Wahrscheinlichkeitstheorie I

Essen, Sommersemester 2005
Prof. Dr. P. L. Davies

Literatur

Wahrscheinlichkeitstheorie:

- Behnen, K. und Neuhaus, G. (1995) *Grundkurs Stochastik*, 3. Auflage,Teubner.
- Dinges, H. und Ost, H. (1982) *Prinzipien der Stochastik*, Teubner.
- Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd Edition, Wiley.
- Feller, W. (1971) *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd Edition, Wiley.
- Gänssler, P. und Stute, W. (1977) *Wahrscheinlichkeitstheorie*, Springer.
- Krengel, U. (1991) *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 3te Auflage, Vieweg.
- Pfanzagl, J. (1988) *Elementare Wahrscheinlichkeitsrechnung*, de Gruyter.
- Williams, D. (2001) *Weighing the Odds*, Cambridge University Press.

Maß- und Integrationstheorie:

- Bauer, H. (1990) *Maß- und Integrationstheorie*, de Gruyter.

Fortgeschrittene Bücher:

- Billingsley, P. (1986) *Probabilty and Measure*, 2nd Edition, Wiley.
- Dudley, R. M. (1989) *Real Analysis and Probability*, Wadsworth.

Inhaltsverzeichnis

1 Stochastische Modelle **4**
- 1.1 Zufall 4
- 1.2 Wahrscheinlichkeitsräume 18
- 1.3 Laplace-Experimente 22

2 Abhängigkeit und Unabhängigkeit **24**
- 2.1 Abhängigkeit 24
- 2.2 Unabhängigkeit 27
- 2.3 Unabhängige Koppelungen von Experimenten 30
- 2.4 Unabhängige Mengensysteme 34

3 Zufallsvariablen **38**
- 3.1 Zufallsvariablen und ihre Verteilungen 38
- 3.2 Einige wichtige eindimensionale Verteilungen 43
  - 3.2.1 Einige diskrete Verteilungen 43
  - 3.2.2 Einige stetige Verteilungen 48
- 3.3 Transformationen von Zufallsvariablen 50
- 3.4 k–dimensionale Zufallsvariablen 53
- 3.5 Transformation von k–dimensionalen Zufallsvariablen 56
- 3.6 Unabhängig verteilte Zufallsvariablen 57

4 Erwartungswerte **61**
- 4.1 Erwartungswert und Varianz 61
- 4.2 Unabhängig verteilte Zufallsvariablen 67
- 4.3 Momente und Ungleichungen 68
- 4.4 Charakteristische Funktionen 72

5 Summen von unabhängig verteilten Zufallsvariablen **80**
- 5.1 Einführung 80
- 5.2 Charakteristische Funktionen 83
- 5.3 Gesetze der großen Zahlen 84

6 Schwache Konvergenz **93**
- 6.1 Einführung 93
- 6.2 Straffheit 98

7 Bedingte Erwartungswerte **106**
- 7.1 Definition 106
- 7.2 Eigenschaften des bedingten Erwartungswertes 109

8 Martingale **114**
8.1 Filtrationen und Stoppzeiten 114
8.2 Sub-, Super- und Martingale 115

1 Stochastische Modelle

1.1 Zufall

Die Wahrscheinlichkeitstheorie ist die mathematische Theorie des Zufalls. Das einfachste Beispiel eines zufälligen Vorgangs ist der Standardmünzwurf. Wir werfen eine faire Münze n-mal nacheinander und kodieren die zwei möglichen Ergebnisse, Kopf und Zahl, mit -1 und 1. Eine von mir durchgeführte Serie von Münzwürfen mit n=100 ergab folgende Ergebnisse:

1 -1 1 -1 -1 -1 -1 -1 -1 1 1 1 1 1 1 -1 1 -1 -1 1

-1 -1 1 1 -1 1 -1 1 -1 -1 1 -1 1 1 -1 1 -1 -1 1 -1

-1 -1 -1 -1 1 -1 1 -1 1 -1 -1 1 -1 1 1 1 -1 -1 -1 1

-1 1 -1 -1 -1 -1 1 1 1 1 1 1 1 -1 1 -1 -1 1 -1 1

1 -1 1 1 -1 1 1 1 1 -1 -1 -1 1 -1 1 -1 1 -1 1 -1

Trotz seiner Einfachheit enthält dieses Beispiel alle wesentlichen Ingredienzen der Wahrscheinlichkeitstheorie, weil alles auf den einfachen Münzwurf zurückgeführt werden kann.

Die Ergebnisse scheinen zuerst nicht besonders auffällig zu sein mit Ausnahme der “runs“ der Länge 6 bei den Würfen 4-9, 10-15 und 68-73. Mit dem mathematischen Modell für den einfachen Münzwurf kann man die Wahrscheinlichkeit dafür bestimmen, dass man bei 100 Münzwürfen mindestens dreimal einen Run der Länge 6 bekommt. Ein einfaches Argument zeigt, dass diese Wahrscheinlichkeit mindestens 0,014 beträgt. Dies ist nicht besonders groß, aber auch nicht verschwindend klein. In anderen Worten, die Ergebnisse sind eher typisch. Die Länge von “runs“ in zufälligen Folgen wird meistens unterschätzt. In der Tat, wenn Testpersonen gebeten werden, die Zahlen -1 und 1 zufällig aufzuschreiben, kann man die Reihen relativ leicht von zufälligen Reihen unterscheiden, weil die Runs der Testpersonen viel zu kurz sind.

Die obigen Ergebnisse können wir als Punkte darstellen (Abbildung 1 oberes Bild) und, noch besser, mit Geraden verbinden (Abbildung 1 unteres Bild).

Bei dem einfachen Münzwurf kommen nur die Werte -1 und 1 vor, und dies ist für die Anwendung viel zu restriktiv. Durch Addition kann man die Anzahl

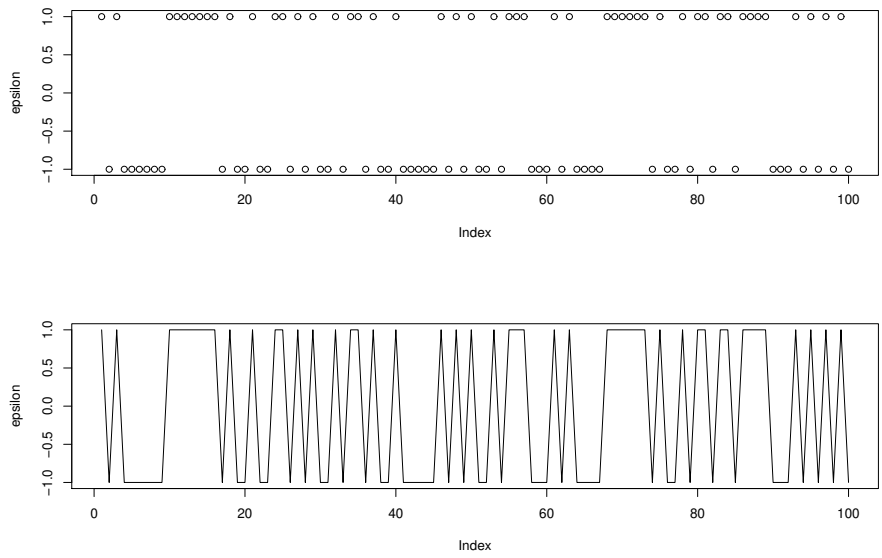

Abbildung 1: Die Ergebnisse von 100 Münzwürfen

der möglichen Werte erhöhen. Wir betrachten nun die Summe der ersten fünf Würfe, der nächsten fünf Würfe u.s.w. und bekommen folgende Ergebnisse:

-1 -3 5 -1 1 -1 1 -1 -5 1 -1 -1 -1 1 3 -1 1 3 -1 -1

Sie sind in Abbildung 2 graphisch dargestellt. Bei Summen von 5 Würfen gibt es nur 6 verschiedene Ergebnisse, nämlich

-5 -3 -1 1 3 5.

Wenn man aber Summen von 50 Würfen betrachtet, hat man nun 51 mögliche Werte, nämlich

-50, -48, ...-2, 0, 2, ...48, 50.

Für die obige empirische Reihe erhalten wir nur zwei Werte -4 und 2. Wenn man mehr Werte will, muss man zum Computer greifen, um die zufälligen Zahlen zu erzeugen. Es ergibt sich spätestens zu diesem Zeitpunkt ein Problem. Kann ein Computer zufällige Zahlen erzeugen? Die Antwort lautet *nein* und die Zahlen werden konsequenterweise *pseudo-zufällig* genannt. Wir werden das Problem aber ignorieren und so tun, als ob die Zahlen zufällig

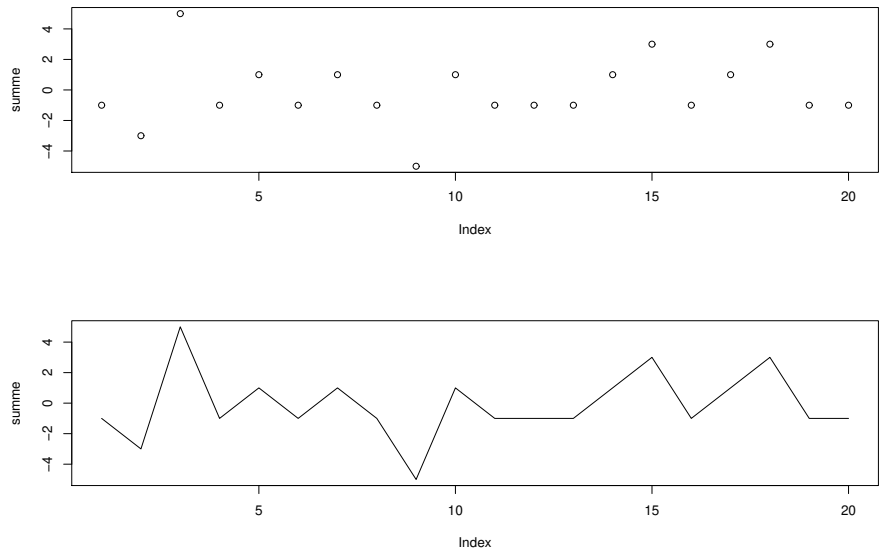

Abbildung 2: Summen über Blöcke von 5 Münzwürfen

sind. Wir erzeugen 10000 zufällige Zahlen, die entweder $+1$ oder -1 sind und bilden die Summen über Blöcke der Länge 50. Die ersten 10 Ergebnisse sind:

4 4 8 -8 2 0 -14 -6 2 -2.

Die Abbildung 3 zeigt die Ergebnisse. Wir wollen nun diese Werte mit echten Daten vergleichen. Abbildung 4 zeigt die Ergebnisse eines Experiments in der Dünnschichtphysik. Sie stammen von Herrn Mergel aus dem Fachbereich Physik und geben die Intensität von Röntgenstrahlen in Abhängigkeit vom Einfallswinkel an. Obwohl es an dem Bild nicht ersichtlich ist, sind die Messwerte ganze Zahlen. Die ersten 10 lauten: 15 17 10 16 9 16 23 13 10 18. Die Abbildung 5 zeigt die ersten 200 Werte von Abbildung 3 und die ersten 200 Werte der Daten von Abbildung 4. Die Daten von Herrn Mergel zeigen einen leichten Trend (siehe Abbildung 4), aber ansonsten sind die Bilder sehr ähnlich. In anderen Worten kann man das Rauschen von der Röntgenintensität durch den einfachen Münzwurf gut modellieren. Dies gilt auch für ganz andere Datensätze. Abbildung 6 zeigt den täglichen Dax-Kurs über 30 Jahre. Von Interesse sind die täglichen prozentuellen Änderungen des Kurses, die in Abbildung 7 graphisch dargestellt sind. Die ersten 200 Werte sind in Abbildung 8 zu sehen.

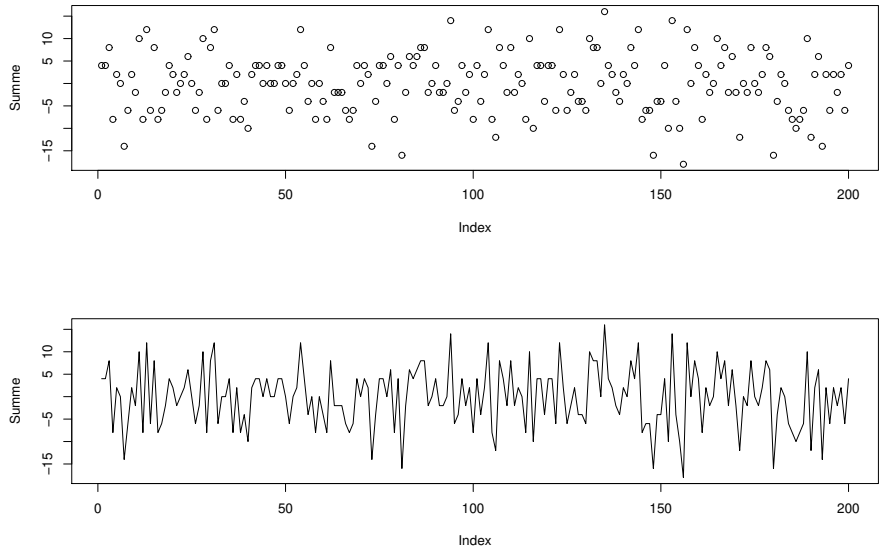

Abbildung 3: Summen über Blöcke von 50 Münzwürfen

Die Ähnlichkeit mit dem unteren Bild von Abbildung 3 ist klar. Das Rauschen vom Dax kann man also auch sehr gut mit dem einfachen Münzwurf modellieren; dies gilt für viele Datensätze aus den verschiedensten Gebieten. Der Zufall, exemplifiziert durch den einfachen Münzwurf, ist charakterisiert durch Unvorhersagbarkeit. Das Ergebnis der ersten Würfe enthält keine Information über das Ergebnis des nächsten Wurfes. Es ist eben diese Unvorhersagbarkeit, die es den oben erwähnten Testpersonen so schwierig macht, sich zufällig zu verhalten. Hat man schon fünfmal die "1 " , so denken die meisten Testpersonen, "Es muß endlich eine -1 kommen ", mit dem Ergebnis, dass die Runlängen zu klein sind.Trotz dieser Regellosigkeit unterliegt der Zufall Gesetzen und läßt sich mathematisch untersuchen. Wir bezeichnen die Ergebnisse der einzelnen Würfe mit $Z_i, i = 1, \ldots, n$, und bilden die partiellen Summen:

$$S_j = \sum_{i=1}^{j} Z_i$$

Für den einfachen Münzwurf erwartet man, dass die Proportion der 1-Würfe und die Proportion der -1-Würfe ungefähr gleich sind, etwas genauer, wir erwarten, dass gilt:

