richtige Zeitpunkt.

Sodann die Finanzierung. Google benötigt riesige Serverfarmen, diese kosten enorme Geldsummen, in Anschaffung und Unterhaltung, auch in Programmierung und fortlaufender Optimierung. Daher machte Google anfangs hohe Verluste, seither generiert Werbung gigantische Einkünfte.

In ihrem Artikel gehen Brin und Page erfreulich offen und detailliert auf die Architektur von Hardware und Software ein. Für die Nutzererfahrung ist nach der Qualität der Ergebnisse vor allem rasche Antwort wichtig. Wenige Sekunden sind tolerierbar, wünschenswert ist nahezu instantan.

Die Katalogisierung des Web und der PageRank-Algorithmus hingegen laufen auf längerer Zeitskala über Tage und Wochen im Hintergrund. Die schnelle Konvergenz verdanken sie Banachs Fixpunktsatz!

Illustration zu Banachs Fixpunktsatz

Beispiel: Legen Sie im Hörsaal eine Karte des Campus auf den Tisch.

Immer fällt genau ein Punkt der Karte auf den geographischen Punkt, den er bezeichnet.

Hat jede Kontraktion $f : (X, d) \to (X, d)$ einen Fixpunkt?

Nein, etwa $f(x) = x/2$ auf $X = \mathbb{R}^n \setminus \{0\}$.

Der Raum (X, d) muss dazu vollständig sein!

Banachs Fixpunktsatz ist ein universelles Werkzeug.

Aufgabe: (1) Illustrieren Sie Iterationen und Banachs Fixpunktsatz.
(2) Sie kennen bereits spektakuläre Anwendungen: Nennen Sie einige!
(3) Wiederholung: Formulieren und beweisen Sie Banachs Fixpunktsatz.

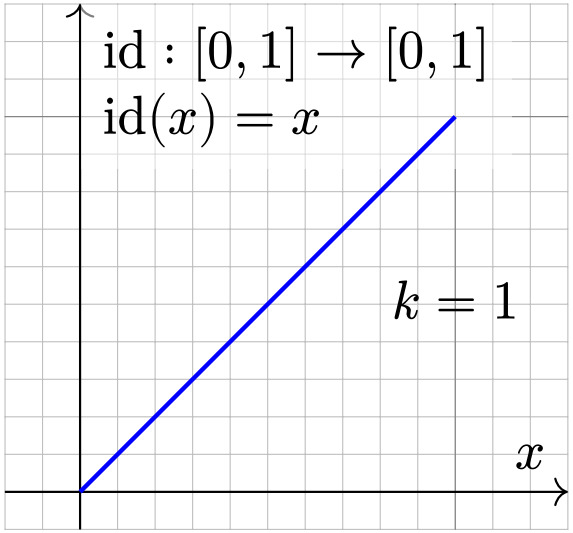
Lösung: (1) Bildhaftes Beispiel: Wenn Sie von Ihrer geographischen Umgebung X eine Landkarte im Maßstab $k = 1 : n$ (mit $n > 1$) vor sich auf den Tisch legen, dann definiert die Zuordnung jedes realen Punktes zu seinem Bildpunkt eine k–kontraktive Abbildung $f : X \to X$. Genau ein Punkt der Karte liegt auf dem geographischen Punkt, den er bezeichnet.

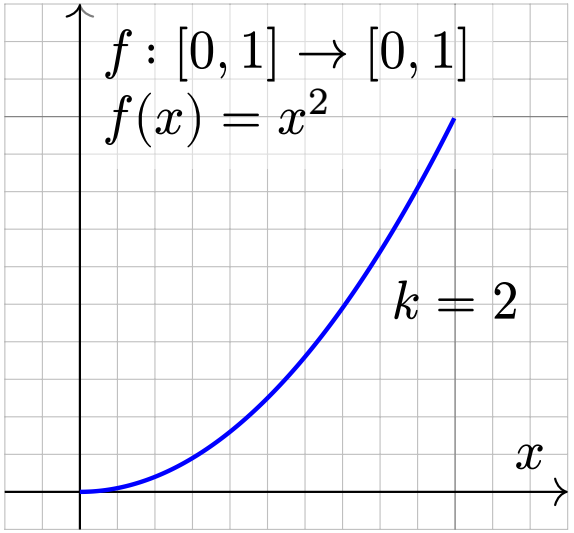
(2) Mit dem Iterationsverfahren können Sie viele Gleichungen numerisch lösen wie $x = \cos(x)$ für $x \in [0, 1]$. Das **Newton–Verfahren** baut darauf auf und verbessert ganz wesentlich die Konvergenzgeschwindigkeit.

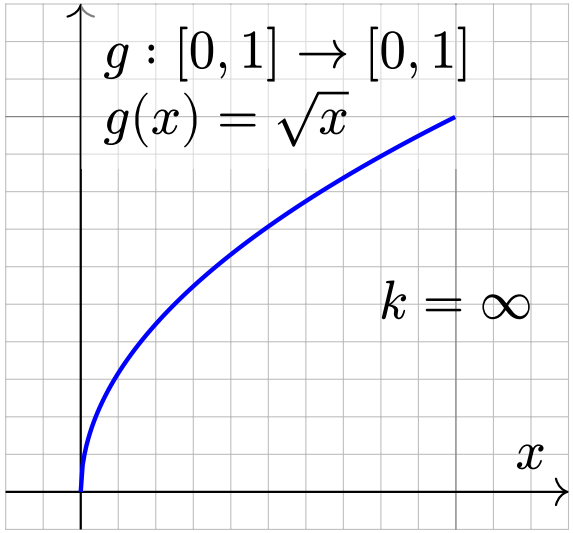
Weitere Anwendungen sind der **Satz von Picard–Lindelöf** zur Lösung von Differentialgleichungen $y' = f(x, y)$ und der **lokale Umkehrsatz** zur Konstruktion lokaler Diffeomorphismen $(f, g) : \mathbb{R}^n \supseteq U \cong V \subseteq \mathbb{R}^n$.

Diese Sätze sind schon spektakulär, doch nur die ersten Anwendungen im Grundstudium, als Spitze des Eisbergs! Viele weitere kommen hinzu.

Dehnungsschranken und Kontraktionen







Seien (X, d_X) und (Y, d_Y) Räume mit endlichen Metriken und $k \in \mathbb{R}_{\geq 0}$. Eine Abbildung $f : (X, d_Y) \to (Y, d_Y)$ heißt **k–lipschitz–stetig**, falls gilt:

$$d_Y(f(a), f(b)) \leq k\, d_X(a, b) \qquad \text{für alle } a, b \in X$$

Im Falle $0 \leq k < 1$ nennen wir f **kontraktiv** oder eine **k–Kontraktion**.

Beispiel: Für $f : \mathbb{R}^n \to \mathbb{R}^m : x \mapsto Ax + b$ mit $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^n$ gilt $|f(x) - f(y)| = |A(x - y)| \leq \|A\| \cdot |x - y|$ bezüglich der Operatornorm.

Beispiel: Sei $X \subseteq \mathbb{R}^n$ konvex und $f : X \to \mathbb{R}^m$ diff'bar mit $\|f'\| \leq k$. Für alle $x, y \in X$ existiert $z \in [x, y]$, sodass $f(x) - f(y) = f'(z)(x - y)$. Demnach gilt $|f(x) - f(y)| \leq \|f'(z)\| \cdot |x - y| \leq k\,|x - y|$.

Dehnungsschranken und Kontraktionen

Wir nennen eine solche Funktion f auch **dehnungsbeschränkt**.
Der **metrische Differenzenquotient** ist hier beschränkt gemäß

$$\frac{d_Y(f(a), f(b))}{d_X(a, b)} \leq k \qquad \text{für alle } a \neq b \text{ in } X.$$

Für $f : (X, d_X) \to (Y, d_Y)$ definieren wir daher die **Lipschitz–Norm**

$$\|f\|_{\mathrm{Lip}} = \mathrm{Lip}(f) := \sup\left\{ \frac{d_Y(f(a), f(b))}{d_X(a, b)} \,\middle|\, a \neq b \text{ in } X \right\}.$$

Die folgende Formulierung vereinheitlicht Ausnahmen und Sonderfälle:

$$\|f\|_{\mathrm{Lip}} := \inf\{ k \in \mathbb{R}_{\geq 0} \mid \forall a, b \in X : d_Y(f(a), f(b)) \leq k\, d_X(a, b) \}$$

Dies entspricht der **Operatornorm** von linearen Abbildungen normierter Vektorräume. Genau dann gilt $\|f\| < \infty$, wenn f lipschitz–stetig ist.

Genau dann gilt $\|f\| = 0$, wenn f konstant ist. Speziell für die Identität $\mathrm{id}_X : X \to X$ gilt $\|\mathrm{id}_X\| = 1$, im Sonderfall $X = \{x\}$ jedoch nur $\|f\| = 0$.

Für die Komposition von Abbildungen gilt $\|g \circ f\| \leq \|g\| \cdot \|f\|$.

Banachs Fixpunktsatz

Satz D2A: Fixpunktsatz von Banach, 1922

Sei (X, d) ein metrischer Raum, nicht-leer und vollständig. Hierauf sei $f : X \to X$ eine k–Kontraktion: Wir haben eine Kontraktionskonstante $k \in [0, 1[$, und für alle $x, y \in X$ gilt $d(f(x), f(y)) \le k\, d(x, y)$. Dann folgt:

(1) Zur Abbildung f existiert genau ein Fixpunkt $a \in X$, mit $f(a) = a$.
(2) Dieser ist Grenzwert jeder Iteration mit $x_0 \in X$ und $x_{n+1} = f(x_n)$.
(3) Diese Approximation $x_n \to a$ erfüllt die beiden Fehlerschranken

$$d(a, x_n) \;\le\; \underbrace{\frac{k}{1-k}\, d(x_n, x_{n-1})}_{\text{a posteriori, nach letztem Schritt}} \;\le\; \underbrace{\frac{k^n}{1-k}\, d(x_1, x_0)}_{\text{a priori, nach erstem Schritt}} \;\searrow\; 0.$$

(4) Allgemeiner genügt $d(f^n(x), f^n(y)) \le k_n\, d(x, y)$ für alle $x, y \in X$ und alle $n \in \mathbb{N}$ sowie $\sum_{n=0}^{\infty} k_n < \infty$. Damit gilt die feinere Fehlerschranke

$$d(a, x_n) \;\le\; d(x_1, x_0) \sum_{j=n}^{\infty} k_j \;\searrow\; 0.$$

☺ Das garantiert Eindeutigkeit, Existenz, Konstruktion und Effizienz.

Banachs Fixpunktsatz: Erläuterung

📖 S. Banach: *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales.* Fund. Math. 3 (1922) 133–181.

Die Aussage (1) garantiert Existenz und Eindeutigkeit des Fixpunktes. Die Konstruktion (2) liefert eine extrem praktische Approximation.

Gemäß (3) ist die Konvergenz $x_n \to a$ hierbei mindestens so schnell wie die Konvergenz der geometrischen Folge $k^n \searrow 0$: Dies nutzen wir bei iterativen Berechnungen als *a priori* Abschätzung des Zeitaufwandes, nach der Rechnung haben wir die (bessere) *a posteriori* Abschätzung.

Dieser wunderbare Satz geht auf Stefan Banach (1892–1945) zurück, der nützliche Zusatz (4) stammt von Johannes Weissinger (1913–1995). Zum Beispiel genügt es, dass eine gewisse Iteration f^m kontraktiv ist, also $d(f^m(x), f^m(y)) \leq k\, d(x,y)$ für eine Konstante $k \in [0,1[$ erfüllt.

Ebenso kommt es vor, dass höhere Iterationen f^n stärker kontrahieren, und die Reihe $\sum k_n$ somit viel kleiner ausfällt als die geometrische $\sum k^n$. Für die *qualitative* Konvergenzaussage (2) ist dies zunächst unwesentlich, doch ungemein hilfreich für die *quantitative* Fehlerabschätzung (3-4).

Banachs Fixpunktsatz: Beweis der Konvergenz

Beweis: *Eindeutigkeit:* Für je zwei Fixpunkte $a = f(a)$ und $b = f(b)$ gilt $d(a,b) = d(f(a), f(b)) \leq k\, d(a,b)$ mit $k < 1$, also $d(a,b) = 0$, somit $a = b$.

Existenz: Wir wählen $x_0 \in X \neq \emptyset$ und setzen $x_{n+1} = f(x_n)$ für alle $n \in \mathbb{N}$.

Per Induktion gilt $d(x_{n+1}, x_n) \leq k^n d(x_1, x_0)$: Für $n = 0$ ist dies trivial, für $n \geq 1$ gilt $d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq k\, d(x_n, x_{n-1}) \leq k^n d(x_1, x_0)$.

Dank Dreiecksungleichung erhalten wir für alle $n \leq p < q$:

$$\begin{aligned} d(x_q, x_p) &\leq d(x_q, x_{q-1}) + \cdots + d(x_{p+2}, x_{p+1}) + d(x_{p+1}, x_p) \\ &\leq (k^{q-p-1} + \cdots + k + 1)\, d(x_{p+1}, x_p) \;=\; \frac{1 - k^{q-p}}{1-k}\, d(x_{p+1}, x_p) \\ &\leq \frac{k^p - k^q}{1-k}\, d(x_1, x_0) \;\leq\; \frac{k^n}{1-k}\, d(x_1, x_0) \;\searrow\; 0 \quad \text{für } n \to \infty. \end{aligned}$$

Demnach ist $(x_n)_{n \in \mathbb{N}}$ eine Cauchy–Folge im metrischen Raum (X, d). Da (X, d) vollständig ist, existiert ein Grenzwert $a \in X$ mit $x_n \to a$.

Da f kontraktiv und somit stetig ist, folgt aus $x_{n+1} = f(x_n)$ für alle $n \in \mathbb{N}$ per Grenzübergang $a = \lim x_{n+1} = \lim f(x_n) = f(\lim x_n) = f(a)$.

Banachs Fixpunktsatz: Beweis der Fehlerschranke

Die Ungleichung für $n = p$ und $q \to \infty$ ergibt die Fehlerabschätzung (3):

$$d(a, x_n) \quad \leq \quad \underbrace{\frac{k}{1-k}\, d(x_n, x_{n-1})}_{\text{a posteriori, nach letztem Schritt}} \quad \leq \quad \underbrace{\frac{k^n}{1-k}\, d(x_1, x_0)}_{\text{a priori, nach erstem Schritt}} \quad \searrow \quad 0$$

Aussage (4) beweisen wir genauso: Wegen $\sum_{n=0}^{\infty} k_n < \infty$ gilt $k_n \to 0$ für $n \to \infty$. Insbesondere existiert $m \in \mathbb{N}$ sodass $k_n \leq 1/2$ für alle $n \geq m$, das heißt f^n ist kontraktiv für alle $n \geq m$. Sind $a = f(a)$ und $b = f(b)$ Fixpunkte von f, so auch von f^n, also folgt $a = b$ wie oben.

Für jede iterative Folge mit $x_0 \in X$ und $x_{n+1} = f(x_n)$ erhalten wir für $n \leq p < q$ wie oben $d(x_q, x_p) \leq d(x_1, x_0) \sum_{j=n}^{\infty} k_j$. Somit ist $(x_n)_{n \in \mathbb{N}}$ eine Cauchy–Folge in (X, d), also existiert ein Grenzwert $a \in X$ mit $x_n \to a$.

Für $n = p$ und $q \to \infty$ erhalten wir die feinere Fehlerabschätzung

$$d(a, x_n) \leq d(x_1, x_0) \sum_{j=n}^{\infty} k_j \quad \searrow \quad 0.$$

Das beinhaltet, für $k_j = k^j$, die beiden Abschätzungen aus (3).
Wie immer bei Konvergenz gilt auch hier: Ende gut, alles gut. $\boxed{\text{QED}}$

Erinnerung: normierte Vektorräume

Zur Erinnerung: Auf dem Raum $X = \mathbb{R}^n$ definieren wir für jedes $x \in \mathbb{R}^n$

$$\begin{aligned}
&\text{die euklidische Norm} && |x|_2 &&:= \sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2}\,,\\
&\text{die Maximumsnorm} && |x|_\infty &&:= \max\{|x_1|, |x_2|, \dots, |x_n|\},\\
&\text{die Taxinorm} && |x|_1 &&:= |x_1| + |x_2| + \cdots + |x_n|.
\end{aligned}$$

Übung: Jede dieser Normen $x \mapsto |x|$ erfreuen sich folgender Eigenschaften für alle Vektoren $x, y \in X$ und Skalare $\lambda \in \mathbb{R}$:

N0: $|x| \geq 0 = |0|$ (Positivität)

N1: $|x| > 0$ für $x \neq 0$ (Definitheit)

N2: $|\lambda x| = |\lambda| \cdot |x|$ (Homogenität)

N3: $|x + y| \leq |x| + |y|$ (Dreiecksungleichung)

Definition D2B: Norm auf einem Vektorram

Eine **Norm** auf einem Vektorraum X ist eine Abbildung $|-| : X \to \mathbb{R}_{\geq 0}$, die (N0–3) erfüllt. Das Paar $(V, |-|)$ heißt dann **normierter Raum** oder **Prä–Banach–Raum**, und bei Vollständigkeit **Banach–Raum** (D2D).

Erinnerung: metrische Räume

Sei $(X, |-|)$ ein **normierter $\mathbb{R}$–Vektorraum**, etwa $X = \mathbb{R}^n$ mit einer der obigen Normen. Allgemeiner genügt eine **Pseudonorm** $|-| : X \to [0, \infty]$ mit Eigenschaften (N0–3). Die zugehörige **Metrik** misst den Abstand:

$$d : X \times X \to [0, \infty] : (x, y) \mapsto |x - y|$$

Übung: Für alle Punkte $x, y, z \in X$ gilt dann:

M0:	$d(x, y) \geq 0 = d(x, x)$	(Positivität)
M1:	$d(x, y) > 0$ für $x \neq y$	(Definitheit)
M2:	$d(x, y) = d(y, x)$	(Symmetrie)
M3:	$d(x, z) \leq d(x, y) + d(y, z)$	(Dreiecksungleichung)

Definition D2c: Metrik und metrischer Raum

Eine **Metrik** auf einer Menge X ist eine Abbildung $d : X \times X \to [0, \infty]$, die (M0–3) erfüllt. Das Paar (X, d) heißt dann ein **metrischer Raum**.

Für Metriken ist es bequem, die Möglichkeit $d(x, y) = \infty$ zuzulassen. Für Normen hingegen verlangen wir stets Endlichkeit, wie oben erklärt.

Vollständige metrische Räume

Eine Folge $(x_n)_{n\in\mathbb{N}}$ in (X,d) **konvergiert** gegen einen Punkt $a \in X$, wenn der Abstand $d(x_n, a)$ schließlich beliebig klein wird. Ausführlich:

$$(x_n)_{n\in\mathbb{N}} \to a \text{ in } (X,d) \quad :\Leftrightarrow \quad d(x_n,a) \to 0 \text{ für } n \to \infty$$
$$:\Leftrightarrow \quad \forall \varepsilon \in \mathbb{R}_{>0}\ \exists m \in \mathbb{N}\ \forall n \in \mathbb{N}_{\geq m} : d(x_n,a) < \varepsilon$$

Wir nennen dann die Folge $(x_n)_{n\in\mathbb{N}}$ **konvergent** und a ihren **Grenzwert**. Konvergenz in (X,d) ist eine **zweistellige Relation** zwischen Folgen $(x_n)_{n\in\mathbb{N}}$ und Punkten a in X. Wir schreiben hierfür kurz $x_n \to a$.

Wir nennen $(x_n)_{n\in\mathbb{N}}$ **Cauchy–Folge** in (X,d), wenn der Durchmesser $\delta_n := \sup_{p,q\geq n} d(x_p,x_q)$ eine Nullfolge ist. In Quantorenschreibweise:

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists n \in \mathbb{N}\ \forall p,q \geq n : d(x_p,x_q) \leq \varepsilon$$

Jede konvergente Folge ist eine Cauchy–Folge, aber nicht umgekehrt.

Definition D2D: vollständiger metrischer Raum

Ein metrischer Raum (X,d) heißt **vollständig**, wenn jede Cauchy–Folge $(x_n)_{n\in\mathbb{N}}$ in (X,d) konvergiert, also ein Grenzwert $x_n \to a \in X$ existiert.

Vollständige metrische Räume

Beispiel: Im Raum $X =]0,1]$ mit euklidischer Metrik ist $x_n = 2^{-n}$ eine Cauchy–Folge, aber nicht konvergent. (Im Raum $[0,1]$ gilt $x_n \to 0$.)

Beispiel: Im Raum $\mathbb{Q}$ mit euklidischer Metrik ist $x_n = \sum_{k=0}^{n} 1/k!$ eine Cauchy–Folge, hat aber in $\mathbb{Q}$ keinen Grenzwert. (In $\mathbb{R}$ hingegen gilt $x_n \to \mathrm{e} = 2.71828\ldots$, aber dieser Grenzwert liegt nicht in $\mathbb{Q}$.)

Beispiel: Das Newton–Verfahren liefert eine effiziente Approximation von $\sqrt{5}$ durch eine rasch konvergente Folge $x_n \to \sqrt{5}$ *rationaler Zahlen*: Wir definieren $(x_n)_{n\in\mathbb{N}}$ rekursiv durch $x_0 = 3$ und $x_{n+1} = \frac{1}{2}(x_n + 5/x_n)$. Diese Folge ist eine Cauchy–Folge in $\mathbb{Q}$, sie konvergiert aber nicht in $\mathbb{Q}$. (In $\mathbb{R}$ gilt wie gewünscht $x_n \to \sqrt{5}$, aber dieser Wert liegt nicht in $\mathbb{Q}$.)

Der Raum $\mathbb{Q}$ ist unvollständig, hat also ganz anschaulich noch Lücken. Jede Cauchy–Folge möchte konvergieren, doch oft fehlt der Grenzwert. In einem vollständigen Raum kann dieses Problem niemals auftreten!

Beispiel: Die reellen Zahlen $\mathbb{R}$ sind vollständig bezüglich der Metrik $d(x,y) = |x-y|$, ebenso $\mathbb{R}^n$ bezüglich jeder beliebigen Norm und jede abgeschlossene Menge $X \subseteq \mathbb{R}^n$ bezüglich der eingeschränkten Metrik.

Vollständigkeit des Raumes $B(X, \mathbb{R})$

Sei X eine Menge. Für $u : X \to \mathbb{R}$ definieren wir die Supremumsnorm:

$$\|u\| = |u|_X := \sup\{\, |u(x)| \,\big|\, x \in X \,\}$$

Dies ist eine Norm auf dem Vektorraum der beschränkten Funktionen:

$$B(X, \mathbb{R}) := \{\, u : X \to \mathbb{R} \,\big|\, \|u\| < \infty \,\}$$

Hier steht „B“ für beschränkt [engl. *bounded*, frz. *borné*].

Satz D2E: Vollständigkeit von $B(X, \mathbb{R})$

Der $\mathbb{R}$–Vektorraum $B(X, \mathbb{R})$ ist vollständig, also ein Banach–Raum.

Beweis: Gegeben sei eine Cauchy–Folge $(u_n)_{n\in\mathbb{N}}$ in $B(X, \mathbb{R})$:
Zu $\varepsilon \in \mathbb{R}_{>0}$ existiert $m \in \mathbb{N}$ sodass für alle $p, q \geq m$ gilt $\|u_p - u_q\| \leq \varepsilon$.
Zu jedem Punkt $x \in X$ ist dann $u_n(x)_{n\in\mathbb{N}}$ eine Cauchy–Folge in $\mathbb{R}$.
Da $\mathbb{R}$ vollständig ist, existiert $u(x) \in \mathbb{R}$ als Grenzwert $u_n(x) \to u(x)$.
Für $p = m$ und $q \to \infty$ folgt $|u_m(x) - u(x)| \leq \varepsilon$, somit $\|u_m - u\| \leq \varepsilon$.
Also ist u beschränkt, und es gilt $u_n \to u$ in $B(X, \mathbb{R})$. QED

☺ Somit ist jede abgeschlossene Teilmenge $E \subseteq B(X, \mathbb{R})$ vollständig.

Äquivalenz aller Normen auf $\mathbb{R}^n$

Im Spezialfall einer endlichen Menge X, etwa $X = n = \{0, 1, \dots, n-1\}$, ist $B(X, \mathbb{R})$ unser Modellraum $\mathbb{R}^n$ mit der Maximumsnorm $\|-\| = |-|_\infty$. Zudem haben wir die Norm $|x|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p}$ für $1 \le p < \infty$. Diese Normen sind äquivalent gemäß $|x|_\infty \le |x|_p \le |x|_1 \le n|x|_\infty$, daher definieren sie dieselbe Topologie, Konvergenz, Cauchy–Folgen, etc.

Satz D2F: Äquivalenz aller Normen auf $\mathbb{R}^n$

Auf jedem $\mathbb{R}$–Vektorraum X endlicher Dimension sind je zwei Normen $|-|$ und $\|-\|$ äquivalent. Ausführlich bedeutet das: Es gibt positive Konstanten $\ell, L \in \mathbb{R}_{>0}$, sodass $\ell\,|x| \le \|x\| \le L\,|x|$ für alle $x \in X$ gilt.

Speziell für unseren Modellraum $X = \mathbb{R}^n$ mit euklidischer Norm $|-|$ ist die Sphäre $\mathbb{S}^{n-1} = \{\, x \in \mathbb{R}^n \mid |x| = 1 \,\}$ kompakt, und es genügen

$$\ell = \min\{\, \|x\| \mid x \in \mathbb{S}^{n-1} \,\} \quad \text{und} \quad L = \max\{\, \|x\| \mid x \in \mathbb{S}^{n-1} \,\}.$$

Übung: Beweisen Sie diese Schranken zur Wiederholung. Mahnendes Gegenbeispiel: Für die ℓ^p–Normen auf $\mathbb{R}^{(\mathbb{N})} \subset \ell^p(\mathbb{N}, \mathbb{R})$ gilt dies nicht!

☺ Je nach Anwendung wählen wir eine bequem passende Norm.

Blackwells hinreichendes Kriterium

Sei X eine Menge und sowie $E \subseteq B(X, \mathbb{R})$ mit Supremumsnorm $\|-\|$.

Für $u, \tilde{u} \in E$ schreiben wir $u \leq \tilde{u}$, falls $u(x) \leq \tilde{u}(x)$ für alle $x \in X$ gilt.
Ein Operator $\Phi : E \to E$ ist **isoton**, falls gilt: Aus $u \leq \tilde{u}$ folgt $\Phi(u) \leq \Phi(\tilde{u})$.

Sei $\delta \in [0, 1]$. Wir nennen $\Phi : E \to E$ **isoton δ–diskontiert**, falls für alle $u, \tilde{u} \in E$ und $c \in \mathbb{R}_{\geq 0}$ gilt: Aus $u \leq \tilde{u} + c$ folgt $\Phi(u) \leq \Phi(\tilde{u}) + \delta c$.

