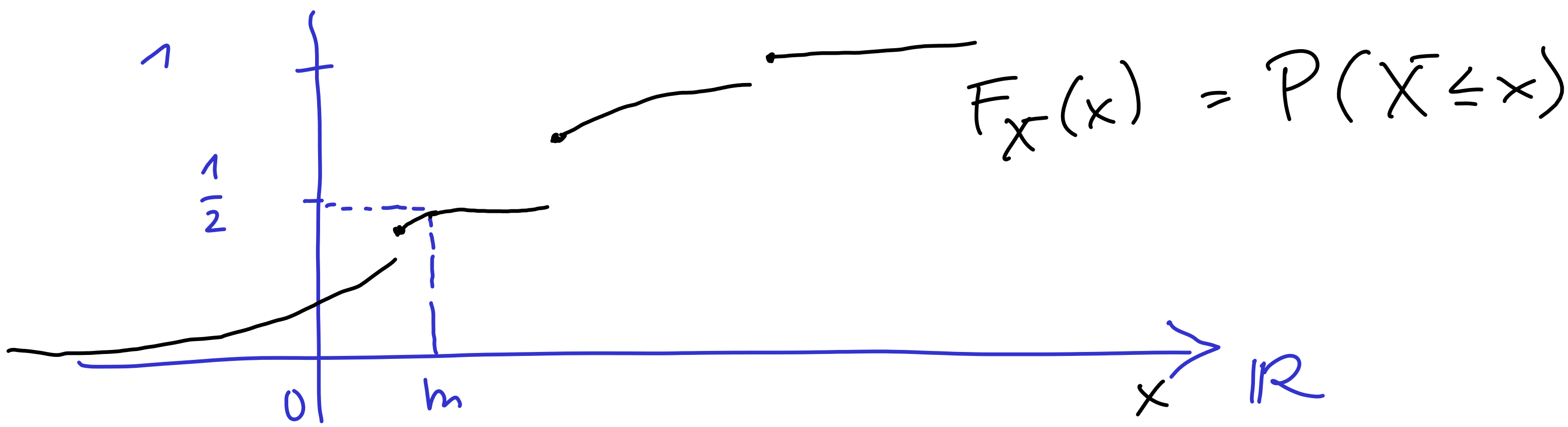
3.4 Median(e)

Definition 3.22. X reelle ZV, m heißt (ein) Median von X (auch „Zentralwert“, manchmal auch m_X geschrieben), wenn gilt

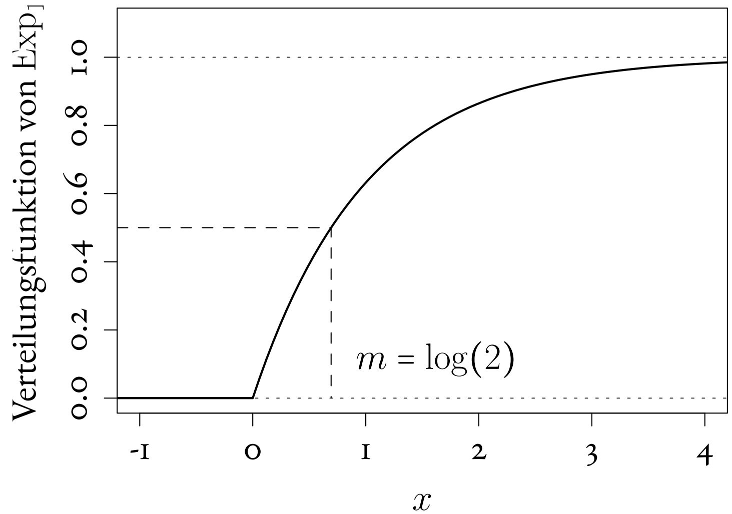
$$P(X \geq m) \geq \frac{1}{2} \quad \text{und} \quad P(X \leq m) \geq \frac{1}{2}.$$

Ein Median existiert stets, auch wenn X keinen Erwartungswert besitzt.

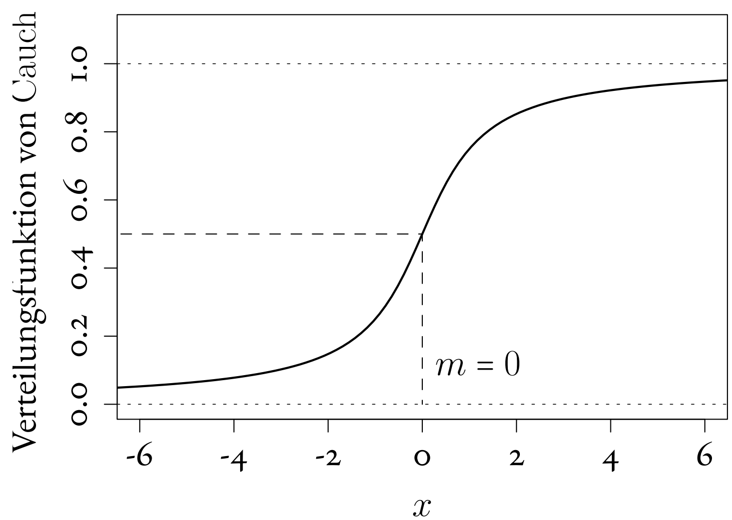
Man kann den Median als eine „robustere“ Antwort auf die Aufgabe, für eine ZV *nur einen* „typischen Wert“ anzugeben, ansehen. Allerdings gibt es für Mediane keine so angenehmen Rechenregeln, wie sie Satz 3.4 für den Erwartungswert liefert.