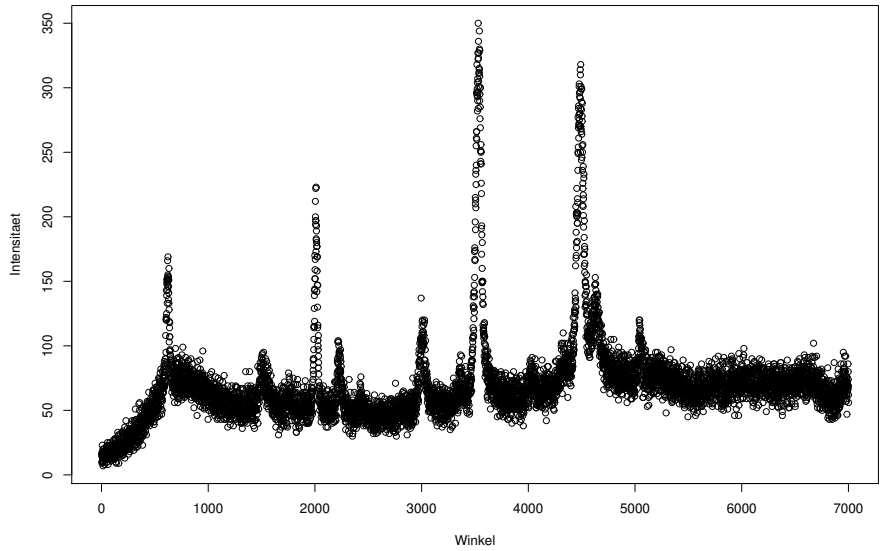

Abbildung 4: Daten aus der Dünnschichtphysik

$$\lim_{n\to\infty} \quad \frac{S_n}{n} = 0 \tag{1.1}$$

Das obere Bild der Abbildung 9 zeigt die Werte $S_j/j, j = 1, 100$ für den empirischen Münzwurf. Das untere Bild zeigt die Werte S_j/j, $j = 1, ...1000$ für die am Computer erzeugten Zahlen. Aus (1.1) folgt: $S_n = \mathrm{o}(n)$, und es fragt sich, was die "korrekte " Größenordnung von S_n ist. Die Antwort lautet $\sqrt{n}$ und Abbildung 10 zeigt die Werte $S_j/\sqrt{j}, j = 1, \cdots 100$ für die empirischen Daten (oberes Bild) sowie die Werte $S_j/\sqrt{j}, j = 1, \cdots 1000$ für die vom Computer erzeugten Zahlen (unteres Bild). Wir wollen uns den Prozess $S_j, j = 1, \cdots n$, selbst betrachten. Er heißt "die einfache Irrfahrt ". S_j gibt den Standpunkt einer Person zum Zeitpunkt j an. Sie wirft nun eine Münze und geht je nach Ergebnis einen Schritt nach links oder einen Schritt nach rechts. Wir fangen zum Zeitpunkt 0 im Nullpunkt an, und setzen $S_0 = 0$. Abbildung 11 zeigt die einfache Irrfahrt für die empirischen Ergebnisse (oberes Bild) sowie für die ersten 500 vom Computer erzeugten Werte (unteres Bild). Es ergeben sich mehrere Fragen über das Verhalten der einfachen Irrfahrt. Läuft sie beliebig weit nach unten und nach oben? - Ja.

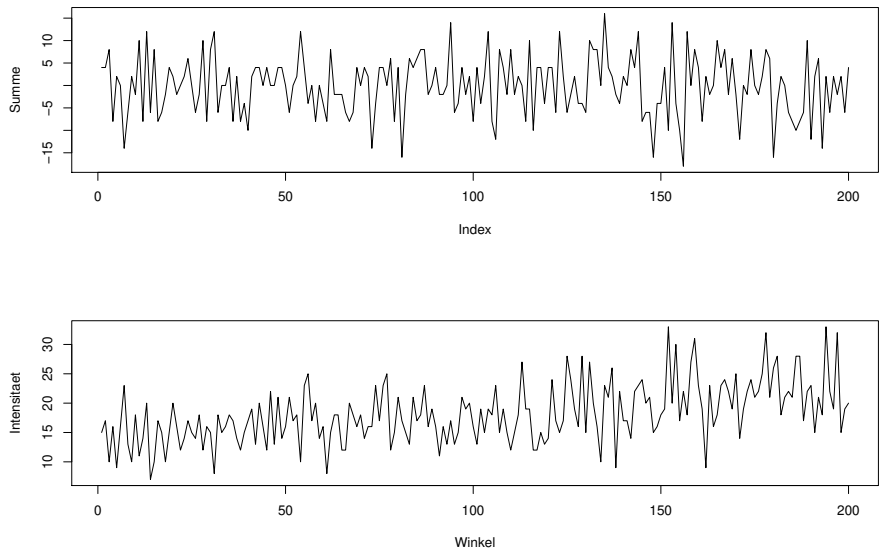

Abbildung 5: Das obere Bild ist das untere Bild von Abbildung 3. Das untere Bild zeigt die ersten 200 Werte von Abbildung 4

Wie lange dauert es zwischen aufeinander folgenden Besuchen im Nullpunkt? - Im Schnitt unendlich lange.

Die zweidimensionale Irrfahrt erhält man, in dem man die Punkte (S_j, S'_j), $j = 1, \cdots n$, von zwei eindimensionalen Irrfahrten plottet. Abbildung 12 zeigt eine zweidimensionale Irrfahrt. Ist die Person zum Zeitpunkt t im Punkt (i, j), so wählt sie zufällig einen der vier benachbarten Punkte $(i-1, j-1)$, $(i+1, j-1), (i-1, j+1), (i+1j+1)$ aus und geht dann dorthin zum Zeitpunkt t+1.
Wir normieren nun die einfache Irrfahrt, indem wir $S_j/\sqrt{n}, j = 0, \cdots n$ betrachten, aber wir plotten die Werte als eine Funktion auf dem Intervall [0,1]. $(j/n, S_j/\sqrt{n}), j = 0, \cdots n$. Abbildung 13 zeigt die Bilder fr n=500 (oberes Bild) und n=5000 (unteres Bild). Man hat den Eindruck einer stetigen Funktion, die aber nicht glatt ist. In der Tat kann man Konvergenz für $n \to \infty$ gegen eine stetige aber nicht differenzierbare Funktion auf [0,1] zeigen. Dazu muss man aber den Konvergenzbegriff richtig erfassen, was nicht ganz leicht ist. Die Grenzfunktion heißt die Brownsche Bewegung. Um die Frage zu beantworten, warum sie so heißt, mssen wir zuerst den zweidimensiona-

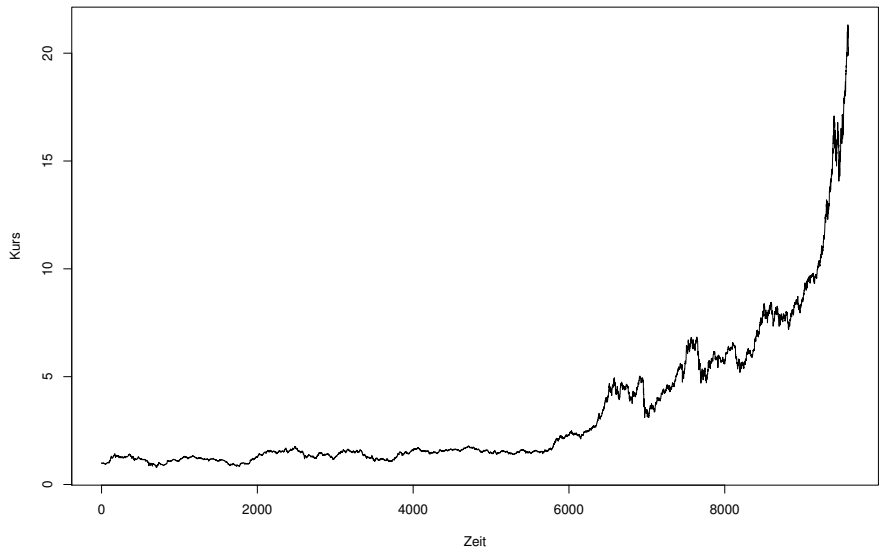

Abbildung 6: Die Tageswerte vom DAX über 30 Jahre

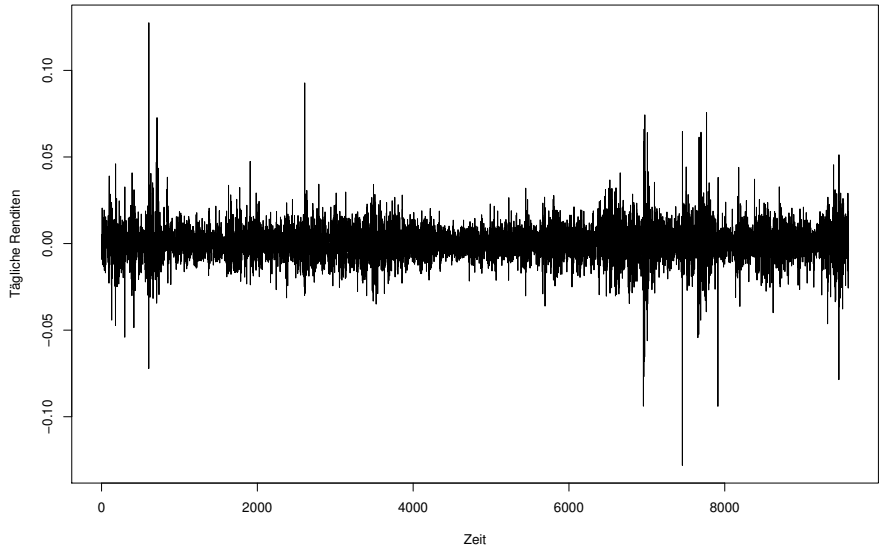

Abbildung 7: Daxrentditen; tägliche Renditen

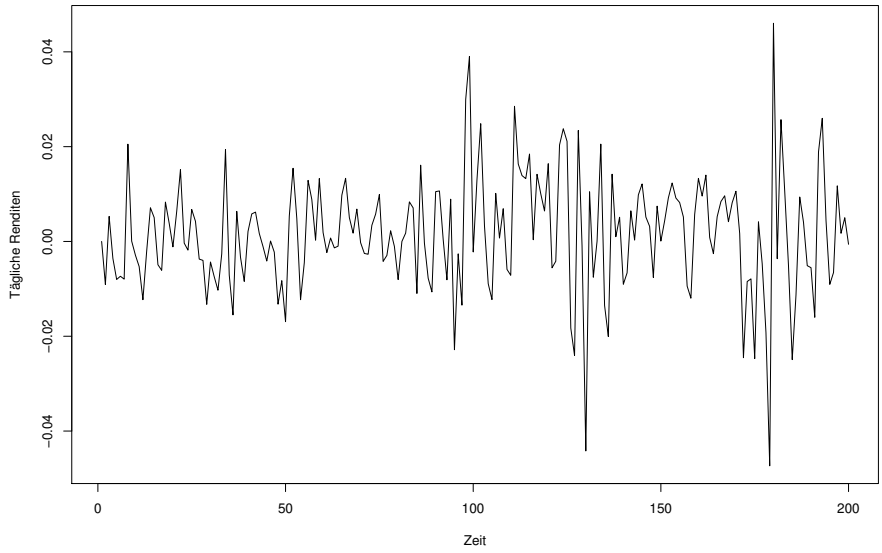

Abbildung 8: Die ersten 200 Werte von Abbildung 7

len Fall betrachten. Das haben wir schon gemacht in Abbildung 12, aber sie zeigt immer noch ganz deutlich die Werte ± 1. Wir benutzen also die Idee, die zur Abbildung 3 führte, und betrachten nun nur jeden fünften Zeitpunkt. Das Ergebnis ist Abbildung 14. So etwas hat der schottische Botaniker Robert Brown in 1828 gesehen, als er die Bewegung von Pollenteilchen auf der Oberfläche einer Flüssigkeit beobachtete. Die mathematische Theorie der Brownschen Bewegung hat erst im 20. Jahrhundert stattgefunden, und daran beteiligt waren viele bekannte Wissenschaftler Bachelier, Einstein, Lebesgue, Wiener, Daniel, Kolmogoroff, Lévy, Ito, Doob und andere.