Satz D2G: Blackwells hinreichendes Kriterium

Ist $\Phi : E \to E$ isoton δ–diskontiert, so auch δ–lipschitz–stetig.

Beweis: Für alle $u, \tilde{u} \in E$ gilt $u - \tilde{u} \leq \|u - \tilde{u}\|$, also $u \leq \tilde{u} + \|u - \tilde{u}\|$.
Dank isotoner δ–Diskontierung folgt daraus $\Phi(u) \leq \Phi(\tilde{u}) + \delta\|u - \tilde{u}\|$.
Wir erhalten $\Phi(u) - \Phi(\tilde{u}) \leq \delta\|u - \tilde{u}\|$, ebenso $\Phi(\tilde{u}) - \Phi(u) \leq \delta\|u - \tilde{u}\|$.
Das bedeutet $\|\Phi(u) - \Phi(\tilde{u})\| \leq \delta\|u - \tilde{u}\|$, wie behauptet. QED

☺ Dieses Kontraktionskriterium ist nicht *notwendig*, aber *hinreichend*:
Wir nutzen die partielle Ordnung auf E. Mehr Struktur vereinfacht.

☺ Ist der betrachtete Teilraum E zudem abgeschlossen in $B(X, \mathbb{R})$,
so ist E vollständig, und wir können Banachs Fixpunktsatz anwenden.

Blackwells hinreichendes Kriterium

Jede Grundvorlesung zur Analysis behandelt Banachs Fixpunktsatz im Laufe des ersten Jahres. Dazu notwendig sind Konvergenz von Folgen, Cauchy–Kriterium und Vollständigkeit. Zum den Grundlagen gehören ebenso Skalarprodukte und Normen sowie Metriken und Topologien.

Das Kriterium von Blackwell wird in der Analysis meist nicht erwähnt, da es für die dortigen Anwendungen nicht unmittelbar benötigt wird. Im Kontext der Ökonomie und Optimierung hilft es jedoch ungemein, und somit rückt es nun in den Mittelpunkt unseres Interesses.

☺ Für Blackwell genügt eine beliebige Teilmenge $E \subseteq B(X, \mathbb{R})$. Für Banach sollte E zudem abgeschlossen in $B(X, \mathbb{R})$ sein.

Gegeben sei eine Zerlegung $X = X^\circ \sqcup \partial X$ sowie $v \in B(\partial X, \mathbb{R})$ als Randbedingung. Dies definiert in $B(X, \mathbb{R})$ den affinen Teilraum

$$E_v := \left\{ u \in B(X, \mathbb{R}) \,\middle|\, u|_{\partial X} = v \right\}.$$

☺ Der $\mathbb{R}$–Vektorraum $B(X, \mathbb{R})$ ist vollständig, also ein Banach–Raum. Hierin ist $E_v \subseteq B(X, \mathbb{R})$ abgeschlossen, also ebenfalls vollständig.

Von Graphen und Markov–Ketten zu Markov–Graphen

Wir betrachten einen **Graphen** $\Gamma = (X, A, \sigma, \tau)$ wie in B1I erklärt. Vereinfachend gelte $A = \bigcup_{x \in X} \{x\} \times A_x$ mit $\sigma = \mathrm{pr}_1 : A \to X : (x, a) \mapsto x$ und der **Transition** $\tau : A \to X : (x, a) \mapsto y$ bzw. lokal $\tau_x : A_x \to X : a \mapsto y$.

Diese determinische Sichtweise als Graph verallgemeinern wir nun zu einem stochastischen Modell mit Zufall / Unsicherheit. Zur Erinnerung:

Definition D2H: (diskrete) Markov–Kette

Eine (diskrete) **Markov–Kette** (X, τ) besteht aus einer (abzählbaren) Zustandsmenge X und Transition $\tau : X \to [X] : x \mapsto \sum_{y \in X} p(x, y)\, y$ mit Wkten $p(x, y) \geq 0$ und $\sum_{y \in X} p(x, y) = 1$ für alle $x \in X$.

Dies entspricht einer **stochastischen Matrix** $P = (p(x, y))_{(x,y) \in X \times X}$. Aus dem Zustand x entsteht mit Übergangswkt $p(x, y)$ der Zustand y. Die Verteilung $\mu \in [X]$ geht über in die Verteilung $\bar{\mu} = \mu\, P \in [X]$ mit

$$\bar{\mu}(y) = \sum_{x \in X} \mu(x)\, p(x, y).$$

Dies entspricht dem **Matrixprodukt**: Zeilenvektor μ mal Matrix P.

Von Graphen und Markov–Ketten zu Markov–Graphen

Angenommen, auf X sind nach dem nächsten Schritt $x \to y$
die **Gewinnerwartungen** gegeben durch $u : X \to \mathbb{R} : y \mapsto u(y)$.
Dann ist die Gewinnerwartung vor dem Schritt gegeben durch

$$\bar{u}(x) = \sum_{y \in X} p(x, y) u(y).$$

Dies entspricht dem **Matrixprodukt**: Matrix P mal Spaltenvektor u.
Die Wkten μ werden nach vorne geschoben gemäß $\mu \mapsto \bar{\mu} = \mu P$.
Die Erwartungen u werden zurück gezogen gemäß $u \mapsto \bar{u} = Pu$.

Das erinnert uns an das Prinzip der Rekursion / Rückwärtsinduktion:
Gespielt wird immer vorwärts, aber optimiert wird leichter rückwärts.

Wir verbinden nun deterministische Spiele (Zustände und Aktionen)
mit stochastischen Prozessen (aus Zuständen und Übergangswkten):
Der Spieler wählt in jedem aktiven Zustand $x \in X^\circ$ eine Aktion $a \in A_x$.
Das System geht dann vom Zustand x über nach y mit Wkt $p(x, a, y)$.

Damit gelangen wir also zu den extrem vielseitigen **Markov–Graphen**,
die wir schon aus unseren bisherigen Beispielen kennen und schätzen.

Markov–Belohnungsprozess / MRP

Bevor wir Markov–Spiele [*Markov decision processes* / MDP] erklären, ist es als Zwischenschritt vielleicht hilfreich, zunächst vereinfachend Belohnugsprozesse [*Markov reward processes* / MRP] zu betrachten: Wir verfeinern Markov–Ketten durch die Zerlegung in aktive Zustände X° und terminale Zustände ∂X und erklären zudem Belohnungen.

Definition D21: Markov–Belohnungsprozess / MRP

Ein **Markov–Belohnungsprozess** (X, τ, r, v) besteht aus einer Menge $X = X^\circ \sqcup \partial X$ mit einer Transition $\tau : X^\circ \to [X] : x \mapsto \sum_{y \in X} p(x,y)\, y$ sowie einer sofortigen Belohnung $r : X^\circ \times X \to \mathbb{R} : (x,y) \mapsto r(x,y)$ und einer terminalen Auszahlung $v : \partial X \to \mathbb{R} : x \mapsto v(x)$. Zum Diskontfaktor $\delta \in [0,1]$ erklären wir die **Erwartungsgleichung** für $u : X \to \mathbb{R}$ durch

$$u(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \sum_{y \in X} p(x,y)\big[r(x,y) + \delta u(y)\big] & \text{für } x \in X^\circ. \end{cases}$$

Wir nennen den Prozess **lösbar**, falls diese Gleichung eine eindeutige Lösung u besitzt. (Insbesondere muss jede Reihe absolut konvergieren.)

Markov–Belohnungsprozess / MRP

Bei einem Belohnungsprozess gibt es noch nichts zu entscheiden:
Er beschreibt eine Markov–Kette, eventuell mit terminalen Zuständen,
und einen Strom von Zahlungen, die wir diskontiert aufsummieren.

Ist jede Trajektorie endlich, $x_0, x_1, \dots, x_n \in X$, so führt sie zur endlichen Summe $r(x_0, x_1) + \delta r(x_1, x_2) + \cdots + \delta^{n-1} r(x_{n-1}, x_n) + \delta^n v(x_n) \in \mathbb{R}$.
Die Berechnung der erwarteten Auszahlung $u : X \to \mathbb{R}$ gelingt dann per Rückwärtsinduktion, wie in B1L erklärt. Wir müssen lediglich absolute Konvergenz sicherstellen, so dass die Erwartung wohldefiniert ist.
Dies gilt zum Beispiel, falls $y \mapsto r(x, y) + \delta u(y)$ beschränkt ist.

Im Allgemeinen gibt es unendliche Trajektorien, insbesondere Zykel.
Wir interpretieren die Erwartungsgleichung D2I als Fixpunktgleichung und nutzen den Banachschen Fixpunktsatz D2A: Sind $r : X^\circ \times X \to \mathbb{R}$ und $v : \partial X \to \mathbb{R}$ beschränkt und $\delta \in [0, 1[$, so existiert genau eine Lösung $u : X \to \mathbb{R}$, und diese lässt sich iterativ berechnen.

Übung: Formulieren Sie dies als Satz und beweisen Sie ihn.
(Wir führen die Rechnungen in Satz D2K allgemein aus.)

Markov–Spiel / MDP und Bellman–Gleichung

Definition D2J: Markov–Spiel und Bellman–Gleichung

Ein **Markov–Graph** $\Gamma = (X, A, \tau)$ besteht aus einer Zustandsmenge X und einer Aktionsmenge $A = \bigcup_{x \in X} \{x\} \times A_x$ zusammen mit Projektion $\sigma : (x, a) \mapsto x$ und Transition $\tau : A \to [X] : (x, a) \mapsto \sum_{y \in X} p(x, a, y)\, y$.

Für alle $(x, a) \in A$ und $y \in X$ gelte $p(x, a, y) \geq 0$ und $\sum_y p(x, a, y) = 1$. Wie üblich zerlegen wir die Zustandsmenge $X = X^\circ \sqcup \partial X$ in aktive Zustände $X^\circ = \operatorname{Im}(\sigma)$ und terminale Zustände $\partial X = X \smallsetminus X^\circ$.

Die **Strategiemenge** ist $S(\Gamma) := \{\, s : X^\circ \to A \mid \sigma \circ s = \mathrm{id}_{X^\circ} \,\} = \prod_{x \in X^\circ} A_x$. Auszahlungen seien terminal $v : \partial X \to \mathbb{R} : x \mapsto v(x)$ oder instantan $r : A \times X \to \mathbb{R} : (x, a, y) \mapsto r(x, a, y)$ mit Diskontfaktor $\delta \in [0, 1]$.

(0) Dieses **Markov–Spiel** (Γ, r, v) definiert die **Bellman–Gleichung**

$$u(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \sup\limits_{a \in A_x} \underbrace{\sum_{y \in X} p(x, a, y) \left[r(x, a, y) + \delta u(y) \right]}_{\text{Hamilton–Funktion } H(x,\, a,\, u)} & \text{für } x \in X^\circ. \end{cases}$$

Markov–Spiel / MDP und Bellman–Gleichung

⚠ Für die Reihe über $y \in X$ verlangen wir absolute Konvergenz, etwa $y \mapsto p(x,a,y)$ endlich getragen oder $y \mapsto r(x,a,y) + \delta u(y)$ beschränkt.

☺ Im Folgenden verlangen wir, dass r und v sowie u beschränkt sind. Damit lösen sich alle Fragen zur Konvergenz in Wohlgefallen auf.

☺ Der deterministische Spezialfall entspricht $\tau : (x,a) \mapsto y$ wie zuvor, also $p(x,a,y) = 1$ für das Ziel $y \in X$ und $p(x,a,y') = 0$ für alle $y' \neq y$.

☺ Ist der (deterministische) Graph $\Gamma = (X, A, \tau)$ zudem artinsch, so löst Rekursion / Rückwärtsinduktion B1L die Bellman–Gleichung.

☺ Jede Aktion $a : x \to y$ hat zwei Auswirkungen, die wir ausbalancieren: die sofortige Belohnung $r(x,a,y)$ und der langfristige Nutzen $u(y)$.

☺ Ist Γ lokal-endlich, so wird das Supremum jeweils angenommen. Im lösbaren Falle gilt also $u(x) = \max_{a \in A_x} H(x,a,u)$ für alle $x \in X^\circ$.

☺ Die Funktion u liefert uns die optimale Auszahlung $x \mapsto u(x)$ sowie Aktionen $s(x) \in \operatorname{Arg\,max}_{a \in A_x} H(x,a,u)$, also eine optimale Strategie s!

☺ Optimale Züge erkennen Sie leicht sobald Sie das Optimum kennen.

Bellman–Operator und Erwartungsoperator

Definition D2J: Gewinnerwartung und Optimalität

(1) Sei $E = \{u \in B(X, \mathbb{R}) \mid u|_{\partial X} = v\}$. Auf diesem affinen Teilraum definieren wir den **Bellman–Operator** $\Phi : E \to E : u \mapsto \bar{u}$ durch

$$\bar{u}(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \sup\limits_{a \in A_x} \underbrace{\sum_{y \in X} p(x,a,y)\,[r(x,a,y) + \delta u(y)]}_{\text{Hamilton–Funktion } H(x,\,a,\,u)} & \text{für } x \in X^\circ. \end{cases}$$

Die Fixpunkte von Φ sind genau die Lösungen der **Bellman–Gleichung**. Wir nennen Φ **eindeutig lösbar**, falls genau ein Fixpunkt $u \in E$ existiert, und **konvergent**, falls zudem $\Phi^n(\tilde{u}) \to u$ für $n \to \infty$ gilt für alle $\tilde{u} \in E$.

(2) Zu $s \in S(\Gamma)$ definieren wir den **Erwartungsoperator** $\Phi_s : u \mapsto \bar{u}$ durch

$$\bar{u}(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \underbrace{\sum_{y \in X} p(x,s(x),y)\,[r(x,s(x),y) + \delta u(y)]}_{\text{Hamilton–Funktion } H(x,\,s(x),\,u)} & \text{für } x \in X^\circ. \end{cases}$$

Fixpunkte $u_s \in E$ von Φ_s sind Lösungen der **Erwartungsgleichung**.

Bellman–Operator und Erwartungsoperator

Aus unserem Markov–Spiel (Γ, r, v), also den Entscheidungsprozess (X, A, τ, r, v) gemäß D2J, wird durch die Festlegung einer Strategie $s \in S(\Gamma)$ ein Belohnungprozess $(X, \tilde{\tau}, \tilde{r}, v)$ gemäß D2I. Ausführlich:

$$\tilde{\tau} : X \to [X] : x \mapsto \textstyle\sum_{y \in X} p(x, s(x), y)\, y$$
$$\tilde{r} : X \times X \to \mathbb{R} : (x, y) \mapsto r(x, s(a), y)$$

Die zugehörige Erwartungsgleichung haben wir in (2) wiederholt, als Fixpunktgleichung für den Erwartungsoperator Φ_s. Umgekehrt ist jeder Belohnungsprozess $(X, \tilde{\tau}, \tilde{r}, v)$ ein Entscheidungsprozess (X, A, τ, r, v) mit $A_x = \{a_x\}$, $p(x, a_x, y) = \tilde{p}(x, y)$, $r(x, a_x, y) = \tilde{r}(x, y)$ für alle $x \in X^\circ$

Um einen direkten, raschen Zugang anzubieten, habe ich hier beide Aspekte in einer gemeinsamen Definition D2J zusammengefasst.

Alternativ kann man die Theorie kleinschrittiger aufbauen und zunächst Markov–Ketten D2H und Belohnugsprozesse D2I als Zwischenetappen untersuchen und ihre Konvergenzfragen klären. Ich formuliere dies hier als Übung und konzentriere mich auf allgemeine Markov–Spiele (D2K).

Bellman–Operator und Erwartungsoperator

Aufgabe: Formalisieren Sie das Spiel vom Kapitelanfang explizit als ein Markov–Spiel und lösen Sie es mit Hilfe der Bellman–Gleichung.

7€								23€

Lösung: Die Zustandsmenge ist $X = \{*, 0, 1, \dots, 8\}$ mit $\partial X = \{*, 0, 8\}$ und den Auszahlungen $v(*) = 0$ sowie $v(0) = 7$ und $v(8) = 23$. Jeder aktive Zustand $x \in X^\circ = \{1, 2, \dots, 7\}$ bietet zwei Züge, $A_x = \{\text{go}, \text{stop}\}$, mit Wkten $p(x, \text{go}, x \pm 1) = 1/2$ und Belohnung $r(x, \text{go}, x \pm 1) = c := -1$ sowie den Spielabbruch mit $p(x, \text{stop}, *) = 1$ und $r(x, \text{stop}, *) = 0$.

(1) Die Strategie $s' : x \mapsto \text{go}$ für $x \in X^\circ$ ergibt $u_{s'} = \Phi_{s'}(u_{s'})$ wie folgt:

7	2	-1	-2	-1	2	7	14	23

(2) Wechsel zur Strategie s mit $s(3) = \text{stop}$ ergibt $u = \Phi_s(u)$ wie folgt:

7	$8/3$	$1/3$	0	$3/5$	$16/5$	$39/5$	$72/5$	23

☺ Diese Funktion u erfüllt zugleich die Bellman–Gleichung $u = \Phi(u)$. Dank Bellmans Optimalitätsprinzip D2M ist dies die optimale Strategie!

Bellman–Operator und Erwartungsoperator

☺ Markov–Spiele und Bellman–Gleichung sind natürlich und einfach. Das erste Anwendungsbeispiel zeigt: Unser Modell passt wunderbar!

☺ Die gezeigten Funktionen sind jeweils die einzigen Lösungen. Jede Strategie $s \in S$ ist randverbunden und somit Φ_s kontraktiv. (D2L)

Insgesamt gibt es $|S| = 2^7 = 128$ Strategien, also exponentiell in $|X^\circ|$. Wir können jede dieser 128 Strategien $s \in S$ untersuchen und jeweils die Gewinnerwarung $u_s : X \to \mathbb{R}$ berechnen. Dann können wir daraus die optimale Strategie auswählen, genauer die optimale Gewinnerwartung $u_* = \max u_s$ und dazu eine optimale Strategie $s \in S$, sodass $u_* = u_s$.

☹ Diese *globale* Optimierung durch *brute force* ist jedoch aufwändig, da die Strategiemenge $S = S(\Gamma)$ sehr groß und unübersichtlich ist.

☺ Die Bellman–Gleichung bietet dagegen eine *lokale* Optimierung, und diese gelingt wesentlich effizienter: Das ist ihr großer Nutzen! Dass beide Rechenwege zum selben Ergebnis führen, ist die Aussage von Bellmans Optimalitätsprinzip D2M, das wir als nächstes beweisen.

Gewinnerwartung und Optimalität

Definition D2J erklärt das grundlegende Modell und alle relevanten Daten: der Markov–Graph $\Gamma = (X, A, \tau)$, das Markov–Spiel (Γ, r, v), die Strategiemenge $S(\Gamma)$ und die Bellman–Gleichung für $u : X \to \mathbb{R}$.

Darauf bauend erklären wir in Definition D2J die Operatoren Φ und Φ_s. Wir wünschen uns, dass sie kontrahieren, oder wenigstens konvergieren, oder jeder einen eindeutigen Fixpunkt hat: $u = \Phi(u)$ bzw. $u_s = \Phi_s(u_s)$.

Das sind zunächst einmal Hoffnungen / Wünsche / Annahmen / Axiome. Für praktische Anwendungen benötigen wir jeweils handfeste Kriterien. Immerhin können wir mit den Begriffen Phänomene präzise benennen.

☺ Der Operator Φ_s beschreibt die **Erwartung** der Strategie $s \in S(\Gamma)$, die zugehörige Fixpunktgleichung $u_s = \Phi_s(u_s)$ heißt dementsprechend **Erwartungsgleichung**. Ihre (hoffentlich eindeutige) Lösung $u_s : X \to \mathbb{R}$ heißt **Erwartungsfunktion** der hier vorgegebenen Strategie $s \in S(\Gamma)$: Vom Zustand $x \in X$ erwarten wir mit Strategie s den Gewinn $u_s(x)$.

☺ Die **Erwartungsoperatoren** Φ_s sind affin-linear, haben bessere Eigenschaften, die stärkere Theorie und sind leichter zu behandeln.

Gewinnerwartung und Optimalität

Die Abbildung $\Phi : E \to E$ heißt **Bellman–(Optimalitäts–)Operator**, die Fixpunktgleichung $u = \Phi(u)$ heißt **Bellman–(Optimalitäts–)Gleichung**. Ihre (hoffentlich eindeutige) Lösung $u : X \to \mathbb{R}$ heißt **Gewinnfunktion**: Vom Zustand $x \in X$ erwarten wir bei optimalem Spiel den Gewinn $u(x)$.

Die Bellman–Gleichung ist vielseitig einsetzbar und daher berühmt. Sie ist eine Funktionalgleichung, denn die hierbei gesuchte Größe ist eine Funktion $u : X \to \mathbb{R}$. (Nun ja, eigentlich ist u auch nur ein Vektor, doch wir erlauben vorsorglich auch unendliche Zustandsmengen X.)

⚠ Hierzu ist noch keine Strategie vorgegeben. Im Gegenteil nutzen wir die Funktion u, um daraus schließlich eine optimale Strategie abzulesen!

Zur Berechnung bzw. Approximation der Funktion u wurden Dutzende Verfahren vorschlagen; uns geht es hier zunächst um ihre Definition, dann um grundlegende Eigenschaften und schließlich die Berechnung.

⚠ Anders als Φ_s ist der Operator Φ nicht-linear und daher schwieriger. Zur Definition der Funktion u nutzen wir zunächst die Fixpunktgleichung $u = \Phi(u)$; anschließend zeigen wir das Optimalitätsprinzip $u = \sup_s u_s$.

Iteration für Gewinnerwartung und Optimalität

Sei (Γ, r, v) ein Markov–Spiel mit beschränkten Belohnungen r und v. Letzteres gilt automatisch auf jedem endlichen Markov–Graphen Γ.

Auf $E = E_v = \{\, u \in B(X, \mathbb{R}) \mid u|_{\partial X} = v \,\}$ nutzen wir die Operatoren

$$\Phi : u \mapsto \bar{u} : \bar{u}(x) = \sup_{a \in A_x} \underbrace{\textstyle\sum_{y \in X} p(x,a,y)\,[r(x,a,y) + \delta u(y)]}_{\text{Hamilton–Funktion } H(x,\,a,\,u)} \quad \text{und}$$

$$\Phi_s : u \mapsto \bar{u} : \bar{u}(x) = H(x, s(x), u) \quad \text{für jede Strategie } s \in S(\Gamma).$$

Satz D2κ: Kontraktion, somit Existenz und Eindeutigkeit

Für jeden Diskontfaktor $\delta \in [0,1]$ sind alle Erwartungsoperatoren Φ_s und der Bellman–Operator Φ isoton δ–diskontiert, somit δ–lipschitz.

Speziell für $\delta \in [0,1[$ können wir Banachs Fixpunktsatz D2A anwenden, wie in unseren Beispielen motiviert: Es gibt genau einen Fixpunkt und diesen können wir durch das Iterationsverfahren effizient annähern.

Beweis: Dies folgt aus den Definitionen durch geduldiges Nachrechnen. Führen Sie dies sorgsam aus, es ist eine gute Übung zur Wiederholung!

Iteration für Gewinnerwartung und Optimalität

Ausführlich: Sei $u \leq \tilde{u} + c$. Für alle $x \in X^\circ$, $a \in A_x$ und $y \in X$ gilt:

$$\begin{aligned}
u(y) &\leq & \tilde{u}(y) + c \\
r(x,a,y) + \delta u(y) &\leq & r(x,a,y) + \delta\tilde{u}(y) + \delta c \\
p(x,a,y)\big[r(x,a,y) + \delta u(y)\big] &\leq & p(x,a,y)\big[r(x,a,y) + \delta\tilde{u}(y) + \delta c\big] \\
\textstyle\sum_y p(x,a,y)\big[r(x,a,y) + \delta u(y)\big] &\leq & \textstyle\sum_y p(x,a,y)\big[r(x,a,y) + \delta\tilde{u}(y)\big] + \delta c \\
H(x,a,u) &\leq & H(x,a,\tilde{u}) + \delta c
\end{aligned}$$

Zur Definition dieser Erwartung fordern wir absolute Konvergenz. Das ist garantiert, falls neben u und $\tilde{u}$ auch $y \mapsto r(x,a,y)$ beschränkt ist.

Ist die Strategie $s \in S(\Gamma)$ vorgegeben, so wählen wir im Zustand $x \in X^\circ$ die Aktion $a = s(x)$. Obige Rechnung garantiert $\Phi_s(u) \leq \Phi_s(\tilde{u}) + \delta c$.

Für den Operator Φ wird optimiert: Wir bilden das Supremum über alle Aktionen $a \in A_x$, erst rechts dann links, und erhalten $\Phi(u) \leq \Phi(\tilde{u}) + \delta c$.

Endlichkeit ist garantiert, falls neben u und $\tilde{u}$ auch für jeden Zustand x die Funktion $a \mapsto \sum_{y \in X} p(x,a,y) r(x,a,y)$ beschränkt ist. Ist zudem A_x endlich, so wird das Supremum angenommen, ist also ein Maximum.

Iteration für Gewinnerwartung und Optimalität

Sei (Γ, r, v) ein Markov–Spiel mit r, v beschränkt und $\delta = 1$. Wir fixieren $s \in S(\Gamma)$ mit Übergangswkten $p : X^\circ \times X \to [0,1] : (x,y) \mapsto p(x, s(x), y)$. Wir schreiben $x \to y$ falls $p(x,y) > 0$. Für $x \in \partial X$ setzen wir $p(x,x) = 1$. Somit ist (X, p) eine Markov–Kette mit endlichem Zustandsraum X.

Wir nennen s **randverbunden**, wenn es zu jedem $x_0 \in X$ einen Weg $x_0 \to x_1 \to \ldots \to x_\ell$ mit $x_\ell \in \partial X$ gibt. Falls X endlich ist, so ist s sogar **stark (ℓ, ε)–randverbunden** für ein geeignetes Paar $(\ell, \varepsilon) \in \mathbb{N} \times \mathbb{R}_{>0}$: Jeder Startzustand führt nach ℓ Schritten mit Wkt $\geq \varepsilon$ in den Rand ∂X.

🙂 Anschaulich: Die Wkt diffundiert in den Rand, langsam aber sicher. Nach ℓ Schritten ist die Gesamtwkt aller aktiven Zustände $\leq k = 1 - \varepsilon$. Auch für $\delta = 1$ kann also Kontraktion vorliegen; wir müssen hinschauen:

Satz D2L: randverbunden impliziert konvergent

Sei (Γ, r, v) ein Markov–Spiel und $\delta = 1$. Die Strategie $s \in S(\Gamma)$ sei stark (ℓ, ε)–randverbunden. Dann ist Φ_s^ℓ kontraktiv mit Konstante $k = 1 - \varepsilon$.

Insbesondere ist $\Phi_s : E \to E$ konvergent: Es existiert genau ein Fixpunkt $u_s \in E$ und zudem konvergiert $\Phi_s^n(\tilde{u}) \to u$ für jeden Startwert $\tilde{u} \in E$.

