Gesetze der großen Zahlen

Anton Klimovsky

Grenzwertsätze für die Summen der ZV. Schwaches Gesetz der großen Zahlen. Konvergenz in Wahrscheinlichkeit (Stochastische Konvergenz). Starkes Gesetz der großen Zahlen. Fast sichere Konvergenz. Indikatorfunktion. Ergodensatz (= Gesetz der großen Zahlen für Markov-Ketten).

Grenzwertsätze spielen in der Stochastik eine zentrale Rolle. Oft besagen solche Sätze, dass viel von scheinbarem Zufall oft viel *Struktur*[1] zeigt.

Hauptsächlich werden wir uns mit den Grenzwertsätzen für die Summen der ZVen beschäftigen. Dabei kann man denken an den Gesamteinfluss (die Summe), der additiv aus vielen zufälligen Einflüssen (Summanden) besteht.

[1] Und diese kann man geschickt ausnutzen.

Summen der ZVen

Sei $\{X_i\}_{i=1}^{\infty}$ eine Folge der ZVen auf einem WRaum $(\Omega, \mathcal{A}, \mathbb{P})$. Betrachte die partielle Summen:

$$S_n := X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i. \tag{1}$$

Solche Summen tauchen in vielen Kontexten auf. Vergleiche die Irrfahrt.

Definition 0.1 (Empirischer Mittelwert)**.** *Der empirische Mittelwert ist die folgende ZV*

$$\bar{X}_n := \frac{1}{n} S_n. \tag{2}$$

Ist $\mathbb{E}[X_1] = \mathbb{E}[X_2] = ... = \mathbb{E}[X_n] = \mu \leq \infty$, so gilt

$$\mathbb{E}[\bar{X}_n] = \mu. \tag{3}$$

Frage: Wie verhält sich S_n für n groß? Kann man erwarten, dass[2]

$$S_n \approx \mu \cdot n \tag{4}$$

ist? Wie groß sind die Schwankungen von S_n?

[2] in einem geeignetem Sinne

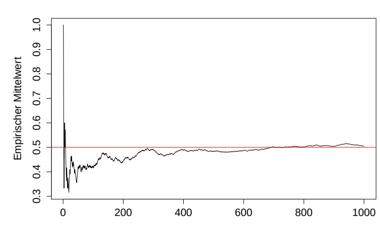

Abbildung 1: Die Grafik einer Realisierung des empirischen Mittelwerts $\bar{X}_n(\omega)$ in Abhängigkeit von der Anzahl der Münzwurfen n. (Dies ist eine Simulation des Bernoulli-Experiments mit der Erfolgswahrscheinlichkeit $0,5$.)

Eine Antwort liefern die Grenzwertsätze für die Summen der ZVen.

Grenzwertsätze sind nützlich weil:

- Sie eine approximative Analyse der ZV S_n erlauben.
- Sie eine wichtige Rolle in der Statistik spielen, sobald man genügend Daten[3] hat.

[3] D.h. n groß genug in (1).

Schwaches Gesetz der großen Zahlen

Wir erinnern uns an das folgende Theorem aus der ersten Hälfte der Vorlesung.

Theorem 0.1 (Schwaches Gesetz der großen Zahlen)**.** [4]

Sei $\{X_i\}_{i=1}^{\infty}$ eine Folge der unabhängigen identisch-verteilten[5] Zufallsvariablen[6] auf dem Wahrscheinlichkeitsraum[7] $(\Omega, \mathcal{A}, \mathbb{P})$ mit $\mathbb{E}\left[X_1\right] = \mu < \infty$.

Dann gilt

$$\boxed{\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_n - \mu\right| > \varepsilon\right\} = 0 \quad \textit{für alle } \epsilon > 0.} \tag{5}$$

[4] = schwaches GGZ wurde von schweizerischen Mathematiker Jakob Bernoulli in seiner Arbeit "Ars Conjectandi" (von lat. Kunst des Vermutens, also Stochastik) bewiesen.

Abbildung 2: Jakob Bernoulli (1655-1705)

[5] = u.i.v.