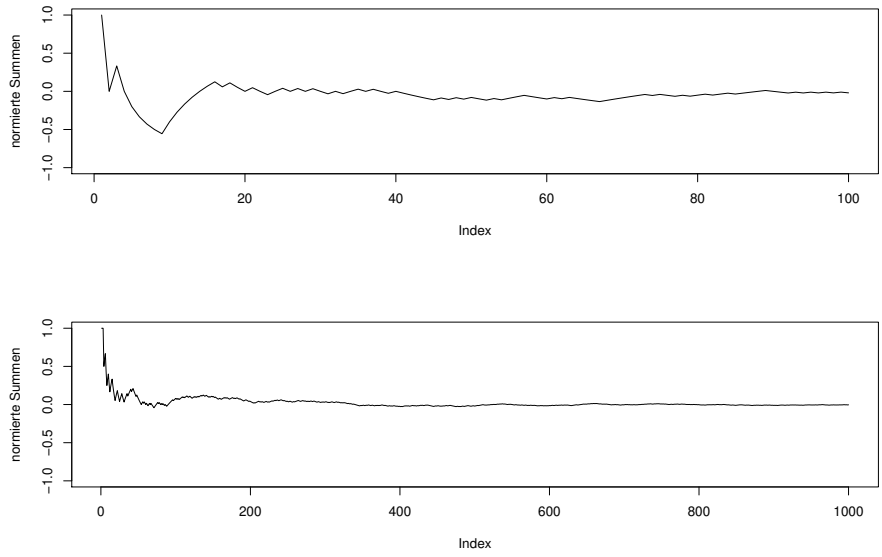

Abbildung 9: Laufende Mittelwerte der Daten von Abbildung 1 (oberes Bild) und von 1000 simulierten ± Werten (unteres Bild)

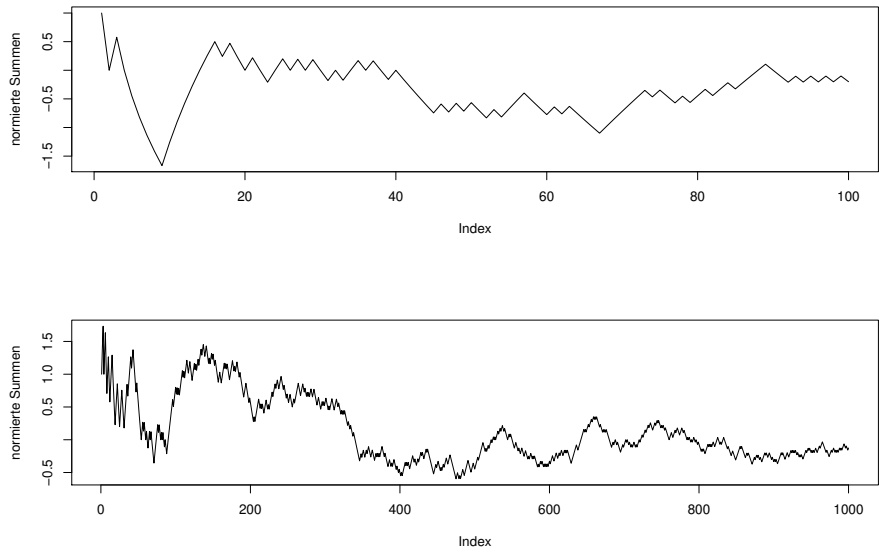

Abbildung 10: Die normierten Summen $S_j/\sqrt{j}$ für die Daten von Abbildung 9

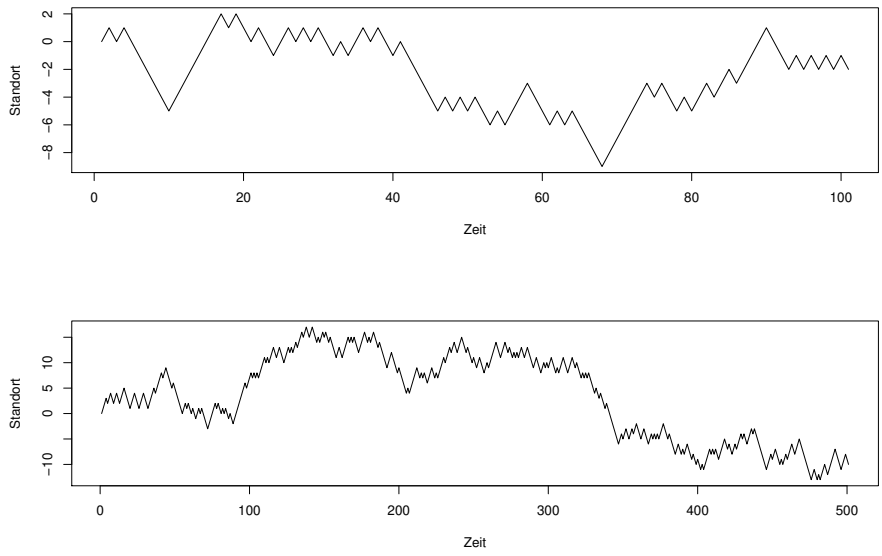

Abbildung 11: Einfache Irrfahrt für die Daten von Abbildung 1 (oberes Bild) sowie für simulante Daten (unteres Bild)

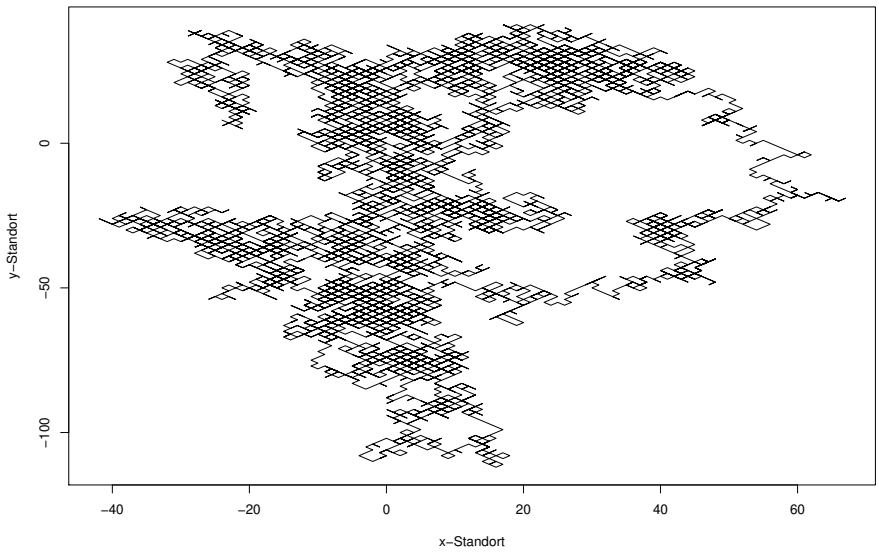

Abbildung 12: Eine zweidimensionale einfache Irrfahrt

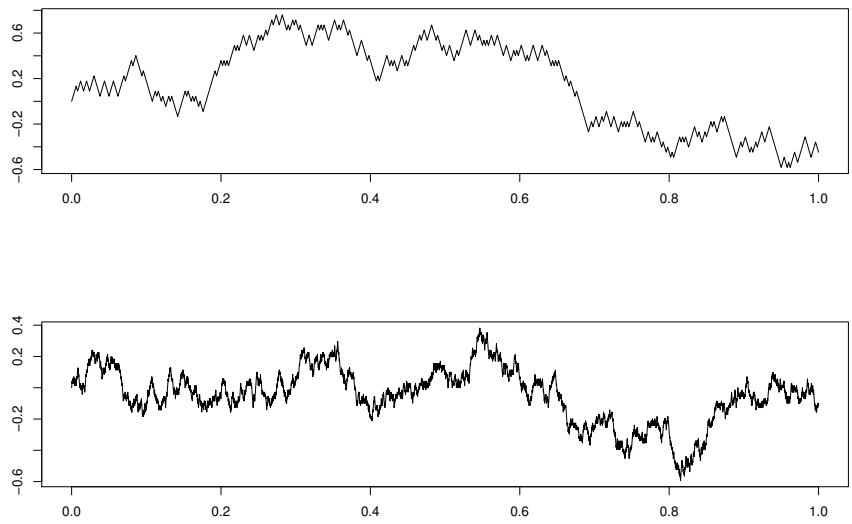

Abbildung 13: Die normierten Summen $S_j/\sqrt{n}, j = 0 \cdots n$. Oberes Bild n=500, unteres Bild n=5000.

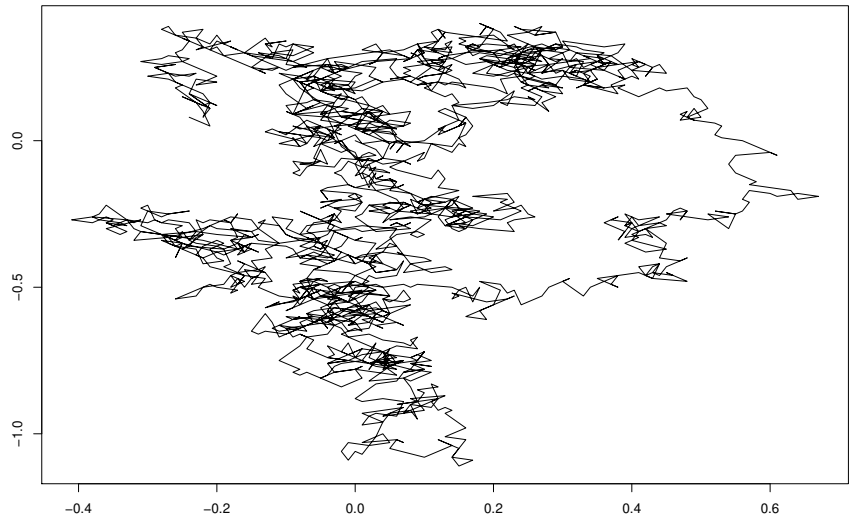

Abbildung 14: Eine zweidimensionale Irrfahrt gebildet aus Summen über Blöcke der Länge 5

1.2 Wahrscheinlichkeitsräume

Unser Modell fr den Zufall besteht aus drei Elementen, einem Stichprobenraum Ω, einer σ-Algebra $\mathcal{F}$ von Teilmengen von Ω und einem Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathcal{F}$.

Definition 1.1.
Das Dreitupel $(\Omega, \mathcal{F}, \mathbb{P})$ heißt ein **Wahrscheinlichkeitsraum**, *wenn folgende Eigenschaften erfüllt sind:*

(a) Ω *ist ein nichtleerer Raum.*

(b) $\mathcal{F}$ *ist eine σ–Algebra, d.h. ein System von Teilmengen von Ω mit folgenden Eigenschaften:*

(i) $\Omega \in \mathcal{F}$

(ii) $F \in \mathcal{F} \Rightarrow \Omega \setminus F \in \mathcal{F}$

(iii) $(F_j)_1^\infty \in \mathcal{F} \Rightarrow \cup_1^\infty F_j \in \mathcal{F}$.

(c) $\mathbb{P}$ *ist ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$, d.h. ist eine Abbildung*

$$\mathbb{P} : \mathcal{F} \to [0, 1]$$

mit folgenden Eigenschaften:

(i) $\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1.$

(ii) $(F_j)_1^\infty \in \mathcal{F}$ *paarweise disjunkt, d.h. $F_i \cap F_j = \emptyset, i \neq j$, so gilt*

$$\mathbb{P}(\cup_1^\infty F_j) = \sum_1^\infty \mathbb{P}(F_j).$$

Falls $(F_j)_1^\infty$ paarweise disjunkt sind, schreiben wir $\sum_1^\infty F_j$ anstelle von $\cup_1^\infty F_j$. Mit dieser Schreibweise können wir (ii) von (c) wie folgt schreiben:

$$\mathbb{P}(\sum_1^\infty F_j) = \sum_1^\infty \mathbb{P}(F_j).$$

Die Idee, Wahrscheinlichkeit als ein σ–additives normiertes Maß zu definieren, geht auf Kolmogoroff, Wiener und andere zurück. Manchmal werden die in (c) ausgeführten Eigenschaften die **Axiome von Kolmogoroff** genannt. Eine partielle Rechtfertigung für (c) liefert die folgende Überlegung. Es sei $n(A) =$ Anzahl des Vorkommens von A in n Versuchen. Wenn n wächst, sieht

es so aus, als ob die relative Häufigkeit $\frac{n(A)}{n}$ von A gegen einen Grenzwert strebt. Wir definieren $\mathbb{P}(A)$ durch

$$\mathbb{P}(A) = \lim_{n\to\infty} \frac{n(A)}{n}$$

Sind A und B disjunkt, so gilt $n(A+B) = n(A) + n(B)$ und somit

$$\mathbb{P}(A+B) = \lim_{n\to\infty} \frac{n(A+B)}{n} = \lim_{n\to\infty} \frac{n(A)+n(B)}{n} = \mathbb{P}(A) + \mathbb{P}(B).$$

Diese Überlegung zeigt nur, dass $\mathbb{P}$ endlich additiv ist. Die σ–Additivität läßt sich auf diese Art und Weise nicht begründen.

Die Wahrscheinlichkeitstheorie hat ihre eigene Sprache, die sich eng an die Alltagssprache hält.

1. Ω heißt der **Stichprobenraum** oder **Grundgesamtheit** und enthält alle möglichen Ergebnisse des Experiments.

2. $\omega \in \Omega$ heißt ein **Ergebnis**.

3. Jede Teilmenge F von Ω mit $F \in \mathcal{F}$ heißt ein **Ereignis**.

4. Für jedes Ereignis F ist $\mathbb{P}(F)$ die Wahrscheinlichkeit für das Eintreten von F .

5. Das Wort 'oder' wird durch Vereinigung beschrieben. $E = F \cup G$ ist das Ereignis, dass entweder F oder G (oder beides) eintritt.

6. Das Wort 'und' wird durch Durchschnitt beschrieben. $E = F \cap G$ ist das Ereignis, dass F und G gleichzeitig eintreten.

7. Das Wort 'nicht' wird durch das Komplement beschrieben. $E = \Omega \setminus F$ ist das Ereignis, dass F nicht eintritt.