Iteration für Gewinnerwartung und Optimalität

Beweis: Wir untersuchen $\Phi_s(u)(x) = \sum_{y\in X} p(x,y)[r(x,y) + u(y)]$.

Wir betrachten $P = (p(x,y))_{(x,y)\in X\times X}$ als zeilen-stochastische Matrix und $u = (u(y))_{y\in X}$ als Spaltenvektor. Damit gilt $\Phi_s(u) = c + Pu$ mit additiver Belohnung $c = (c(x))_{x\in X}$ und $c(x) = \sum_{y\in X} p(x,y)r(x,y)$.

Per Induktion erhalten wir $\Phi_s^n(u) = c + Pc + P^2c + \dots P^{n-1}c + P^n u$.
Die Potenz $P^n = (p_n(x,y))_{(x,y)\in X\times X}$ berechnet die Wkt $p_n(x,y)$, in n Schritten von x nach y zu gehen, als Summe aller Wege der Länge n.

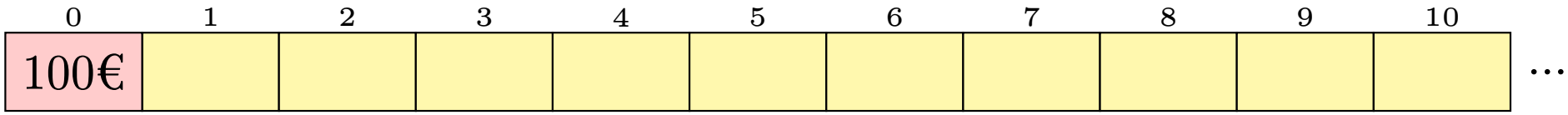
Für alle $u, \tilde{u} \in E$ und $n \in \mathbb{N}$ gilt somit $\Phi_s^n(u) - \Phi_s^n(\tilde{u}) = P^n(u - \tilde{u})$.
Wir zeigen nun $|P^\ell(u - \tilde{u})| \le k|u - \tilde{u}|$. Wir untersuchen $\hat{u} = P^\ell(u - \tilde{u})$:
Für $x \in \partial X$ gilt $p_\ell(x,x) = 1$, also $\hat{u}(x) = u(x) - \tilde{u}(x) = v(x) - v(x) = 0$.
Für jeden aktiven Zustand $x \in X^\circ$ hingegen finden wir:

$$\hat{u}(x) = \sum_{y\in X} p_\ell(x,y)\big[u(y) - \tilde{u}(y)\big] \;= \sum_{y\in X^\circ} p_\ell(x,y)\big[u(y) - \tilde{u}(y)\big]$$

$$|\hat{u}(x)| \le \sum_{y\in X^\circ} p_\ell(x,y)\big|u(y) - \tilde{u}(y)\big| \le \sum_{y\in X^\circ} p_\ell(x,y)\big|u - \tilde{u}\big| \le k\big|u - \tilde{u}\big|$$

Somit ist Φ_s^ℓ kontraktiv und Banachs Fixpunktsatz D2A anwendbar. $\boxed{\text{QED}}$

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph



Aufgabe: Wir untersuchen die Irrfahrt auf $X = \mathbb{N}$ mit Rand $\partial X = \{0\}$; mit Wkt $p \in \,]0,1[$ geht es nach rechts, mit Wkt $q = 1 - p$ nach links. Die Belohnung sei $r \in \mathbb{R}$ in jedem Schritt mit Diskontfaktor $\delta \in \,]0,1[$.

(1) Finden Sie alle $u : \mathbb{N} \to \mathbb{R}$ mit $\Phi(u) = u$. Welche sind beschränkt?
(2) Wie verhalten sich beschränkte Lösungen für $r = 0$ und $\delta \nearrow 1$?
(3) Lösen Sie den Fall $\delta = 1$. (4) Ist Φ kontraktiv? (5) konvergent?

Lösung: (1) Für $x \in \mathbb{N}_{\geq 1}$ lösen wir die Fixpunktgleichung $u = \Phi(u)$:

$$\boxed{u(x) = r + \delta[pu(x+1) + qu(x-1)]}$$

Durch $u(0)$ und $u(1)$ sind alle Werte $u(2), u(3), \ldots$ rekursiv festgelegt. Der Ansatz $u(x) = ab^x + r/(1-\delta)$ führt zu $b^x = \delta[pb^{x+1} + qb^{x-1}]$, also

$$b \in \{b_1, b_2\} \quad \text{mit} \quad b_{1/2} = \frac{1 \mp \sqrt{1 - 4pq\delta^2}}{2p\delta} \quad \text{und} \quad 0 < b_1 < 1 < b_2.$$

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

Damit erhalten wir alle Lösungen der Erwartungsgleichung $u = \Phi(u)$:

$$u(x) = a_1 b_1^x + a_2 b_2^x + \frac{r}{1-\delta} \qquad \text{mit} \qquad a_1, a_2 \in \mathbb{R}$$

Die beschränkten Lösungen sind demnach $u(x) = a_1 b_1^x + r/(1-\delta)$.
Der Startwert $u(0) = v(0)$ bestimmt die Konstante $a_1 = u(0) - r/(1-\delta)$.

$$\boxed{u(x) = \left[u(0) - \frac{r}{1-\delta}\right]\left[\frac{1-\sqrt{1-4pq\delta^2}}{2p\delta}\right]^x + \frac{r}{1-\delta}}$$

☺ Dank Diskont $\delta < 1$ erhalten wir in jedem Falle genau eine Lösung!
Das entspricht dem Kontraktionssatz D2K, hier nun wunderbar konkret.
Obwohl der Graph unendlich ist, greifen unsere Werkzeuge bestens.

☺ Der gegebene Anfangswert klingt exponentiell ab gegen $r/(1-\delta)$.
Konkretes Beispiel: Für $p = q = 1/2$ und $\delta = 0.8$ finden wir $b_1 = 1/2$,
und somit die Gewinnfunktion $u : \mathbb{N} \to \mathbb{R} : x \mapsto [u(0) - 5r]2^{-x} + 5r$.

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

(2) Für $r = 0$ finden wir den Grenzwert $\lim_{\delta \nearrow 1} u(x) = u(0)\, b^x$ mit

$$\begin{aligned} b &= \frac{1 \mp \sqrt{1 - 4p(1-p)}}{2p} = \frac{1 \mp \sqrt{(1-2p)^2}}{2p} \\ &= \frac{1 - |1-2p|}{2p} = \begin{cases} 1 & \text{für } 0 < p \leq 1/2, \\ (1-p)/p & \text{für } 1/2 < p < 1. \end{cases} \end{aligned}$$

☺ Das ist tatsächlich eine beschränkte Lösung des Falls $\delta = 1$.
Für $1/2 < p < 1$ gilt $b < 1$ und somit $u(x) \to 0$ für $x \to \infty$.
Unten in (3) finden wir weitere beschränkte Lösungen,
die diese beiden Eigenschaften nicht haben.

Nochmal anders gesagt: Für $\delta = 1$ sind selbst beschränkte Lösungen nicht eindeutig. Unter den vielen „mathematischen“ Lösungen finden wir genau eine „natürliche“ oder „physikalisch-plausible“ Lösung wie oben. Allein die Erwartungsgleichung / Bilanzgleichung genügt hier also noch nicht zur Charakterisierung der „richtigen“ Lösung.

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

Die hier gefundene Diskrepanz ist erhellend für unsere Modellbildung und unsere Lösungsmethoden: Bilanzgleichung vs Mikrofundierung, mathematische Fixpunktgleichung vs physikalische Betrachtung.

Dahinter steckt folgendes Argument: Die zufällige Irrfahrt auf $\mathbb{N}$ mit Wkten (q, p) ist ein wohldefinierter stochastischer Prozess. Der erwartete Gewinn $u(x) \in \bar{\mathbb{R}}$ bei Start in $x \in \mathbb{N}$ ist eine wohldefinierte Erwartung:

$$\begin{aligned} r = 0 \quad &\Longrightarrow \quad u : \mathbb{N} \to [0, 100] \\ r > 0 \quad &\Longrightarrow \quad u : \mathbb{N} \to [0, +\infty] \\ r < 0 \quad &\Longrightarrow \quad u : \mathbb{N} \to [-\infty, 100] \end{aligned}$$

Mit anderen Worten, die Lösung u existiert und ist eindeutig, wir wollen sie effizient berechnen. Wenn hierzu die Bilanzgleichung alleine nicht genügt, so extrahieren wir weitere Einschränkungen / Rechenregeln. Notfalls müssen wir auf das Mikroniveau der Trajektorien absteigen, also den stochastischen Prozess untersuchen und Erwartungen ausrechnen.

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

(3) Wir setzen $\delta = 1$ und lösen die Fixpunktgleichung $u = \Phi(u)$:

$$\boxed{u(x) = r + pu(x+1) + qu(x-1).}$$

Für $p = q = 1/2$ haben wir bereits zuvor alle Lösungen bestimmt:
Diese sind Parabeln der Form $u(x) = u(0) + ax - rx^2$ mit $a \in \mathbb{R}$.
Im Folgenden sei daher $p \neq 1/2$ und weiterhin $0 < p < 1$.

Der Ansatz $u(x) = ab^x + rx/(1-2p)$ führt zu $b^x = pb^{x+1} + qb^{x-1}$, also

$$b \in \left\{ \frac{1 \mp \sqrt{1-4pq}}{2p} \right\} = \left\{ \frac{1 \mp |1-2p|}{2p} \right\} = \left\{ 1, \frac{1-p}{p} \right\}.$$

Für $b = (1-p)/p$ und $a \in \mathbb{R}$ erhalten wir also die allgemeine Lösung

$$\boxed{u(x) = u(0) - a + ab^x + \frac{rx}{1-2p}}$$

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

Für $0 < p < 1/2$ gilt $b > 1$: Beschränkte Lösungen existieren nur für $a = r = 0$, also nur die konstante Lösung $u(x) = u(0)$ für alle $x \in \mathbb{N}$.

Für $1/2 < p < 1$ gilt $0 < b < 1$: Beschränkte Lösungen existieren nur für $r = 0$; dann gilt $u(x) = u(0) - a + ab^x$, wobei $a \in \mathbb{R}$ beliebig wählbar ist.

Für $x \to \infty$ erfüllt diese Lösung $u(x) \to u(0) - a$. Nur für $a = u(0)$ erfüllt unsere Lösung $u(x) = u(0)\, b^x$ zudem $u(x) \to 0$ für $x \to \infty$.

☺ Das ist die „natürliche“ oder „physikalisch-plausible“ Lösung, die wir oben in (2) als den Grenzwert für $\delta \nearrow 1$ gefunden haben. Alle anderen Lösungen erfüllen ebenfalls die Erwartungsgleichung, doch diese alleine genügt hier noch nicht zur Eindeutigkeit.

☺ Diese schöne Illustration ist ein heilsames Gegenbeispiel: Für $\delta = 1$ ist Eindeutigkeit keinesfalls selbstverständlich!

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

(4) Weiter sei $\delta = 1$ und $r = 0$. Für $0 < p \leq 1/2$ hat $\Phi(u) = u$ genau eine beschränkte Lösung $u \in E$, nämlich $u(x) = u(0)$ für alle $x \in \mathbb{N}$.

Dennoch ist $\Phi : E \to E$ **nicht kontraktiv**, auch keine Potenz Φ^n: Wir vergleichen $u, \tilde{u} : \mathbb{N} \to \mathbb{R}$ mit $u(0) = \tilde{u}(0) = 0$ sowie $u(x) = 0$ und $\tilde{u}(x) = 1$ für $x \in \mathbb{N}_{\geq 1}$. Dann gilt $\Phi^n(u) = u$ und $\Phi^n(\tilde{u})(x) = 1$ für $x > n$.

Dies zeigt, dass $\Phi : E \to E$ **nicht gleichmäßig konvergent** ist, denn es gilt nicht $\Phi^n(\tilde{u}) \to u$ bezüglich der Supremumsnorm auf E.

☺ Das ist ein weiteres schönes Beispiel für unser Repertoire: Der Bellman–Operator $\Phi : E \to E$ ist hier nicht kontraktiv, dennoch existiert genau eine Lösung $u = \Phi(u)$.

Die vorsichtige Begriffsbildung in Definition D2J nimmt so langsam konkrete Gestalt an und füllt sich nachträglich mit Leben.

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

(5) Für $0 < p \leq 1/2$ ist Φ immerhin noch **punktweise konvergent**:
Es existiert genau ein Fixpunkt $u \in E$ und für jeden Startwert $\tilde{u} \in E$ gilt Konvergenz $\Phi^n(\tilde{u})(x) \to u(x)$ für $n \to \infty$ in jedem Punkt $x \in X$.

Wir zeigen dies für $v(0) = 0$. Sei $u_0 \in E$ gegeben durch $u_0(0) = 0$ und $u_0(x) = 1$ für $x \in \mathbb{N}$. Für $u_n = \Phi^n(u_0)$ gilt $u_0 \geq u_1 \geq u_2 \geq \cdots \searrow u \geq 0$. Aus $u_{n+1}(x) = pu_n(x+1) + qu_n(x-1)$ wird $u(x) = pu(x+1) + qu(x-1)$ für $n \to \infty$, also $u = \Phi(u)$. Somit gilt $u = 0$ dank Eindeutigkeit.

Für jeden Startwert $\tilde{u} \in E$ gilt $-cu_0 \leq \tilde{u} \leq cu_0$ mit $c = |\tilde{u}|$. Daraus folgt $-cu_n \leq \Phi^n(\tilde{u}) \leq cu_n$ für alle $n \in \mathbb{N}$, also punktweise $\Phi^n(\tilde{u})(x) \to 0$.

☺ Auch dies ist ein weiteres schönes Beispiel für unser Repertoire.
Satz D20 erklärt ein allgemeines Kriterium für punktweise Konvergenz.

Bellmans Optimalitätsprinzip $u_* = u$

Gegeben sei ein Markov–Spiel (Γ, r, v) und $u : X \to \mathbb{R}$ mit $u = \Phi(u)$.
Wie kann die Auszahlung u realisiert werden? Gar noch höhere?
Wir vergleichen mit $u_* := \sup\{\, u_s \in E \mid s \in S(\Gamma),\ \Phi_s(u_s) = u_s \,\}$.
Zur Optimierung gilt dann folgendes Lokal-Global-Prinzip:

Satz D2м: Bellmans Optimalitätsprinzip $u_* = u$

(1) Wenn Φ für jeden Startwert gegen u konvergiert, so gilt $u_* \le u$.
Das bedeutet, lokale Optimierung ist mindestens so gut wie globale.
Das gilt insbesondere, wenn Φ isoton δ–diskontiert ist, also δ–kontraktiv.

(2) Wird in der Bellman–Gleichung überall das Supremum angenommen, etwa weil Γ lokal-endlich ist, so existieren optimale Strategien $s \in S(\Gamma)$ mit $s(x) \in \operatorname{Arg\,max}_{a \in A_x} H(x, a, u)$, also $\Phi_s(u) = u$, und es folgt $u_* \ge u$.

Beweis: (1) Für jede Strategie $s \in S$ und jeden Fixpunkt $u_s = \Phi_s(u_s)$ gilt in $x \in X^\circ$ stets $\Phi(u_s)(x) = \sup_{a \in A_x} H(x, a, u_s) \ge H(x, s(x), u_s) = u_s(x)$.
Zudem ist Φ isoton: $u_s \le \Phi(u_s) \le \Phi^2(u_s) \le \ldots \nearrow u$. Somit gilt $u_* \le u$.

(2) Für s gilt $u(x) = H(x, s(x), u)$, also $u = \Phi_s(u)$, somit $u_* \ge u$. $\boxed{\text{QED}}$

Bellmans Optimalitätsprinzip $u_* = u$

Bellmans Optimalitätsprinzip $u_* = u$ kommt recht unscheinbar daher, daher möchte ich seine anschaulich-praktische Bedeutung erläutern.

Lokale Endlichkeit des Markov–Graphen Γ vereinfacht alle Argumente. Den allgemeinen Fall behandeln wir gleich anschließend in Satz D2N.

Rückwärts / lokale Optimierung: Ein rationaler Spieler wählt in jedem aktiven Zustand $x \in X^\circ$ eine optimale Aktion $a \in A_x$, die seine erwartete Auszahlung $H(x, a, u)$ maximiert. Dies führt zur Bellman–Gleichung:

$$u(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \max\limits_{a \in A_x} \underbrace{\sum_{y \in X} p(x,a,y)\,[r(x,a,y) + \delta u(y)]}_{\text{Hamilton–Funktion } H(x,\,a,\,u)} & \text{für } x \in X^\circ. \end{cases}$$

In jedem Zustand $x \in X^\circ$ stellen wir uns einen **lokalen Optimierer** vor. Dieser handelt lokal, kurzsichtig, egoistisch: Jeder optimiert für sich!

Das erinnert uns an das Prinzip der Rekursion / Rückwärtsinduktion: Optimiert wird rückwärts, aber gespielt wird immer vorwärts.

Bellmans Optimalitätsprinzip $u_* = u$

Vorwärts / globale Optimierung: Jede Strategie $s \in S = S(\Gamma)$ definiert eine Gewinnerwartung $u_s : X \to \mathbb{R}$ durch die Gleichung $u_s = \Phi_s(u_s)$:

$$u_s(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ H(x, s(x), u_s) & \text{für } x \in X^\circ. \end{cases}$$

Eine **globale Optimiererin** kann somit zu jeder Strategie $s \in S(\Gamma)$ ihre Auszahlung u_s berechnen und das Supremum $u_* = \sup u_s$ bilden. Dies geschieht global, weitblickend, kooperativ: alle $x \in X^\circ$ gemeinsam!

In jedem Zustand $x \in X$ können wir demnach bestenfalls den Gewinn $u_*(x) = \sup_{s \in S} u_s(x)$ erwarten bzw. als Auszahlung realisieren.

Das Optimalitätsprinzip $u_* = u$ besagt, dass die Bellman–Gleichung tatsächlich tut, was sie soll: Die globale Optimierung u_* und die lokale Optimierung u stimmen überein! Beide Sichtweisen werden versöhnt.

⚠ Das Optimalitätsprinzip $u_* = u$ erfordert gewisse Voraussetzungen; diese versuche ich hier möglichst schwach und allgemein zu halten.

Bellmans Optimalitätsprinzip $u_* = u$

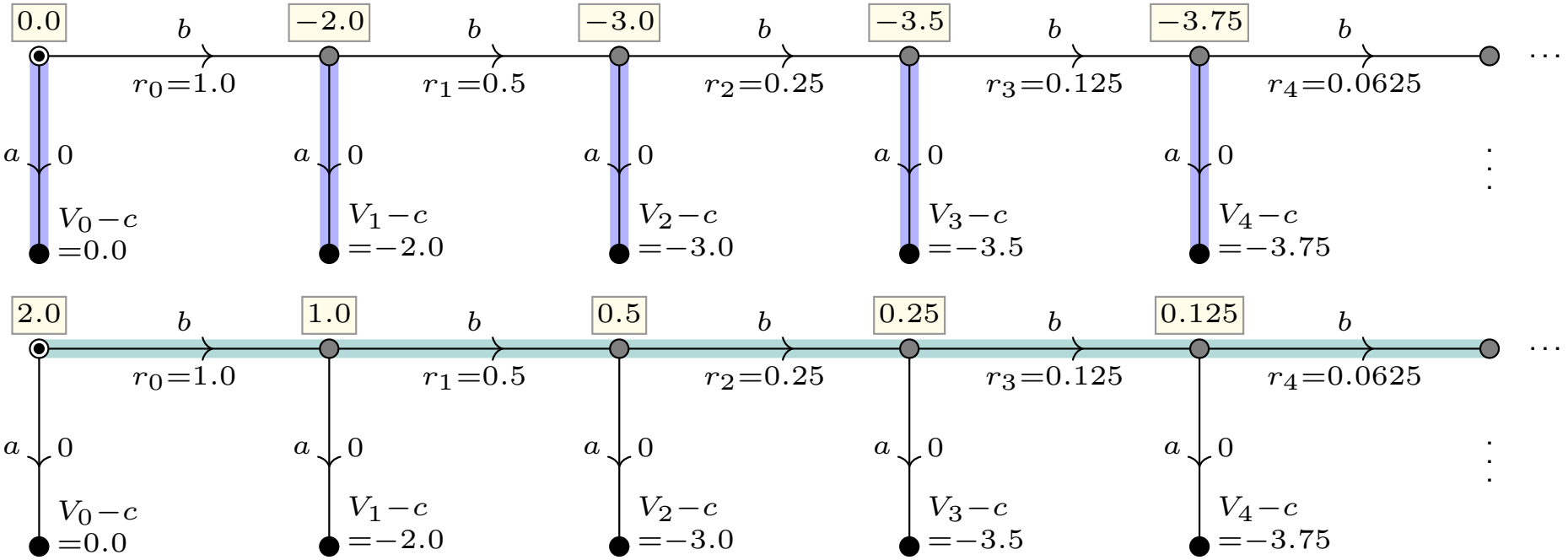
Vielleicht scheint Ihnen auf den ersten Blick das Optimalitätsprinzip $u_* = u$ intuitiv-selbstverständlich und daher kaum der Rede wert. Das wäre ein Jammer und ein Irrtum! Ist hier etwas zu beweisen? Ja, sicher, und zur Illustration helfen Gegenbeispiele wie B1N.

Drastische Gegenbeispiele wie das folgende sind überaus lehrreich, und ein großes Beispielrepertoire bewahrt Sie vor voreiligen Trugschlüssen. Die gute mathematische Vorgehensweise ist wie immer komplementär: Sätze und Gegen/Beispiele ergänzen und erklären sich gegenseitig.

In D3 skizzieren wir eine wunderschöne Anwendung aus dem Bereich des Maschinellen Lernens: Robi, der Saugroboter, versucht seine Route optimal zu planen. Meist wählt man dort einen Diskontfaktor $\delta \in [0, 1[$, da die theoretische Grundlage dann besonders einfach und solide ist.

Die Robustheit der Rechnungen ist für die Anwendung extrem wichtig, gerade deshalb diskutieren wir auch den kritischen Randfall $\delta = 1$. Gute Beispiele bewahren Sie vor naivem Irrglauben und illustrieren eindrücklich den Nutzen möglichst starker theoretischer Werkzeuge.

Infinipede und frustrierte Bellman–Lösung



♦ **Beispiel B1N:** exotische Lösungen der Bellman–Gleichung

Unser Graph Γ habe die aktiven Zustände $x \in X^\circ = \{\, b^k \mid k \in \mathbb{N} \,\}$ mit Aktionen $A_x = \{a, b\}$ und Belohnungen $r(b^k, a) = 0$ und $r(b^k, b) = r_k$. Terminal sind $\partial X = \{\, b^k a \mid k \in \mathbb{N} \,\}$ mit Auszahlungen $v(b^k a) = V_k - c$.

Gegeben seien hierzu $0 < r_k < v_k$ in $\mathbb{R}$ für $k \in \mathbb{N}$ mit $\sum_{k=0}^{\infty} v_k < \infty$. Wir setzen $R_k = \sum_{i=k}^{\infty} r_i$ und $V_k = \sum_{i=k}^{\infty} v_i$ und wählen $c > V_0 - R_0 > 0$. Die Skizze zeigt $r_k = 2^{-k}$, $R_k = 2^{1-k}$, $v_k = 2^{1-k}$, $V_k = 2^{2-k}$ und $c = -4$.

Infinipede und frustrierte Bellman–Lösung

Aufgabe: Zeigen Sie, dass die abgebildeten Funktionen $u_0, u_1 : X \to \mathbb{R}$ die Bellman–Gleichung lösen: Einerseits die pessimistische Lösung $u_0(b^k) = V_k - c$, andererseits die optimistische Lösung $u_1(b^k) = R_k$.

Lösung: Jede der beiden gezeigten Strategien ist lokal optimal:
(1) Für u_0 wird immer $s_0(x) = a$ gespielt, und dies ist lokal optimal.
(2) Für u_1 wird immer $s_1(x) = b$ gespielt, und dies ist lokal optimal.

Sie kennen das Problem *lokale vs globale Optimierung* aus der Analysis: Hier gilt $u_0 < u_1$; bei **globaler Optimierung** würde man also u_1 wählen. Bei **lokaler Optimierung** erlaubt s_0 keine Möglicheit der Verbesserung: Dies geschieht lokal, kurzsichtig, egoistisch: Jeder optimiert für sich! Das globale Optimum erreichen wir hier nur weitblickend, kooperativ.

⚠ In Fällen wie diesem gilt Bellmans Optimalitätsprinzip nicht! Unsere allgemeinen Werkzeuge versagen hier, wir müssen genauer hinschauen.

Frustration: Das obige Beispiel kennen Sie aus dem Alltag: „Ich würde gerne etwas verbessern, doch alleine kann ich gar nichts ausrichten.” So auch hier, wenn jeder lokale Optimierer in $x \in X^\circ$ allein handelt.

Infinipede und frustrierte Bellman–Lösung

Bemerkung: Weitere Bellman–Lösungen sind $u(b^k) = R_k + v(b^\infty)$. Die Konstante $v(b^\infty) \in \mathbb{R}$ ist dabei die (fiktive) Auszahlung im „unendlich fernen“ Randpunkt b^∞. Wir vereinbaren hier kurzerhand $v(b^\infty) = 0$.

Aufgabe: Gibt es weitere Lösungen neben (u_0, s_0) und (u_1, s_1)?

Lösung: Nein. Sei $u : X \to \mathbb{R}$ eine Lösung der Bellman–Gleichung und $s : X^\circ \to \{a, b\}$ eine (lokal-optimale) Strategie mit $u(x) = H(x, s(x), u)$. Gilt $s(b^j) = a$, so folgt $s(b^i) = a$ für alle $i \le j$. Somit wird entweder $s = s_0$ gespielt oder es existiert ein $k \in \mathbb{N}$ mit $s(b^i) = a$ für $i < k$ und $s(b^i) = b$ für alle $i \ge k$. Der Fall $k > 0$ ist nicht lokal-optimal, da die Alternative $s(b^{k-1}) = b$ lukrativer ist. Also muss $k = 0$ gelten und somit $s = s_1$.

Das Beispiel B1N ist raffiniert gebaut und notgedrungen nicht-artinsch. Für jeden artinschen Graphen können wir dank Rückwärtsinduktion B1L die Bellman–Gleichung rekursiv lösen, und diese Lösung ist eindeutig! Ist unser artinscher Graph Γ zudem lokal-endlich, so ist diese Lösung tatsächlich optimal dank B1L. In diesen günstigen Fällen geht alles gut. Es gibt einen zweiten Ausweg aus unserer Misslage: Diskontierung!

Infinipede und frustrierte Bellman–Lösung

Aufgabe: Analysieren Sie das obige Beispiel B1N mit Diskont $\delta \in [0,1[$. Bestimmen Sie die (dank Satz D2K) eindeutige beschränkte Lösung $u_\delta : X \to \mathbb{R}$ der Bellman–Gleichung $u_\delta = \Phi_\delta(u_\delta)$. Was gilt für $\delta \nearrow 1$?