[6] = ZV

[7] = WRaum

In Worten bedeutet Theorem 0.1 Folgendes. Unter den Voraussetzungen von Theorem 0.1 konzentriert sich die Verteilung des *empirischen Mittelwerts* stark in der ε-Umgebung um den *theoretischen Erwartungswert* μ. Sie tut dies für alle $\varepsilon > 0$.

Definition 0.2 (Konvergenz in Wahrscheinlichkeit (= stochastische Konvergenz))**.** *Sei $\{Y_i\}_{i=1}^{\infty}$ eine Folge der ZV und Y_∞ eine ZV auf dem WRaum $(\Omega, \mathcal{A}, \mathbb{P})$. Die Folge $\{Y_i\}_{i=1}^{\infty}$ konvergiert in Wahrscheinlichkeit[8], falls*

$$\boxed{\forall\varepsilon > 0\colon \lim_{n\to\infty} \mathbb{P}\{|Y_n - Y_\infty| > \varepsilon\} = 0}. \tag{6}$$

[8] alternativ sagt man: "konvergiert stochastisch"

Notation. Falls (6) gilt, schreibt man

$$Y_n \xrightarrow[n\to\infty]{\mathbb{P}} Y_\infty. \tag{7}$$

Eine recht nützliche ZV[9] ist wie folgt definiert.

[9] Sie erlaubt uns Aussagen über Ereignisse mit den Aussagen über ZVen kompatibel zu machen.

Definition 0.3 (Indikatorfunktion)**.** *Sei $A \subset \Omega$, wobei $(\Omega, \mathcal{A}, \mathbb{P})$ ein WRaum ist. Die Funktion $\mathbb{1}_A\colon \Omega \to \mathbb{R}$ heißt **Indikatorfunktion** des Ereignisses A, falls*

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \in \Omega \setminus A. \end{cases} \tag{8}$$

Bemerkung. Die Indikatorfunktion ist eine handliche ZV. Man kann z.B. leicht den Erwartungswert für sie ausrechnen:

$$\mathbb{E}[\mathbb{1}_A] = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(\Omega \setminus A) = \mathbb{P}(A). \tag{9}$$

Example 0.1 (Wahrscheinlichkeiten und Häufigkeiten)**.** *Sei $A \in \mathcal{A}$ ein Ereignis im Kontext eines zufälligen Experiments. Sei $p := \mathbb{P}(A)$ die*

Wahrscheinlichkeit dieses Ereignisses. Betrachte eine Folge aus n unabhängigen Wiederholungen des Experiments $A_1, A_2, \ldots, A_n \in \mathcal{A}$ *und sei* $\bar{X}_n$ *der Zeitanteil, in dem das Ereignis A vorkommt. Es gilt*

$$\bar{X}_n = \frac{1}{n}\left(\mathbb{1}_{A_1} + \mathbb{1}_{A_2} + \ldots + \mathbb{1}_{A_n}\right). \tag{10}$$

Dank (9) *gilt* $\mathbb{E}[\mathbb{1}_{A_i}] = p$. *Das schwache Gesetz der großen Zahlen ist anwendbar und zeigt, dass für großes n der empirische Mittelwert sehr wahrscheinlich in der* ε*-Umgebung von p liegt. Sehe Abbildung 1.*

Grob gesagt erlaubt uns dies zu schließen, dass der empirische Mittelwert eine Abschätzung von p liefert[10].

Alternativ ist dies ein Schritt in die Richtung der Interpretation der Wahrscheinlichkeit von A als die Frequenz von A in einer Experimentenserie.

[10] Dies ist ein Link zur Statistik – ein Teil der Stochastik, der sich mit den empirischen Daten beschäftigt. Statistik entwickelt Methoden die (unter den Modelannahmen) einige Schlussfolgerungen aus den Daten ermöglichen, wie z.B. Abschätzungen für die unbekannte Parametern.