Beispiel 1.2.
Ein Wahrscheinlichkeitsraum für den einmaligen Münzwurf läßt sich einfach konstruieren. Wir setzen

$$\Omega = \{0, 1\}$$

und

$$\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$$

$$\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\{1\}) = \frac{1}{2}, \quad \mathbb{P}(\{0\}) = \frac{1}{2}, \quad \mathbb{P}(\{0,1\}) = 1.$$

Die Interpretation ist naheliegend. Wir identifizieren 0 mit einem Kopfwurf und 1 mit einem Zahlwurf. Das Ereignis $\{0\}$ ist das Ereignis, dass ein Kopfwurf erzielt wird. Bei einer fairen Münze, die anständig geworfen wird, erwartet man, dass 'Kopf' mit Wahrscheinlichkeit $\frac{1}{2}$ erzielt wird. Wir setzen also $\mathbb{P}(\{0\}) = \frac{1}{2}$ und $\mathbb{P}(\{1\}) = \frac{1}{2}$.

Beispiel 1.3.
Der kanonische Wahrscheinlichkeitsraum für das einmalige Würfeln ist

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\mathcal{F} = \mathcal{P}(\Omega) = \{\phi, \{1\}, \ldots, \{6\}, \{1, 2\}, \{1, 3\}, \ldots$$
$$\{1, 2, 3\}, \ldots \{1, 2, 3, 4\}, \ldots \{1, 2, 3, 4, 5\}, \ldots \{1, 2, 3, 4, 5, 6\}\}.$$

Wir gehen davon aus, dass der Würfel fair ist und dass fair gewürfelt wird. Hieraus ergeben sich die Wahrscheinlichkeiten $\mathbb{P}(\{1\}) = \cdots = \mathbb{P}(\{6\}) = \frac{1}{6}$. Alle anderen Wahrscheinlichkeiten lassen sich hieraus bestimmen, so ist z.B. das Ereignis, dass die Augenzahl eine gerade Zahl ist,

$$F = \{2, 4, 6\} = \{2\} + \{4\} + \{6\}.$$

Es gilt somit

$$\begin{aligned} \mathbb{P}(\{2, 4, 6\}) &= \mathbb{P}(\{2\} + \{4\} + \{6\}) \\ &= \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{2}. \end{aligned}$$

Beispiel 1.4.
Ein Wahrscheinlichkeitsraum für den zweimaligen Münzwurf ist gegeben durch

$$\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$
$$\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{(0, 0)\}, \ldots \{(0, 0), \ldots (1, 1)\}\}$$

Die Interpretation ist wieder naheliegend. Das Element $\omega = (0, 1)$ bedeutet, dass der erste Wurf ein Kopfwurf und der zweite ein Zahlwurf ist. Es gibt insgesamt vier Möglichkeiten: $(0, 0), (0, 1), (1, 0), (1, 1)$ und vorausgesetzt, dass die Münze und die Art des Wurfes fair sind, ist es plausibel,

$$\mathbb{P}(\{(0, 0)\}) = \cdots = \mathbb{P}(\{(1, 1)\}) = \frac{1}{4}$$

zu setzen.

Beispiel 1.5.
Die Verallgemeinerung vom zweifachen zum n–fachen Münzwurf geschieht wie folgt. Wir setzen $\Omega = \{0,1\}^n, \mathcal{F} = \mathcal{P}(\Omega)$ und definieren $\mathbb{P}$ durch

$$\mathbb{P}(F) := \frac{|F|}{|\Omega|},$$

wobei $|F|$ die Anzahl von Elementen in F bezeichnet. Ein Element ω von Ω ist nun von der Form $\omega = (\epsilon_1, \ldots, \epsilon_n)$, wobei jedes ϵ_j entweder 0 oder 1 ist. Ferner ist ϵ_j genau dann 0 , wenn der $j$te Wurf ein Kopfwurf ist. Somit gibt ω die genaue Wurffolge an. Jede Teilmenge von Ω ist nun ein Ereignis. Die Zuordnung der Wahrscheinlichkeiten kann wie folgt begründet werden. Es gibt insgesamt $2^n = |\Omega|$ mögliche Reihenfolgen. Ist die Münze fair und wird sie auf eine faire Art und Weise geworfen, so haben alle Reihenfolgen dieselbe Wahrscheinlichkeit. Wir setzen also $\mathbb{P}(\{\omega\}) = 2^{-n} = 1/|\Omega|$. Für ein Ereignis F mit k Elementen $F = \{\omega_1, \ldots \omega_k\}$ gilt

$$\begin{aligned} \mathbb{P}(F) &= \mathbb{P}\left(\sum_{j=1}^{k} \{\omega_j\}\right) = \sum_{1}^{k} \mathbb{P}(\{\omega_j\}) \\ &= \sum_{1}^{k} 1/|\Omega| = k/|\Omega| = \frac{|F|}{|\Omega|}. \end{aligned}$$

Beispiel 1.6.
Dieses Beispiel ist grundsätzlich anders. Wir wollen einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ konstruieren, der die zufällige Wahl einer Zahl zwischen 0 und 1 beschreibt. Die vorangegangenen Beispiele beschreiben Experimente, die durchgeführt werden können. Es ist aber klar, dass es aufgrund der Unendlichkeit der Wahlmöglichkeiten nicht möglich ist, eine Zahl zwischen 0 und 1 zufällig zu ziehen. Wir können dieses Experiment nur in einem approximativen Sinne durchführen. Eine Möglichkeit ist die folgende. Wir werfen eine faire Münze 100 mal und setzen dann

$$\omega = \sum_{j=1}^{100} \epsilon_j 2^{-j}, \tag{1.2}$$

wobei $\epsilon_j = 0$ bzw 1 , falls das Ergebnis des j-ten Wurfes Kopf bzw. Zahl ist. Dies ist etwas mühsam und liefert nicht alle möglichen Zahlen zwischen 0 und 1. Sie können sich überlegen, ob Sie eine bessere Methode finden können. Die Darstellung (1.2) legt es nahe, einen unendlichen Münzwurf zu betrachten und

$$\omega = \sum_{j=1}^{\infty} \epsilon_j / 2^j \tag{1.3}$$

zu setzen. Es stellt sich aber heraus, dass, einen Wahrscheinlichkeitsraum für den unendlichen Münzwurf zu konstruieren, gleichbedeutend mit dem Problem ist, einen Wahrscheinlichkeitsraum für das ursprüngliche Problem zu konstruieren. Statt gemäß (1.3) gehen wir wie folgt vor: Wir setzen aus naheliegenden Gründen $\Omega = [0, 1]$. Wir gehen davon aus, dass die Teilmenge $(a, b]$ von $[0, 1]$ ein Ereignis ist. In der Tat: Ein Wahrscheinlichkeitsmodell, das es nicht zuläßt, von der Wahrscheinlichkeit dafür zu reden, dass die ausgewählte Zahl zwischen den Zahlen a und b liegt, wre für praktische Zwecke nutzlos. Ist nun das Intervall $(a, b]$ ein Ereignis, welche Wahrscheinlichkeit sollte diesem Ereignis zugeordnet werden? Wenn die Wahl rein zufällig ist, so sollte ein Intervall von der Länge l halb so wahrscheinlich sein, wie ein Intervall der Länge $2l$. Dies führt zu der Zuordnung

$$\mathbb{P}((a, b]) = b - a. \tag{1.4}$$

Gibt es nun eine σ-Algebra $\mathcal{F}$ von Teilmengen in $[0, 1]$, die alle Intervalle enthält und ein Wahrscheinlichkeitsmaß

$$\mathbb{P} : \mathcal{F} \to [0, 1]$$

das (1.4) erfüllt? Die Antwort ist ja. Wir setzen

$$\Omega = [0, 1], \quad \mathcal{F} = \mathcal{B}([0, 1]) = \text{ Borel } -\sigma - \text{Algebra auf } [0, 1]$$

und

$$\mathbb{P} = \text{ Lebesguemaß auf } [0, 1].$$

Es läßt sich unter Hinzunahme des Auswahlaxioms zeigen, dass es nicht möglich ist, $\mathbb{P}$ auf allen Teilmengen von $[0, 1]$ zu definieren und weiterhin (1.4) zu verlangen. Somit sind nicht alle Teilmengen von $[0, 1]$ Ereignisse.

1.3 Laplace-Experimente

Wir wollen nun die einfachsten stochastischen Modelle betrachten, die sogenannten Laplace- Experimente.

Definition 1.7.
$(\Omega, \mathcal{F}, \mathbb{P})$ *heißt ein* **Laplace-Experiment**, *falls*

(a) $|\Omega| < \infty$

(b) $\mathcal{F} = \mathcal{P}(\Omega)$

(c) $\mathbb{P}(F) = \frac{|F|}{|\Omega|}$

Einigen Beispielen sind wir schon begegnet: Dem einmaligen Würfeln sowie dem n-fachen Münzwurf mit einer fairen Münze. Bei einem Laplace-Experiment haben alle Elementarereignisse $\{\omega\}$ dieselbe Wahrscheinlichkeit, nämlich $\frac{1}{|\Omega|}$.
Die Schwierigkeit bei Laplace-Experimenten besteht darin, $|F|$ und $|\Omega|$ zu bestimmen. Dies führt zu kombinatorischen berlegungen:

Bemerkung 1.8.
Es seien $\{O_1, \ldots, O_n\}$ verschiedene Objekte. Es gilt folgendes:

1. Es gibt n^r geordnete Stichproben vom Umfang r mit Zurücklegen
2. Es gibt $n(n-1)\ldots(n-r+1)$ geordnete Stichproben vom Umfang r ohne Zurücklegen
3. Es gibt $\binom{n+r-1}{r}$ ungeordnete Stichproben vom Umfang r mit Zurücklegen
4. Es gibt $\binom{n}{r}$ ungeordnete Stichproben vom Umfang r ohne Zurücklegen.

Beispiel 1.9.
Eine Sendung von N Artikeln enthält genau L fehlerhafte Stücke. Eine Stichprobe vom Umfang n wird ohne Zurücklegen der Sendung entnommen. Wie groß ist die Wahrscheinlichkeit dafür, dass die Stichprobe genau l fehlerhafte Stücke enthält?

$$\mathbb{P}\ (\text{ genau } l \text{ fehlerhaft }) = \frac{\binom{N-L}{n-l}\binom{L}{l}}{\binom{N}{n}}\{0 \le l \le L\}$$

N, n, l sind bekannt. Wie würden Sie L schätzen?

Beispiel 1.10.
Ein Skatspieler hat 5 Trumpfkarten auf der Hand (einschließlich Buben). Wie groß ist die Wahrscheinlichkeit dafür, dass der Skat mindestens 1 weitere Trumpfkarte enthält? Der Spieler hat 10 Karten auf der Hand, davon sind 5 Trumpfkarten. Es bleiben 22 Karten übrig, wovon 6 Trumpfkarten sind. Es gilt

$$\begin{aligned}
\mathbb{P}(\text{mindestens 1 Trumpfkarte}) &= \mathbb{P}\ (\text{genau 1}) + \mathbb{P}\ (\text{genau 2}) \\
&= \frac{\binom{6}{1}\binom{16}{1}}{\binom{22}{2}} + \frac{\binom{6}{2}\binom{16}{0}}{\binom{22}{2}} \qquad (\text{warum?}) \\
&= \frac{6 \cdot 16 + \frac{6\cdot 5}{2}}{\frac{22\cdot 21}{2}} = \frac{96+15}{231} = \frac{111}{231} \approx 0.48
\end{aligned}$$

2 Abhängigkeit und Unabhängigkeit

2.1 Abhängigkeit

Der Begriff von Abhängigkeit unterscheidet die Wahrscheinlichkeitstheorie von der Maß– und Integrationstheorie. Um den Begriff anschaulich zu machen, fangen wir mit einem bekannten Beispiel an, dem Zusammenhang zwischen Lungenkrebs und Rauchen. Dazu werden n Personen untersucht. Es seien

- n_L = Anzahl Personen, die unter Lungenkrebs leiden
- n_R = Anzahl Personen, die rauchen
- $n_{L\cap R}$ = Anzahl Personen, die rauchen und unter Lungenkrebs leiden.

Die bedingte Wahrscheinlichkeit dafür, dass ein Raucher unter Lungenkrebs leidet, d.h. dass jemand unter Lungenkrebs leidet gegeben, dass er raucht, kann man wie folgt heuristisch herleiten.

$$\mathbb{P}(L|R) = \lim_{n\to\infty} \frac{n_{L\cap R}}{n_R} = \lim_{n\to\infty} \left(\frac{n_{L\cap R}}{n}\right) / \left(\frac{n_R}{n}\right) = \mathbb{P}(L\cap R)/\mathbb{P}(R)$$

Dies führt zu folgender Definition.

Definition 2.1.
Es seien A und B Ereignisse mit $\mathbb{P}(B) > 0$. Die bedingte Wahrscheinlichkeit für A gegeben B, $\mathbb{P}(A|B)$, wird definiert durch

$$\mathbb{P}(A|B) := \mathbb{P}(A\cap B)/\mathbb{P}(B).$$

Der Nutzen von bedingten Wahrscheinlichkeiten besteht darin, dass sie sehr oft unmittelbar interpretierbar sind.

Satz 2.2. *(Satz der totalen Wahrscheinlichkeit)*
Es seien $(B_j)_{j\in J}$, J abzählbar viele, disjunkte Ereignisse mit $\mathbb{P}(B_j) > 0, j \in J$, und $\sum_{j\in J} B_j = \Omega$. Dann gilt

$$\mathbb{P}(A) = \sum_{j\in J} \mathbb{P}(A|B_j)\cdot\mathbb{P}(B_j).$$

Beweis: Es gilt

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}(A \cap \sum_{j \in J} B_j) = \mathbb{P}(\sum_{j \in J} A \cap B_j)$$

und aus der σ–Additivität von Wahrscheinlichkeitsmaßen folgt nun

$$\mathbb{P}(A) = \sum_{j \in J} \mathbb{P}(A \cap B_j) = \sum_{j \in J} \mathbb{P}(A|B_j)\mathbb{P}(B_j). \quad \square$$

Beispiel 2.3.
Als einfaches Beispiel betrachten wir 6 Urnen, die Kugeln wie folgt enthalten.