Lösung: (1) Ist die Strategie $s_0(b^k) = a$ weiterhin lokal optimal? Nein!
Wir vergleichen die Auszahlung $\alpha_k := r(b^k, a) + \delta v(b^k a) = \delta(V_k - c)$
mit $\beta_k := r(b^k, b) + \delta[r(b^{k+1}, a) + \delta v(b^{k+1} a)] = r_k + \delta^2(V_{k+1} - c)$.
Die Differenz $\beta_k - \alpha_k = r_k + \delta(1-\delta)c - \delta(V_k - \delta V_{k+1})$ ist positiv
für hinreichend große $k \in \mathbb{N}$, denn $r_k, c > 0$ und $V_k - \delta V_{k+1} \searrow 0$.
Es lohnt sich also (lokal!), von $s_0(b^k) = a$ auf b zu wechseln.

Im Extremfall $\delta = 1$ gilt $\beta_k - \alpha_k = r_k - v_k < 0$ für alle $k \in \mathbb{N}$, daher ist ausgehend von der Strategie s_0 keine lokale Verbesserung möglich.

(2) Ist die Strategie $s_1(b^k) = b$ weiterhin lokal optimal? Ja!
Das Argument ist für alle Diskontfaktoren $\delta \in [0,1]$ gleich:
Jede lokale Abweichung verschlechtert die Auszahlung.
Für alle $\delta \in [0,1[$ erhalten wir also dieselbe Strategie $s_\delta = s_1$
mit der optimistischen Gewinnfunktion $u_\delta \nearrow u_1$ für $\delta \nearrow 1$.

Bellmans Optimalitätsprinzip $u_* = u$

☺ Bei Diskontierung $\delta < 1$ gilt das Optimalitätsprinzip ganz allgemein für alle beschränkten Markov–Spiele auf beliebigen Markov–Graphen:

Satz D2N: Optimalitätsprinzip, allgemeiner diskontierter Fall

Sei (Γ, r, v) ein Markov–Spiel mit r, v beschränkt und $0 \le \delta < 1$.

Auf $E = E_v = \{u \in B(X, \mathbb{R}) \mid u|_{\partial X} = v\}$ nutzen wir die Operatoren $\Phi(u)(x) = \sup_{a \in A_x} H(x, a, u)$ und $\Phi_s(u)(x) = H(x, s(x), u)$ für $x \in X^\circ$.

Diese konvergieren gegen $u = \Phi(u)$ bzw. $u_s = \Phi_s(u_s)$ für $s \in S(\Gamma)$.
Wir setzen punktweise $u_*(x) = \sup_{s \in S} u_s(x)$. Dann gilt $u_* = u$.

Beweis: (1) Wie in Satz D2M gilt $u_* \le u$. (2) Wir zeigen noch $u \le u_*$:

Sei $\varepsilon \in \mathbb{R}_{>0}$ vorgegeben. Zu jedem aktiven Zustand $x \in X^\circ$ wählen wir eine ε–optimale Aktion $s(x) \in A_x$, sodass $H(x, s(x), u) \ge u(x) - \varepsilon$ gilt.

Das bedeutet $u \le \Phi_s(u) + \varepsilon$, dank Diskontierung $\Phi_s(u) \le \Phi_s^2(u) + \delta\varepsilon$, per Induktion $u \le \Phi_s^n(u) + \varepsilon \sum_{k=0}^{n-1} \delta^k = \Phi_s^n(u) + \varepsilon(1 - \delta^n)/(1 - \delta)$.

Für $n \to \infty$ erhalten wir $u \le u_s + \varepsilon/(1 - \delta) \le u_* + \varepsilon/(1 - \delta)$.
Letzteres gilt für alle $\varepsilon \in \mathbb{R}_{>0}$. Daraus folgt $u \le u_*$ QED

Bellmans Optimalitätsprinzip $u_* = u$

Im diskontierten Fall $0 \leq \delta < 1$ sind alle Operatoren Φ_s und Φ kontraktiv. Für jede Strategie $s \in S = S(\Gamma)$ können wir Φ_s iterieren und erhalten:

$$\Phi_s^n(\tilde{u}) \to u_s$$

Dies können wir anschließend global optimieren und gewinnen so:

$$u_* = \sup_{s \in S} u_s$$

Umgekehrt können wir auch erst lokal optimieren und dann iterieren. Nach Definition gilt $\Phi = \sup_{s \in S} \Phi_s$, das heißt punktweise für alle $x \in X$:

$$\Phi(\tilde{u})(x) = \sup_{s \in S} \Phi_s(\tilde{u})(x) \overset{x \in X^\circ}{=} \sup_{a \in A_x} H(x, a, \tilde{u})$$

Wenn wir nun diesen Bellman–Operator Φ iterieren, so erhalten wir:

$$\Phi^n(\tilde{u}) \to u$$

Bellmans Optimalitätsprinzip besagt, unter der Voraussetzung $0 \leq \delta < 1$, dass wir auf beiden Wegen dasselbe Ergebnis erhalten, kurz $u_* = u$.

Zusammenfassung zum Optimalitätsprinzip

Bellmans Optimalitätsprinzip betrachten wir rückblickend als

0 Vertauschung von Iteration/Kontraktion und Maximum/Supremum.

Aus der Analysis kennen Sie viele wichtige Sätze zur Vertauschung:

1 Folgengrenzwerte $\lim_{k\to\infty} \lim_{n\to\infty} a_{k,n} \overset{?}{=} \lim_{n\to\infty} \lim_{k\to\infty} a_{k,n}$

2 Vertauschung von Ableitungen $\partial_x \partial_y f(x,y) \overset{?}{=} \partial_y \partial_x f(x,y)$

3 Ableitung und Grenzwert $\lim_{n\to\infty} \partial_x f_n(x) \overset{?}{=} \partial_x \lim_{n\to\infty} f_n(x)$

4 Vertauschung von Reihen $\sum_{k=0}^{\infty} \sum_{n=0}^{\infty} a_{k,n} \overset{?}{=} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} a_{k,n}$

5 Reihen und Grenzwert $\lim_k \sum_{n=0}^{\infty} a_{k,n} \overset{?}{=} \sum_{n=0}^{\infty} \lim_k a_{k,n}$

6 Reihen und Ableitung $\partial_x \sum_{n=0}^{\infty} f_n(x) \overset{?}{=} \sum_{n=0}^{\infty} \partial_x f_n(x)$

7 Integrale $\int_{x\in X} \int_{y\in Y} f(x,y)\,\mathrm{d}y\,\mathrm{d}x \overset{?}{=} \int_{y\in Y} \int_{x\in X} f(x,y)\,\mathrm{d}x\,\mathrm{d}y$

8 Integral und Ableitung $\partial_x \int_{y\in Y} f(x,y)\,\mathrm{d}y \overset{?}{=} \int_{y\in Y} \partial_x f(x,y)\,\mathrm{d}y$

9 Integral und Grenzwert $\lim_n \int_{x\in X} f_n(x)\,\mathrm{d}x \overset{?}{=} \int_{x\in X} \lim_n f_n(x)\,\mathrm{d}x$

Übung: Geben Sie jeweils zwei interessante Gegenbeispiele.
Anschließend wiederholen Sie Voraussetzungen, Satz und Beweis.

Konvergenz des Bellman–Operators

Wir nutzen hier möglichst einfache, zumeist hinreichende Bedingungen: r, v beschränkt und $\delta \in [0, 1[$. Für $\delta = 1$ müssen wir genauer hinschauen!

Die Übungen geben konkretes und motivierendes Anschauungsmaterial. Dort können Sie in konkreten Beispielen auch $\delta = 1$ durchrechnen. Auch der folgende Satz D20 kann im Falle $\delta = 1$ genutzt werden.

Die Konvergenzfrage ist ähnlich zur Theorie komplexer **Potenzreihen**: Für jede Potenzreihe $\sum_{n=0}^{\infty} a_n z^n$ definieren wir den **Konvergenzradius**

$$\boxed{\rho := 1/\limsup \sqrt[n]{|a_n|}.}$$

Für $|z| < \rho$ konvergiert die Reihe absolut, für $|z| > \rho$ divergiert sie. Auf dem Rand $|z| = \rho$ ist alles möglich; dazu gibt es keine allgemeine Aussage. Antworten über das subtile Verhalten auf dem Kreisrand gibt die Theorie der Fourier–Reihen. Allgemein gehört die Konvergenz solcher Reihen zu den schwierigsten Fragen der Analysis.

Auf hoher See und am Konvergenzkreisrand sind wir alle in Gottes Hand.

Punktweise Konvergenz des Bellman–Operators

Satz D2o: punktweise Konvergenz des Bellman–Operators

Sei (Γ, r, v) ein Markov–Spiel, Γ lokal-endlich, r, v beschränkt und $r \leq 0$.

(1) Sei $F = \{\, u_s \in E \mid s \in S,\ \Phi_s(u_s) = u_s \,\}$. Ist $u_* = \sup F$ beschränkt, dann ist u_* der größte Fixpunkt des Bellman–Operators $\Phi : E \to E$.

(2) Sei $s \in S(\Gamma)$ eine Strategie mit $\Phi_s(u_*) = u_*$. Konvergiert $\Phi_s^n(u) \to u_*$ für jeden Startwert $u \in E$, dann konvergiert auch Φ gegen u_*.

☺ Anders als im vorigen Satz D2N verlangen wir keine Diskontierung mit $\delta \in [0, 1[$. Der Satz D2O behandelt vielmehr den Randfall $\delta = 1$.

☺ Aussage (1) ist eine schwache Fassung des Optimalitätsprinzips. Hat der Operator $\Phi : E \to E$ zudem nur einen Fixpunkt u, so gilt $u = u_*$.

☺ Aussage (2) ist eine starke Fassung analog zum obigen Satz D2M: Φ konvergiert für jeden Startwert gegen den einzigen Fixpunkt u_*.

☺ Die Einschränkung $r \leq 0$ scheint zunächst nicht besonders elegant, doch in manchen Anwendungen wie D3 ist sie recht natürlich.

Punktweise Konvergenz des Bellman–Operators

Beweis: (0) Zu jedem Fixpunkt $u = \Phi(u)$ in E existiert eine Strategie $s \in S = S(\Gamma)$ mit $\Phi_s(u) = u$. Demnach gilt $u \in F$ und somit $u \leq u_*$

(1) Für jedes $u_s \in F$ gilt $u_* \geq u_s$, also $\Phi(u_*) \geq \Phi(u_s) \geq \Phi_s(u_s) = u_s$. Daraus folgt $\Phi(u_*) \geq u_*$, also $u_* \leq \Phi(u_*) \leq \Phi^2(u_*) \leq \ldots \nearrow u : X \to \bar{\mathbb{R}}$. Wir zeigen nun, dass $u : X \to \mathbb{R} \cup \{+\infty\}$ beschränkt ist, also $u \in E$ gilt.

Für $c \geq \max u_*$ und $\hat{u} = v\,\mathbf{1}_{\partial X} + c\,\mathbf{1}_{X^\circ}$ gilt $u_* \leq \hat{u}$, also $\Phi(u_*) \leq \Phi(\hat{u}) \leq \hat{u}$, letzteres dank $r \leq 0$. Per Induktion erhalten wir $\Phi^n(u_*) \leq \hat{u}$, also $u \leq \hat{u}$. Dank Stetigkeit folgt $\Phi(u) = u$. Dank (0) gilt $u \leq u_*$, also $u = u_* = \Phi(u_*)$.

(2a) Aus $\hat{u} \geq \Phi(\hat{u}) \geq \Phi(u_*) = u_*$ folgt $\hat{u} \geq \Phi(\hat{u}) \geq \Phi^2(\hat{u}) \geq \ldots \searrow \bar{u} \geq u_*$. Dank Stetigkeit folgt $\Phi(\bar{u}) = \bar{u}$. Dank (1) folgt $\bar{u} = u_*$.

(2b) Gegeben sei $u \in E$. Wir wählen $\hat{u} = v\,\mathbf{1}_{\partial X} + c\,\mathbf{1}_{X^\circ} \geq u, u_*$ wie oben. Dann gilt $\Phi_s(u) \leq \Phi(u) \leq \Phi(\hat{u})$, per Induktion $\Phi_s^n(u) \leq \Phi^n(u) \leq \Phi^n(\hat{u})$. Nach Voraussetzung gilt $\Phi_s^n(u) \to u_*$. Dank (2a) gilt auch $\Phi^n(\hat{u}) \to u_*$. Daraus folgt die behauptete Konvergenz $\Phi^n(u) \to u_*$. $\boxed{\text{QED}}$

Punktweise Konvergenz des Bellman–Operators

☺ Der Fixpunktsatz von Banach fordert eine δ–Kontraktion, also eine starke Voraussetzung, garantiert dafür aber gleichmäßige Konvergenz mit expliziter Fehlerschranke als extrem nützliche Schlussfolgerung. Das ist der Idealfall, an dem wir uns orientieren möchten.

Bellmans Optimalitätsprinzip formulieren wir entsprechend, je nach Anwendung, unter stärkeren oder schwächeren Voraussetzungen. Der bequemste Fall ist die Diskontierung mit $\delta \in [0, 1[$ (Satz D2N). Diese strenge Voraussetzung ist leider nicht immer gegeben.

Für Bellmans Optimalitätsprinzip D2M, genauer für die Ungleichung $u_* \leq u$, genügt uns bescheidener bereits die punktweise Konvergenz $\Phi^n(\tilde{u}) \to u$. Genau hierfür bietet Satz D2O ein hinreichendes Kriterium. Das kann in Anwendungen mit $\delta = 1$ ein praktisches Hilfsmittel sein.

Aus Erfahrung empfehle ich, diese schönen Sätze zunächst einmal zur Kenntnis zu nehmen und dann konkrete Anwendungen zu untersuchen. In einem zweiten Durchgang möchten Sie dann stärkere Werkzeuge und werden gerne auf hilfreiche Sätze und Beweise zurückkommen.

Ausblick auf erfolgreiche Anwendungen

☺ In Markov–Spielen steckt viel schöne Mathematik!
Sie sind sehr einfach gebaut und doch vielseitig einsetzbar.
Zur Programmierung nutzen wir vor allem endliche Markov–Graphen,
doch auch der unendliche Fall ist mathematisch reizvoll und nützlich.

Markov–Spiele (aka Markov–Entscheidungsprozesse) finden Sie daher
in zahlreichen Anwendungen der Ökonomik und Finanzmathematik:

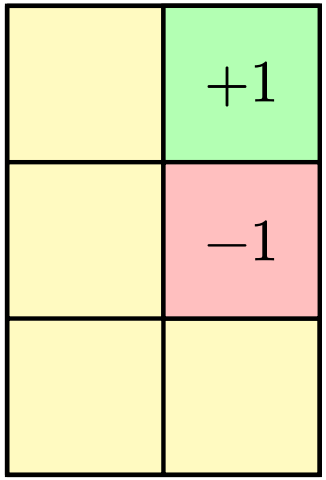
Investition vs Konsum optimieren: Geld sparen oder ausgeben?
Anbau vs Abbau optimieren: Bäume pflanzen oder fällen?
Anlage optimieren: Aktienportfolio optimal steuern?
Allgemein: kurzfristiger oder langfristiger Nutzen?

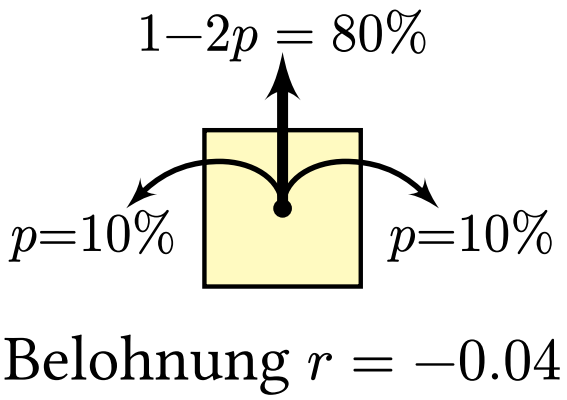
Markov–Spiele finden Sie ebenso in der künstlichen Intelligenz,
insbesondere maschinellem Lernen, etwa bestärkendem Lernen:

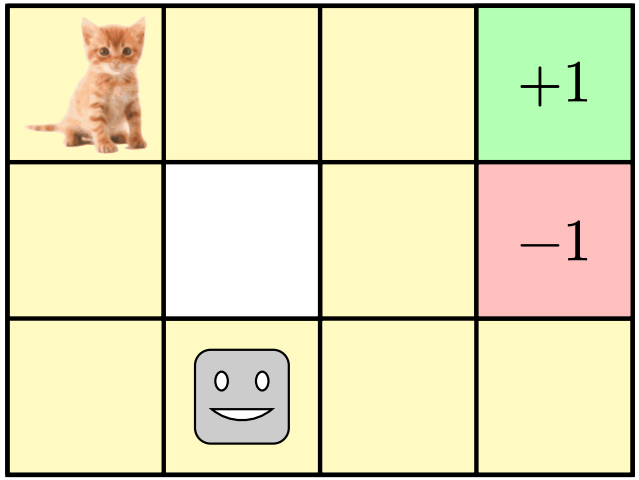
Logistik / Routenplanung: Wege minimieren, Nutzen maximieren.
Strategischer Akteur / autonomes Fahrzeug in stochastischer Umwelt:
Aus Erfahrung lernen, Strategie optimieren, Exploration vs Exploitation.
Aufzugsteuerung: Ja, es soll auch intelligente Aufzüge geben!

Robi the Robot und Machine Learning

Robi, der Saugroboter, lebt in einer 2×3–Wohnung (links), später in einer 4×3–Wohnung mit einem Rundgang (rechts).







Nutzen $\sum_{t=0}^{T-1} \delta^t r_t + \delta^T v(x_T)$ mit $x_T \in \partial X$ und Diskont $\delta \in [0, 1]$.

In jedem Zeitschritt wählt Robi eine Richtung; diese fährt er mit Wkt 80%, mit Wkt $p = 10\%$ links oder rechts dazu, weil z.B. die Katze ihn schubst. Falls sich in Fahrtrichtung eine Wand befindet, bleibt er einfach stehen.

Jeder Zeitschritt bringt eine Belohnung $r \in \mathbb{R}$, etwa $r < 0$ für Verbrauch, egal ob oder wohin er fährt. Wenn er die Ladestation oder die Treppe erreicht, endet seine Fahrt mit der Auszahlung $+1$ bzw. -1.

Robi the Robot und Machine Learning

Dieses beliebte Lehrbeispiel des maschinellen Lernens stammt aus
📖 S. Russell, P. Norvig: *Artificial Intelligence: A Modern Approach.* Addison Wesley (3rd ed.) 2016 (Kapitel 17: Making complex decisions)

In den letzten Jahren hat dieser Ansatz großen Zulauf erhalten: Es gibt einen Massenmarkt, denn Geräte sind preiswert und alltagstauglich. Zudem genügen in vielen einfachen Projekten bereits die Grundlagen der Theorie für beachtliche Anwendungserfolge. Dank Softwarepaketen ist auch die Hürde in der Programmierung inzwischen recht gering.

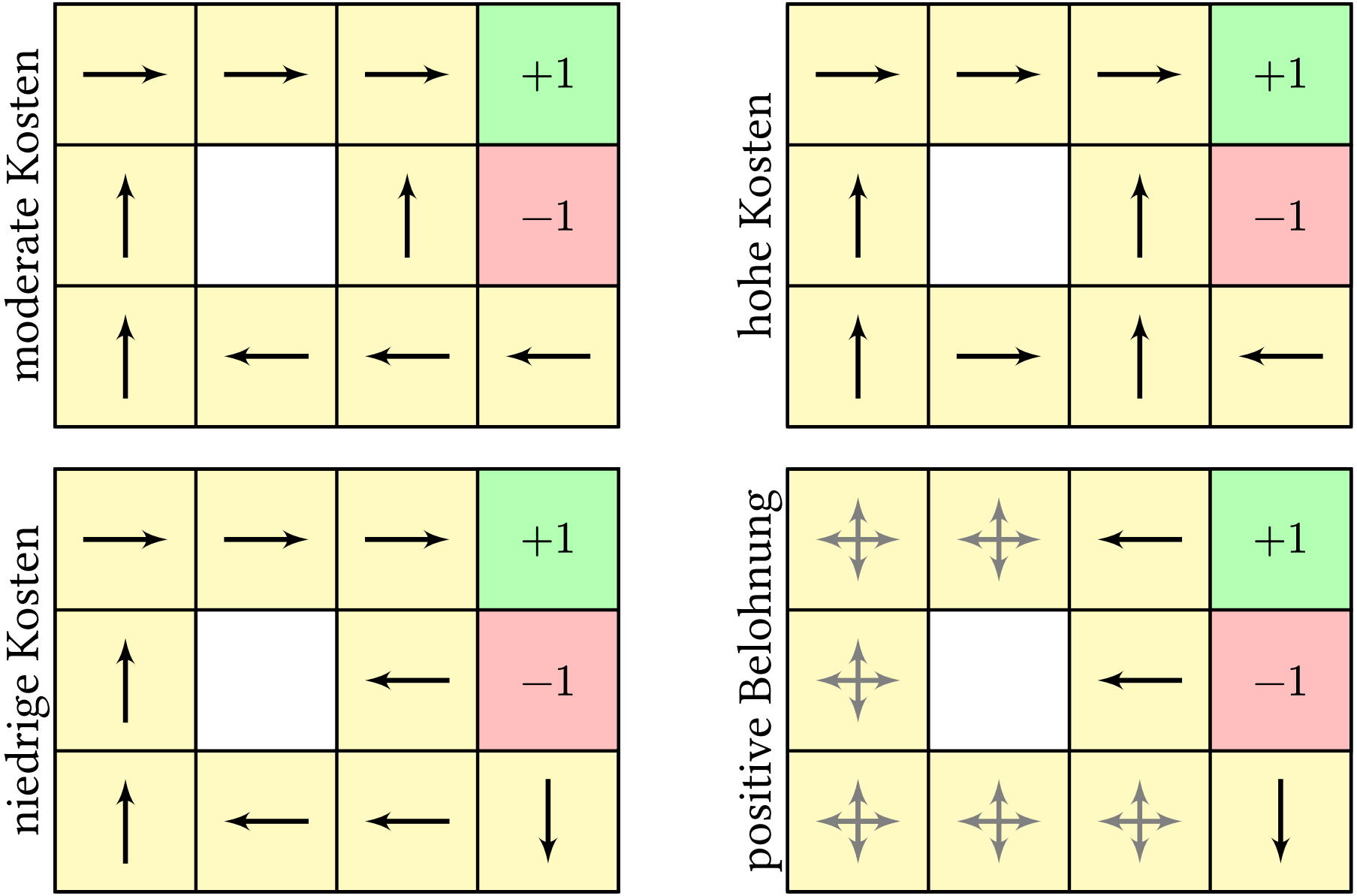
🙂 Sie können und sollen mit diesem Übungsbeispiel konkret rechnen und numerische Erfahrungen sammeln: Sie finden eine Umsetzung als Tabellenkalkulation unter `eiserm.de/lehre/Spieltheorie/Robi.ods`.

Selbst in diesem einfachen Beispiel sind die Strategien überraschend. Spielen Sie etwas mit den Parametern, Belohnung und Auszahlungen! Wie ändert das Robis Verhalten? Begründung? Interpretation? Wann und warum konvergiert der Bellman–Operator?

Übung: Berechnen / implementieren Sie ebenso die 4×3–Wohnung.

Robi the Robot und Machine Learning

Robis Verhalten ist erstaunlich komplex, abhängig von den Kosten r:



Robi the Robot und Machine Learning

🙂 Robi handelt strategisch, er sucht die Balance zwischen Belohnung und Risiko. Das ist durchaus typisch für viele realistische Anwendungen.

Bei moderaten Kosten lohnt sich für Robi der Umweg des Rundgangs. Bei hohen Kosten ist der direkte Weg besser. Probieren Sie es aus!

Bei geringen Kosten, r knapp unter Null, kann Robi sogar garantieren, nie die Treppe hinunterzufallen: Er kann ihr ganz vorsichtig ausweichen, auch wenn er sich öfter die Nase an der Wand stößt und länger fährt.

Der Fall $r = 0$ bietet die ersten Überraschungen: Probieren Sie es!

Bei positiver Belohnung $r > 0$ will Robi nie aufhören, auch nie aufladen, da er beliebig Nutzen generieren kann, indem er fröhlich umherfährt. Er ist wie auf Drogen, *high on reward*, als gäbe es kein morgen mehr.

Bei zu hohen Kosten jedoch ist Robis Leben so miserabel, dass er stets den nächsten Ausgang wählt, notfalls stürzt er sich die Treppe hinunter. Das ist zwar traurig, aber unter diesen Bedingungen das beste für ihn.

⚠ Die Wahl des Belohnugssystems entscheidet über das Verhalten. Alle Trainer / Lehrer / Eltern wissen das: *Choose rewards wisely!*

Gewinniteration / *value iteration*

Zu (Γ, r, v) betrachten wir $\Phi : E \to E$ sowie $\Phi_s : E \to E$ für $s \in S(\Gamma)$.
Wir nehmen an, dass wir zu Φ eine Kontraktionskonstante k kennen.

Algo D3A: Gewinniteration / *value iteration*

Global: Markov–Spiel (Γ, r, v), Kontraktionskonstante $k \in [0, 1[$
Eingabe: eine initiale Erwartung $u \in E$ und eine Toleranz $\varepsilon \in \mathbb{R}_{>0}$
Ausgabe: eine ε–optimale Erwartung $u \in E$ und Strategie $s \in S(\Gamma)$

1: repeat
2: kopiere $u' \leftarrow u$ und aktualisiere $u \leftarrow \Phi(u)$
3: until $|u - u'| \leq \frac{1-k}{k}\varepsilon$
4: for $x \in X^\circ$ do wähle $s(x) \in \operatorname{Arg\,max}_{a \in A_x} H(x, a, u)$
5: return (u, s)

Aufgabe: Warum endet dieser Algorithmus und ist korrekt?

Lösung: Beides verdanken wir Banachs Fixpunktsatz D2A!

⚠ Die abgelesene Strategie s ist im Allgemeinen noch nicht optimal!
Um dies zu garantieren, muss die Näherung hinreichend gut sein.

Strategie vs Gewinniteration

Aufgabe: Wie klein sollten wir $\varepsilon \in \mathbb{R}_{>0}$ wählen, damit wir aus jeder ε–Näherung $\tilde{u}$ alle optimalen Strategien $s \in S(\Gamma)$ ablesen können?