Ergodensatz

Für Markov-Ketten gilt GGZ auch und heißt Ergodensatz.[11]

[11] von griech. *ergon* ("Energie, Arbeit") + *odos* ("Pfad, Weg"). Diese (etwas unglückliche) Terminologie stammt von bedeutenden österreichischen physiker:

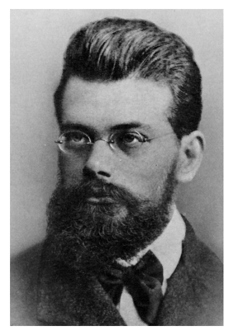

Abbildung 3: Ludwig Boltzmann (1844-1906)

Markov-Ketten sind Folgen der Abhängigen ZVen, deswegen ist das schwache GGZ für sie nicht direkt anwendbar!

Definition 0.4 (Ergodische Markov-Kette)**.** *Eine Markov-Kette* $\{X_i\}_{i=1}^{\infty}$ *mit einer Übergangsmatrix P heißt ergodisch, falls die entsprechende n-Schritt Übergangswahrscheinlichkeiten konvergieren gegen eine nicht-triviale Verteilung konvergieren*[12]*:*

$$\forall j\colon \lim_{n\to\infty} P_{i,j}^{(n)} = \pi_j > 0. \tag{11}$$

[12] Nicht-trivial heißt hier, dass für alle j: $\pi_j > 0$ gilt.

Theorem 0.2 (schwacher Ergodensatz = schwacher GGZ für Markov-Ketten)**.** *Sei* $\{X_i\}_{i=1}^{\infty}$ *eine ergodische Markov-Kette auf dem Zustandsraum S mit der Übergangsmatrix P. Definiere den empirischen Zeitanteil, den die Markov-Kette im Zustand* $A \subset S$ *nach* $n \in \mathbb{N}$ *Schritten verbracht hat:*

$$\nu_n(A) := \frac{1}{n}\left(\mathbb{1}_A(X_1) + \ldots + \mathbb{1}_A(X_n)\right), \quad A \subset S. \tag{12}$$

Dann gilt

$$\nu_n(A) \xrightarrow[n\to\infty]{\mathbb{P}} \pi(A). \tag{13}$$

Insbesondere impliziert (13)*, dass für alle beschränkte Funktionen* $f\colon S \to \mathbb{R}$ *die Konvergenz*

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow[n\to\infty]{\mathbb{P}} \sum_{s\in S} f(s)\pi\{s\}} \tag{14}$$

gilt.

NOTATION. Für die rechte Seite von (14) benutzt man die Notation $\pi[f]$.

IN WORTEN ist die rechte Seite von (14) der Erwartungswert von f bezüglich π. Die linke Seite von (14) ist ein *Mittel über die Zeitperiode* von $i = 1$ bis $i = n$ von den Beobachtungen[13] $f(X_i)$ unserer Markov-Kette.

[13] Eine Beobachtung ist eine Funktion $f \colon S \to \mathbb{R}$ auf dem Zustandsraum. Denn oft beobachtet man in den Experimenten nicht den Gesamtzustand vom System, sondern eine Funktion (ein Parameter) davon.

ERGODENSATZ BESAGT: "das *Zeitmittel* (linke Seite) konvergiert gegen das *Zustandsmittel* (rechte Seite)", s. (14).

EINE LANGE STICHPROBE unserer Markov-Kette (linke Seite) approximiert das Gleichgewicht (rechte Seite), s. (14).

MARKOV-KETTE ist eine Folge der *abhängigen* ZV. Deswegen ist Theorem 0.1 nicht direkt anwendbar!