Urne	Anzahl Kugeln
1	3 weiße, 1 rote Kugel
2 und 3	1 weiße, 1 rote Kugel
4, 5 und 6	2 weiße, 1 rote Kugel

Eine Urne wird zufällig ausgewählt und eine Kugel der Urne zufällig entnommen. Wie groß ist die Wahrscheinlichkeit dafür, dass es sich um eine weiße Kugel handelt? Wir setzen

B_1	Urne 1
B_2	Urne 2 oder 3
B_3	Urne 4, 5 oder 6

Nach dem Satz der totalen Wahrscheinlichkeit gilt

$$\begin{aligned}
\mathbb{P}(\text{weiß}) &= \sum \mathbb{P}(\text{weiß}|B_j)\mathbb{P}(B_j) \\
&= \mathbb{P}(\text{weiß}|B_1)\mathbb{P}(B_1) + \mathbb{P}(\text{weiß}|B_2)\mathbb{P}(B_2) + \mathbb{P}(\text{weiß}|B_3)\mathbb{P}(B_3) \\
&= \frac{3}{4} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{2} \\
&= \frac{3+4+8}{24} = \frac{15}{24}
\end{aligned}$$

Der nächste Satz spielt eine große, aber auch sehr umstrittene Rolle in der Statistik. Als mathematischer Satz ist er aber hiervon unberührt.

Satz 2.4. *(Satz von Bayes)*
Es seien $(B_j)_{j \in J}, J$ abzählbar viele, disjunkte Ereignisse mit $\sum_{j \in J} B_j = \Omega$ und $\mathbb{P}(B_j) > 0, \forall j$. Dann gilt für jedes Ereignis A mit $\mathbb{P}(A) > 0$

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}{\sum_{i \in J} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

Beweis: Aus der Definition der bedingten Wahrscheinlichkeit folgt

$$\mathbb{P}(B_j|A) = \mathbb{P}(B_j \cap A)/\mathbb{P}(A).$$

Aus dem Satz der totalen Wahrscheinlichkeit leiten wir nun ab

$$\begin{aligned}\mathbb{P}(B_j|A) &= \mathbb{P}(B_j \cap A)/\sum_{i\in J}\mathbb{P}(A|B_i)\mathbb{P}(B_i) \\ &= \mathbb{P}(A|B_j)\mathbb{P}(B_j)/\sum_{i\in J}\mathbb{P}(A|B_i)\mathbb{P}(B_i)\end{aligned}$$

wobei wir nochmals die Definition der bedingten Wahrscheinlichkeit benutzt haben. □

Manchmal führt die Anwendung des Satzes von Bayes zu nicht erwarteten Ergebnissen.

Beispiel 2.5.
Wir betrachten 3 Truhen, die eine enthält 1 Goldmünze und 1 Silbermünze, die zweite 2 Goldmünzen und die dritte 2 Silbermünzen. Eine Truhe wird zufällig gewählt und der Truhe eine Münze zufällig entnommen. Es handelt sich um eine Goldmünze. Wie groß ist die Wahrscheinlichkeit dafür, dass die zweite Münze auch eine Goldmünze ist? Wir setzen

B_1	Truhe 1: Silbermünze, Goldmünze
B_2	Truhe 2: Goldmünze, Goldmünze
B_3	Truhe 3: Silbermünze, Silbermünze
A	erste Münze ist eine Goldmünze

Aus dem Satz von Bayes folgt

$$\begin{aligned}\mathbb{P}(B_2|A) &= \frac{\mathbb{P}(A|B_2)\cdot\mathbb{P}(B_2)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1)+\mathbb{P}(A|B_2)\mathbb{P}(B_2)+\mathbb{P}(A|B_3)\mathbb{P}(B_3)} \\ &= \frac{1\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3}+1\cdot\frac{1}{3}+0\cdot\frac{1}{3}} = \frac{2}{1+2} = \frac{2}{3}\end{aligned}$$

Hätten Sie eher auf $\frac{1}{2}$ getippt?

Beispiel 2.6.
Ein Medikament zur Linderung bestimmter Beschwerden wird in einem kleinen Versuch mit folgenden Ergebnissen getestet.

	Verbesserung	Keine Verbesserung	Erfolgsrate
mit Medikament	20	20	50%
ohne Medikament	16	24	40%

Auf den ersten Blick sieht es so aus, als ob das Medikament geholfen hat. Die Versuchspersonen werden nun nach Geschlecht mit folgenden Ergebnissen aufgeteilt.

Mnner

	Verbesserung	Keine Verbesserung	Erfolgsrate
mit Medikament	18	12	60%
ohne Medikament	7	3	70%

Frauen

	Verbesserung	Keine Verbesserung	Erfolgsrate
mit Medikament	2	8	20%
ohne Medikament	9	21	30%

Danach hat das Experiment zu einer Verschlechterung sowohl bei den Frauen als auch bei den Mnnern gefhrt, insgesamt aber zu einer Verbesserung. Sollte das Medikament eingefhrt werden oder nicht?

2.2 Unabhängigkeit

Wir wollen nun den Begriff von Unabhängigkeit definieren. Dazu betrachten wir wieder das Beispiel von Rauchen und Lungenkrebs. Ist Lungenkrebs unabhängig vom Rauchen, so würden wir erwarten, dass gilt $\mathbb{P}(L|R) = \mathbb{P}(L)$, d.h., die Wahrscheinlichkeit dafür, dass man unter Lungenkrebs leidet, hängt nicht vom Rauchen ab. Hieraus folgt

$$\frac{\mathbb{P}(L \cap R)}{\mathbb{P}(R)} = \mathbb{P}(L)$$

und somit

$$\mathbb{P}(L \cap R) = \mathbb{P}(L)\mathbb{P}(R).$$

Dies führt zu folgender Definition.

Definition 2.7.
Zwei Ereignisse A und B heißen unabhängig, falls gilt

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Bemerkung 2.8.
Es ist nicht notwendig, dass $\mathbb{P}(A) > 0$ oder $\mathbb{P}(B) > 0$.

Beispiel 2.9.
Ein einfaches Beispiel liefert das zweimalige Würfeln. Wir setzen

$$\Omega = \{(1,1), \ldots, (6,6)\}.$$

Es gilt

$$\begin{aligned} A &= \{ \text{ 1. Augenzahl 6 } \} = \{(6,1), (6,2), \ldots, (6,6)\} \\ B &= \{ \text{ 2. Augenzahl 6 } \} = \{(1,6), (2,6), \ldots, (6,6)\} \end{aligned}$$

Hieraus folgt

$$\mathbb{P}(A) = \frac{6}{36} = \frac{1}{6} \text{ und } \mathbb{P}(B) = \frac{6}{36} = \frac{1}{6}.$$

Wegen $A \cap B = \{(6,6)\}$ gilt

$$\mathbb{P}(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(B).$$

Somit sind A und B unabhängig.

Bemerkung 2.10.
Sind die Ereignisse A, B unabhängig, so folgt, dass $(\Omega \setminus A)$ und B unabhängig sind. Denn es gilt

$$\begin{aligned} \mathbb{P}((\Omega \setminus A) \cap B) &= \mathbb{P}(B \setminus (A \cap B)) = \mathbb{P}(B) - \mathbb{P}(A \cap B), (A \cup B \subset B) \\ &= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \quad (A \text{ und } B \text{ unabhängig}) \\ &= (1 - \mathbb{P}(A))\mathbb{P}(B) \\ &= \mathbb{P}(\Omega \setminus A)\mathbb{P}(B). \end{aligned}$$

Man kann sich leicht überlegen, dass $(\Omega \setminus A)$ und $(\Omega \setminus B)$ auch unabhängig sind.
Wie definiert man Unabhängigkeit für drei oder mehr Ereignisse? Dies ist etwas schwieriger, als man zuerst denkt.

Beispiel 2.11.
Betrachten wir den zweifachen Münzwurf und setzen $A =$ erster Wurf 'Kopf', $B =$ zweiter Wurf 'Kopf', $C =$ Würfe sind verschieden. Die Ereignisse A und B sind unabhängig, denn es gilt

$$A = \{(0,1), (0,0)\}, \qquad B = \{(0,0), (1,0)\}$$

und somit

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}.$$

Wegen $A \cap B = \{(0,0)\}$ folgt $\mathbb{P}(A \cap B) = \frac{1}{4}$ und somit

$$\mathbb{P}(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Somit sind A und B unabhängig. Ganz analog kann man zeigen, dass A und C und B und C unabhängig sind. Die naheliegende Interpretation, dass A keine Information über C enthält und B auch keine Information über C enthält, ist mit Vorsicht zu genießen. Denn A und B zusammen enthalten Informationen über C. Es gilt nämlich $A \cap B \cap C = \emptyset$. Die folgende Definition hat sich als 'richtig' herausgestellt.

Definition 2.12.

1. *Es seien $F_1, \ldots, F_n$ Ereignisse. Sie heißen stochastisch unabhängig, falls gilt*
$$\mathbb{P}(F_{i_1} \cap \cdots \cap F_{i_k}) = \Pi_{j=1}^{k} \mathbb{P}(F_{i_j})$$
für jede Teilmenge $\{i_1, \ldots, i_k\}$ von $\{1, \ldots, n\}$.

2. *Es seien $(F_i)_{i \in I}$ Ereignisse, I eine beliebige Indexmenge. Die $(F_i)_{i \in I}$ heißen stochastisch unabhängig, falls gilt*
$$\mathbb{P}(F_{i_1} \cap \cdots \cap F_{i_k}) = \Pi_{j=1}^{k} \mathbb{P}(F_{i_j})$$
für jede endliche Teilmenge $\{i_1, \ldots, i_k\}$ von I.

Mit Hilfe des Begriffes der Unabhängigkeit kann man Situationen modellieren, in denen man meint, dass es keine gegenseitige Beeinflussung gibt, oder man wissen will, ob es eine gegenseitige Beeinflussung gibt.

Beispiel 2.13.
Eine faire Münze wird n–mal nacheinander geworfen, die einzelnen Würfe seien voneinander unabhängig, und die Wahrscheinlichkeit für Kopf sei stets p, $0 < p < 1$. Wir betrachten nun eine fest vorgegebene Reihenfolge, z.B. $(K, K, Z, Z, Z, K, \ldots, Z, K)$. Unter der Annahme der Unabhängigkeit beträgt die Wahrscheinlichkeit für genau diese Reihenfolge

$$p \cdot p \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot p \ldots (1-p) \cdot p.$$

Enthält diese Reihe genau k 'Kopfwürfe', so folgt, dass die Wahrscheinlichkeit $p^k(1-p)^{n-k}$ beträgt. Wir setzen

$$\Omega = \{0,1\}^n = \{(0,\ldots,0),(1,0,\ldots,0),(1,\ldots,1)\}$$

und

$$\mathbb{P}(\{(\epsilon_1,\ldots,\epsilon_n)\}) := p^k q^{n-k}, \quad k = \sum_1^n \epsilon_j.$$

Wir können nun die Wahrscheinlichkeit für genau k 'Kopfwürfe' berechnen. Wir setzen $F = \{$ genau k–mal Kopf $\}$. Jede Reihe mit k 'Kopfwürfen' hat die Wahrscheinlichkeit $p^k(1-p)^{n-k}$ und es gibt $\binom{n}{k}$ solcher Reihen. Somit folgt

$$\mathbb{P}(F) = \binom{n}{k} p^k q^{n-k}$$

.

2.3 Unabhängige Koppelungen von Experimenten

Wir wollen nun ein mathematisches Modell für die unabhängige Koppelung von Experimenten konstruieren. Als Beispiel betrachten wir einen Münzwurf gefolgt von Würfeln. Wie kann man dieses Experiment unabhängig koppeln? Wir setzen

$$\text{Experiment } 1 = (\Omega_1, \mathcal{F}_1, \mathbb{P}_1) - \text{ Münzwurf}$$

und

$$\text{Experiment } 2 = (\Omega_2, \mathcal{F}_2, \mathbb{P}_2) - \text{ Würfeln.}$$

Betrachten wir den Produktraum $\Omega = \Omega_1 \times \Omega_2$ mit Elementen $\omega = (\omega_1, \omega_2)$, wobei $\omega_1 \in \Omega_1$ und $\omega_2 \in \Omega_2$.
Es sei nun

$$F = F_1 \times F_2 = \{(\omega_1, \omega_2) : \omega_1 \in F_1, \omega_2 \in F_2\}.$$

Dies kann man so interpretieren, dass beim ersten Experiment F_1 eintritt und dass beim zweiten Experiment F_2 eintritt. Mengen der Gestalt $F_1 \times F_2$ mit $F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2$ heißen Zylindermengen. Wir bezeichnen mit $\mathcal{F}$ die kleinste σ –Algebra, die alle Zylindermengen enthält.

Definition 2.14.
Es sei $\mathcal{C}$ eine Familie von Teilmengen von Ω. Die von $\mathcal{C}$ erzeugte σ-Algebra, $\sigma - \mathcal{C}$ wird definiert als die kleinste σ-Algebra, die $\mathcal{C}$ enthält. Es gilt

$$\sigma - \mathcal{C} = \cap\{\mathcal{G} : \mathcal{C} \subset \mathcal{G}, \mathcal{G} \text{ eine } \sigma - \text{Algebra}\}. \tag{2.1}$$

Die rechte Seite enthält $\mathcal{C}$ und, da alle $\mathcal{G}$ $\sigma-$Algebren sind, ist die rechte Seite auch eine σ-Algebra. Ferner ist nach Konstruktion die linke Seite in jeder σ-Algebra enthalten, die $\mathcal{C}$ enthält.
Wir wollen nun ein Wahrscheinlichkeitsmaß $\mathbb{P} : \mathcal{F} \to [0,1]$ mit folgenden Eigenschaften

$$\mathbb{P}(F_1 \times F_2) := \mathbb{P}_1(F_1)\mathbb{P}_2(F_2) \tag{2.2}$$

konstruieren.