Lösung: Sei $u = \Phi(u)$ die exakte Lösung. Gegeben ist $|\tilde{u} - u| \leq \varepsilon$.
Zu $x \in X^\circ$ sei $\{\, u(x) - H(x,a,u) \mid a \in A_x \,\} = \{0 = \lambda_x^0 < \lambda_x^1 < \dots\}$.
Wir nennen $\lambda := \min\{\, \lambda_x^1 \mid x \in X^\circ \,\}$ die **Spektrallücke** von (Γ, r, v).
Für jede Aktion $a \in A_x$ ist $\tilde{u}(x) - H(x,a,\tilde{u})$ gegeben durch

$$[\tilde{u}(x) - u(x)] + [u(x) - H(x,a,u)] + [H(x,a,u) - H(x,a,\tilde{u})]$$

Für $a \in A_x$ sub/optimal gilt $\tilde{u}(x) - H(x,a,\tilde{u}) \geq \lambda - 2\varepsilon$ bzw. $\leq 2\varepsilon$.
Für $4\varepsilon < \lambda$ können wir so aus $\tilde{u}$ alle optimalen Strategien s ablesen.

Aufgabe: Die Schranke λ nutzt die exakte, unbekannte Lösung u.
Können Sie eine Schranke finden, die nur die Näherung $\tilde{u}$ nutzt?

Lösung: Zu $x \in X^\circ$ und $A_x = \{a_0, \dots, a_\ell\}$ sei $\mu_x^i = \tilde{u}(x) - H(x, a_i, \tilde{u})$ sortiert gemäß $\mu_x^0 \leq \mu_x^1 \leq \dots \leq \mu_x^\ell$. Wir setzen $\mu := \min\{\, \mu_x^1 \mid x \in X^\circ \,\}$.
Gilt $2\varepsilon < \mu$, so können wir die optimale Strategie s aus $\tilde{u}$ ablesen:
Für $a \in A_x$ sub/optimal gilt $\tilde{u}(x) - H(x,a,\tilde{u}) \geq \mu > 2\varepsilon$ bzw. $\leq 2\varepsilon$.

Strategieiteration / *policy iteration*

Algo D3B: Strategieiteration / *policy iteration*

Global: das endliche Markov–Spiel (Γ, r, v)
Eingabe: eine initiale Strategie $s \in S(\Gamma)$
Ausgabe: eine optimale Strategie $s \in S(\Gamma)$ und ihre Erwartung $u \in E$

1: repeat
2: löse $u = \Phi_s(u)$ und setze done $\leftarrow$ true
3: for $x \in X^\circ$ do
4: if $\max_{a \in A_x} H(x, a, u) > H(x, s(x), u)$ then
5: wähle $s(x) \in \operatorname{Arg\,max}_{a \in A_x} H(x, a, u)$ und setze done $\leftarrow$ false
6: until done
7: return (s, u)

Aufgabe: Warum endet dieser Algorithmus und ist korrekt?

Lösung: Die Strategiemenge $S = S(\Gamma)$ ist endlich, und es gilt $u_0 \lneq u_1 \lneq u_2 \lneq u_3 \lneq \dots$ bis schließlich $u = \Phi_s(u) = \Phi(u)$.

Ist Φ zudem kontraktiv, so ist u die eindeutige Lösung.

kombinierte Strategie-und-Gewinniteration

Oft kombiniert man die Gewinniteration mit der Strategieiteration:

Algo D3c: kombinierte Strategie-und-Gewinniteration

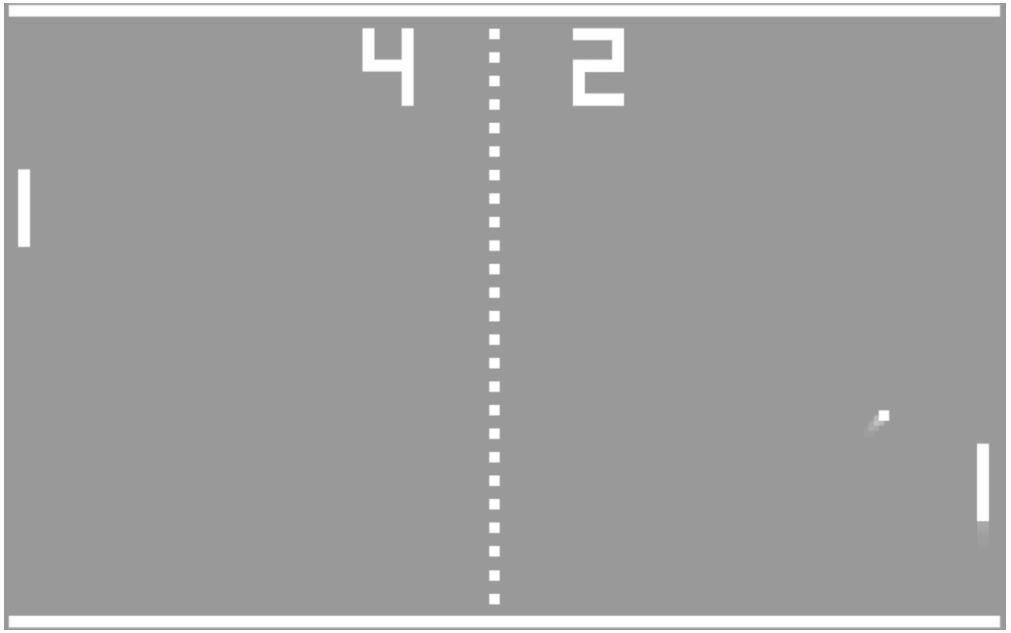
Global: das Markov–Spiel (Γ, r, v), Kontraktionskonstante $k \in [0, 1[$
Eingabe: eine initiale Erwartung $u \in E$ und eine Toleranz $\varepsilon \in \mathbb{R}_{>0}$
Ausgabe: eine optimale Strategie $s \in S(\Gamma)$ und ihre Erwartung $u \in E$

1: Gewinniteration D3A: berechne ein ε–optimales Paar (u, s)
2: Strategieiteration D3B: prüfe und nötigenfalls optimiere (s, u)
3: return (s, u)

Übung: Wenn Sie möchten, können Sie diese Methode (und Varianten) mathematisch untersuchen, implementieren und experimentell erproben. Wird die Theorie schwächer, so muss praktische Erfahrung ausgleichen.

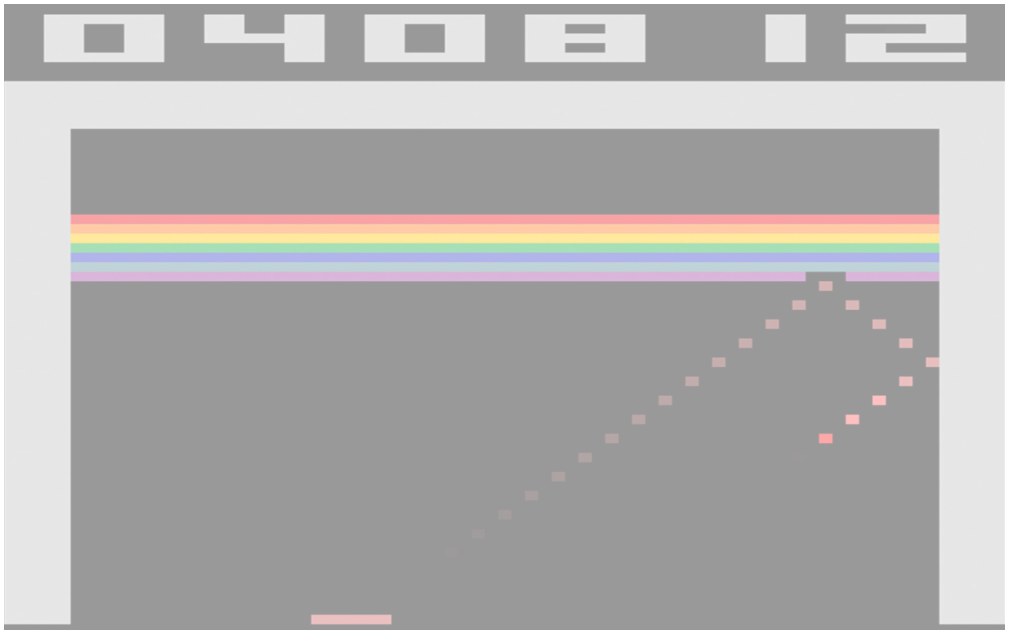
☺ Sie bearbeiten das schöne Beispiel *Robi the Robot* in unserer Übung. Solch handfeste Erfahrungen müssen Sie unbedingt selbst erleben; dann stellen Sie die richtigen Fragen und sind bereit für Antworten.

Pinball Wizard (frei nach *The Who*, 1969)



Pong
(Atari 1972)

zwei Spieler
deterministisch



Breakout
(Atari 1976)

ein Spieler
probabilistisch

Pinball Wizard (frei nach *The Who*, 1969)

Die **Künstliche Intelligenz** feiert inzwischen beachtliche Erfolge in Alltagsprodukten, von Spracherkennung bis autonomen Fahrzeugen. Furore machten zuletzt generative Sprachmodelle, kurz *LLM*s.

Durch **bestärkendes Lernen** [*reinforcement learning*] und ähnlichen Algorithmen kann ein strategischer Akteur (Agent, Computer, Roboter) *selbständig* aus Erfahrung lernen und immer bessere Strategien finden. Diese Eigenständigkeit ist eine ganz wesentliche, innovative Technik.

Ein amüsant-spektakuläres Beispiel sind Arcade-Spiele: Das Startup **DeepMind** (`en.wikipedia.org/wiki/DeepMind`) gehört inzwischen zu Google und hat diese grundlegende Idee seit 2013 sehr erfolgreich umgesetzt.

📖 Mnih et al: *Human-level control through deep reinforcement learning.* Nature 518 (2015) 529–533, `www.nature.com/articles/nature14236`

▶ `youtu.be/V1eYniJ0Rnk` als zweiminütiges Video oder ausführlicher als schön geschriebener Artikel `doi.org/10.1126/science.aaa7903`.

Arcade-Spiele sind ein gutes Testfeld: weder zu einfach noch zu schwer. Die Belohnungen sind zumeist dicht genug, um das Lernen zu fördern.

Pinball Wizard (frei nach *The Who*, 1969)

Einige Besonderheiten des **bestärkenden Lernens**:

- **Unsupervised:** Es gibt keine Anleitung und keinen Trainer
  Der Spieler lernt nur aus den Signalen seiner Belohnungen.
- **Exploration:** Aktionen beeinflussen zukünftige Informationen.
  Der Spieler muss daher aktiv interagieren und erforschen.
- **Delayed Feedback:** Aktionen bewirken spätere Belohnungen.
  Der Zeitablauf des Spiels ist ein wesentlicher Faktor.

Der Spieler muss seine Aktionen wählen und hierzu einen Kompromiss finden zwischen der **Erkundung** [*exploration*] neuer Möglichkeiten und der **Nutzung** [*exploitation*] bewährter Wege. Wie im richtigen Leben!

Die **sofortigen Belohnungen** leiten den Spieler in seinem Lernprozess: Er sucht die Balance zwischen kurzfristigem und langfristigem Nutzen. Nur mit gutem **Belohnungssignal** macht das Lernen gute Fortschritte.

Bei **allzu seltenen Belohnung** r oder endlastigen Auszahlungen v sind die Fortschritte recht langsam. In großen Umgebungen (Spielgraphen) müssen die Zustände geeignet zusammengefasst / abstrahiert werden.

Pinball Wizard (frei nach *The Who*, 1969)

Das grundlegende Lehrbuch zum bestärkenden Lernen:
📖 Richard S. Sutton, Andrew G. Barto: *Reinforcement Learning*.
The MIT Press (2nd ed.) 2018 (hier speziell §6.5: Q–learning),
online verfügbar unter `incompleteideas.net/book/RLbook2020.pdf`

Das folgende Vorlesungsskript ist deutlich kürzer und knapper:
📖 Csaba Szepesvári: *Algorithms for Reinforcement Learning*,
online `sites.ualberta.ca/~szepesva/papers/RLAlgsInMDPs-lecture.pdf`

Es gibt exzellente Online-Kurse zu diesem Thema, zum Beispiel von
▶ David Silver: *Intro to Reinforcement Learning*. `youtu.be/2pWv7GOvuf0`
Silver hat 2008 bei Sutton promoviert und ist seither extrem erfolgreich.
Hierzu ein Interview mit Hannah Fry zu Geschichte und Zukunft:
▶ Google DeepMind: *Is Human Data Enough?* `youtu.be/zzXyPGEtseI`

Ich kann der Versuchung nicht widerstehen und will Ihnen einen ganz kurzen Ausblick dieses aktuellen und faszinierenden Gebiets skizzieren: Es verbindet Informatik und Mathematik, Lerntheorie und Spieltheorie, und ist im Ingenieurwesen und seinen Anwendungen angekommen.

Bestärkendes Lernen / *reinforcement learning*

☺ Angenommen, wir kennen zu (Γ, r, v) die Gewinnfunktion $u = \Phi(u)$. Im Zustand $x \in X^\circ$ bewerten wir die Qualität jeder Aktion $a \in A_x$ durch

$$q(x,a) = H(x,a,u) := \sum_{y \in X} p(x,a,y)\,\big[r(x,a,y) + \delta u(y)\big].$$

Optimalitätsprinzip: Jede optimale Aktion $a \in A_x$ maximiert $q(x,a)$.

Angenommen, der Spieler kennt anfangs nur den Markov–Graphen Γ. Wie kann er die Bewertungen q und u spielend-explorativ erlernen?

☺ Als Näherung für q nutzt er sein bisheriges **Erfahrungswissen**:

$$Q : A \to \mathbb{R} : (x,a) \mapsto Q(x,a)$$

Jeden aktiven Zustand $x \in X^\circ$ bewertet er näherungsweise durch

$$U : X^\circ \to \mathbb{R} : x \mapsto U(x) = \max_{a \in A_x} Q(x,a).$$

Im aktiven Zustand $x \in X^\circ$ wählt der Spieler eine Aktion $a \in A_x$ und gelangt zum Zustand y mit Belohnung $R = r(x,a,y)$. Er aktualisiert

$$Q(x,a) \leftarrow (1-\alpha)\,\underbrace{Q(x,a)}_{\text{alte und…}} + \alpha\,\underbrace{\big[R + \delta U(y)\big]}_{\text{neue Erfahrung}} \quad \text{und} \quad U(x) \leftarrow \max_{a \in A_x} Q(x,a).$$

Bestärkendes Lernen / *reinforcement learning*

Der Spieler lernt durch seine Erfahrung mit der **Lernrate** $\alpha \in]0,1[$.
Sein **Erfahrungswissen** Q und U wird dabei wie folgt aktualisiert:

Algo D3D: Aktualisierung von Q und U

Global: die Bewertungen $Q : A \to \mathbb{R}$ und $U : X \to \mathbb{R}$
Eingabe: die Aktion $a : x \to y$ mit Belohnung $R \in \mathbb{R}$.

1: aktualisiere $Q(x,a) \leftarrow (1-\alpha)Q(x,a) + \alpha[R + \delta U(y)]$
2: aktualisiere $U(x) \leftarrow \max\{U(x), Q(x,a)\}$

Ist y terminal, so erhält er zudem $V = v(y)$ und aktualisiert $U(y) \leftarrow V$.
Damit endet diese **Episode**, die Trajektorie dieses Spieldurchgangs.
Zur weiteren Verbesserung werden noch weitere Episoden gespielt.

Der hier beschriebene Algorithmus heißt **Q–Lernen** [Q–*learning*].
Der Buchstabe „Q“ steht für die Funktion $Q : A \to \mathbb{R}$, die die Qualität der Aktionen bewertet und in diesem Algorithmus die Hauptrolle spielt.

☺ Erhofft ist die rasche **Konvergenz** $Q_t \to q$ und $U_t \to u$ für $t \to \infty$.
Dazu gibt es mathematische Sätze und eindrückliche praktische Erfolge.

Bestärkendes Lernen / *reinforcement learning*

Maschinelles Lernen nutzt Algorithmen, die aus Erfahrungen lernen, mit Hilfe statistischer Methoden auf Trainingsdaten, hier Spieldaten.

Zwei grundlegende Algorithmen sind **Gewinniteration** [*value iteration*] und **Strategieiteration** [*policy iteration*], wie oben zusammengefasst. Sie setzen allerdings voraus, dass das ganze Spiel (Γ, r, v) bekannt ist; dann stehen alle genannten mathematischen Werkzeuge zur Verfügung.

Beim **bestärkenden Lernen** [*reinforcement learning*] erlernt der Agent selbständig eine Strategie. Ihm wird anfangs nur der Graph Γ gegeben, aber keine Hinweise, welche Aktion in welcher Situation die beste wäre. Noch realistischer: Er muss auch den Graphen Γ erst selbst erkunden!

Reines Beobachten genügt in diesem Lernprozess nicht, Informationen gewinnt der Spieler nur durch **Interaktion**. Alle seine Aktionen $a : x \to y$ führen zu Belohnungen, aus diesen approximiert er die Qualität $Q(x, a)$ der Aktion und den Nutzen $U(x) = \max_{a \in A_x} Q(x, a)$ des Zustands x.

Der **Algorithmus** stammt von Watkins (1989) und Bozinovski (1981). Seine **Konvergenz** wurde bewiesen von Watkins und Dayan (1992).

Bestärkendes Lernen / *reinforcement learning*

Die Grundidee stammt aus der **Psychologie** und wurde bereits seit den Anfängen der Kybernetik verwendet, so etwa von Marvin Minsky in seiner Dissertation: *Neural Nets and the Brain Model Problem.* (1954)

Bestärkendes Lernen ist inzwischen ein großes interdisziplinäres Gebiet und verbindet Informatik, Optimierung und Ökonomik mit Psychologie und Neurowissenschaften. Verfolgt werden dabei zwei Ziele:

(1) Bei realen Lebewesen soll das beobachtete (Lern)Verhalten durch geeignete Modelle möglichst gut erklärt werden (deskriptiv, explikativ).

(2) Künstliche Agenten sollen mit ihrer Umwelt strategisch interagieren und daraus möglichst effizient lernen (normativ, konstruktiv).

Übung: Wenn Sie gerne programmieren, dann können Sie unsere Miniaturbeispiele implementieren und durch bestärkendes Lernen lösen.

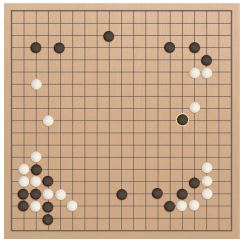
Allgemein lohnt sich hierbei eine möglichst generische Problemlösung, die allgemein Markov–Spiele behandelt. Sie können dabei viel lernen!

Wenn Sie mit den Grundideen (und auch Problemen) vertraut sind, dann lohnt sich ein Blick auf die umfangreichen Softwarepakete.

Anwendungen im maschinellen Lernen

Google DeepMind ist spezialisiert auf maschinelles Lernen:

2013: Deep Q-Learning meistert Arcade-Spiele wie Pong, Breakout, etc.



2015: AlphaGo schlägt den mehrfachen Europameister Fan Hui.
2016: AlphaGo schlägt den südkoreanischen Profi Lee Sedol.
2017: AlphaGo schlägt den Weltranglistenersten Ke Jie.

2017: AlphaGo Zero verbessert AlphaGo. Es lernt selbständig anhand der Spielregeln und benötigt sonst keinerlei Vorwissen über das Spiel. *Self-Play* ist eine entscheidende Erweiterung zum *Human Feedback*.

2018: AlphaFold sagt Proteinstrukturen vorher. Chemie-Nobelpreis 2024. Es ist wichtig, und notorisch schwierig, anhand der Aminosäuresequenz die dreidimensionale Faltung eines Proteins vorherzusagen.

2022: AlphaTensor optimiert schnelle Matrixmultiplikation.

2024: AlphaProof erreicht Silbermedaille in der IMO. Auch Mathematik, insbesondere das Entwickeln mathematischer Beweise, ist ein Spiel: Hier dient LEAN als Beweisassistent und prüft die vorgeschlagenen Beweisen. Dieses objektive Kriterium schützt insbesondere vor Halluzinationen.

Anwendungen im maschinellen Lernen

Anfangs trainierte man Modelle mit menschlicher Supervision / *Human Feedback*, doch nun zeigt selbständiges Spielen / *Self-Play* sein Potential.

In theory, computers could learn new skills at incredible speed if they didn't have to wait for human teachers to give feedback on whether they are on the right track. But this approach – known as unsupervised learning – has rarely worked [...]. The video playing behind Hassabis showed what seemed impossible: A computer learning on its own to play complicated video games like Breakout [...]. After exploring the game by playing it, the computer discovered advanced strategies that few humans know about, such as digging a hole to bounce the ball along the back side of the wall.

Science (2015), `doi.org/10.1126/science.aaa7903`

Diese Idee ist revolutionär, universell einsetzbar und höchst erfolgreich. Demis Hassabis und John Jumper erhielten 2024 den Chemie-Nobelpreis für ihre Forschung zur Proteinvorhersage. Ihr erstaunlicher Weg führte von Biologie über das Computerspiel *Foldit* zu maschinellem Lernen.

The Most Useful Thing AI Has Ever Done. `youtu.be/P_fHJIYENdI`

Anwendungen im maschinellen Lernen

Inzwischen ist die Technik des bestärkenden Lernens längst zu einem bewährten Universalwerkzeug avanciert, sowohl in der Forschung als auch in Alltagsprodukten. In vielen Haushalten verrichten Saugroboter erfolgreich ihren Dienst, in vielen Gärten arbeiten Mähroboter, uvm.

Bislang haben wir nur einen einzelnen Akteur betrachtet. Der nächste Schritt ist die Interaktion mehrerer Akteure, etwa die Kooperation einer Rudels von Mährobotern oder eines Schwarms fliegender Drohnen. Bei divergierenden Zielen wird Optimierung zur Spieltheorie!

Im Februar 2022 befahl der russische Präsident Wladimir Putin seinen schändlichen Angriffskrieg gegen die Ukraine. In den Kriegshandlungen spielen erstmals Flugdrohnen aller Größen eine überaus wichtige Rolle. Noch werden sie meist von menschlichen Pilot:innen gesteuert.

Maschinelles Lernen wird selbstverständlich auch hier eingesetzt werden. In Anbetracht des allseits großen Drucks wird es in den kommenden Jahren autonome Flugdrohnen mit künstlicher Intelligenz versehen. Wo immer Anwendungen locken, da zum Guten wie zum Bösen.

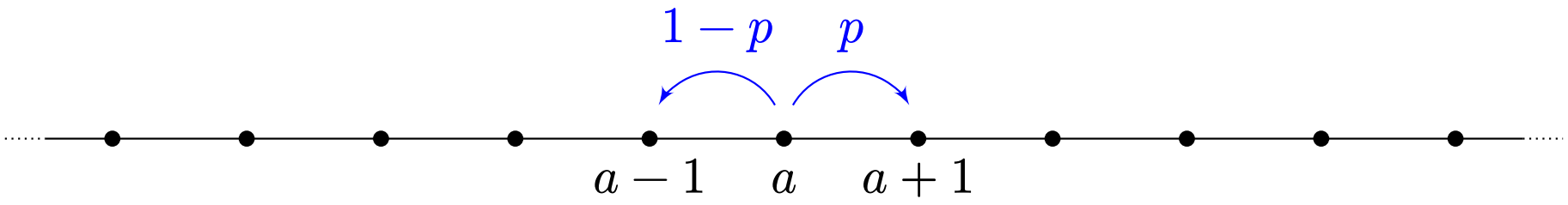
Anwendungen im maschinellen Lernen

„Sentre's Neurocaster: Winning Design Awards. And Wars."

The Electric State ist ein illustrierter Roman des schwedischen Autors Simon Stålenhag aus dem Jahr 2018. Er spielt in den 1990er Jahren einer alternativen Realität, nachdem ein Drohnenkrieg die USA verwüstet hat.

Markov–Ketten: Konstruktion dank Pfadregel

Eine **Markov–Kette** ist ein Modell für eine Abfolge zufälliger Ereignisse. Typisches Beispiel ist die zufällige Irrfahrt, ebenso einfach wie berühmt:



Zufällige Irrfahrt in $\mathbb{Z}$: Zur Zeit $t=0$ starten Sie im Punkt $X_0=0$. Von X_t nach X_{t+1} gehen Sie mit Wkt $p \in [0,1]$ nach rechts, sonst links.

Allgemein: Sei S_t eine diskrete Zustandsmenge. Der Start sei gegeben als eine WVerteilung $q_0 : S_0 \to [0,1]$, mit $\sum_a q_0(a) = 1$, der Übergang von t zu $t+1$ durch $p_t : S_t \times S_{t+1} \to [0,1]$ mit $\sum_b p_t(a,b) = 1$ für alle $a \in S_t$. Dies definiert eine Markov–Kette $(X_t)_{t\in\mathbb{N}}$ gemäß der **Pfadregel**:

$$\boxed{\mathbf{P}(\, X_0{=}a_0,\, X_1{=}a_1,\, \dots,\, X_n{=}a_n\,) = q_0(a_0)\, p_0(a_0,a_1) \cdots p_{n-1}(a_{n-1},a_n)}$$

Bei obiger Irrfahrt ist $S_t = \mathbb{Z}$, darauf $q_0(0) = 1$, somit $q_0(a) = 0$ für $a \neq 0$, sowie $p_t(a,a+1) = p$ und $p_t(a,a-1) = 1-p$, sonst folglich $p_t(a,b) = 0$.

Markov–Ketten: Konstruktion dank Pfadregel

☺ Dadurch sind alle relevanten Wkten für $(X_t)_{t\in\mathbb{N}}$ eindeutig bestimmt. Auf dieser Grundlage können wir also im Folgenden konsistent rechnen. Wir fordern genau so viel, wie wir benötigen – nicht mehr, nicht weniger.