Beispiel 3.23. 1. $X \sim \mathrm{Exp}_\theta$ hat Dichte $\theta e^{-\theta x}\mathbf{1}_{[0,\infty)}(x)$, Verteilungsfunktion $(1-e^{-\theta x})\mathbf{1}_{[0,\infty)}(x)$, demnach ist der (eindeutig bestimmte) Median $m = \frac{1}{\theta}\log 2$.



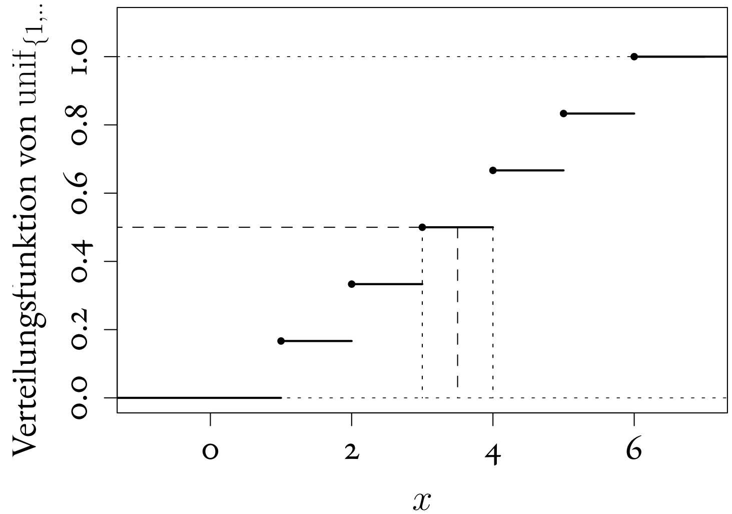
2. X Cauchy-verteilt mit Dichte $\frac{1}{\pi}\frac{1}{1+x^2}$, Verteilungsfunktion $\frac{1}{2}\frac{1}{\pi}\arctan(x)$, der (eindeutig bestimmte) Median ist $m = 0$.



(Wegen der Symmetrie der Dichte, es gibt keinen Erwartungswert, vgl. Bsp. 3.10, 3.)

denn $\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{|x|}{1+x^2}\,dx = \infty$

3. $X \sim \mathrm{unif}_{\{1,2,\ldots,6\}}$

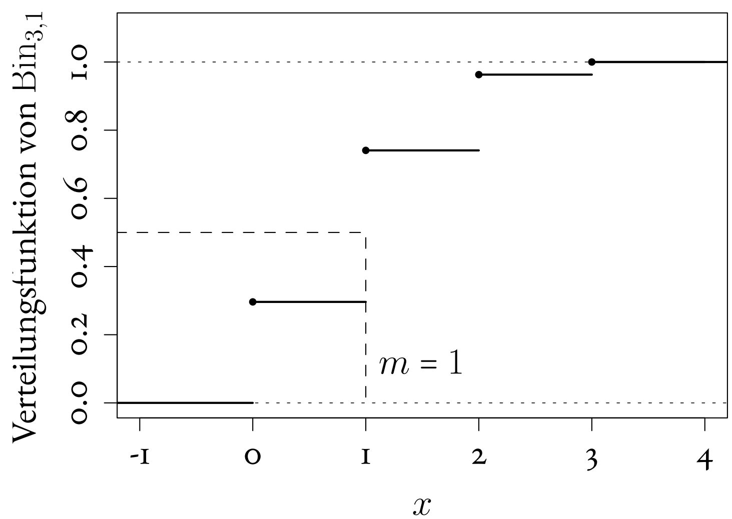


$P(X \geq 4) = 1/2$

$P(X \leq 4) \geq 1/2$

Jeder Wert $m \in [3,4]$ ist ein Median (und die vielleicht „kanonischste“ Wahl wäre $m = 3{,}5$).

4. $X \sim \mathrm{Bin}_{3,1/3}$ hat Median 1



Bemerkung 3.24. Sei $X \in \mathcal{L}^1$.