Satz 2.15. *Es existiert ein eindeutig definiertes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathcal{F}$, das (2.2) erfüllt.*

Der Raum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt der Produktraum, $\Omega = \Omega_1 \times \Omega_2$, die σ-Algebra $\mathcal{F}$ schreiben wir als $\mathcal{F}_1 \otimes \mathcal{F}_2$. Sie heißt die Produkt σ-Algebra. Das Wahrscheinlichkeitsmaß in Satz 2.15 heißt das Produktmaß, und wir schreiben $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$.
Der Produktraum $(\Omega, \mathcal{F}, \mathbb{P})$ ist ein Modell für die unabhängigige Koppelung der Experimente 1 und 2. Um dies zu sehen, betrachten wir $F_1 \in \mathcal{F}_1$ und $F_2 \in \mathcal{F}_2$. Wir schreiben

$$\tilde{F}_1 = F_1 \times \Omega_2, \qquad \tilde{F}_2 = \Omega_1 \times F_2.$$

$\tilde{F}_1$ und $\tilde{F}_2$ haben folgende Interpretationen. $\tilde{F}_1$ ist das Ereignis im gekoppelten Experiment, dass das Ereignis F_1 beim ersten Experiment eintritt. $\tilde{F}_2$ ist das Ereignis in gekoppeltem Experiment, dass das Ereignis F_2 beim zweiten Experiment eintritt. Es gilt mit dem Produktmaß $\mathbb{P}$

$$\mathbb{P}(\tilde{F}_1) = \mathbb{P}(F_1 \times \Omega_2) = \mathbb{P}_1(F_1)\mathbb{P}_2(\Omega_2) = \mathbb{P}_1(F_1).$$

Analog gilt $\mathbb{P}(\tilde{F}_2) = \mathbb{P}_2(F_2)$ und es folgt

$$\mathbb{P}(\tilde{F}_1 \cap \tilde{F}_2) = \mathbb{P}(F_1 \times F_2) = \mathbb{P}_1(F_1)\mathbb{P}_2(F_2) = \mathbb{P}(\tilde{F}_1)\mathbb{P}(\tilde{F}_2)$$

In der Tat sind also $\tilde{F}_1$ und $\tilde{F}_2$ unabhängig. $(\Omega, \mathcal{F}, \mathbb{P})$ ist somit ein Modell für die unabhängige Koppelung der beiden Experimente.
Wir können dies nun leicht verallgemeinern auf endlich viele Experimente. Es seien $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, $1 \leq i \leq n$, Wahrscheinlichkeitsräume. Wir setzen $\Omega = \Pi_1^n \Omega_j$ und $\mathcal{F} = \otimes_1^n \mathcal{F}_j$, d.h. die kleinste σ-Algebra, die alle Mengen der Gestalt $\Pi_1^n F_j$ mit

$$F_j \in \mathcal{F}_j, \quad 1 \leq j \leq n,$$

enthält.

Satz 2.16. *Es existiert ein eindeutig definiertes Wahrscheinlichkeitsmaß* $\mathbb{P}$ *auf*

$$\mathcal{F} = \otimes_1^n \mathcal{F}_j$$

mit

$$\mathbb{P}(\Pi_1^n F_j) = \Pi_1^n \mathbb{P}_j(F_j)$$

für alle Zylindermengen $(F_j)_l^n$.

Wir schreiben $\mathbb{P} = \otimes_1^n \mathbb{P}_j$ und $(\Omega, \mathcal{F}, \mathbb{P}) = \otimes_1^n (\Omega_j, \mathcal{F}_j, \mathbb{P}_j)$.

Wir zeigen, dass $(\Omega, \mathcal{F}, \mathbb{P})$ ein Modell für die unabhängige Koppelung der Experimente $(\Omega_j, \mathcal{F}_j, \mathbb{P}_j)$ ist. Wir schreiben

$$\tilde{F}_j = \Omega_1 \times \cdots \times \Omega_{j-1} \times F_j \times \Omega_{j+1} \times \cdots \times \Omega_n$$

mit $F_j \in \mathcal{F}_j$. $\tilde{F}_j$ ist das Ereignis in den gekoppelten Experimenten, dass das Ereignis F_j beim j-ten Experiment eintritt. Es gilt

$$\begin{aligned} \mathbb{P}(\tilde{F}_j) &= \mathbb{P}(\Omega_1 \times \cdots \times \Omega_{j-1} \times F_j \times \Omega_{j+1} \times \cdots \times \Omega_n) \\ &= \mathbb{P}_1(\Omega_1) \cdots \mathbb{P}_{j-1}(\Omega_{j-1}) \mathbb{P}_j(F_j) \mathbb{P}_{j+1}(\Omega_{j+1}) \cdots \mathbb{P}_n(\Omega_n) \\ &= \mathbb{P}_j(F_j) \end{aligned}$$

Es sei nun $\{j_1, \ldots, j_k\}$ eine Teilmenge von $\{1, \ldots, n\}$. Es gilt

$$\tilde{F}_{j_1} \cap \cdots \cap \tilde{F}_{j_k} = \Pi_1^n G_j,$$

wobei

$$G_j = \begin{cases} F_j, & j_{\in}\{j_1, \ldots, j_k\} \\ \Omega_j, & j \notin \{j_1, \ldots, j_k\} \end{cases}$$

$\tilde{F}_j$ hat genau wie vorher die folgende Interpretation: das Ereignis $\tilde{F}_j$ ist das Ereignis im gekoppelten Raum, dass beim j-ten Experiment F_j eintritt. Es gilt

$$\mathbb{P}(\tilde{F}_{j_1} \cap \cdots \cap \tilde{F}_{j_k}) = \mathbb{P}(\Pi_1^n G_j) = \Pi_1^n \mathbb{P}_j(G_j) = \Pi_{i=1}^k \mathbb{P}_{j_i}(F_{j_i}) = \Pi_{i=1}^k \mathbb{P}(\tilde{F}_{j_i})$$

Somit sind die Ereignisse $\tilde{F}_1, \ldots, \tilde{F}_n$ unabhängig.

Wir können dies fortsetzen zu abzählbar unendlichen Produkten. Das folgende Beispiel zeigt, dass dies notwendig ist, um auch einfache Situationen zu beschreiben.

Beispiel 2.17.
Eine faire Münze wird so lange geworfen, bis zum ersten Mal 'Kopf' vorkommt. Dieses Experiment kann nicht durch ein Modell, das nur endlich viele Würfe zuläßt, beschrieben werden.

Die Experimente sind durch die Räume $(\Omega_j, \mathcal{F}_j, \mathbb{P}_j), 1 \leq j < \infty$, beschrieben. Wir setzen $\Omega = \Pi_1^\infty \Omega_j$ und $\mathcal{F} = \otimes_1^\infty \mathcal{F}_j$ d.h. die kleinste σ-Algebra, die alle Zylindermengen enthält. Was ist aber eine Zylindermenge? Eine Zylindermenge ist eine Menge F von der Gestalt

$$F = \Pi_1^\infty F_j$$

mit $F_j = \Omega_j$ für alle bis auf endlich viele j-Werte. Dies ist wichtig. Wie im endlichen Fall setzen wir

$$\tilde{F}_j = \Pi_1^{j-1} \Omega_i \times F_j \times \Pi_{j+1}^\infty \Omega_i, \quad F_j \in \mathcal{F}_j$$

$\tilde{F}_j$ ist nach wie vor das Ereignis in dem gekoppelten Experiment, dass im j-ten Experiment F_j eintritt.

Satz 2.18. *Es existiert ein eindeutig definiertes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathcal{F} = \otimes_1^\infty \mathcal{F}_j$ mit*

$$\mathbb{P}(\Pi_1^\infty F_j) = \Pi_1^\infty \mathbb{P}(F_j)$$

für alle Zylindermengen $\Pi_1^\infty F_j$.

Die Ereignisse $(\tilde{F}_j)_1^\infty$ sind stochastisch unabhängig. Denn es gilt

$$\tilde{F}_{j_1} \cap \cdots \cap \tilde{F}_{j_k} = \Pi_1^\infty G_j$$

mit

$$G_j = \begin{cases} F_j, & j \in \{j_1, \ldots, j_k\} \\ \Omega_j, & j \notin \{j_1, \ldots, j_k\}. \end{cases}$$

Hieraus folgt

$$\mathbb{P}(\tilde{F}_{j_1} \cap \cdots \cap \tilde{F}_{j_k}) = \mathbb{P}(\Pi_1^\infty G_j) = \Pi_1^\infty \mathbb{P}_j(G_j) = \Pi_1^k \mathbb{P}_{j_i}(F_{j_i}) = \Pi_1^k \mathbb{P}(\tilde{F}_{j_i}),$$

denn für alle bis auf endlich viele j-Werte gilt $\mathbb{P}_j(G_j) = 1$. Somit sind die Ereignisse $(\tilde{F}_j)_1^\infty$ unabhängig. $(\Omega, \mathcal{F}, \mathbb{P})$ ist ein Modell für die unabhängige Koppelung der Experimente $(\Omega_j, \mathcal{F}_j, \mathbb{P}_j), \quad 1 \leq j \leq \infty$. Wir schreiben $\mathbb{P} = \otimes_1^\infty \mathbb{P}_j$ und

$$(\Omega, \mathcal{F}, \mathbb{P}) = \otimes_1^\infty (\Omega_j, \mathcal{F}_j, \mathbb{P}_j).$$

Beispiel 2.19.
Wir können den unendlichen Münzwurf wie folgt modellieren. Wir setzen

$$\Omega_j = \{0, 1\}, \quad \mathcal{F}_j = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}, \quad \mathbb{P}_j(\{0\}) = \mathbb{P}_j(\{1\}) = \frac{1}{2}$$

Der Produktraum ist gegeben durch

$$\begin{aligned} \Omega &= \{\delta : \delta = (\delta_1, \delta_2, \cdots), \quad \delta_j \in \{0,1\}\} \\ \mathcal{F} &= \otimes_1^\infty \mathcal{P}(\{0,1\}) \end{aligned}$$

und

$$\mathbb{P}(\{\delta_1\} \times \cdots \times \{\delta_n\} \times \Pi_{n+1}^\infty \{0,1\}) = 2^{-n}.$$

2.4 Unabhängige Mengensysteme

In Kapitel 2.2 haben wir gesehen, dass aus F und G unabhängig folgt, dass $\Omega \setminus F$ und G,F und $\Omega \setminus G$ und $\Omega \setminus F$ und $\Omega \setminus G$ unabhängig sind. Wir können dies so formulieren, dass die beiden Mengensysteme (Ereignissysteme)

$$\{\emptyset, F, \Omega \setminus F, \Omega\} \text{ und } \{\emptyset, G, \Omega \setminus G, \Omega\}$$

unabhängig sind.

Definition 2.20.
Es seien $\{\mathcal{C}_i\}_{i\in I}$ Ereignissysteme, d.h. $C_i \in \mathcal{F}$ für alle $C_i \in \mathcal{C}_i, i \in I$. Die Ereignissysteme $\{\mathcal{C}_i\}_{i\in I}$ heißen stochastisch unabhängig, falls gilt

$$\mathbb{P}(C_{i_1} \cap \cdots \cap C_{i_n}) = \Pi_{j=1}^n \mathbb{P}(C_{i_j})$$

für jede endliche Teilmenge $\{i_1, \ldots, i_n\}$ von I und jede Wahl von $C_{i_j} \in \mathcal{C}_{i_j}, 1 \leq j \leq n$.

Sind nun zwei Ereignissysteme $\mathcal{C}_1$ und $\mathcal{C}_2$ unabhängig, wollen wir die Unabhängigkeit auf die erzeugten σ –Algebren $\sigma - \mathcal{C}_1$ und $\sigma - \mathcal{C}_2$ "hochziehen". Dass dies manchmal möglich ist, folgt aus dem einfachen Beispiel $\mathcal{C}_1 = \{F\}$ und $\mathcal{C}_2 = \{G\}$. Aus der Unabhängigkeit von F und G folgt die Unabhängigkeit von $\sigma - \{F\} = \{\emptyset, F, \Omega \setminus F, \Omega\}$ und $\sigma - \{G\} = \{\emptyset, G, \Omega \setminus G, \Omega\}$. Dass es aber nicht immer geht, zeigt folgendes Beispiel.

Beispiel 2.21.
Zweifacher Münzwurf. Es seien A = erster Wurf "Kopf", B = zweiter Wurf "Kopf" und C = die Würfe seien verschieden. Wir setzen $\mathcal{C}_1 = \{A, B\}$, und $\mathcal{C}_2 = \{C\}$. Dann sind $\mathcal{C}_1$ und $\mathcal{C}_2$ unabhängig (Beispiel 2.11). Aber $\sigma - \mathcal{C}_1$ und $\sigma - \mathcal{C}_2$ sind nicht unabhängig. Denn $A \cap B \in \sigma - \mathcal{C}_1$ und es gilt

$$\mathbb{P}(A \cap B \cap C) = 0 \neq \mathbb{P}(A \cap B) \cdot \mathbb{P}(C) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}.$$

Zusätzliche Bedingungen an die Ereignissysteme müssen gestellt werden.

Definition 2.22.

1. *Ein Ereignissystem $\mathcal{C}$ heißt ein π–System, falls gilt:*
$$C_1, C_2 \in \mathcal{C} \Rightarrow C_1 \cap C_2 \in \mathcal{C}$$
2. *Ein Ereignissystem $\mathcal{C}$ heißt ein Δ–System (Dynkin-System), falls gilt:*
$$C \in \mathcal{C} \Rightarrow \Omega \setminus C \in \mathcal{C}$$
$$(C_j)_1^\infty \in \mathcal{C} \Rightarrow \sum_1^\infty C_j \in \mathcal{C}$$

Ist $\mathcal{C}$ ein Ereignissystem, so bezeichnen wir mit $\mathcal{D}(\mathcal{C})$ das von $\mathcal{C}$ erzeugte Δ–System. Analog zu (2.1) wird $\mathcal{D}(\mathcal{C})$ durch
$$\mathcal{D}(\mathcal{C}) = \cap\{\mathcal{H} : \mathcal{C} \subset \mathcal{H}, \mathcal{H} \text{ ein } \Delta\text{–System}\} \tag{2.3}$$
gegeben.