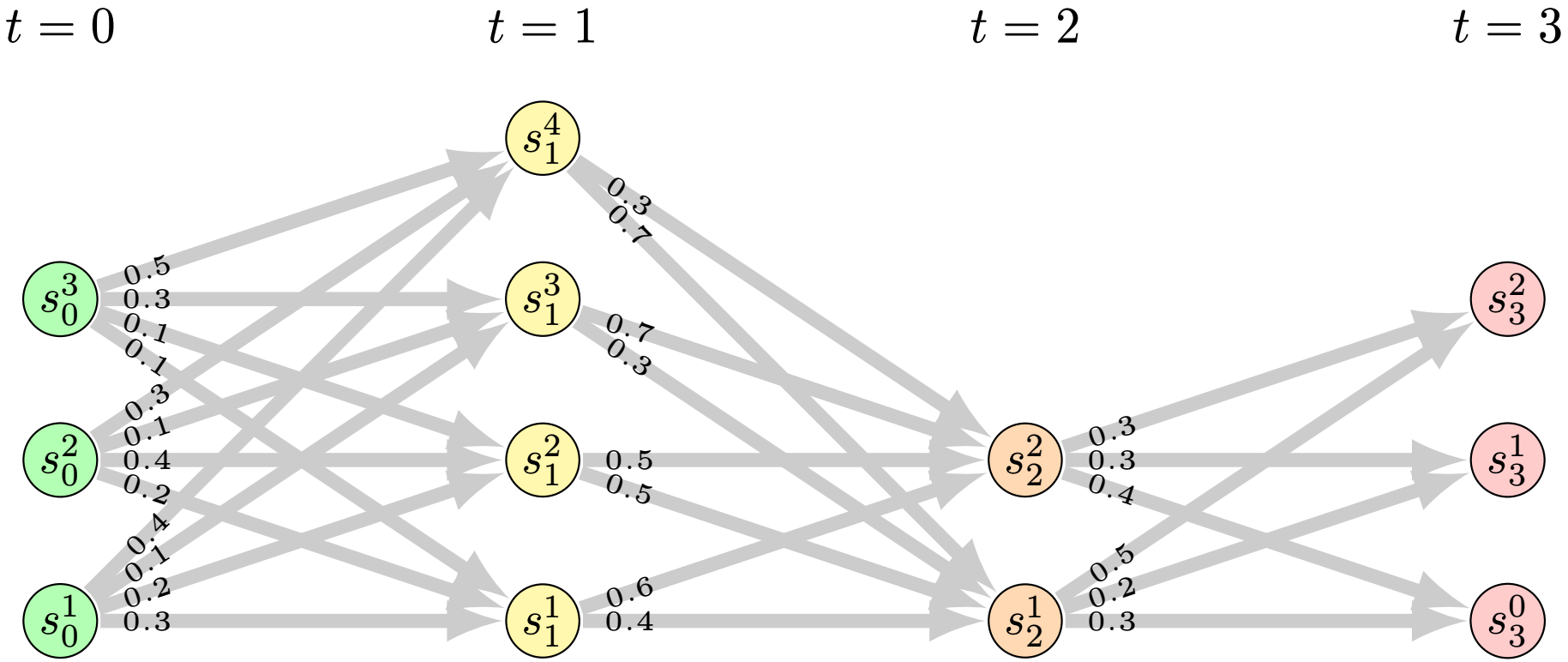
? Unser Vorgehen ist elegant und allgemein. Müssen wir dabei etwaige Einschränkungen beachten oder gar logische Widersprüche befürchten? Wie können wir sicher sein, dass solch ein Modell existiert, das heißt ein WRaum $(\Omega, \mathscr{A}, \mathbf{P}, (X_t)_{t\in\mathbb{N}})$, der all unsere obigen Forderungen erfüllt?

? Für *endliche* Pfade ist alles klar. Wie behandeln wir *unendliche* Pfade? Wir betrachten die Menge $\Omega = \prod_{t\in\mathbb{N}} S_t$ aller Pfade $\omega = (\omega_0, \omega_1, \omega_2, \ldots)$. Die ZVariable $X_t : \Omega \to S_t : \omega \mapsto \omega_t$ projiziert $\omega \in \Omega$ auf $X_t(\omega) = \omega_t \in S_t$. Diese Familie $(X_t)_{t\in\mathbb{N}}$ erzeugt die σ–Algebra $\mathscr{A}$ auf Ω. Auf $(\Omega, \mathscr{A})$ existiert genau ein WMaß $\mathbf{P} : \mathscr{A} \to [0,1]$, das unsere ersehnte Pfadregel erfüllt.

⚠ Diese Konstruktion ist plausibel, aber keineswegs trivial. Den letzten Schritt verdanken wir dem **Maßerweiterungssatz** von Carathéodory.

📖 Jürgen Elstrodt: *Maß- und Integrationstheorie.* Springer 2018.
Heinz Bauer: *Wahrscheinlichkeitstheorie.* Walter de Gruyter 1991.

Markov–Ketten: Pfadregel & Summenregel = Matrixprodukt



$$[S_0] \xrightarrow{P_0} [S_1] \xrightarrow{P_1} [S_2] \xrightarrow{P_2} [S_3]$$

$$\begin{bmatrix} 0.3 & 0.2 & 0.1 & 0.4 \\ 0.2 & 0.4 & 0.1 & 0.3 \\ 0.1 & 0.1 & 0.3 & 0.5 \end{bmatrix} \quad \begin{bmatrix} 0.4 & 0.6 \\ 0.5 & 0.5 \\ 0.3 & 0.7 \\ 0.7 & 0.3 \end{bmatrix} \quad \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.4 & 0.3 & 0.3 \end{bmatrix}$$

Markov–Ketten: Pfadregel & Summenregel = Matrixprodukt

Als konkretes Beispiel betrachten wir die endlichen Zustandsmengen $S_0 = \{s_0^1, s_0^2, s_0^3\}$, $S_1 = \{s_1^1, s_1^2, s_1^3, s_1^4\}$, $S_2 = \{s_0^1, s_0^2\}$, $S_3 = \{s_3^1, s_3^2, s_3^3\}$, usw.

Auf der Menge S_0 ist jede WVerteilung $w_0 : S_0 \to \mathbb{R}_{\geq 0} : s_0^i \mapsto w_0^i$ ein stochastischer Vektor $w_0 = (w_0^1, w_0^2, w_0^3)$ bestehend aus den Wkten $w_0^1, w_0^2, w_0^3 \in \mathbb{R}_{\geq 0}$ mit Summe $w_0^1 + w_0^2 + w_0^3 = 1$. Wir schreiben kurz $[S_0] \subset \mathbb{R}^{S_0}$ für die Menge dieser stochastischen Vektoren.

Die Übergangswkten $p_0(s_0^i, s_1^j)$ schreiben wir als Matrix $P_0 \in \mathbb{R}^{3\times 2}$. Der Eintrag $P_0^{ij} = p_0(s_0^i, s_1^j)$ ist die Wkt, von s_0^i nach s_1^j überzugehen. Alle Einträge sind nicht-negativ, und die Summe jeder Zeile ergibt 1.

Ist $w_0 = (w_0^1, w_0^2, w_0^3)$ die Verteilung zur Zeit $t = 0$, so ist $w_1 = w_0 P_0$ die WVerteilung zur Zeit $t = 1$, dank Matrixprodukt $w_1^j = \sum_{i=1}^3 w_0^i P_0^{ij}$. Das entspricht der Pfadregel mit Summenregeln für disjunkte Ereignisse.

Wir erhalten so die Abbildung $P_0 : [S_0] \to [S_1] : w_0 \mapsto w_1 = w_0 P_0$. Diese Situation ist oben graphisch durch Pfeile veranschaulicht. Die Komposition $P_0 P_1 : [S_0] \to [S_2]$ ist das Matrixprodukt.

Markov–Ketten: Definition dank Markov–Eigenschaft

Eine **Markov–Kette** ist ein Modell für eine Abfolge zufälliger Ereignisse. Das Zeitintervall ist diskret $I = \mathbb{N} = \mathbb{Z}_{\geq 0}$, später kontinuierlich $I = \mathbb{R}_{\geq 0}$. Zu jedem Zeitpunkt $t \in I$ beschreibt X_t den Zustand. Die Menge S_t der möglichen Zustände sei zunächst diskret, etwa $S_t \subseteq \mathbb{Z}^n$, später $S_t \subseteq \mathbb{R}^n$. Vereinfachend betrachtet man oft eine feste Menge $S = S_t$ für alle $t \in I$.

Definition D4A: Markov–Kette $(\Omega, \mathscr{A}, \mathbf{P}, (X_t)_{t\in\mathbb{N}})$

Sei $(\Omega, \mathscr{A}, \mathbf{P})$ ein WRaum. Zu jedem $t \in \mathbb{N}$ sei $X_t : \Omega \to S_t$ eine ZVariable. Die Familie $(X_t)_{t\in\mathbb{N}}$ auf $(\Omega, \mathscr{A}, \mathbf{P})$ nennen wir **Markov–Kette**, falls gilt:

$$\boxed{\mathbf{P}(\, X_{n+1}{=}a_{n+1} \mid X_n{=}a_n,\, \dots,\, X_0{=}a_0\,) = \mathbf{P}(\, X_{n+1}{=}a_{n+1} \mid X_n{=}a_n\,)}$$

für alle Zeitpunkte $n \in \mathbb{N}$ und Zustände $a_0 \in S_0, \dots, a_n \in S_n, a_{n+1} \in S_{n+1}$ mit $\mathbf{P}(\, X_n{=}a_n,\, \dots,\, X_0{=}a_0\,) > 0$, so dass beide Seiten definiert sind.

In Worten: Unter der Bedingung des gegenwärtigen Zustands $X_n = a_n$ ist die Zukunft X_{n+1} unabhängig von der Vergangenheit $X_{n-1}, \dots, X_0$. Das klingt poetisch und umschreibt anschaulich, die Formel präzisiert!

Markov–Ketten: Definition dank Markov–Eigenschaft

Aufgabe: Rechnen Sie nach, dass unsere vorige Konstruktion der Kette $(\Omega, \mathscr{A}, \mathbf{P}, (X_t)_{t\in\mathbb{N}})$ tatsächlich die Definition einer Markov–Kette erfüllt.

Lösung: (1) Dank Pfadregel finden wir für die linke Seite:

$$\mathbf{P}(X_{n+1}{=}a_{n+1}, X_n{=}a_n, \dots, X_0{=}a_0) = q_0(a_0) \cdots p_{n-1}(a_{n-1}, a_n)\, p_n(a_n, a_{n+1})$$

$$\mathbf{P}(X_n{=}a_n, \dots, X_0{=}a_0) = q_0(a_0) \cdots p_{n-1}(a_{n-1}, a_n)$$

$$\mathbf{P}(X_{n+1}{=}a_{n+1} \mid X_n{=}a_n, \dots, X_0{=}a_0) = \qquad p_n(a_n, a_{n+1})$$

Gemäß Definition setzen wir hierzu $\mathbf{P}(\, X_n{=}a_n, \dots, X_0{=}a_0\,) > 0$ voraus.

(2) Rechts nutzen wir die Summenregeln für disjunkte Ereignisse:

$$\mathbf{P}(\, X_{n+1}{=}a_{n+1}, X_n{=}a_n\,) = \sum_{b_0} \cdots \sum_{b_{n-1}} q_0(b_0) \cdots p_{n-1}(b_{n-1}, a_n)\, p_n(a_n, a_{n+1})$$

$$\mathbf{P}(\, X_n{=}a_n\,) = \sum_{b_0} \cdots \sum_{b_{n-1}} q_0(b_0) \cdots p_{n-1}(b_{n-1}, a_n)$$

$$\mathbf{P}(\, X_{n+1}{=}a_{n+1} \mid X_n{=}a_n\,) = \qquad p_n(a_n, a_{n+1})$$

Die Division ist möglich, da $\mathbf{P}(\, X_n{=}a_n\,) \geq \mathbf{P}(\, X_n{=}a_n, \dots, X_0{=}a_0\,) > 0$.

Markov–Ketten: Vorsicht, falsche Definition!

Zu Markov–Ketten bot die Seite `lehrerfortbildung-bw.de` bis Mai 2025 Arbeitsmaterial für den Vertiefungskurs Mathematik an. Dort stand:

> ***Definition:*** *Eine (endliche, homogene) Markov–Kette ist eine Folge $(X_n)_{n\geq 0}$ von Zufallsgrößen, für die gilt:*
> *(1) Die X_n nehmen Werte aus einer endlichen Menge an.*
> *(2) Zu jeweils zwei Zuständen i und j existiert $p_{ij} \in [0,1]$, so dass für jedes $n \in \mathbb{N}$ gilt: $\mathbf{P}(\, X_{n+1}{=}j \mid X_n{=}i\,) = p_{ij}$.*

Wie so oft wird auch hier schamlos durch 0 dividiert. Na gut, sei's drum, das könnte man noch reparieren. Ich beklage ein viel ernsteres Problem:

Aufgabe: Zeigen Sie durch Beispiele, dass diese Bedingung nicht genügt!

Hier werden zwei Sichtweisen unglücklich vermischt: Für die *Definition* einer Markov–Kette ist dies nicht genug. Für die *Konstruktion* benötigen wir, wie oben ausgeführt, die Startwkten q_i und die Übergangswkten p_{ij}. Zusammen mit der Pfadregeln können wir daraus unsere Markov–Kette $(\Omega, \mathscr{A}, \mathbf{P}, (X_t)_{t\in\mathbb{N}})$ bauen. Genau diese Pfadregel fehlt uns jedoch hier!

Markov–Ketten: Vorsicht, falsche Definition!

Wie zeigen wir, dass eine gegebene Erklärung unvollständig ist? Am besten durch ein konretes Gegenbeispiel, in dem sie fehlschlägt!

Beispiel: Wir betrachten die Zustandsmenge $S = \{0, 1\}$ und dazu den Raum $\Omega = S^{\mathbb{N}}$ aller Folgen mit Projektionen $X_n = \mathrm{pr}_n : \Omega \to S : \omega \mapsto \omega_n$.

Wie üblich nennen wir $\omega = (\omega_0, \omega_1, \omega_2, \ldots)$ eine **Fibonacci–Folge**, falls $\omega_{n+2} = \omega_{n+1} + \omega_n \mod 2$ für alle $n \in \mathbb{N}$ gilt. In Ω gibt es genau 4 solcher Folgen, je eine zu den Startwerten $(0, 0, \ldots)$, $(0, 1, \ldots)$, $(1, 0, \ldots)$, $(1, 1, \ldots)$.

Wir setzen $\mathbf{P}(\omega) = 1/4$, falls ω eine Fibonacci–Folge ist, sonst $\mathbf{P}(\omega) = 0$. Damit finden wir $\mathbf{P}(\, X_{n+1}{=}j \mid X_n{=}i \,) = 1/2$. Hingegen gilt

$$\mathbf{P}(\, X_{n+1}{=}j \mid X_n{=}i,\ X_{n-1}{=}h \,) = \begin{cases} 1 & \text{falls } j = i + h, \\ 0 & \text{falls } j \neq i + h. \end{cases}$$

Die Übergangswkten alleine genügen nicht! Die Markov–Eigenschaft ist unverzichtbar: Unter der Bedingung des gegenwärtigen Zustands X_n ist die Zukunft X_{n+1} unabhängig von der Vergangenheit $X_{n-1}, \ldots, X_0$.

Markov–Ketten: Bauanleitung

? Können wir Markov–Ketten nach unserem Belieben bauen? Ja!

Satz D4B: Zu vorgegebenen Daten existiert eine Markov–Kette.

Zu jedem $t \in \mathbb{N}$ sei S_t eine diskrete Zustandsmenge. Der Start sei gegeben als eine WVerteilung $q_0 : S_0 \to [0,1]$, mit $\sum_a q_0(a) = 1$, der Übergang von t zu $t+1$ durch $p_t : S_t \times S_{t+1} \to [0,1]$ mit $\sum_b p_t(a,b) = 1$ für alle $a \in S_t$.

Zu diesen Daten existiert eine Markov–Kette $(\Omega, \mathscr{A}, \mathbf{P}, (X_t)_{t\in\mathbb{N}})$ mit

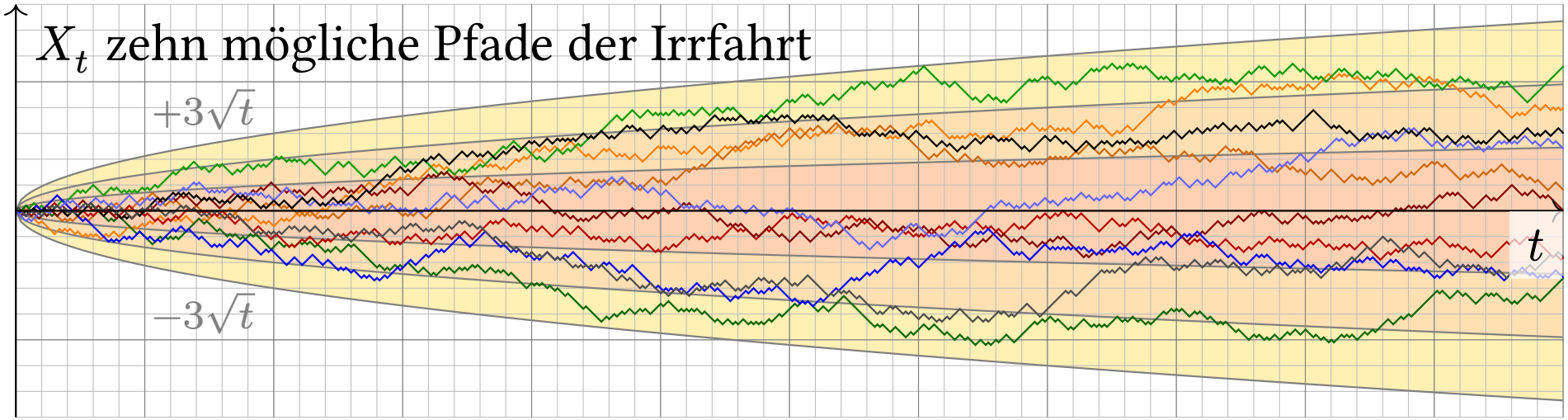
$$\boxed{\mathbf{P}(\, X_0{=}a_0,\, X_1{=}a_1,\, \dots,\, X_n{=}a_n\,) = q_0(a_0)\, p_0(a_0,a_1) \cdots p_{n-1}(a_{n-1},a_n)}$$

für alle Zeiten $n \in \mathbb{N}$ und alle Zustände $a_0 \in S_0$, $a_1 \in S_1$, ..., $a_n \in S_n$.

Übung: Wiederholen Sie die Konstruktion D402 eines geeigneten Modells $(\Omega, \mathscr{A}, \mathbf{P}, (X_t)_{t\in\mathbb{N}})$ und den Nachweis D406 seiner Markov–Eigenschaft.

Bemerkung: Dieser schöne Satz sichert die *Existenz* von Markov–Ketten. *Eindeutigkeit* können wir hingegen nicht erwarten. Im Gegenteil ist es oft vorteilhaft, den WRaum $(\Omega, \mathscr{A}, \mathbf{P})$ recht frei wählen zu können, etwa um weitere Eigenschaften zu erhalten oder Platz für zusätzliche ZVariablen.

Markov–Ketten: spektakuläre Anwendungen!



Wir betrachten erneut die **Irrfahrt auf $\mathbb{Z}$**, ein ebenso einfaches wie wichtiges Modell. Bei dieser Markov–Kette $(X_t)_{t\in\mathbb{N}}$ ist der Zustandsraum $\mathbb{Z}$, darauf $q_0(0) = 1$, somit $q_0(a) = 0$ für $a \neq 0$, sowie $p_t(a, a+1) = p_t(a, a-1) = 1/2$, sonst folglich $p_t(a, b) = 0$. Typische Anwendung: Kontostand bei zufälligen Gewinnen und Verlusten, etwa **Aktienkurse** (Louis Bachelier, 1900).

Ähnlich entsteht die **Brownsche Bewegung** durch Wärmebewegung. Der schottische Botaniker Robert Brown (1773–1858) entdeckte 1827 unter dem Mikroskop das unregelmäßige Zittern von Pollen in Wasser. Anfangs hielt er Pollen für belebt, doch er fand dasselbe bei Staubteilchen.

Albert Einstein (1879–1955) erklärte dies 1905 durch die ungeordnete **Wärmebewegung** der Wassermoleküle, die aus allen Richtungen in großer Zahl gegen die Pollen stoßen. Quantitativ konnte er so die Größe von Atomen bestimmen und die Anzahl pro Mol, die **Avogadro–Zahl**.

Experimentell bestätigte dies Jean Perrin (1870–1942): *Mouvement Brownien et Réalité Moléculaire* (1909). Für diesen Nachweis der Teilchennatur der Materie erhielt er 1926 den Physik-Nobelpreis.

Brownsche Bewegung aka Wiener–Prozess

Stochastische Prozesse $(X_t)_{t\in I}$ gibt es neben zeitdiskret mit $I = \mathbb{N}$ auch zeitkontinuierlich auf einem reellen Zeitintervall $I \subseteq \mathbb{R}$. Zudem muss der Zustandsraum wie $S \subseteq \mathbb{R}^n$ nicht diskret sein. Auch hier können wir die Markov–Eigenschaft untersuchen und nutzen. Berühmtestes Beispiel:

Definition D4c: Brownsche Bewegung 1827, Wiener–Prozess 1920

Ein **Wiener–Prozess** $(\Omega, \mathscr{A}, \mathbf{P}, (W_t)_{t\geq 0})$ besteht aus einem WRaum $(\Omega, \mathscr{A}, \mathbf{P})$ mit Zufallsvariablen $W_t : \Omega \to \mathbb{R}$ für $t \in \mathbb{R}_{\geq 0}$, für die gilt:

0. Standard: Als Startpunkt gilt $\mathbf{P}(W_0{=}0) = 1$, also $W_0 = 0$ fast-sicher.
1. Zu je endlich vielen Zeitpunkten $t_0 < t_1 < \ldots < t_n$ in $\mathbb{R}_{\geq 0}$ sind die Zuwächse $W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ stochastisch unabhängig.
2. Für alle $s < t$ in $\mathbb{R}_{\geq 0}$ gilt $W_t - W_s \sim \mathrm{N}(0, t-s)$. Die Zuwächse sind also stationär und normalverteilt mit Erwartungswert 0 und Varianz $t - s$.
3. Für fast-alle $\omega \in \Omega$ ist der Pfad $\gamma_\omega : \mathbb{R}_{\geq 0} \to \mathbb{R} : t \mapsto W_t(\omega)$ stetig.

☺ Schön kurz zusammengefasst: Ein Wiener–Prozess $(W_t)_{t\geq 0}$ ist reell, zeitstetig, standardisiert mit unabhängig normalverteilten Zuwächsen.

Brownsche Bewegung aka Wiener–Prozess

? Damit fordern wir vielleicht zuviel. Lässt sich auch dies noch erfüllen? Ja, Wiener–Prozesse $(\Omega, \mathscr{A}, \mathbf{P}, (W_t)_{t\geq 0})$ existieren tatsächlich! Genauer:

Satz D4D: das Wiener–Maß auf $\mathscr{C}(\mathbb{R}_{\geq 0}, \mathbb{R})$

Als Grundgesamtheit $\Omega := \mathscr{C}(\mathbb{R}_{\geq 0}, \mathbb{R})$ betrachten wir die Menge aller stetigen Funktionen $\omega : \mathbb{R}_{\geq 0} \to \mathbb{R} : t \mapsto \omega_t$. Zu jedem Zeitpunkt $t \in \mathbb{R}_{\geq 0}$ sei $X_t(\omega) := \omega_t$ die Auswertung in t. Die Familie $(X_t)_{t\geq 0}$ erzeugt eine σ–Algebra $\mathscr{A}$ auf Ω. Auf $(\Omega, \mathscr{A})$ existiert genau ein WMaß $\mathbf{P}$, genannt **Wiener–Maß**, das $(\Omega, \mathscr{A}, \mathbf{P}, (W_t)_{t\geq 0})$ zu einem Wiener–Prozess macht.

Besonders anschaulich ist Donskers Konstruktion: Wir betrachten eine Irrfahrt $(X_t)_{t\in\mathbb{N}}$ in $\mathbb{Z}$ mit unabhängigen gleichverteilten Zuwächsen ± 1. Diese interpolieren wir stückweise affin-linear zu dem stetigen Prozess $(X_t)_{t\geq 0}$. Der skalierte Prozess $(X_{nt}/\sqrt{n})_{t\geq 0}$ konvergiert in Verteilung für $n \to \infty$. Die Grenzverteilung $\mathbf{P}$ ist ein / das Wiener–Maß auf $\mathscr{C}(\mathbb{R}_{\geq 0}, \mathbb{R})$.

Die obige Skizze illustriert die Idee. Der Beweis mobilisiert die WTheorie, allen voran den zentralen Grenzwertsatz. Versuchen Sie es einmal selbst!

Satz D4E: zentraler Grenzwertsatz, ZGS

Sei $(\Omega, \mathscr{A}, \mathbf{P})$ ein WRaum und $X_1, X_2, X_3, \ldots : \Omega \to \mathbb{R}$ unabhängig mit

1. endlichen Erwartungen $\mu_k = \mathbf{E}(X_k)$ dank $\mathbf{E}(|X_k|) < \infty$,
2. strikt positiven Varianzen $\sigma_k^2 = \mathbf{E}(|X_k - \mu_k|^2) \geq \sigma_0^2 > 0$,
3. beschränkten dritten Momenten $\rho_k^3 = \mathbf{E}(|X_k - \mu_k|^3) \leq \rho_0^3 < \infty$.

Die Summe $S = X_1 + \ldots + X_n$ hat die Erwartung $\mu = \mu_1 + \ldots + \mu_n$ und die Varianz $\sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2$. Es gilt $\mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2)$, genauer

$$\mathbf{P}(a \leq S \leq b) = \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}} \, \mathrm{d}\xi + \delta,$$

$$|\delta| \leq \frac{\rho_1^3 + \ldots + \rho_n^3}{(\sigma_1^2 + \ldots + \sigma_n^2)^{3/2}} \leq \frac{\rho_0^3}{\sigma_0^3 \sqrt{n}} \searrow 0.$$

Zentriert und normiert zu $S^* := (S - \mu)/\sigma$ erhalten wir $\mathbf{P}_{S^*} \approx \mathrm{N}(0, 1)$.

⚠ Alle drei Voraussetzungen (1–3) werden wirklich benötigt.

Was sagt uns der zentrale Grenzwertsatz?

In Wissenschaft, Wirtschaft und Gesellschaft treten **zufällige Größen** X auf. Sie sind meist unvermeidbar, genauere Informationen sind oft zu teuer oder manchmal gar nicht zugänglich. Die Stochastik liefert Werkzeuge zum sicheren Umgang mit unsicheren Ereignissen, allgemein zum rationalen Entscheiden unter Unsicherheit. Dazu ist von zentraler Bedeutung, die **Verteilung** $\mathbf{P}_X$ zu kennen, oder zumindest so gut abzuschätzen wie es realistisch möglich oder nötig ist.

Oft ist unsere Zufallsgröße $S = X_1 + \ldots + X_n$ die Summe vieler unabhängiger Einflüsse $X_1, \ldots, X_n$, jeder einzelne ist klein, alle sind ähnlich groß. Der zentrale Grenzwertsatz (ZGS) besagt: Ihre Summe S ist annähernd normalverteilt, als Faustregel kurz $\mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2)$.

Beispiel: Für unabhängige und identisch verteilte 0–1–Zufallsvariablen $X_k : \Omega \to \{0, 1\}$ ist $S \sim \mathrm{B}(n, t)$ exakt binomialverteilt. Der lokale Grenzwertsatz (LGS) erklärt die Näherung $\mathrm{B}(n, t) \approx \mathrm{N}(\mu, \sigma^2)$ mit $\mu = nt$ und $\sigma^2 = nt(1-t)$ und garantiert Fehlerschranken.

Im allgemeineren ZGS dürfen $X_1, \ldots, X_n$ beliebig verteilt sein, unter milden Vorkehrungen:
(1) Die Beiträge $X_1, X_2, \ldots, X_n$ sollen unabhängig sein – nicht $S = X + X + \ldots + X$.
(2) Die Beiträge $X_1, X_2, \ldots, X_n$ sollen alle etwas streuen – nicht $S = 1 + 1 + \ldots + 1$.
(3) Die Beiträge $X_1, X_2, \ldots, X_n$ sollen „ähnlich groß“ sein – nicht $S = \sum_k 2^{-k} X_k$.