1. Jeder Median von X ist ein Minimierer von $a \mapsto \mathbb{E}[|X-a|]$.

2. Für jeden Median m ist $|\mathbb{E}[X] - m| \leq \sqrt{\mathrm{Var}[X]}$.

(Zum Vergleich $a \mapsto \mathbb{E}[(X-a)^2]$ wird durch $a = \mathbb{E}[X]$ minimiert)

Zu 1.: Sei m ein Median,

falls $a > m$:

$$|X-a| - |X-m| \geq \mathbb{1}_{\{X \leq m\}} \underbrace{((a-X)-(m-X))}_{=(a-m)} + \mathbb{1}_{\{X>m\}}(-(a-m)),$$

somit

$$\mathbb{E}[|X-a|] - \mathbb{E}[|X-m|] \geq (a-m)(\underbrace{\mathbb{P}(X \leq m)}_{\geq 1/2} - \underbrace{\mathbb{P}(X>m)}_{\leq 1/2}) \geq 0$$

analog, falls $a < m$.

2. $$|\mathbb{E}[X] - m| \leq \mathbb{E}[|X-m|] \overset{1.}{\leq} \mathbb{E}[|X - \mathbb{E}[X]|] \leq \sqrt{\mathbb{E}[(X-\mathbb{E}[X])^2]}$$

(mit Jensen- oder Cauchy-Schwarz-Ungl.

$a > m$:

$$|X-a| - |X-m|$$

$$= \mathbb{1}_{\{X \leq m\}}(a-m) + \mathbb{1}_{\{m < X < a\}} \underbrace{\overbrace{(a - X - (X-m))}^{= a+m-2X}}_{\geq m-a \text{ (in diesem Fall)}}$$

$$+ \mathbb{1}_{\{X \geq a\}}(m-a)$$

$$\geq \mathbb{1}_{\{X \leq m\}}(a-m) - (a-m)\left(\mathbb{1}_{\{m<X<a\}} + \mathbb{1}_{\{X \geq a\}}\right)$$

3.5 Erzeugende Funktionen*

Definition 3.25. Sei X eine ZV mit Werten in $\mathbb{N}_0$,

$$\varphi_X(s) = \mathbb{E}\big[s^X\big] = \sum_{n=0}^{\infty} s^n P(X = n), \quad s \in [0,1].$$

heißt die *erzeugende Funktion* von X.
Analog ist für ein Wahrscheinlichkeitsmaß μ auf $\mathbb{N}_0$ die erzeugende Funktion

$$\varphi_\mu(s) = \sum_{n=0}^{\infty} s^n \mu(\{n\}).$$

Beobachtung. φ_X (und genauso φ_μ) ist zumindest für $s \in [0,1]$ wohldefiniert mit $0 \le \varphi_X(s) \le 1 = \varphi_X(1)$, ist auf $[0,1)$ glatt, ist konvex (strikt konvex, sofern $P(X > 1) > 0$ bzw. $\mu(\{2,3,\dots\}) > 0$).

Satz 3.26 (Momentenbestimmung mittels der erzeugenden Funktion)**.**

1. $P(X=n) = \frac{\varphi_X^{(n)}(0)}{n!}$ *für* $n \in \mathbb{N}_0$, *insbesondere ist* $\mathscr{L}(X)$ *durch* φ_X *eindeutig bestimmt.*
2. $\mathbb{E}[X] = \varphi_X'(1)$ *(sofern existent)*
3. $\mathbb{E}[X(X-1)] = \varphi_X''(1)$, *insbesondere ist* $\mathrm{Var}[X] = \varphi_X''(1) + \mathbb{E}[X] - \big(\mathbb{E}[X]\big)^2$ *(sofern existent)*

Beweis. 1. Wegen $0 \leq P(X=k)$ und $\sum_{k=0}^\infty P(X=k) = 1$ hat die Potenzreihe

$$\sum_{k=0}^\infty s^k P(X=k) = \varphi_X(s)$$

mindestens Konvergenzradius 1. Aus der Analysis ist bekannt, dass sie somit an jeder Stelle s mit $|s| < 1$ (n-mal) gliedweise differenziert werden kann und es gilt

$$\frac{d^n}{ds^n}\varphi_X(s) = \sum_{k=n}^\infty k(k-1)\cdots(k-n+1)s^{k-n}P(X=k),$$

also für $s=0$

$$\frac{d^n}{ds^n}\varphi_X(0) = n!P(X=n).$$

2. Für $0 \leq s < 1$ ist nach obigem

$$\varphi_X'(s) = \sum_{n=1}^\infty n s^{n-1} P(X=n) = \mathbb{E}\big[X s^{X-1}\big].$$

Für $s \nearrow 1$ steigt die rechte Seite monoton auf gegen $\sum_{n=1}^\infty nP(X=n)$, wenn φ_X in $s=1$ differenzierbar ist, konvergiert die linke Seite gegen $\varphi'(1-) = \varphi'(1)$.

(Da φ_X konvex ist in $[0,1)$, existiert

$$\varphi_X'(1-) = \lim_{s \nearrow 1} \varphi_X'(s) \in (-\infty, \infty]$$

stets, und wir sehen aus obigem Argument, dass $\mathbb{E}[X] < \infty \iff \varphi_X'(1-) < \infty$.)