Proposition 0.1 (Ergodische Markov-Ketten konvergieren geometrisch schnell ins Gleichgewicht)**.** *Sei* $\{X_n\}_{n=1}^{\infty}$ *eine ergodische Markov-Kette. Dann gilt*

$$\exists \rho \in (0,1) \colon |p_{i,j}^{(n)} - \pi_j| \le C\rho^n. \tag{15}$$

Proof of Theorem 0.2. Sei $s_i, s_j \in S$. Wir möchten zeigen, dass

$$\forall \varepsilon > 0 \colon \mathbb{P}\{|\nu_n\{j\} - \pi_j| > \varepsilon \mid X_1 = s_j\} \underset{n\to\infty}{\longrightarrow} 0. \tag{16}$$

Die Ungleichung von Tschebyscheff liefert

$$\mathbb{P}\{|\nu_n\{j\} - \pi_j| > \varepsilon X_1 = s_i\} \le \frac{\mathbb{E}[|\nu_n\{j\} - \pi_j|^2 \mid X_1 = s_i]}{\varepsilon^2} \tag{17}$$

Es bleibt z.Z., dass

$$\mathbb{E}[|\nu_n\{j\} - \pi_j|^2 \mid X_1 = s_i] \underset{n\to\infty}{\longrightarrow} 0. \tag{18}$$

In der Tat gilt

$$\begin{aligned} \mathbb{E}[|\nu_n\{j\} - \pi_j|^2 \mid X_1 = s_i] &= \frac{1}{n^2}\mathbb{E}\left[|\sum_{k=1}^{n} \mathbb{1}_{\{s_j\}}(X_k) - \pi_j|^2 \mid X_1 = s_i \right] \\ &= \frac{1}{n^2} \sum_{k=1}^{n}\sum_{l=1}^{n} m_{i,j}^{(k,l)}, \end{aligned} \tag{19}$$

wobei

$$\begin{aligned} m_{i,j}^{(k,l)} &= \mathbb{E}[\mathbb{1}_{s_j}(X_k)\mathbb{1}_{s_j}(X_l) \mid X_0 = s_i] - \pi_j \mathbb{E}[\mathbb{1}_{s_j}(X_k) \mid X_0 = s_i] \\ &\quad - \pi_j \mathbb{E}[\mathbb{1}_{s_j}(X_l) \mid X_0 = s_i] + \pi_j^2 \\ &= p_{i,j}^{(s)} p_{j,j}^{(t)} - \pi_j p_{i,j}^{(k)} - \pi_j p_{i.j}^{(l)} + \pi_j^2, \end{aligned} \tag{20}$$

wobei $s = \min\{k, l\}$ und $t = |k - l|$.

Nach Proposition 0.1 gilt

$$p_{i,j}^{(n)} = \pi_j + \varepsilon_{i,j}^{(n)} \quad |\varepsilon_{i.j}^{(n)}| \leq C\rho^n \tag{21}$$

Deswegen gilt

$$|m_{i,j}^{(k,l)}| \leq C(\rho^s + \rho^t + \rho^k + \rho^l). \tag{22}$$

Somit

$$\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} m_{i,j}^{(k,l)} \leq \frac{C}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}(\rho^s + \rho^t + \rho^k + \rho^l) = O(n^{-1}) \xrightarrow[n\to\infty]{} 0. \tag{23}$$

□

Theorem 0.3 (Ergodensatz, starke Formulierung)**.** *Sei* $\{X_n\}_{n=1}^{\infty}$ *eine irreduzieble und aperiodische Markov-Kette mit der stationärer Verteilung* $\pi \in \mathcal{M}_1(S)$*. Für jede Funktion* $f : S \to \mathbb{R}$ *gilt*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(X_i) = \pi[f], \quad f.s. \tag{24}$$

MAN KANN ZEIGEN, dass für eine ergodische Markov-Kette die ZVen X_k, X_l mit $|k - l|$ groß “fast unabhängig” sind. Dies ist eine Begründung, warum der GGZ für ergodische Markov-Ketten gilt.

Wie kann man numerisch einen Erwartungswert ausrechnen?

PROBLEM. Wie kann man für $\pi \in \mathcal{M}_1(S)$ und $f \colon S \to \mathbb{R}$ den Erwartungswert $\pi[f]$ numerisch ausrechnen?