Den Namen prägte George Pólya: *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung* (1920). In der Mathematik geht man gerne zum **Grenzwert** über, um die Ergebnisse bequem und übersichtlich zu formulieren. Für praktische Anwendungen ist eine **Fehlerschranke** notwendig für den Wechsel von $\mathbf{P}_S$ zu $\mathrm{N}(\mu, \sigma^2)$. Satz D4E stellt hierzu alles bereit. Unsere Formulierung folgt A.C. Berry 1941 und C.G. Esséen 1944 und präzisiert diese Voraussetzungen in Form des 2. und 3. Moments und quantifiziert den Approximationsfehler durch eine explizite Schranke.

Lobeshymne auf den zentralen Grenzwertsatz

☺ Der zentrale Genzwertsatz ist ein phantastisches Ergebnis: allgemein gültige Näherung mit verlässlicher Fehlerabschätzung! Das ermöglicht die Modellierung, Berechnung und Erklärung vieler stochastischer Phänomene, die uns täglich umgeben. Implizit oder explizit wird er in praktisch allen naturwissenschaftlich-technischen Modellen oder industriellen Prozessen benutzt.

☺ Die Voraussetzungen des ZGS sind milde und oft erfüllt. Daher ist der ZGS nahezu überall anwendbar (anders als der LGS). Zunächst setzen wir die Unabhängigkeit der Zufallsvariablen $X_1, X_2, X_3, \ldots$ voraus, was für viele Anwendungen realistisch ist. (Man kann sogar dies noch weiter abschwächen, und ein kontrolliertes Maß an schwacher Abhängigkeit erlauben.)

☺ Die oben für den ZGS vorausgesetzten Schranken $\sigma_k \geq \sigma_0 > 0$ und $\rho_k \leq \rho_0 < \infty$ stellen sicher, dass alle Beiträge $X_1, X_2, X_3, \ldots$ ähnlich stark streuen. Andernfalls könnte zum Beispiel X_1 alle anderen Summanden $X_2, X_3, \ldots$ überwiegen, und in der Summe würde keine Normalverteilung entstehen, sondern im Wesentlichen die Verteilung von X_1.

☺ Eine allgegenwärtige Anwendung sind Konfidenzintervalle. Diese werden in Wissenschaft und Technik überall verwendet, um die Genauigkeit einer Messung oder Rechnung zu quantifizieren (Fehlerrechnung). Diese wichtige Information muss man als Teil unserer Gesellschaft mindestens passiv intuitiv verstehen, je nach Rolle auch aktiv präzise nutzen (Datenkompetenz).

☺ Sehr wenige Informationen reichen für den ZGS bereits aus! Für die Näherung genügen Erwartungen μ_k und Varianzen $\sigma_k^2 \geq \sigma_0^2 > 0$, für die Fehlerschranke die dritten Momente $\rho_k^3 \leq \rho_0^3 < \infty$, notfalls nach oben abgeschätzt. Das gelingt oft problemlos und effizient.

☺ Wir können den ZGS durch Fourier–Transformation verstehen / illustrieren / nachrechnen. Mit dem Werkzeugkasten der Analysis transformiert sich der Beweis in eine einfache Rechnung.

Lobeshymne auf den zentralen Grenzwertsatz

☺ Ein besonders wichtiger Spezialfall ist die unabhängige Wiederholung von Messungen bzw. Experimenten: Hier sind die Zufallsvariablen $X_1, \dots, X_n$ unabhängig und identisch verteilt, insbesondere gilt für ihre Kenngrößen $\mu_k = \mu_0$, $\sigma_k = \sigma_0$, $\rho_k = \rho_0$ für alle $k = 1, 2, \dots, n$. Zur Existenz dieser Integrale benötigen wir lediglich $X_k \in L^3(\Omega; \mathbb{R})$, denn $L^3 \supseteq L^2 \supseteq L^1$.

⚠ Ohne die geforderten Momente gilt der ZGS im Allgemeinen nicht, wie Gegenbeispiele zeigen. Die Cauchy–Verteilungen zum Beispiel erfüllen nicht die Voraussetzungen des ZGS und auch nicht seine Schlussfolgerung, denn sie konvergieren nicht gegen die Normalverteilung.

☺ Der ZGS liefert eine explizite Fehlerschranke $|\delta| \leq \rho_0^3/\sigma_0^3\sqrt{n}$. Der Grenzwertsatz in der Form $\mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2)$ oder besser $\mathbf{P}_S \to \mathrm{N}(\mu, \sigma^2)$ für $n \to \infty$ ist eine qualitative Aussage: Er garantiert die Konvergenz der Verteilungen, sagt aber streng genommen noch rein gar nichts über die Konvergenzgeschwindigkeit oder den Approximationsfehler für ein vorgegebenes n.

☺ Oft verwendet man den ZGS nicht für $n \to \infty$, sondern als $\mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2)$ für ein festes n. Für die konkrete Anwendung ist es dann wichtig zu wissen: Wie gut ist die Approximation? Obige Formulierung von Berry–Esséen ist eine quantitative Präzisierung: Sie besagt, wie schnell die Konvergenz ist (nämlich wie $1/\sqrt{n} \to 0$) und garantiert sogar explizite Fehlerschranken!

☺ Die Näherung ist oft besser als die pessimistische Fehlerschranke! Die im ZGS angegebene Fehlerschranke $|\delta| \leq \rho_0^3/\sigma_0^3\sqrt{n}$ ist ein bequemer erster Schritt. Sie gilt allgemein, daher muss sie vom schlimmsten ausgehen. Sie lässt sich allgemein nicht weiterverbessern, wie Gegenbeispiele belegen. In günstigen Fällen ist die Approximation tatsächlich besser, wie Beispiele illustrieren.

☺ Der lokale Grenzwertsatz ist bei Bedarf genauer: Dort geht es nur um Binomialverteilungen, dafür bietet der LGS nützliche Korrekturterme mit wesentlich schärferen Fehlerschranken.

Was ist Entropie?

The most misunderstood concept in physics

Veritasium, youtu.be/DxL2HoqLbyA

Hauptsätze der Thermodynamik

1) *Die Energie der Welt ist constant.*
2) *Die Entropie der Welt strebt einem Maximum zu.*

Rudolf Clausius (1822–1888), *Mechanische Wärmetheorie* (1865)

You may see a cup of tea fall off a table and break into pieces on the floor… But you will never see the cup gather itself back together and jump back on the table. The increase of disorder, or entropy, is what distinguishes the past from the future, giving a direction to time.

Stephen Hawking (1942–2018), *A Brief History of Time* (1988)

The fact that you can remember yesterday but not tomorrow is because of entropy. […] So I think that entropy is underappreciated as something that has a crucial role in how we go through life.

Sean M. Carroll (1966–), *The Passage of Time and the Meaning of Life*

youtu.be/KFXYCrqUTmI, ausführlich /EntaPoSEXog

Energie und Entropie in der Thermodynamik

Die **Wärmelehre** ist eine wissenschaftlich-technische Errungenschaft des 19. Jahrhunderts und prägt unsere hochtechnisierte Zivilisation.

(1) Die **technische Thermodynamik** beschreibt **makroskopische Phänomene** durch Temperatur, Wärme, Druck, Volumen, Arbeit, etc. und ihre Änderung in Prozessen (Hauptsätze, Zustandsgleichungen). Sie ist grundlegend zur Konstruktion und Analyse von Maschinen.

Ihre grobe Sichtweise (aka Abstraktion) ist didaktische Herausforderung und methodische Stärke; dadurch wird sie extrem vielseitig anwendbar auf große Klassen von Systemen und Prozessen.

(2) Die **statistische Physik** untersucht die **mikroskopischen Zustände** der Teilchen (Moleküle). Sie erklärt die makroskopischen Phänomene durch klassische Mechanik (Energie, Impuls, etc.) und dient so als theoretisches Fundament der Thermodynamik (Mikrofundierung).

Analogie in der Wirtschaftslehre: Die **Makroökonomik** untersucht übergeordnete Größen und Kennzahlen: Investition & Konsum, Export & Import, Staatsausgaben & Steuern, Geldpolitik, etc. Die **Mikroökonomik**, insbesondere die Spieltheorie, untersucht das Verhalten einzelner Akteure (Menschen, Haushalte, Unternehmen). Auch dies dient der Mikrofundierung, also dem detaillierten mikroökonomischen Verständnis der makroökonmisch beobachtbaren Phänomene.

Energie und Entropie in der Thermodynamik

0. Hauptsatz: Die Temperatur $T > 0$ eines Systems (im thermischen Gleichgewicht) beschreibt die mittlere Energie ungeordneter Bewegung.

Die Thermodynamik handelt von Temperatur, Wärme und Entropie. Allgemein beschreibt sie Vielteilchensysteme und hat daher eine fundamentale Stellung in Naturwissenschaft und Technik. Gegenüber Systemen mit wenigen Teilchen zeigen sich hier völlig neue Erscheinungen wie die Irreversibilität fast aller makroskopischen Prozesse. Die Entropie kennt nur eine Richtung!

1. Hauptsatz: Die Energie U bleibt erhalten gemäß $\mathrm{d}U = \mathrm{d}Q + \mathrm{d}W$.

Es gibt kein Perpetuum Mobile erster Art: Keine Maschine kann aus nichts Arbeit verrichten. Die Thermodynamik wirkt als Systemtheorie: Jedes Teilsystem wird als Black Box durch nur wenige Parameter beschrieben (Temperatur T, Druck p, Volumen V, etc.) und wechselwirkt durch Ströme (Energie $\mathrm{d}U$, Arbeit $\mathrm{d}W = p\,\mathrm{d}V +$ weitere, Wärme $\mathrm{d}Q$, Entropie $\mathrm{d}S$, etc.)

2. Hauptsatz: Die Entropie S wächst gemäß $\mathrm{d}S = \mathrm{d}Q/T \geq 0$.

In jedem abgeschlossenen System kann die Entropie nur anwachsen: $S_1 - S_0 = \int_{t_0}^{t_1} \mathrm{d}Q/T \geq 0$. Bestenfalls bleibt sie konstant, bei reversiblen Prozessen. Das hat weitreichende Konsequenzen:
Carnot: Wärmeenergie kann niemals vollständig in mechanische Arbeit umgewandelt werden.
Kelvin, Planck: Es gibt kein Perpetuum Mobile zweiter Art. Keine Maschine kann mechanische Arbeit verrichten allein durch Abkühlung eines Wärmespeichers.

3. Hauptsatz: Im idealen Grundzustand gilt $T = 0$ und $U = 0$ und $S = 0$.

Was ist Entropie? … in mikroskopischer Sicht?

Wir betrachten einen endlichen WRaum (Ω, P) mit $\Omega = \{\omega_1, \dots, \omega_n\}$ als M „Lostopf“, P „Zustandsraum“ oder I „Quelle zufälliger Nachrichten“.
Das Ergebnis ω_i hat Wkt $p_i \in [0,1]$ und Überraschung $-\log_2 p_i \in [0,\infty]$.

p	1	$1/2$	$1/4$	$1/8$	…	0
$-\log_2 p$	0	1	2	3	…	∞

Das misst den Informationsgehalt der Nachricht ω_i in Bit, zur Basis 2.
Jede andere Basis b skaliert die Information gemäß $\log_b p = \ln(p)/\ln(b)$.
Unabhängigkeit heißt Wkten multiplizieren, Überraschungen addieren.

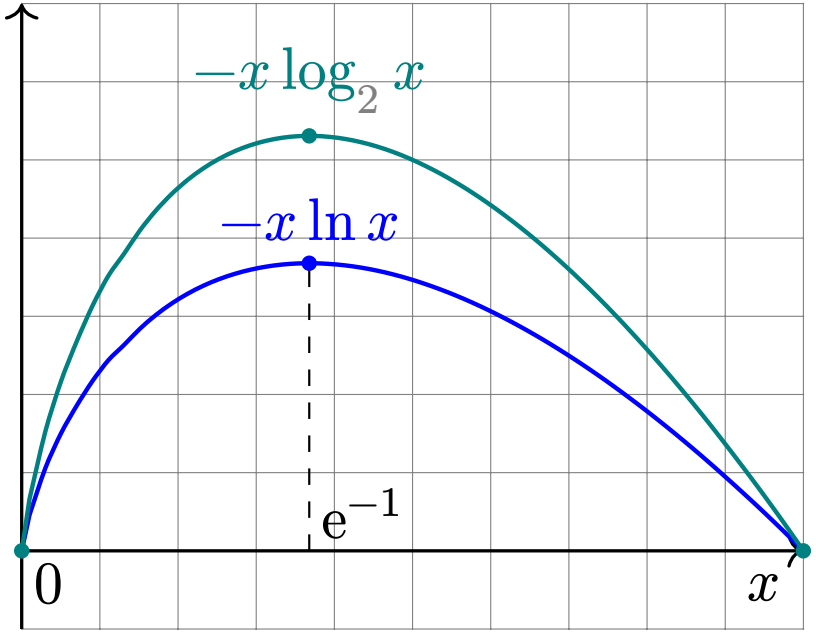
Die **Entropie** ist die erwartete Überraschung oder mittlere Information:

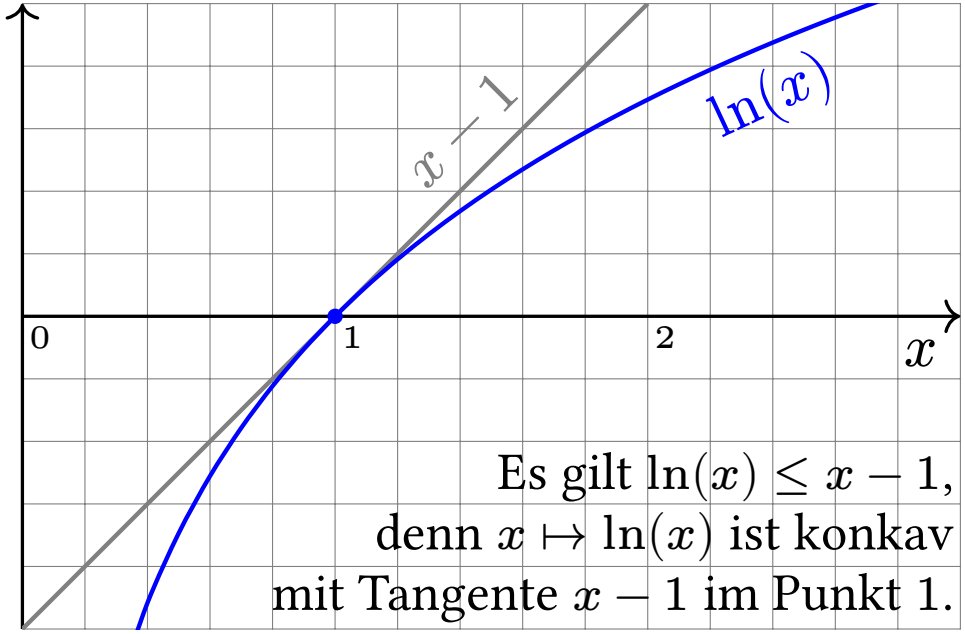
$$\mathrm{H}_2(P) := \mathbf{E}_P[\omega_i \mapsto -\log_2 p_i] = -\sum_{i=1}^{n} p_i \log_2 p_i \quad \in \mathbb{R}_{\geq 0}$$

Beispiel: Münzwurf $\Omega = \{\, 0 = \text{Kopf},\ 1 = \text{Zahl},\ 2 = \text{Rand} \,\}$
mit den Wkten $q = (\,0.50,\ 0.50,\ 0.00\,)$ und $p = (\,0.49,\ 0.49,\ 0.02\,)$.

$$\begin{aligned} \mathrm{H}(Q) &\overset{\text{Def}}{=} 0.50 \cdot 1.00 + 0.50 \cdot 1.00 + 0.00 \cdot \infty &&= 1.00 \\ \mathrm{H}(P) &\approx 0.49 \cdot 1.03 + 0.49 \cdot 1.03 + 0.02 \cdot 5.64 &&\approx 1.12. \end{aligned}$$

Eintropie nach Clausius, Boltzmann, Gibbs, Shannon, ...





Eingeführt wurde die Entropie 1865 von Rudolf Clausius im 2. Hauptsatz der Thermodynamik. Ludwig Boltzmann und Willard Gibbs erklärten sie 1875 in der statistischen Mechanik durch $S = -\mathrm{k} \sum_i p_i \ln(p_i)$ mit den Wkten p_i aller Mikrozustände i, skaliert durch die Boltzmann–Konstante $\mathrm{k} = 1.38 \times 10^{-23}$ J/K. Claude Shannon klärte 1948 optimale Codierung von Information durch Entropie und gründete die Informationstheorie.

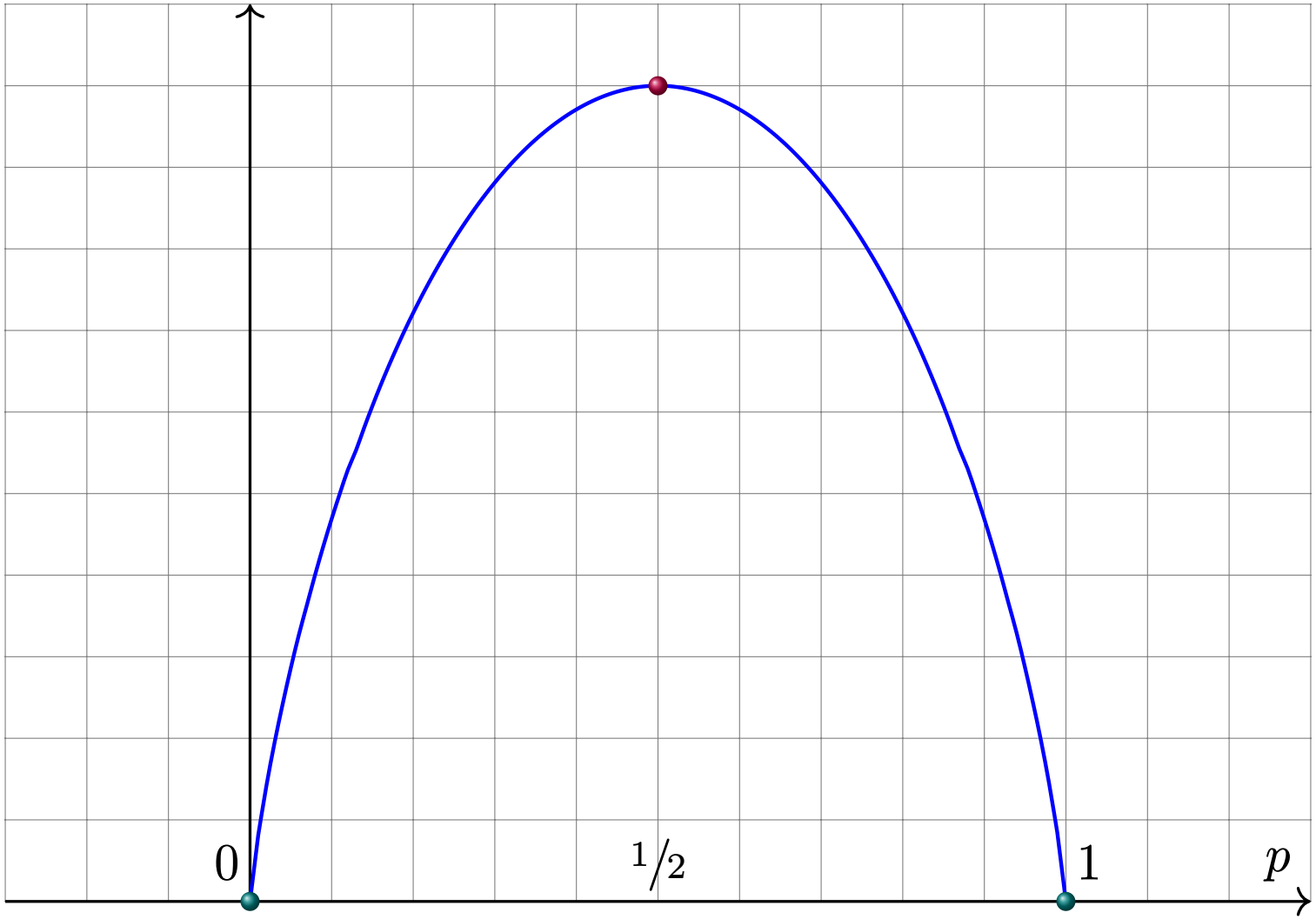
📖 Thomas M. Cover, Joy A. Thomas: *Information Theory.* Wiley 2006

Entropie spielt überall eine zentrale Rolle, von Quanten über Strömung zu schwarzen Löchern, von Kommunikation zu maschinellem Lernen.

Welche WVerteilung hat maximale Entropie?

Auf $\Omega = \{1, 2\}$ hat die WVerteilung $P := (p, 1-p)$ die Entropie

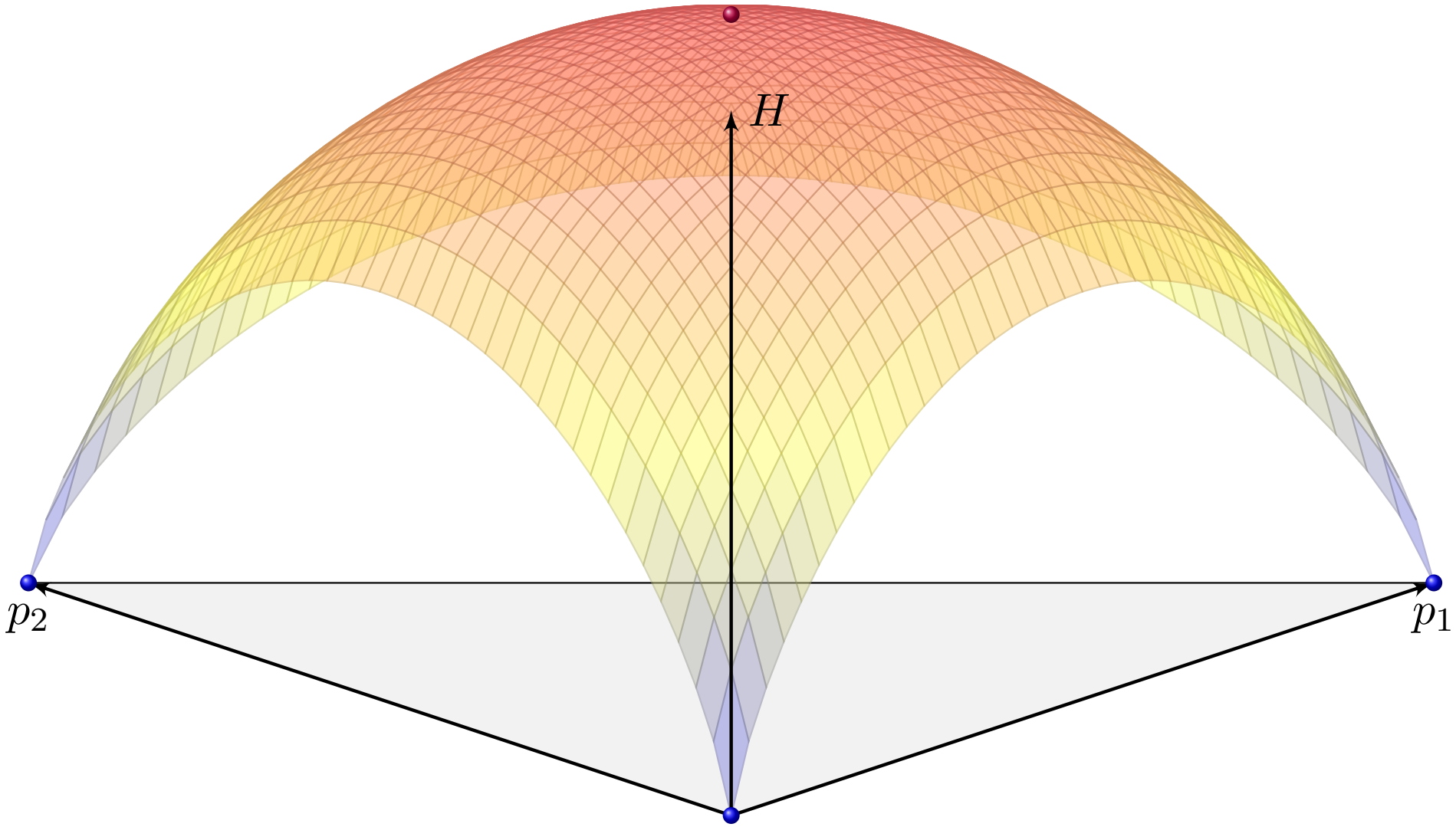
$$\mathrm{H}(P) = -p\log(p) - (1-p)\log(1-p).$$



Welche WVerteilung hat maximale Entropie?

Auf $\Omega = \{1, 2, 3\}$ hat die WVerteilung $P := (p_1, p_2, p_3)$ die Entropie

$$\mathrm{H}(P) = -p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3.$$



Die Kreuzentropie $\mathrm{H}(P, Q)$ misst die Überraschung.

Oft schätzen wir die wahre Verteilung P (engl. *ground truth*) durch ein Modell Q. Das Ergebnis ω_i hat Wkt p_i, aber die Überraschung $-\log q_i$ als unser Maß der Information. Erstere ist objektiv, also physikalisch real, zweitere subjektiv, eine Eigenschaft unseres Modells der Wirklichkeit.

$$\mathrm{H}(P, Q) := \underbrace{\mathbf{E}_P[\omega_i \mapsto -\log q_i]}_{\text{experimentell messbar}} = -\sum_{i=1}^{n} p_i \log q_i \quad \in [0, \infty]$$

Die **Kreuzentropie** von P und Q ist auch hier die „mittlere Information“ oder die „erwartete Überraschung“ des Modells Q in der Wirklichkeit P. Diese kann durch Stichprobe geschätzt werden (Monte Carlo Methode).

The Key Equation Behind Probability. youtu.be/KHVR587oW8I

Beispiel: Weiter sei $q = (\,0.50,\ 0.50,\ 0.00\,)$ und $p = (\,0.49,\ 0.49,\ 0.02\,)$.

$$\mathrm{H}(P, Q) \;=\; 0.49 \cdot 1.00 + 0.49 \cdot 1.00 + 0.02 \cdot \infty \;=\; \infty$$

Ist das Modell $q' = (\,0.46,\ 0.46,\ 0.08\,)$ näher an der Realität?

$$\mathrm{H}(P, Q') \;\approx\; 0.49 \cdot 1.12 + 0.49 \cdot 1.12 + 0.02 \cdot 3.64 \;\approx\; 1.17$$

$$\mathrm{H}(P, P) \;\approx\; 0.49 \cdot 1.03 + 0.49 \cdot 1.03 + 0.02 \cdot 5.64 \;\approx\; 1.12$$

Die relative Entropie $\mathrm{D}(P,Q)$ misst die Abweichung.

Die **relative Entropie** des Modells Q bezüglich der Wirklichkeit P, oder **Kullback–Leibler–Divergenz**, ist die „zusätzliche Überraschung“:

$$\mathrm{D}(P,Q) \;:=\; \underbrace{\mathrm{H}(P,Q)}_{\text{exp. messbar}} - \underbrace{\mathrm{H}(P)}_{\text{konstant}} \;=\; \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \quad \in [0,\infty]$$

Divergenz: Gilt $\mathrm{D}(P,Q') < \mathrm{D}(P,Q)$, so weicht Q' weniger ab als Q.