3. Analog ist für $0 \leq s < 1$

$$\varphi_X''(s) = \sum_{n=2}^{\infty} n(n-1)s^{n-2}P(X=n) = \mathbb{E}\left[X(X-1)s^{X-2}\right]$$

und sofern φ_X in $s = 1$ zweimal differenzierbar ist, folgt die Behauptung mit $s \nearrow 1$.
(Wie oben ist $\varphi_X''(1-) < \infty \iff \mathbb{E}\left[X(X-1)\right] < \infty$.) $\square$

Satz 3.27. *$X_1, \dots, X_m$ unabhängige, $\mathbb{N}_0$-wertige ZVn, so ist*

$$\varphi_{X_1+\dots+X_m}(s) = \varphi_{X_1}(s) \cdot \varphi_{X_2}(s) \cdot \dots \cdot \varphi_{X_m}(s)$$

Analog gilt für W'maße $\mu_1, \mu_2, \dots, \mu_n$ auf $\mathbb{N}_0$

$$\varphi_{\mu_1 * \dots * \mu_m}(s) = \varphi_{\mu_1}(s) \cdot \dots \cdot \varphi_{\mu_m}(s)$$

Beweis. Für $s \in [0,1]$ ist

$$\mathbb{E}\left[s^{X_1+\dots+X_m}\right] = \mathbb{E}\left[s^{X_1} \cdot s^{X_2} \dots s^{X_m}\right] = \mathbb{E}\left[s^{X_1}\right] \cdot \mathbb{E}\left[s^{X_2}\right] \dots \mathbb{E}\left[s^{X_m}\right]$$

mit Satz 3.4, 4. □

Beispiel 3.28.

1. $X \sim \mathrm{Ber}_p$, $\varphi_X(s) = ps + 1 - p$
2. $X \sim \mathrm{Bin}_{n,p}$, $\varphi_X(s) = (ps + 1 - p)^n = (1 - p(1-s))^n$ (explizite Rechnung oder verwende Satz 3.27)
3. $X \sim \mathrm{Poi}_\lambda$, $\varphi_X(s) = \sum_{n=0}^\infty e^{-\lambda} \frac{\lambda^n}{n!} s^n = e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)}$
4. $X \sim \mathrm{Geom}_p$, $\varphi_X(s) = \sum_{n=0}^\infty p(1-p)^n s^n = \frac{p}{1-(1-p)s}$

Beispiel 3.29 (Eine Skizze zu Galton-Watson-Prozessen)**.** Wir beobachten zunächst: Sei X $\mathbb{N}_0$-wertig, $X_1, X_2, \dots$ u.i.v. Kopien von X, davon unabhängig Y $\mathbb{N}_0$-wertig, $Z := \sum_{i=1}^{Y} X_i$ (mit Interpretation $Z = 0$ auf $\{Y = 0\}$), dann ist $\varphi_Z(s) = \varphi_Y(\varphi_X(s))$, denn

$$\begin{aligned}
\mathbb{E}\big[s^Z\big] &= \sum_{z=0}^{\infty} s^z P(Z = z) = \sum_{z=0}^{\infty}\sum_{y=0}^{\infty} s^z P(Z = z, Y = y) \\
&= \sum_{z=0}^{\infty}\sum_{y=0}^{\infty} \sum_{\substack{x_1,\dots,x_y \in \mathbb{N}_0 \\ x_1+\dots+x_y = z}} s^z P(Y = y, X_1 = x_1, \dots, X_y = x_y) \\
&= \sum_{z=0}^{\infty}\sum_{y=0}^{\infty} \sum_{\substack{x_1,\dots,x_y \in \mathbb{N}_0 \\ x_1+\dots+x_y = z}} s^{x_1+\dots+x_y} P(Y = y) P(X_1 = x_1) \cdots P(X_y = x_y) \\
&= \sum_{y=0}^{\infty} P(Y = y) \sum_{x_1=0}^{\infty} \cdots \sum_{x_y=0}^{\infty} s^{x_1} P(X = x_1) \cdots s^{x_y} P(X = x_y) \\
&= \sum_{y=0}^{\infty} P(Y = y) \big(\varphi_X(s)\big)^y = \varphi_Y(\varphi_X(s)).
\end{aligned}$$

Man findet daraus (zusammen mit Satz 3.26, 2.) insbesondere die Wald'sche Identität (zumindest im Fall $\mathbb{N}_0$-wertiger Summanden):

$$\mathbb{E}[Z] = \varphi_Z'(1) = \varphi_Y'(\varphi_X(1))\varphi_X'(1) = \varphi_Y'(1)\varphi_X'(1) = \mathbb{E}[Y]\mathbb{E}[X].$$

Seien nun $X_{n,i}, n, i \in \mathbb{N}$ u.i.v. Kopien einer $\mathbb{N}_0$-wertigen ZV X (mit $m := \mathbb{E}[X] = \varphi_X'(1) < \infty$ und $P(X = 1) < 1$), setze

$$Z_0 = 1, \quad Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i} \quad \text{für } n \in \mathbb{N}.$$

Interpretation: Z_n ist die Größe der n-ten Generation in einer Population, in der jedes Individuum unabhängig eine zufällige, wie X verteilte Anzahl Nachkommen hat. (Man nennt $(Z_n)_{n\in\mathbb{N}_0}$ auch einen Verzweigungsprozess oder Galton-Watson-Prozess.)