ANTWORT: Mittels MCMC und Ergodensatzes! Man kann z.B. mit dem Metropolis-Hastings-Algorithmus eine Realisierung[14]

$$\{x_i := X_i(\omega)\}_{i=1}^{n}, \quad \text{für irgendein } \omega \in \Omega \tag{25}$$

der Markov-Kette $\{X_i\}_{i=1}^{n}$ erzeugen. Um dann $\pi[f]$ zu schätzen, kann man einfach die linke Seite von (14) an der erzeugten Realisierung $\{x_i\}_{i=1}^{n}$ auswerten und der Ergodensatz liefert[15]

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} f(x_i) \approx \pi[f].} \tag{26}$$

[14] Eine Realisierung einer ZV X ist ein Wert $x = X(\omega)$ für ein gegebenes $\omega \in \Omega$.

[15] Eine schwierige Frage hier ist: Wie groß soll n in (26) sein?

Starkes Gesetz der großen Zahlen

Starkes GGZ ist auch eine Aussage über die Konvergenz vom empirischen Mittel gegen den Erwartungswert. Allerdings ist dabei die Konvergenzart eine andere.

Theorem 0.4 (Starkes Gesetz der großen Zahlen)**.** *Unter den Voraussetzungen vom Theorem 0.1 gilt*

$$\boxed{\mathbb{P}\left\{\omega \in \Omega\colon \lim_{n\to\infty} \bar{X}_n(\omega) = \mu\right\} = 1} \tag{27}$$

In Worten bedeutet starkes Gesetz der großen Zahlen, dass für fast alle Realisierungen des Zufalls der empirische Mittelwert einer Folge der unabhängigen identisch verteilten ZV konvergiert gegen den Erwartungswert von einem einzelnen Summand.[16]

Definition 0.5 (Fast sichere Konvergenz)**.** *Sei $\{Y_i\}_{i=1}^{\infty}$ eine Folge der ZV und Y_∞ eine ZV auf dem WRaum $(\Omega, \mathcal{A}, \mathbb{P})$. Die Folge $\{Y_i\}_{i=1}^{\infty}$ konvergiert fast sicher (f.s.)*[17]*, falls es ein sicheres Ereignis*[18] *$A \in \mathcal{A}$ existiert, so dass*

$$\boxed{\forall \omega \in A\colon \lim_{n\to\infty} Y_n(\omega) = Y_\infty(\omega).} \tag{28}$$

Notation. Falls (6) gilt, schreibt man[19]

$$Y_n \xrightarrow[n\to\infty]{\text{f.s.}} Y_\infty. \tag{29}$$

In dieser Notation kann man (27) so umschreiben:

$$\bar{X}_n \xrightarrow[n\to\infty]{\text{f.s.}} \mu. \tag{30}$$

Die fast sichere Konvergenz ist nichts anderes als eine *punktweise Konvergenz*[20] auf einer Menge der vollen Wahrscheinlichkeit.

Die Konvergenz für fast alle Realisierungen des Zufalls ist eigentlich das, was wir in der Simulation auf der Abbildung 1 beobachten.

Man kann zeigen, dass

$$\text{f.s. Konvergenz} \Rightarrow \text{Konvergenz in Wahrscheinlichkeit} \tag{31}$$

$$\text{Konvergenz in Wahrscheinlichkeit} \not\Rightarrow \text{f.s. Konvergenz} \tag{32}$$

Damit ist die f.s. Konvergenz eine stärkere Konvergenzart als die Konvergenz in Wahrscheinlichkeit.

[16] Philosophisch bringt diese Botschaft die ExperimentatorInnen und TheoretikerInnen zusammen: der Mittelwert von Beobachtungen aus einem Experiment konvergiert gegen den theoretischen Erwartungswert für fast alle Realisierungen des Experiments als wir mehr und mehr Beobachtungen machen. So kann man von den Beobachtungen auf die Wirklichkeit schließen.

[17] alternativ sagt man: “konvergiert mit Wahrscheinlichkeit eins”.

[18] D.h. $\mathbb{P}(A) = 1$.

[19] Statt “f.s.” schreibt man auch “$\mathbb{P}$-f.s.”, wenn es nicht klar ist, welches WMaß in (28) gemeint ist.

[20] Punktweise Konvergenz soll aus der Analysis bekannt sein.