Satz D4F: Gibbs' Informationsungleichung

Es gilt $\mathrm{D}(P,Q) \geq 0$, also $\mathrm{H}(P,Q) \geq \mathrm{H}(P)$, mit Gleichheit nur für $P = Q$.

Beweis: Für $x \in \mathbb{R}_{\geq 0}$ gilt $\ln(x) \leq x - 1$, mit Gleichheit gdw $x = 1$. Also:

$$-\textstyle\sum_i p_i \ln(q_i/p_i) \overset{(1)}{\geq} -\sum_i p_i(q_i/p_i - 1) \overset{(2)}{\geq} \sum_i p_i - q_i \overset{\text{Wkt}}{=} 0$$

In den ersten beiden Summen setzen wir $0 \cdot (\pm\infty) := 0$. Hier zählen also nur Indizes i mit $p_i > 0$. In der dritten Summe zählen alle Indizes i, auch mit $p_i = 0$, daher die Ungleichung (2). Gleichheit gilt genau dann, wenn $p_i = q_i$ für alle i gilt: Für $p_i > 0$ folgt dies aus (1), für $p_i = 0$ aus (2). QED

Maximale Entropie: die diskrete Gleichverteilung

☺ Auf $\Omega = \{1, \dots, n\}$ maximiert $Q := (1/n, \dots, 1/n)$ die Entropie:

Satz D4G: Die diskrete Gleichverteilung maximiert die Entropie.

Jedes WMaß P auf $\Omega = \{1, \dots, n\}$ erfüllt $0 \le \mathrm{H}(P) \le \log n$.
(1) Maximum: Genau dann gilt $\mathrm{H}(P) = \log n$, wenn P gleichverteilt ist.
(2) Minimum: Genau dann gilt $\mathrm{H}(P) = 0$, wenn $p_i = 1$ für ein $i \in \Omega$ gilt.

Beweis: (1) Dank Gibbs' Informationsungleichung $\mathrm{H}(P) \le \mathrm{H}(P, Q)$ gilt:

$$\mathrm{H}(P) \overset{\text{D4F}}{\le} \mathrm{H}(P, Q) \overset{\text{Def}}{=} -\sum_{i=1}^n p_i \log q_i \overset{Q}{=} \sum_{i=1}^n p_i \log n \overset{P}{=} \log n$$

Gleichheit gilt genau dann, wenn $p_i = q_i$ für alle $i \in \Omega$, also $P = Q$.

(2) Für alle $x \in [0, 1]$ gilt $-x \log x \ge 0$, also $\mathrm{H}(P) = -\sum_i p_i \log p_i \ge 0$.
Gleichheit $-x \log x = 0$ gilt nur für $x \in \{0, 1\}$. Aus $\mathrm{H}(P) = 0$ folgt also $p_1, \dots, p_n \in \{0, 1\}$. Mit $\sum_i p_i = 1$ folgt $p_i = 1$ für genau ein $i \in \Omega$. QED

Prinzip der Indifferenz (Laplace 1812): Wenn keine Gründe vorliegen, irgendein Ergebnis zu begünstigen, so sind alle gleich wahrscheinlich.
Dieses Modell maximiert die Symmetrie… und zugleich die Entropie!

Maximale Entropie: die diskrete Gleichverteilung

☺ Für $n = 2$ und $n = 3$ sehen wir das sehr schön in obigen Graphiken. Allgemein hilft uns hier die Informationsungleichung! Oder Lagrange:

Aufgabe: Maximieren Sie die Entropie $H(p_1, \dots, p_n) = \sum_{i=1}^n -p_i \ln p_i$ unter der Nebenbedingung $g(p) := p_1 + \dots + p_n = 1$ und $p_1, \dots, p_n \geq 0$.

Lösung: Wir erinnern uns an den Satz über Lagrange–Multiplikatoren und lösen $\nabla H = -\lambda \nabla g$. Aus $-\ln p_i - 1 = -\lambda$ folgt erneut $p_i = q_i := {}^1\!/_n$.

Ist dies das globale Maximum? Der Definitionsbereich Δ ist kompakt, hierauf ist $H : \Delta \to \mathbb{R}$ stetig, nimmt also Minimum und Maximum an. Im Inneren Δ° ist q die einzige kritische Stelle, mit Wert $H(q) = \ln(n)$. Auf dem Rand $\partial\Delta$ gilt $H(\partial\Delta) = [0, \ln(n-1)]$ per Induktion über n. Also gilt $H(\Delta) = [0, \ln(n)]$ mit q als einziger Maximalstelle. Raffiniert!

☺ Gibbs' Ungleichung / Konkavität erledigt alles auf einen Schlag!

? Ist dieses Ergebnis nicht offensichtlich? Nein, erst im Rückblick!

☺ Das Maximum zu bestimmen, übersichtlich und nachvollziehbar, ist nicht trivial, insbesondere in höherer Dimension. Es ist eine erste, effiziente und elegante Anwendung von Gibbs' Informationsungleichung.

Maximale Entropie: die Boltzmann–Verteilung

Ein Teilchen gerät in Zustand $\omega_i \in \Omega$ mit Energie $u : \Omega \to \mathbb{R} : \omega_i \mapsto u_i$.
Zustände hoher Energie sind unwahrscheinlicher, aber wie genau?

Zur Wkt macht Boltzmann den Exponentialansatz $q_i = \exp(-\alpha - \beta u_i)$. Hier ist $\beta := 1/(\mathrm{k}T)$ mit Temperatur $T > 0$ und Boltzmann–Konstante $\mathrm{k} = 1.38 \times 10^{-23}$ J/K. Normierung: Wir bestimmen die Zustandssumme $Z := \sum_i \exp(-\beta u_i) > 0$ und erhalten $\alpha := \ln Z \in \mathbb{R}$; damit gilt $\sum_i q_i = 1$.

Sie hat Energie $\mathrm{E}(Q) = \sum_i q_i u_i$ und Entropie $\mathrm{S}(Q) = \mathrm{k} \sum_i q_i(\alpha + \beta u_i)$.
Die Shannon–Entropie H skalieren wir zur Gibbs–Entropie $\mathrm{S} = \mathrm{k} \ln(2)\, \mathrm{H}$.

Satz D4H: Boltzmann maximiert die Entropie.

Jedes WMaß P auf Ω mit der Energieschranke $\mathrm{E}(P) \le \mathrm{E}(Q)$ erfüllt auch die Entropieschranke $\mathrm{S}(P) \le \mathrm{S}(Q)$, und Gleichheit gilt nur für $P = Q$.

Beweis: Dank Gibbs' Informationsungleichung $\mathrm{S}(P) \le \mathrm{S}(P, Q)$ gilt:

$$\mathrm{S}(P) \overset{\text{D4F}}{\le} \mathrm{k} \sum_i p_i(\alpha + \beta u_i) \overset{\text{Lin}}{=} \mathrm{k}[\alpha + \beta \mathrm{E}(P)] \overset{\text{Vor}}{\le} \mathrm{k}[\alpha + \beta \mathrm{E}(Q)]$$

Für $P = Q$ gilt Gleichheit, sonst die strikte Ungleichung „<“. QED

Maximale Entropie: geometrische Verteilung

Beispiele: (1) Bei n Zuständen $\omega_1, \dots, \omega_n$ gleicher Energie $u_i = \text{const}$ erhalten wir erneut die Gleichverteilung $q_i = 1/n$ wie oben in Satz D4G.

(2) Auf $\Omega = \mathbb{N}$ ist die geometrische Verteilung $q_i = (1-q)q^i$ mit $q \in [0,1[$ die Boltzmann–Verteilung zu $u_i = i$ und $\beta = -\ln q$ und $\alpha = -\ln(1-q)$. Sie hat Erwartung / Energie $\mathrm{E}(Q) = \sum_{i=0}^{\infty} q_i \cdot i = q/(1-q)$ und Entropie

$$\begin{aligned} \mathrm{H}(Q) &= -\textstyle\sum_{i=0}^{\infty} q_i \big[\ln(1-q) + i\ln(q)\big] \\ &= -\ln(1-q) - \ln(q) \cdot q/(1-q). \end{aligned}$$

Übung: Jedes WMaß P auf $\mathbb{N}$ mit Erwartung $\mathrm{E}(P) \le \mathrm{E}(Q) = q/(1-q)$ erfüllt die Entropieschranke $\mathrm{H}(P) \le \mathrm{H}(Q)$, mit Gleichheit nur für $P = Q$.

☺ Unter allen WMaßen P auf $\Omega = \mathbb{N}$ mit Erwartung $\le \mu < \infty$ maximiert die obige geometrische Verteilung Q mit $\mu = q/(1-q)$ die Entropie; dabei gilt $q = \mu/(1+\mu)$ und $\mathrm{H}(Q) = (1+\mu)\ln(1+\mu) - \mu\ln\mu$.

Methode der maximalen Entropie (Jaynes 1957): Erlauben die Daten mehrere Verteilungen, so wählen wir diejenige mit maximaler Entropie.

Jaynes' Methode der maximalen Entropie

Wir suchen WVerteilungen mit vorgegebenen Eigenschaften, wie z.B. Erwartung und Varianz, und wählen diejenige mit maximaler Entropie. Jede mit kleinerer Entropie enthält mehr Information / Struktur als unsere Messdaten stützen. Beispiel: Laplace' Prinzip der Indifferenz!

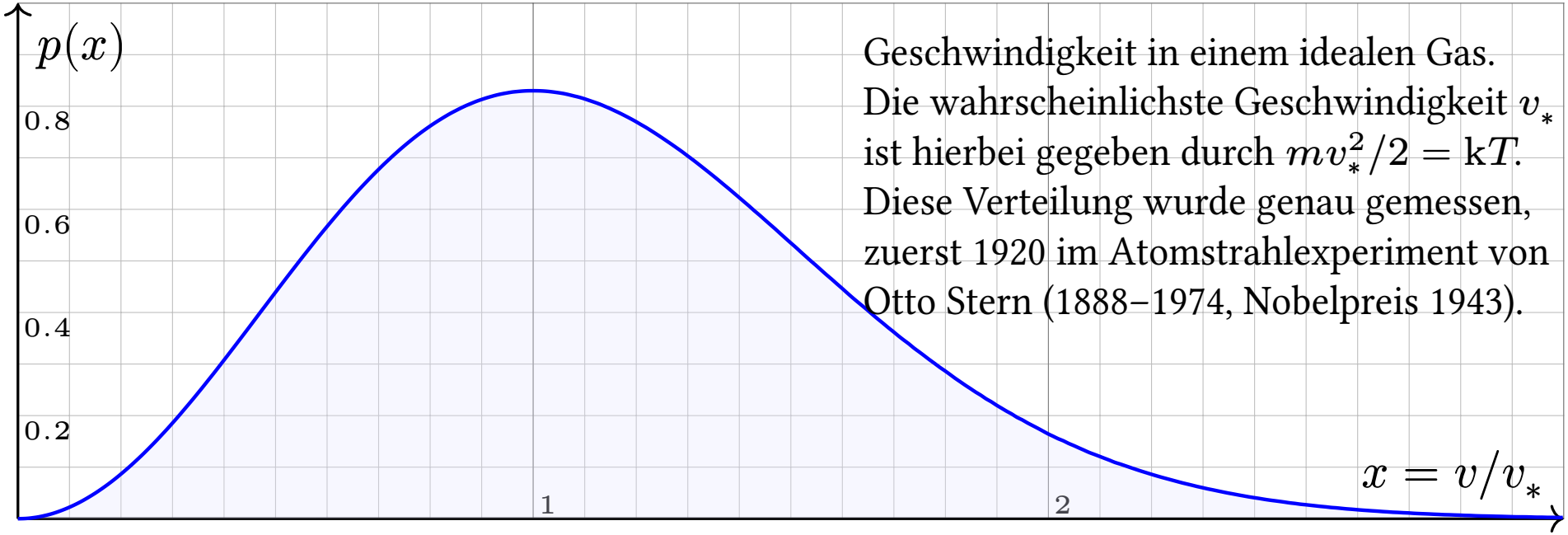
Aufgabe: Ein fairer Würfel $\Omega = \{1, \dots, 6\}$ mit $P = (1/6, \dots, 1/6)$ hat die erwartete Augenzahl $\mu = 3.5$. Sie messen $\mu_1 = 4.7$, andermal $\mu_2 = 2.8$. Welches Modellmaß Q auf Ω wählen Sie? Die Frage ist unterbestimmt!

Lösung: Gemäß der Methode der maximalen Entropie macht Jaynes den Ansatz $q_i = \exp(-\alpha - \beta i)$ mit $\alpha, \beta \in \mathbb{R}$. Numerisch bestimmen wir den Parameter $\beta \in \mathbb{R}$ mit $\mu(\beta) = (\sum_{i=1}^6 \mathrm{e}^{-\beta i} \cdot i)/(\sum_{i=1}^6 \mathrm{e}^{-\beta i})$ und finden $\beta_1 \approx -0.463$ und $Q_1 \approx (0.039, 0.062, 0.098, 0.157, 0.249, 0.395)$ ebenso $\beta_2 \approx +0.249$ und $Q_2 \approx (0.284, 0.221, 0.172, 0.134, 0.104, 0.081)$. *Whaaa?*

Für Jaynes war dies zuerst eine Frage der Logik, dann erst der Physik. Die statistische Mechanik liefert überzeugende weitere Argumente und begründet so den 2. Hauptsatz: Jedes physikalische System strebt zum thermischen Gleichgewicht, das heißt zum Zustand maximaler Entropie.

Maxwell–Verteilung der Geschwindigkeit im idealen Gas

Zustände hoher Energie werden seltener angenommen. Mit dem obigen Exponentialansatz konnten Boltzmann und Gibbs dies quantifizieren!



Geschwindigkeit in einem idealen Gas. Die wahrscheinlichste Geschwindigkeit v_* ist hierbei gegeben durch $mv_*^2/2 = \mathrm{k}T$. Diese Verteilung wurde genau gemessen, zuerst 1920 im Atomstrahlexperiment von Otto Stern (1888–1974, Nobelpreis 1943).

In einem idealen Gas ist die Teilchenenergie rein kinetisch, $E = mv^2/2$. Im thermischen Gleichgewicht ist die Wkt proportional zu $\exp(-E/\mathrm{k}T)$.

In 3 Dimensionen gilt $v^2 = v_x^2 + v_y^2 + v_y^2$. Die Zustände gleicher Energie füllen also die (Energie-)Kugelschale $[v, v + \mathrm{d}v]$ vom Volumen $4\pi v^2\,\mathrm{d}v$.

Für $x = v/v_*$ mit $mv_*^2/2 = \mathrm{k}T$ ist die WDichte $p(x) = \mathrm{e}^{-x^2} \cdot x^2 \cdot 4/\sqrt{\pi}$.

Kontinuierliche WVerteilungen auf $\Omega \subseteq \mathbb{R}^n$ und ihre Entropie

Auf $\Omega \subseteq \mathbb{R}^n$ sei P ein WMaß mit Dichte $p : \Omega \to \mathbb{R}_{\geq 0}$ und $\int_\Omega p(x)\,\mathrm{d}x = 1$. Ihre (natürliche) **Entropie** definieren wir (nun zur Basis e) durch

$$\mathrm{H}(P) := -\int_{x\in\Omega} p(x) \ln p(x)\,\mathrm{d}x \quad \in \bar{\mathbb{R}}$$

gemessen in Nat (*natural unit of information*), alternativ Bit zur Basis 2. Die **Kreuzentropie** kontinuierlicher WVerteilungen P und Q ist demnach

$$\mathrm{H}(P,Q) := -\int_{x\in\Omega} p(x) \ln q(x)\,\mathrm{d}x \quad \in \bar{\mathbb{R}}.$$

Die **relative Entropie** des Modells Q bezüglich der Wirklichkeit P ist

$$\mathrm{D}(P,Q) \;:=\; \mathrm{H}(P,Q) - \mathrm{H}(P) \;=\; \int_{x\in\Omega} p(x) \ln \frac{p(x)}{q(x)}\,\mathrm{d}x \quad \in \bar{\mathbb{R}}.$$

Aufgabe: Weiterhin gilt Gibbs' Informationsungleichung (Satz D4F), also $\mathrm{D}(P,Q) \geq 0$, somit $\mathrm{H}(P,Q) \geq \mathrm{H}(P)$. Gleichheit gilt genau dann, wenn $p(x) = q(x)$ für fast alle $x \in \Omega$ gilt, somit $P = Q$. Sind die Dichten $p, q : \Omega \to \mathbb{R}_{\geq 0}$ sogar stetig, so gilt $p(x) = q(x)$ in jedem Punkt $x \in \Omega$.

Maximale Entropie: kontinuierliche Gleichverteilung

Auf $\Omega \subseteq \mathbb{R}^n$ mit $0 < \operatorname{vol}_n(\Omega) < \infty$ hat die Gleichverteilung Q die Dichte

$$q : \Omega \to \mathbb{R}_{\geq 0} : q(x) = 1/\operatorname{vol}_n(\Omega).$$

Satz D41: Die Gleichverteilung maximiert die Entropie.

(0) Für jedes kontinuierliche WMaß P auf Ω gilt die Schranke

$$\mathrm{H}(P) \leq \mathrm{H}(Q) = \ln \operatorname{vol}_n(\Omega).$$

Gleichheit gilt genau für $P = Q$, die Gleichverteilung auf Ω.

Beweis: Dank Gibbs' Informationsungleichung $\mathrm{H}(P) \leq \mathrm{H}(P, Q)$ gilt:

$$\mathrm{H}(P) \quad \leq \quad \int_{x \in \Omega} p(x) \ln \operatorname{vol}_n(\Omega) \, \mathrm{d}x \quad \overset{P}{=} \quad \ln \operatorname{vol}_n(\Omega)$$

Für $P = Q$ gilt Gleichheit, sonst die strikte Ungleichung „$<$". QED

Beispiel: Die Gleichverteilung auf $[0, 1]^n$ hat die Entropie $\ln 1 = 0$.

⚠ Die kontinuierliche Entropie kann negativ sein, die diskrete niemals.

Maximale Entropie: Exponentialverteilung $\mathrm{E}(\mu^{-1})$

Auf $\Omega = \mathbb{R}_{\geq 0}$ hat die Exponentialverteilung $Q := \mathrm{E}(\mu^{-1})$ die Dichte

$$q : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} : q(x) = \mathrm{e}^{-x/\mu} \cdot \mu^{-1}$$

mit Erwartung $\mu \in \mathbb{R}_{>0}$.Diese können wir wie folgt charakterisieren:

Satz D41: Die Exponentialverteilung maximiert die Entropie.

(1) Für jedes kontinuierliche WMaß P auf $\mathbb{R}_{\geq 0}$ mit Erwartung $\leq \mu$ gilt

$$\mathrm{H}(P) \leq \mathrm{H}(Q) = 1 + \ln \mu.$$

Gleichheit gilt genau für $P = Q$, die Exponentialverteilung $\mathrm{E}(\mu^{-1})$.

Beweis: Dank Gibbs' Informationsungleichung $\mathrm{H}(P) \leq \mathrm{H}(P, Q)$ gilt:

$$\mathrm{H}(P) \quad \leq \quad \int_{x=0}^{\infty} p(x) \left[\frac{x}{\mu} + \ln \mu \right] \mathrm{d}x \quad \overset{P}{\leq} \quad 1 + \ln \mu$$

Für $P = Q$ gilt Gleichheit, sonst die strikte Ungleichung „$<$“. $\boxed{\text{QED}}$

Beispiel: Die Verteilung $\mathrm{E}(1)$ hat die Dichte $x \mapsto \mathrm{e}^{-x}$ und die Entropie 1.

Maximale Entropie: Normalverteilung $\mathrm{N}(\mu, \sigma^2)$

Auf $\Omega = \mathbb{R}$ hat die Normalverteilung $Q := \mathrm{N}(\mu, \sigma^2)$ die Dichte

$$q : \mathbb{R} \to \mathbb{R}_{\geq 0} : q(x) = \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot \frac{1}{\sigma\sqrt{2\pi}}$$

mit Erwartung $\mu \in \mathbb{R}$ und Varianz $\sigma^2 \in \mathbb{R}_{>0}$. Entropie-Charakterisierung:

Satz D41: Die Normalverteilung maximiert die Entropie.

(2) Für jedes kont. WMaß P auf $\mathbb{R}$ mit Erwartung μ und Varianz $\leq \sigma^2$ gilt

$$\mathrm{H}(P) \leq \mathrm{H}(Q) = \ln\sqrt{2\pi\,\mathrm{e}} + \ln\sigma$$

Gleichheit gilt genau für $P = Q$, die Normalverteilung $\mathrm{N}(\mu, \sigma^2)$.

Beweis: Dank Gibbs' Informationsungleichung $\mathrm{H}(P) \leq \mathrm{H}(P, Q)$ gilt:

$$\mathrm{H}(P) \quad \leq \quad \int_{\mathbb{R}} p(x)\left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 + \ln(\sigma\sqrt{2\pi})\right]\mathrm{d}x \quad \overset{P}{\leq} \quad \frac{1}{2} + \ln(\sigma\sqrt{2\pi})$$

Für $P = Q$ gilt Gleichheit, sonst die strikte Ungleichung „$<$". $\boxed{\text{QED}}$

Beispiel: Die Normalverteilung $\mathrm{N}(0, 1)$ hat Entropie $\ln\sqrt{2\pi\,\mathrm{e}} \approx 1.419$.

Entropie im zentralen Grenzwertsatz (ZGS)

Die obigen Rechnungen zeigen eindrucksvolle Beispiele und lassen die große Bedeutung der Entropie erahnen. Als Ausblick skizziere ich drei Anwendungen: Grenzwertsatz, maschinelles Lernen, Thermodynamik.

Gegeben seien unabhängige ZVariablen $X_1, X_2, \ldots : \Omega \to \mathbb{R}$.
Die Summe $S_n = X_1 + \ldots + X_n$ ist annähernd normalverteilt,
also $S_n \approx \mathrm{N}(\mu_n, \sigma_n^2)$ mit $\mu_n = \sum_{i=1}^n \mathbf{E}(X_i)$ und $\sigma_n^2 = \sum_{i=1}^n \mathbf{V}(X_i)$.
Zentriert und normiert erhalten wir $S_n^* := (S_n - \mu_n)/\sigma_n \approx \mathrm{N}(0,1)$.
Die Entropie strebt für $n \to \infty$ zum Maximum, $\mathrm{H}(S_n^*) \nearrow \mathrm{H}(\mathrm{N}(0,1))$.
Die Entropie treibt so die Konvergenz der Verteilungen, $S_n^* \to \mathrm{N}(0,1)$.

Physikalische Anschauung: Die vielen unabhängigen Summanden in S_n^* sorgen für hohe Entropie. Mit wachsendem n nähert sich die Entropie dem Maximum… der Normalverteilung, wie in Satz D41 gesehen.

Oft wird der ZGS durch Fourier–Transformation bewiesen. Das ist genial, aber auch etwas unintuitiv. Das Entropie-Argument hingegen entspricht direkt unserer physikalischen Anschauung aus dem 2. Hauptsatz der Thermodynamik, ist in diesem Sinne also intuitiv näherliegend.

Entropie im zentralen Grenzwertsatz (ZGS)

☺ Mit diesem letzten Höhepunkt endet dieses Kapitel und insgesamt unsere kurze Einführung in die wunderbare Welt der Wahrscheinlichkeit. Unser Weg war lang, auch anstrengend, doch er hat sich wieder gelohnt!

Unsere Beispiele illustrieren eindrücklich: Wahrscheinlichkeit begegnet Ihnen überall, ob Sie wollen oder nicht, in alltäglichen Entscheidungen ebenso wie in Mathematik, Physik und technischen Anwendungen bis hin zu gesellschaftlichen und ökonomischen Belangen.

Unser Ziel sind rationale Entscheidungen unter Unsicherheit.
Die Erfahrung zeigt: Auch mit dem Zufall können wir rechnen!
Wir wollen nachvollziehbar begründete, quantitative Aussagen.
Die Mathematik stellt Ihnen dafür alle Werkzeuge bereit.

Das gilt allgemein für jede physikalisch-technische Betrachtung und insbesondere für raffinierte, fortgeschrittene Konzepte wie die Entropie. Damit können Sie nun der vagen Idee vom „Maß für die Unordnung“ einen präzisen mathematischen Sinn geben.

Divergenz und Entropie im maschinellen Lernen (ML)

Shannon verband 1948 die fundamentalen Begriffe „Information“ und „Entropie“. Jüngst feiert dies im maschinellen Lernen große Erfolge:

Wie kann ein Computer aus Daten lernen? Die Grundidee ist einfach: Die Wirklichkeit entspricht einem WMaß P (*ground truth*). Das Training anhand von Beispieldaten erzeugt ein approximatives WMaß Q (*model*). Wie gut ist es? Die Divergenz $\mathrm{D}(P,Q) = \mathrm{H}(P,Q) - \mathrm{H}(P) \geq 0$ misst die Abweichung des Modells Q von der Wirklichkeit P. Diese Zielgröße wollen wir minimieren: Die Entropie misst den Lernfortschritt!

Problem: Leider kennen wir die Entropie $\mathrm{H}(P)$ der Wirklichkeit P nicht. Geniale Lösung: Es genügt, die Kreuzentropie $\mathrm{H}(P,Q)$ zu minimieren. Sie kann durch Stichprobe geschätzt werden (Monte Carlo Methode). So kann die Güte verbessert und durch Training optimiert werden.

Aus diesem Grund ist die Kreuzentropie $\mathrm{H}(P,Q)$ die zentrale Zielgröße im maschinellen Lernen (ML). Die Erfolge der generativen künstlichen Intelligenz (genAI) beruhen maßgeblich auf dieser fundamentalen Idee und ihrer effizienten Implementierung für gigantisch große Modelle.

Zum guten Schluss

So jung und mit so geringer Entropie
kommen wir nie wieder zusammen.

Tocotronic, `youtu.be/CpqeCZB3sMQ`

In Ihrem Leben, Studium und Karriere, werden diese zentralen Themen eine Rolle spielen. Bewährte Techniken Ihres Studiums werden erweitert durch Künstliche Intelligenz, demnächst Quantum Computing, zukünftig weitere Ideen, die derzeit noch niemand ahnt. Bleiben Sie neugierig!

Ihre mathematische Ausbildung wird Ihnen überall gute Dienste leisten.

Ich wünsche Ihnen, die hier dargelegten Werkzeuge bieten Ihnen eine gute Grundlage im Studium und einen Einstieg zur weiteren Vertiefung. Ich hoffe, damit bleiben die vielen wichtigen Anwendungen nicht „most misunderstood“, sondern führen Sie zu einem besseren Verständnis.

Habe Mut, dich deines eigenen Verstandes zu bedienen!

Much to learn, you still have.
This is just the beginning.