Aus der Definition (zusammen mit obiger Beobachtung und Satz 3.26) findet man induktiv

$$\varphi_{Z_n}(s) = \big(\underbrace{\varphi_X \circ \varphi_X \circ \cdots \circ \varphi_X}_{n\text{-fache Hintereinanderausführung}}\big)(s) \quad \text{und} \quad \mathbb{E}[Z_n] = \varphi_{Z_n}'(1) = m^n.$$

$Z = (Z_n)_n$ heißt *subkritisch*, wenn $m < 1$, *kritisch*, wenn $m = 1$, *superkritisch*, wenn $m > 1$.

Sei $q_n = P(Z_n = 0) = \varphi_{Z_n}(0)$, offenbar ist $q_n \le q_{n+1}$ (denn $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$) und

$$q_n \nearrow_{n\to\infty} q = P\Big(\bigcup_{n=1}^{\infty}\{Z_n = 0\}\Big) = P(Z \text{ stirbt aus}).$$

Tatsächlich gilt

$$P(Z \text{ stirbt aus}) = 1 \iff m \le 1. \tag{3.1}$$

Wegen $q_{n+1} = \varphi_X(q_n)$ folgt mit $n \to \infty$ und Stetigkeit von φ_X, dass $q = \varphi_X(q)$, d.h. $q = \lim_{n\to\infty} q_n$ ist ein Fixpunkt von φ_X.

Sei $\tilde{q} = \varphi_X(\tilde{q})$ ein (möglicherweise anderer) Fixpunkt von φ_X, dann ist $0 = q_0 \le \tilde{q}$, n-malige Anwendung von φ_X liefert

$$q_n \le \varphi_X(\tilde{q}) = \tilde{q} \text{ für alle } n, \qquad \text{d.h. } q \text{ ist der kleinste Fixpunkt von } \varphi_X$$

Da φ_X konvex ist (und n. Vor. $\varphi_X(s) \not\equiv s$), gilt:

$$\text{wenn } m = \varphi_X'(1) \le 1, \text{ so ist } q = 1 \text{ der einzige Fixpunkt in } [0,1]$$

$$\text{wenn } m = \varphi_X'(1) > 1, \text{ so ist } 0 \le q < 1 \text{ und die Fixpunkte sind } \{q, 1\}$$

(Details als Übung)

Historische Anmerkung. Diese Art von zufälligen Populationsprozessen wird üblicherweise in der mathematischen Literatur nach Sir Francis Galton (1822–1911) und Henry William Watson (1827–1903) benannt, die die Eigenschaft (3.1) untersuchten: Galton stellte dazu eine Aufgabe, „Problem 4001", in der Ausgabe vom April 1873 der Zeitschrift *Educational times*, in der er nach der Wahrscheinlichkeit des Aussterbens eines Familiennamens fragte (man lese Z_n als die Anzahl der männlichen Nachkommen eines, sagen wir adeligen, Stammvaters in der n-ten Generation, in Galtons gesellschaftlich-historischem Kontext war klar, dass es für die Namensfrage genügt, die Anzahl der Söhne zu untersuchen). Watson sandte einen – fast richtigen – Lösungsvorschlag, den die beiden dann gemeinsam in dem Artikel „On the probability of the extinction of families", *J. Roy. Anthropol. Inst.*, 4 (1874), 138–144 veröffentlichten.

Die Geschichte des Studiums der Verzweigungsprozesse ist nicht allzu geradlinig verlaufen: Wie sich im Nachhinein herausstellte, hatten Galton und Watson zwar den Fall $m \leq 1$ richtig behandelt, nicht aber den Fall $m > 1$, obwohl der französische Mathematiker und Statistiker Irénée-Jules Bienaymé (1796–1878) das Problem bereits 1845 korrekt gelöst hatte – sein Artikel darüber war in Vergessenheit geraten. Die Frage, unter welchen Bedingungen $P(Z \text{ stirbt aus}) = 1$ gilt, ist wohl derart natürlich, dass sie im Lauf der Zeit von verschiedenen Autoren unabhängig und mit verschiedenen Interpretationen „entdeckt" und auch beantwortet wurde; siehe dazu David Kendall, The genealogy of genealogy: branching processes before (and after) 1873. With a French appendix containing Bienaymé's paper of 1845, *Bull. London Math. Soc.*, 7 (1975), no. 3, 225–253 und die lesenswerten Vortragsnotizen von Peter Jagers, Some Notes on the History of Branching Processes, from my Perspective. Lecture at the Oberwolfach Symposium on Random Trees, 18–24 January, 2009 `http://www.math.chalmers.se/~jagers/BranchingHistory.pdf`