Wie man leicht nachrechnet, ist die rechte Seite von (2.3) ein Δ–System, und es ist nach Konstruktion in jedem Δ–System $\mathcal{H}$ enthalten, das $\mathcal{C}$ enthält.
Der folgende Satz ist außerordentlich nützlich, obwohl er etwas "trocken" aussieht.

Satz 2.23. *($\pi - \Delta$–Satz)*
Es sei $\mathcal{C}$ ein π–System. Dann gilt
$$\sigma - \mathcal{C} = \mathcal{D}(\mathcal{C}).$$

Wir können nun folgenden Satz beweisen.

Satz 2.24. *Es seien $\{\mathcal{C}_i\}_{i\in I}$ unabhängige π–Systeme. Dann sind die erzeugten σ–Algebren $\{\sigma - \mathcal{C}_i\}_{i\in I}$ auch unabhängig.*

Beweis: Die Definition von Unabhängigkeit bezieht sich nur auf endliche Teilmengen von I. Eine kurze Überlegung zeigt, dass es genügt, den Fall $I = \{1, \ldots, n\}$ zu betrachten. Mit $\mathcal{C}_2, \ldots, \mathcal{C}_n$ fest definieren wir
$$\mathcal{I}_1 = \{F \in \mathcal{F} : \{F\}, \mathcal{C}_2, \ldots, \mathcal{C}_n \text{ unabhängig }\}.$$
Es gilt laut Voraussetzung $\mathcal{C}_1 \subset \mathcal{I}_1$. Wir werden zeigen, dass $\mathcal{I}_1$ ein Δ–System ist. Gilt dies, dann folgt wegen Satz 2.23
$$\sigma - \mathcal{C}_1 = \mathcal{D}(\mathcal{C}_1) \subset \mathcal{I}_1.$$
Somit sind $\sigma - \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n$ unabhängig. Setzen wir dieses Argument fort, so erhalten wir zum Schluß, dass $\sigma - \mathcal{C}_1, \sigma - \mathcal{C}_2, \ldots, \sigma - \mathcal{C}_n$ unabhängig sind. Es genügt also zu zeigen, dass $\mathcal{I}_1$ ein Δ–System ist.

1. Aus $F \in \mathcal{I}_1$ folgt, dass $F, C_2, \ldots, C_n$ für jede Wahl von $C_i \in \mathcal{C}_i$, $2 \leq i \leq n$, unabhängig sind. In anderen Worten

$$\mathbb{P}(F \cap C_2 \cap \cdots \cap C_n) = \mathbb{P}(F)\Pi_2^n \mathbb{P}(C_j).$$

Es gilt

$$\begin{aligned}
\mathbb{P}((\Omega \setminus F) \cap C_2 \cap \cdots \cap C_n) &= \mathbb{P}(C_2 \cap \cdots \cap C_n \setminus (F \cap C_2 \cap \cdots \cap C_n)) \\
&= \mathbb{P}(C_2 \cap \cdots \cap C_n) - \mathbb{P}(F \cap C_2 \cdots \cap C_n) \\
&= \Pi_2^n \mathbb{P}(C_j) - \mathbb{P}(F) \cdot \Pi_2^n \mathbb{P}(C_j) \\
&= (1 - \mathbb{P}(F))\Pi_2^n \mathbb{P}(C_j).
\end{aligned}$$

Somit folgt aus $F, C_2, \ldots, C_n$ unabhängig, dass $(\Omega \setminus F), C_2, \ldots, C_n$ unabhängig sind, und zwar fr jede Wahl von C_i aus $\mathcal{C}_i$, $2 \leq i \leq n$. Dies zeigt $\Omega \setminus F \in \mathcal{I}_1$, falls $F \in \mathcal{I}_1$.

2. Es seien $(F_i)_1^\infty \in \mathcal{I}_1$ paarweise disjunkt. Es gilt für $C_j \in \mathcal{C}_j, 2 \leq j \leq n$,

$$\begin{aligned}
\mathbb{P}((\sum_{i=1}^{\infty} F_i) \cap C_2 \cap \cdots \cap C_n) &= \mathbb{P}(\sum_{i=1}^{\infty}(F_i \cap C_2 \cap \cdots \cap C_n)) \\
&= \sum_{1}^{\infty} \mathbb{P}(F_i \cap C_2 \cap \cdots \cap C_n) \\
&= \sum_{1}^{\infty} \mathbb{P}(F_i)\mathbb{P}(C_2) \ldots \mathbb{P}(C_n) \\
&= \sum_{1}^{\infty} \mathbb{P}(F_i))\mathbb{P}(C_2) \ldots \mathbb{P}(C_n) \\
&= \mathbb{P}(\sum_{1}^{\infty} F_i)\mathbb{P}(C_2) \ldots \mathbb{P}(C_n).
\end{aligned}$$

Somit ist $\mathcal{I}_1$ ein Δ–System, und der Satz ist bewiesen. □

Es seien $\mathcal{F}_1, \mathcal{F}_2$ und $\mathcal{F}_3$ unabhängige σ–Algebren. Es liegt nun nahe, dass $\sigma - (\mathcal{F}_1 \cup \mathcal{F}_2)$ und $\mathcal{F}_3$ auch unabhängig sind, denn $\mathcal{F}_1$ und $\mathcal{F}_2$ alleine enthalten keine Information über $\mathcal{F}_3$ und somit sollten beide zusammen keine Information enthalten. Dies stimmt; wir können sogar einen etwas allgemeineren Satz formulieren und beweisen.

Satz 2.25. *Es seien $\{\mathcal{F}_i\}_{i \in I}$ unabhängige σ–Algebren und $I = \sum_{j \in J} I_j$ eine Zerlegung von I. Für jedes $j \in J$ sei $\mathcal{G}_j = \sigma - (\cup_{i \in I_j} \mathcal{F}_i)$. Dann sind die σ–Algebren $(\mathcal{G}_j)_{j \in J}$ auch unabhängig.*

Beweis: Es sei $\mathcal{C}_j$ das System von endlichen Schnitten von Mengen der Form $F_{i_j} \in \mathcal{F}_{i_j}$ mit $i_j \in I_j$. In anderen Worten

$$\mathcal{C}_j = \{F_{i_j 1} \cap \cdots \cap F_{i_j k} : i_j \in I_j, F_{i_j l} \in \mathcal{F}_{i_j l} \quad 1 \leq \ell \leq k, k = 1, 2, \dots\}.$$

Dann ist jedes $\mathcal{C}_j$ ein π –System, denn der Schnitt von zwei Mengen in $\mathcal{C}_j$ ist auch ein endlicher Schnitt von Mengen in $F_{i_j} \in \mathcal{F}_{i_j}$ mit $i_j \in I_j$. Ferner sind die Mengensysteme $\mathcal{C}_j, j \in J$, unabhängig, da die $\mathcal{F}_i, i \in I$, unabhängig sind. Aus Satz 2.23 folgt nun, dass die σ–Algebren $\{\sigma - \mathcal{C}_j\}_{j \in J}$ unabhängig sind. Es ist aber klar, dass gilt

$$\sigma - \mathcal{C}_j = \sigma(\cup_{i \in I_j} \mathcal{F}_i)$$

und somit ist der Satz bewiesen. □

3 Zufallsvariablen

3.1 Zufallsvariablen und ihre Verteilungen

Wir betrachten den n–fachen Münzwurf mit Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Die Anzahl der Kopfwürfe lsst sich als eine Funktion auf Ω darstellen. Wir setzen $X(\omega) = k$, falls die Wurfreihe $\omega = (\delta_1, \cdots, \delta_n)$ k –mal 'Kopf' enthält. X ist eine Abbildung von Ω in $\mathbb{N}_0$.

$$X : \Omega \to \mathbb{N}_0 = \{0, 1, 2, \cdots\}.$$

Damit man von der Wahrscheinlichkeit reden kann, dass $X = k$ gilt, muß gelten

$$\{\omega : X(\omega) = k\} \in \mathcal{F}.$$

Diese kurze Überlegung führt zur folgenden Definition.

Definition 3.1.
Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Abbildung $X : \Omega \to \mathbb{R}^k$ heißt eine $\mathbb{R}^k$–wertige Zufallsvariable, falls X meßbar ist, d.h. $X^{-1}(\mathcal{B}(\mathbb{R}^k)) \subset \mathcal{F}$.

In anderen Worten, eine Zufallsvariable ist eine meßbare Funktion.

Bemerkung 3.2.
Für $b \in \mathbb{R}^k, b = (b_1, \cdots, b_k)$ definieren wir das k–dimensionale Intervall

$$(-\infty, b] = \{x \in \mathbb{R}^k : x_j \leq b_j, 1 \leq j \leq k\}.$$

$X : \Omega \to \mathbb{R}^k$ ist genau dann meßbar, wenn gilt

$$X^{-1}((-\infty, b]) \in \mathcal{F}, \quad \forall b \in \mathbb{R}^k$$

oder, etwas ausfühlicher,

$$\{\omega : X(\omega) \in (-\infty, b]\} \in \mathcal{F}, \quad \forall b \in \mathbb{R}^k.$$

Dies ist aus der Maß– und Integrationstheorie bekannt, läßt sich aber auch mit dem $\pi - \Delta-$Satz beweisen. Wir werden uns meistens mit dem Fall $k = 1$ befassen.

Beispiel 3.3.
Wir betrachten wieder den n–fachen Münzwurf mit Modell

$$(\Omega, \mathcal{F}, \mathbb{P}) = \otimes_1^n (\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$$

wobei $\Omega_i = \{0,1\}, \mathcal{F}_i = \mathcal{P}(\Omega_i)$ und $\mathbb{P}_i(\{0\}) = \mathbb{P}_i(\{1\}) = \frac{1}{2}$. Für $\omega \in \Omega$ gilt $\omega = (\delta_1, \cdots, \delta_n)$ mit $\delta_j \in \{0,1\}, 1 \leq j \leq n$, und der Zuordnung

$$\delta_j = \begin{cases} 1 & \text{Kopf} \\ 0 & \text{Zahl.} \end{cases}$$

Wir setzen

$$X(\omega) := \frac{1}{n} \sum_1^n \delta_j, \quad \omega = (\delta_1, \cdots, \delta_n).$$

d.h. $X(\omega)$ gibt die Proportion von Kopfwürfen an. Die Zufallsvariable

$$S(\omega) = \left| X(\omega) - \frac{1}{2} \right|$$

mißt die Abweichung dieser Proportion von $\frac{1}{2}$.

Der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ ist nie eindeutig festgelegt. Wir können z.B. den einfachen Münzwurf durch den folgenden Wahrscheinlichkeitsraum beschreiben.

$$\begin{aligned} \tilde{\Omega} &= \{-11, 0, 25\}, \quad \tilde{\mathcal{F}} = \mathcal{P}(\tilde{\Omega}) \\ \tilde{\mathbb{P}}(\{-11\}) &= \frac{1}{2}, \quad \tilde{\mathbb{P}}(\{0\}) = 0, \quad \tilde{\mathbb{P}}(\{25\}) = \frac{1}{2}. \end{aligned}$$

Wir identifizieren -11 mit einem Kopfwurf und 25 mit einem Zahlwurf. Es sei $\tilde{X} : \tilde{\Omega} \to \{0,1\}$, die die Anzahl der Kopfwürfe angibt, d.h. $\tilde{X}(-11) = 1,\ \tilde{X}(0) = 0,\ \tilde{X}(25) = 0$. Es sei

$$\Omega = \{0,1\}, \quad \mathcal{F} = \mathcal{P}(\Omega), \quad \mathbb{P}(\{0\}) = \mathbb{P}(\{1\}) = \frac{1}{2}$$

und $X : \Omega \to \{0,1\}, \quad X(\omega) = \omega$ die Anzahl der Kopfwürfe. Obwohl $X \neq \tilde{X}$ ist, gilt

$$\tilde{\mathbb{P}}(\tilde{X} = 1) = \tilde{\mathbb{P}}(\{-11\}) = \frac{1}{2}$$

und

$$\mathbb{P}(X = 1) = \mathbb{P}(\{1\}) = \frac{1}{2}.$$

Für alle k gilt

$$\tilde{\mathbb{P}}(\tilde{X} = k) = \mathbb{P}(X = k),$$

$\tilde{X}$ und X haben dieselbe Verteilung.
Hieraus kann man schließen, dass im Allgemeinen der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ von untergeordnetem Interesse ist.

Definition 3.4.
Es sei $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ eine Zufallsvariable. Die Verteilung $\mathbb{P}^X$ von X ist ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R}^k)$ und wird durch

$$\mathbb{P}^X(B) := \mathbb{P}(\{\omega : X(\omega) \in B\}) = \mathbb{P}(X^{-1}(B))$$

definiert.

Bemerkung 3.5.
$\mathbb{P}^X$ ist in der Tat ein Wahrscheinlichkeitsmaß, denn es gilt

(i) $\mathbb{P}^X(\mathbb{R}^k) = \mathbb{P}(X^{-1}(\mathbb{R}^k)) = \mathbb{P}(\Omega) = 1$

(ii)
$$\begin{aligned} \mathbb{P}^X(\sum_1^\infty B_j) &= \mathbb{P}(X^{-1}(\sum_1^\infty B_j)) = \mathbb{P}(\sum_1^\infty X^{-1}(B_j)) \\ &= \sum_1^\infty \mathbb{P}(X^{-1}(B_j)) \\ &= \sum_1^\infty \mathbb{P}^X(B_j) \end{aligned}$$

Im Fall $k = 1$ läßt sich die Verteilung $\mathbb{P}^X$ von X eindeutig durch die sogenannte Verteilungsfunktion bestimmen.