Die Konvergenz in Wahrscheinlichkeit (Definition 0.2) beinhaltet den Grenzübergang $n \to \infty$ außerhalb des $\mathbb{P}$-Symbols. Allerdings steht der Grenzwert in der f.s. Konvergenz (Definition 0.5) innerhalb des $\mathbb{P}$-Symbols, was (intuitiv) eine stärkere Aussage darstellt.

Sicheres Ergeignis A von der Definition 0.5 hängt von unendlich vielen ZVen $\{X_n\}_{n=1}^{\infty}$ ab. Allerdings nich ganz stark. Der Grenzwert $\lim_{n\to\infty} X_n$ hängt nur davon ab, was für großes n geschiet. Also hängt das Ereignis A nur davon ab, was mit der Folge $\{X_n\}_{n=1}^{\infty}$ im Unendlichen passiert.

Proof of Theorem 0.4. Einfachheitshalber führen wir den Beweis unter einer stärkeren Voraussätzung als angekündigt[21]:

$$\mathbb{E}\left[X_1^4\right] < \infty. \tag{33}$$

o.B.d.A. können wir annehmen, dass $\mathbb{E}[X_1] = 0$.[22] Wir möchten zeigen, dass

$$\mathbb{E}\left[\sum_{i=1}^{\infty} \frac{S_n^4}{n^4}\right] < \infty. \tag{34}$$

Es gilt

$$\mathbb{E}\left[S_n^4\right] = \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{\infty}\sum_{i_3=1}^{\infty}\sum_{i_4=1}^{\infty} \mathbb{E}\left[X_{i_1}X_{i_2}X_{i_3}X_{i_4}\right] \tag{35}$$

□

Beachte, dass es viele Summanden in (35) gibt, die gleich 0 sind. Z.b. falls i_1 is nicht gleich i_2, i_3 und i_4, dann wegen der Unabhängigkeit

$$\mathbb{E}[X_{i_1}X_{i_2}X_{i_3}X_{i_4}] = \mathbb{E}[X_{i_1}]\mathbb{E}\left[X_{i_2}X_{i_3}X_{i_4}\right] = 0. \tag{36}$$

Deswegen muss man in (35) nur die Terme von der folgenden zwei Bauarten betrachten: $\mathbb{E}[X_{i_1}^4]$ (es gibt n Stück davon) und

$$\mathbb{E}[X_{i_1}^2 X_{i_2}^2]. \tag{37}$$

Summanden (37) bekommt man in den folgenden Fällen: $i_1 = i_2 \neq i_3 = i_4$, $i_1 = i_3 \neq i_2 = i_4$ oder $i_1 = i_4 \neq i_2 = i_3$. Deswegen gibt es insgesamt $3n(n-1)$ Termen dieser Bauart. Somit

$$\mathbb{E}\left[S_n^4\right] = n\mathbb{E}[X^4] + 3n(n-1)\mathbb{E}[X_1^2X_2^2] = O(n^2). \tag{38}$$

Deswegen gilt $\frac{1}{n^4}\mathbb{E}[S_n^4] = O(n^{-2})$. Demzufolge gilt (34). Insb. heißt (34), dass[23]

[21] Diese annahme heißt: "Existenz von den vierten Momenten". Ohne diese Annahme ist der Beweis etwas länger.

[22] Betrachte sonst statt X_i die *zentrierten* ZVen $Y_i := X_i - \mu$. Für diese gilt $\mathbb{E}[Y_i] = 0$.

[23] Warum?

$$\sum_{i=1}^{\infty} S_n^4 < \infty \quad \text{f.s.} \tag{39}$$

Somit $S_n^4 \xrightarrow[n\to 0]{} 0$ f.s. und schließlich $S_n \xrightarrow[n\to 0]{} 0$ f.s.

Literaturhinweise: Kapitel 11 aus Ref.[24]

[24] Götz Kersting and Anton Wakolbinger. *Elementare Stochastik*. Springer, 2010

Literatur

Götz Kersting and Anton Wakolbinger. *Elementare Stochastik*. Springer, 2010.