Ohne Zweifel hat die Frage nach dem Aussterben (adeliger) Familiennamen Menschen schon lange vor Galton und Watson beschäftigt, so schreibt Jane Austen (1775–1817) in ihrem Roman *Persuasion* über den Vater der Protagonistin (der sich offenbar gerne mit dem in durch (3.1) im Fall $m \leq 1$ beschriebenen Phänomen beschäftigte, auch wenn er es sicherlich nicht als mathematische Frage auffasste):

> "Sir Walter Elliot, of Kellynch Hall, in Somersetshire, was a man who, for his own amusement, never took up any book but the Baronetage; ... there his faculties were roused into admiration and respect, by contemplating the limited remnant of the earliest patents ..."

Kapitel 4

Gesetz der großen Zahlen

Satz 4.1. *X reelle ZV, $f : [0, \infty) \to [0, \infty)$ monoton wachsend.*

1. Für $a > 0$ mit $f(a) > 0$ gilt

$$P\big(|X| \geq a\big) \leq \frac{1}{f(a)}\mathbb{E}\big[f(|X|)\big] \qquad \text{(Markov[1]-Ungleichung).} \tag{4.1}$$

2. Für $X \in \mathcal{L}^2$ gilt

$$P\big(|X - \mathbb{E}[X]| \geq a\big) \leq \frac{\operatorname{Var}[X]}{a^2} \qquad \text{(Chebyshev[2]-Ungleichung).} \tag{4.2}$$

1. $Y := \mathbb{1}_{\{|X| \geq a\}} f(a)$, dann ist $Y \leq f(|X|)$

$\Rightarrow \quad \mathbb{E}[Y] = f(a) P(|X| \geq a) \leq \mathbb{E}[f(|X|)]$

2. Betr. $\tilde{X} := X - \mathbb{E}[X]$ und $f(a) = a^2$:

$P(|X - \mathbb{E}[X]| \geq a) = P(|\tilde{X}| \geq a) \leq \frac{1}{a^2} \overbrace{\mathbb{E}[\tilde{X}^2]}^{= \operatorname{Var}[X]}$

[1] Andrei Andrejewich Markov, 1856–1922.
[2] Pafnuty Lvovich Chebyshev, 1821–1894.

Korollar 4.2. *1.* $X_1, X_2, \dots \in \mathcal{L}^2$ *seien paarweise unkorreliert mit*

$$\sup_n \mathrm{Var}[X_n] \le \theta < \infty,$$

dann gilt für $\varepsilon > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mathbb{E}[X_i]\right)\right| > \varepsilon\right) \le \frac{\theta}{\varepsilon^2 n} \left(\xrightarrow[n\to\infty]{} 0\right). \tag{4.3}$$

2. Speziell gilt für $X_1, X_2, \dots \in \mathcal{L}^2$ *u.i.v. mit* $\mu := \mathbb{E}[X_1]$ *und* $\sigma^2 := \mathrm{Var}[X_1]$ $(< \infty)$

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \varepsilon\right) \le \frac{\sigma^2}{\varepsilon^2 n} \xrightarrow[n\to\infty]{} 0 \tag{4.4}$$

(schwaches Gesetz der großen Zahlen).

1. $Y := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])$, $\mathbb{E}[Y] = 0$,

$\mathrm{Var}[Y] = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\mathrm{Var}[X_i - \mathbb{E}[X_i]]}_{= \mathrm{Var}[X_i] \le \theta} + 0 \le \frac{\theta}{n}$,

dann verwende Chebyshev-Ungl.

2. Ist Spezialfall von 1. (Unabh. $\Rightarrow$ unkorr.)

Definition 4.3. Seien $X_1, X_2, \dots, X$ relle ZVn (auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, P)$ definiert).

1. $(X_n)_{n\in\mathbb{N}}$ *konvergiert stochastisch* gegen X, auch geschrieben

$$X_n \xrightarrow[n\to\infty]{\text{stoch.}} X,$$

(auch $X_n \xrightarrow[n\to\infty]{} X$ stoch. oder $X_n \xrightarrow[n\to\infty]{P} X$) wenn gilt

$$\forall\ \varepsilon > 0 : \lim_{n\to\infty} P\big(|X_n - X| > \varepsilon\big) = 0.$$

2. $(X_n)_{n\in\mathbb{N}}$ *konvergiert fast sicher* gegen X, auch geschrieben

$$X_n \xrightarrow[n\to\infty]{\text{f.s.}} X,$$

(oder $X_n \xrightarrow[n\to\infty]{} X$ fast sicher / f.s.) wenn gilt

$$P\Big(X_n \xrightarrow[n\to\infty]{} X\Big) = 1.$$

(Erinnerung: $(x_n)_n$ Folge von reellen Zahlen, $x_n \xrightarrow[n\to\infty]{} x$ $(\in\mathbb{R})$, wenn $\forall \varepsilon>0\ \exists n_0: |x_n - x| \leq \varepsilon\ \forall n \geq n_0$)

Beachte:

$$\Big\{X_n \xrightarrow[n\to\infty]{} X\Big\} = \bigcap_{\varepsilon\in Q\cap(0,\infty)} \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \big\{|X_m - X| \leq \varepsilon\big\} \quad \textit{ist } \text{ein Ereignis.}$$

$\varepsilon \in \mathbb{Q} \cap (0,\infty)$

Wir können Kor. 4.2, 2. aussprechen als

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{stoch.}} \mu$$

Beobachtung 4.4. $X_n \to X$ f.s. $\Rightarrow X_n \to X$ stoch., die Umkehrung gilt i.A. nicht.

Sei $P(X_n \xrightarrow[n\to\infty]{} X) = 1$, $\varepsilon > 0$

$$P(|X_n - X| > \varepsilon) \leq P\Big(\bigcup_{m \geq n} \{|X_m - X| > \varepsilon\}\Big)$$

$$\Big\downarrow n \to \infty$$

$$P(|X_m - X| > \varepsilon \text{ für unendlich viele } m) = 0$$

„Gegenbeispiel" zur Umkehrung:

$X_1, X_2, X_3, \ldots$ seien u.a., $X_n \sim \mathrm{Ber}_{1/n}$:

$P(|X_n - 0| > \varepsilon) = P(X_n = 1) = \frac{1}{n} \xrightarrow[n\to]{} 0$, d.h. $X_n \xrightarrow[n\to\infty]{\text{stoch.}} 0$

(falls $0 < \varepsilon < 1$)

$P(X_n = 1$ für unendl. viele $n) = 1$ mit Borel-Cantelli ($\sum_1^\infty 1/n = \infty$),

d.h. $\cancel{X_n \xrightarrow[n\to\infty]{\text{f.s.}} 0}$

Man kann mit konvergenten Folgen „wie gewohnt“ rechnen:

Lemma 4.5. *$X_1, X_2, \dots, Y_1, Y_2, \dots, X, Y$ reelle ZVn, $(a_n)_{n\in\mathbb{N}}$ Folge von reellen Zahlen mit $a_n \to a \in \mathbb{R}$, es gelte $X_n \to X$ stochastisch (bzw. fast sicher) und $Y_n \to Y$ stochastisch (bzw. fast sicher). Dann gilt auch*

$$X_n + Y_n \xrightarrow[n\to\infty]{} X + Y \text{ stochastisch (bzw. fast sicher)}, \tag{4.5}$$

$$a_n X_n \xrightarrow[n\to\infty]{} aX \text{ stochastisch (bzw. fast sicher)}. \tag{4.6}$$

(Übungen: $X_n \xrightarrow[n\to\infty]{} X$ und $Y_n \xrightarrow[n\to\infty]{} Y$, (f.s. oder stoch.)

$g: \mathbb{R}^2 \to \mathbb{R}$ stetig

$g(X_n, Y_n) \xrightarrow[n\to\infty]{} g(X, Y)$)

Die Aussage von Korollar 4.2 gilt tatsächlich auch für den stärkeren Begriff der fast sicheren Konvergenz:

Satz 4.6 (Eine Version des starken Gesetzes der großen Zahlen). *In der Situation von Kor. 4.2, 1. gilt*

$$\frac{1}{n}\sum_{i=1}^{n}\big(X_i - \mathbb{E}[X_i]\big) \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{fast sicher}, \tag{4.7}$$

insbesondere gilt für $X_1, X_2, \dots \in \mathcal{L}^2$ *u.i.v. mit* $\mathbb{E}[X_1] = \mu$

$$\frac{X_1 + \dots + X_n}{n} \underset{n\to\infty}{\longrightarrow} \mu \quad \textit{fast sicher}. \tag{4.8}$$

Den Beweis betrachten wir als Ergänzung in Abschnitt 4.1.

Beispiel 4.7 (Borels[3] Gesetz über normale Zahlen, 1909). Sei $U \sim \mathrm{Unif}_{[0,1]}$, X_i die i-te Ziffer in der (nicht-abbrechenden) Dezimaldarstellung von U (d.h. $U = \sum_{i=1}^{\infty} X_i\, 10^{-i}$).
Dann gilt für $q = 0, 1, \dots, 9$:

$$\frac{1}{n}\big|\{1 \le i \le n : X_i = q\}\big| \underset{n\to\infty}{\longrightarrow} \frac{1}{10} \quad \text{fast sicher.}$$

Beweis. $X_1, X_2, \dots$ sind u.a., $X_i \sim \mathrm{Unif}_{\{0,1,\dots,9\}}$, also sind $\mathbf{1}_{\{X_i=q\}}$, $i = 1, 2, \dots$ u.i.v., $\mathrm{Ber}_{1/10}$, die Beh. folgt aus Satz 4.6. □

Bericht 4.8. Das starke Gesetz der großen Zahlen gilt tatsächlich auch ohne $\mathcal{L}^2$-Annahme:
Für $X_1, X_2, \dots \in \mathcal{L}^1$ u.i.v. mit $\mathbb{E}[X_1] = \mu$ gilt

$$\frac{X_1 + \dots + X_n}{n} \underset{n\to\infty}{\longrightarrow} \mu \quad \text{fast sicher.}$$

[3] Émile Borel, 1871–1956

4.1 Beweis von Satz 4.6*

Die Aussage von Satz 4.6 (das starke Gesetz der großen Zahlen für unkorrelierte ZVn) lautete: Sind $X_1, X_2, \dots \in \mathcal{L}^2$ seien paarweise unkorrelierte ZVn mit

$$\sup_n \operatorname{Var}[X_n] \le \theta < \infty$$

dann gilt

$$\frac{1}{n}\sum_{i=1}^{n}\big(X_i - \mathbb{E}[X_i]\big) \xrightarrow[n\to\infty]{} 0 \quad \text{fast sicher} \tag{4.9}$$

Wir wollen hier als Ergänzung den Beweis betrachten.

Beweis von Satz 4.6. Sei o.E. $\mathbb{E}[X_i] = 0$, sonst betrachte $X_i' := X_i - \mathbb{E}[X_i]$, setze

$$S_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

1. Schritt: Zeige

$$S_{n^2} \xrightarrow[n\to\infty]{} 0 \text{ f.s.}$$

Für $\varepsilon > 0$ zeigt Kor. 4.2, 1.

$$P\big(|S_{n^2}| > \varepsilon\big) \le \frac{\theta}{\varepsilon^2 n^2},$$

also

$$P\big(|S_{n^2}| > \varepsilon \text{ für } \infty\text{-viele } n\big) = 0$$

(mit Borel-Cantelli-Lemma (Satz 2.30), da $\sum_n \frac{\theta}{\varepsilon^2 n^2} < \infty$) und somit

$$P\big(\{S_{n^2} \to 0\}^c\big) \le P\Bigg(\bigcup_{\varepsilon\in\mathbb{Q}\cap(0,\infty)} \big\{|S_{n^2}| > \varepsilon \text{ für } \infty\text{-viele } n\big\}\Bigg) = 0.$$

2. Schritt: Zu $m \in \mathbb{N}$ wähle $n = n(m)$ mit $n^2 \leq m \leq (n+1)^2$:

$$P\Big(\big|\underbrace{mS_m - n^2 S_{n^2}}_{=\sum_{n^2<i\leq m} X_i}\big| > \varepsilon m\Big) \leq \frac{1}{\varepsilon^2 m^2}\mathrm{Var}\Big[\sum_{n^2<i\leq m} X_i\Big] \leq \frac{\theta(m-n^2)}{\varepsilon^2 m^2},$$

somit

$$\begin{aligned}
&\sum_{m=1}^{\infty} P\Big(\big|mS_m - n(m)^2 S_{n(m)^2} > \varepsilon m\Big) \\
&\leq \frac{\theta}{\varepsilon^2}\sum_{n=1}^{\infty}\sum_{m=n^2}^{(n+1)^2-1}\frac{m-n^2}{m^2} \leq \frac{\theta}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{1}{n^4}\underbrace{\sum_{j=1}^{(n+1)^2-1-n^2} j}_{=\frac{2n(2n+1)}{2}} \leq \frac{\theta}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{2n(2n+1)}{n^4} < \infty,
\end{aligned}$$

demnach gilt (wie im 1. Schritt)

$$P\Big(S_m - \frac{n(m)^2}{m} S_{n(m)^2} \underset{m\to\infty}{\longrightarrow} 0\Big) = 1,$$

dies zusammen mit dem 1. Schritt und Lemma 4.5 zeigt $S_m \to 0$ f.s. für $m \to \infty$.
Es ist nämlich für $\varepsilon > 0$ wegen $\frac{n(m)^2}{m} \to_{m\to\infty} 1$

$$\begin{aligned}
A_\varepsilon &:= \big\{|S_m| > \varepsilon \text{ für } \infty\text{-viele } m\big\} \\
&\subset \Big\{|S_{n^2}| > \frac{\varepsilon}{4} \text{ für } \infty\text{-viele } n\Big\} \cup \Big\{\Big|S_m - \frac{n(m)^2}{m} S_{n(m)^2}\Big| > \frac{\varepsilon}{2} \text{ für } \infty\text{-viele } m\Big\},
\end{aligned}$$

also $P(A_\varepsilon) = 0$ und damit auch $P\Big(\bigcap_{\varepsilon\in\mathbb{Q},\varepsilon>0} A_\varepsilon\Big) = 0$. $\square$