Definition 3.6.
Es sei X eine reelle Zufallsvariable. Die Verteilungsfunktion F_X von X wird durch

$$F_X(u) = \mathbb{P}^X((-\infty, u]) = \mathbb{P}(\{\omega : X(\omega) \leq u\})$$

definiert.

Satz 3.7.
Es sei F_X die Verteilungsfunktion der Zufallsvariablen X. Dann gilt

(a) F_X ist isoton

(b) $\lim_{u\to-\infty} F_X(u) = 0, \quad \lim_{u\to\infty} F_X(u) = 1$

(c) F_X ist rechtsseitig stetig, d.h.

$$\lim_{h\downarrow 0} F_X(u+h) = F_X(u), \quad u \in \mathbb{R}$$

Beweis:

(a) Ist $x \leq y$, so gilt $\{\omega : X(\omega) \leq x\} \subset \{\omega : X(\omega) \leq y\}$
und somit

$$F_X(x) = \mathbb{P}(\{\omega : X(\omega) \leq x\}) \leq \mathbb{P}(\{\omega : X(\omega) \leq y\}) = F_X(y)$$

d.h. $F_X$ist isoton.

(b) Wegen (a) genügt es zu zeigen, dass gilt

$$\lim_{n\to\infty} F_X(-n) = 0.$$

Es sei $G_n = \{\omega : X(\omega) \leq -n\}$. Dann gilt $G_n \supset G_{n+1} \supset \cdots$ und $\bigcap_{n=1}^{\infty} G_n = \emptyset$. Wegen der Stetigkeit von Wahrscheinlichkeitsmaßen gilt

$$\lim_{n\to\infty} F_X(-n) = \quad \lim_{n\to\infty} \mathbb{P}(G_n) = \mathbb{P}(\cap_{n=1}^{\infty} G_n) = \mathbb{P}(\emptyset) = 0.$$

Wieder wegen (a) genügt es zu zeigen, dass gilt $\lim_{n\to\infty} F_X(n) = 1$.
Es sei $E_n = \{\omega : X(\omega) \leq n\}$. Es gilt $E_n \subset E_{n+1} \subset \cdots$ und $\bigcup_{j=1}^{\infty} E_n = \Omega$. Wegen der Stetigkeit von Wahrscheinlichkeitsmaßen gilt $\lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} \mathbb{P}(E_n) = \mathbb{P}(\bigcup_{n=1}^{\infty} E_n) = \mathbb{P}(\Omega) = 1$.

(c) Wieder wegen (a) genügt es zu zeigen, dass gilt $\lim_{n\to\infty} F_X\left(x + \frac{1}{n}\right) = F_X(x)$. Es sei $H_n = \left\{\omega : X(\omega) \leq x + \frac{1}{n}\right\}$. Es gilt $H_n \supset H_{n+1} \supset \cdots$ und

$$\bigcap_{n=1}^{\infty} H_n = \{\omega : x(\omega) \leq x\}.$$

Aus der Stetigkeit von Wahrscheinlichkeitsmaßen folgt

$$\lim_{n\to\infty} F_X\left(x + \frac{1}{n}\right) = \lim_{n\to\infty} \mathbb{P}(H_n) = \mathbb{P}(\bigcap_{n=1}^{\infty} H_n) = F_X(x).$$

□

Der nächste Satz zeigt, dass die Verteilungsfunktion die ganze Verteilung bestimmt. Zuerst aber eine Bemerkung:

Satz 3.8.
Es seien X und Y zwei reelle Zufallsvariablen mit $F_X = F_Y$. Dann gilt $\mathbb{P}^X = \mathbb{P}^Y$.

Beweis: Es sei $\mathcal{D} = \{B \in \mathcal{B}(\mathbb{R}) : \mathbb{P}^X(B) = \mathbb{P}^Y(B)\}$. Dann ist $\mathcal{D}$ ein Δ–System. Denn:

(i) Es gilt für $B \in \mathcal{D}$, $\mathbb{P}^X(\mathbb{R} \setminus B) = 1 - \mathbb{P}^X(B) = 1 - \mathbb{P}^Y(B)$ wegen $B \in \mathcal{D}$. Somit erhalten wir

$$\mathbb{P}^X(\mathbb{R} \setminus B) = 1 - \mathbb{P}^Y(B) = \mathbb{P}^Y(\mathbb{R} \setminus B)$$

d.h. $\mathbb{R} \setminus B \in \mathcal{D}$.

(ii) Es seien $(B_j)_1^\infty \in \mathcal{D}$ paarweise disjunkte Mengen in $\mathcal{D}$. Es gilt

$$\mathbb{P}^X(\sum_1^\infty B_j) = \sum_1^\infty \mathbb{P}^X(B_j) = \sum_1^\infty \mathbb{P}^Y(B_j) = \mathbb{P}^Y(\sum_1^\infty B_j).$$

Somit liegt $\sum_1^\infty B_j$ in $\mathcal{D}$. Damit ist $\mathcal{D}$ ein Δ–System.

Es sei $\mathcal{I} = \{(-\infty, u] : u \in \mathbb{R}\}$. Dann ist ein $\mathcal{I}$ ein π–System. Es gilt aber für $(-\infty, u] \in \mathcal{I}$

$$\mathbb{P}^X((-\infty, u]) = F_X(u) = F_Y(u) = \mathbb{P}^Y((-\infty, u]).$$

Hieraus folgt $\mathcal{I} \subset \mathcal{D}$. Mit dem $\pi - \Delta$ –Satz erhalten wir nun $\sigma - \mathcal{I} \subset \mathcal{D}$. Aber $\sigma - \mathcal{I} = \mathcal{B}(\mathbb{R})$ und $\mathcal{D} \subset \mathcal{B}(\mathbb{R})$, woraus folgt $\mathcal{D} = B(\mathbb{R})$ d.h. $\mathbb{P}^X = \mathbb{P}^Y$. □

Der nächste Satz zeigt, dass es zu jeder Funktion F, die (a), (b) und (c) von Satz 3.7 erfüllt, eine Zufallsvariable X mit $F_X = F$ gibt.

Satz 3.9.
Es sei $F : \mathbb{R} \to \mathbb{R}$ eine Funktion, die (a), (b) und (c) von Satz 3.7 erfüllt. Dann gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine darauf definierte Zufallsvariable $X : \Omega \to \mathbb{R}$ mit Verteilungsfunktion $F_X = F$.

Beweis: Wir setzen $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{Q}), X =$ Identität. $\mathbb{Q}$ sei das durch F erzeugte Wahrscheinlichkeitsmaß auf $\mathbb{R}$, d.h. $\mathbb{Q}((-\infty, u]) = F(u)$. Dass es ein solches Wahrscheinlichkeitsmaß $\mathbb{Q}$ gibt, folgt aus dem Maßfortsetzungssatz. Man weist nun einfach nach, dass $F_X = F$ gilt. □

3.2 Einige wichtige eindimensionale Verteilungen

Wir wollen nun den Fall $k = 1$ etwas ausführlicher behandeln. In den meisten Fällen wird $\mathbb{P}^X$ über eine sogenannte Dichtefunktion definiert. Ist ν das Zählmaß auf $\mathbb{Z}$, d.h. $\nu(\{j\}) = 1$ für alle $j \in \mathbb{Z}$, so hat X eine Dichtefunktion f_X bezüglich ν, falls gilt

$$\mathbb{P}^X(B) = \int_B f_X d\nu = \sum_{j \in B} f_X(j), \quad B \in \mathcal{B}(\mathbb{R}).$$

Eine solche Verteilung bzw. Zufallsvariable heißt diskret.

3.2.1 Einige diskrete Verteilungen

Wir definieren nun einige wichtige Verteilungen.

Definition 3.10.
Die Zufallsvariable X heißt gleichverteilt über $\{1, \cdots, n\}$, falls

$$f_X(j) = \frac{1}{n}\{1 \leq j \leq n\},$$

d.h.

$$\mathbb{P}\left(\{X = j\}\right) = \mathbb{P}^X\left(\{j\}\right) = \begin{cases} \frac{1}{n}, & 1 \leq j \leq n \\ 0, & \textit{sonst.} \end{cases}.$$

Beispiel 3.11.
Das einfachste Beispiel ist das einmalige Würfeln, wobei $n = 6$ ist. Die Zufallsvariable X ist die Augenzahl.

Definition 3.12.
Die Zufallsvariable X heißt binomialverteilt mit Parametern n und p, falls

$$f_X(j) = \binom{n}{j} p^j (1-p)^{n-j} \ \{0 \leq j \leq n\}$$

d.h.

$$\mathbb{P}(\{X = j\}) = \mathbb{P}^X(\{j\}) = \begin{cases} \binom{n}{j} p^j (1-p)^{n-j} & , 0 \leq j \leq n \\ 0 & , \ \textit{sonst} \end{cases}.$$

Wir schreiben $X \sim b(n, p)$.

Beispiel 3.13.
n Versuche werden unabhängig voneinander durchgeführt. Bei jedem Versuch gibt es nur zwei Möglichkeiten, 'Erfolg' oder 'Mißerfolg'. Die Wahrscheinlichkeit für einen Erfolg beträgt stets $p, 0 \leq p \leq 1$. Es sei X die Anzahl von Erfolgen bei den n Versuchen. Dann gilt $X \sim b(n,p)$. Denn es sei $(E, E, M, E, \dots,)$ eine bestimmte Reihenfolge mit j–mal Erfolg. Wegen der Unabhängigkeit der Versuche beträgt die Wahrscheinlichkeit für genau diese Reihenfolge

$$pp(1-p)p\dots(1-p) = p^j(1-p)^{n-j}.$$

Es gibt aber $\binom{n}{j}$ Reihenfolgen mit genau j–mal $'E'$. Hieraus folgt
$\mathbb{P}(X=j) = \binom{n}{j}p^j(1-p)^{n-j}$. Es gilt

$$\sum_{j=0}^{n}\mathbb{P}(X=j) = \sum_{j=0}^{n}\binom{n}{j}p^j(1-p)^{n-j} = (p+(1-p))^n = 1$$

nach dem Binomialsatz.

Definition 3.14.
Die Zufallsvariable X heißt Poisson–verteilt mit Parameter λ, falls gilt

$$f_X(j) = \frac{\lambda^j e^{-\lambda}}{j!}\ \{0 \leq j < \infty\}.$$

In anderen Worten

$$\mathbb{P}\left(\{X=j\}\right) = \mathbb{P}^X(j) = \begin{cases} \frac{\lambda^j e^{-\lambda}}{j!} & 0 \leq j < \infty \\ 0 & \text{sonst} \end{cases}.$$

Es gilt

$$\sum_{j=0}^{\infty}\mathbb{P}(\{X=j\}) = \sum_{j=0}^{\infty}\frac{\lambda^j e^{-\lambda}}{j!} = e^{-\lambda}.\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = e^{-\lambda}\cdot e^{\lambda} = 1.$$

Wir geben nun eine heuristische Herleitung der Poissonverteilung. Wir betrachten Ereignisse in Raum oder Zeit. Es sei $p(k,V)$ die Wahrscheinlichkeit für k Ereignisse in einem Gebiet vom Volumen V. Als Beispiel denken Sie vielleicht an die Anzahl von Käfern auf einer Fläche von V Quadratmetern. Wir nehmen an, dass $p(k,V)$ nur vom Volumen (oder Flächeninhalt) und nicht von der Gestalt abhängt.
Wir treffen folgende Annahmen:

(i) $p(0,V) = 1 - \lambda V + o(V), \quad V$ klein,

(ii) $p(1,V) = \lambda V + o(V), \quad V$ klein,

(iii) Ereignisse in disjunkten Gebieten seien voneinander unabhängig. Es seien V und δV disjunkte Gebiete.

Aus dem Satz der totalen Wahrscheinlichkeit und (iii) folgt

$$p(k+1, V+\delta V) = \sum_{j=0}^{k+1} p(j,V) \cdot p(k+1-j, \delta V).$$

Wegen (i) und (ii) gilt $\sum_{j=2}^{\infty} p(j,\delta V) = o(\delta V)$. Wir können also für $k \geq 1$ schreiben

$$\begin{aligned} p(k+1, V+\delta V) &= p(k,V)p(1,\delta V) + p(k+1,V)p(0,\delta V) + o(\delta V) \\ &= \lambda p(k,V)\delta V + p(k+1,V)(1-\lambda\delta V) + o(\delta V) \end{aligned}$$

wobei wir (i) benutzt haben. Hieraus folgt

$$\frac{p(k+1,V+\delta V) - p(k+1,V)}{\delta V} = \lambda p(k,V) - \lambda p(k+1,V) + o(1)$$

und somit $\delta V \downarrow 0$

$$\frac{d}{dV} p(k+1,V) = \lambda(p(k,V) - p(k+1,V)), \quad k = 1,2,\cdots \tag{3.1}$$

Für $k = 0$ ergibt ein ähnliches Argument

$$p(0, V+\delta V) = p(0,V) \cdot p(0,\delta V)(1 - \lambda\delta V + o(\delta V))$$

und somit

$$\frac{d}{dV} p(0,V) = \lambda p(0,V). \tag{3.2}$$

Die Lösung von (3.2) lautet

$$p(0,V) = c \cdot \exp(-\lambda V)$$

mit $c \in \mathbb{R}$. Die Konstante c bestimmen wir wie folgt. Aus (i) ergibt sich

$$c \cdot \exp(-\lambda\delta V) = p(0,\delta V) = 1 - \lambda\delta V + o(\delta V)$$

und mit $\delta V \downarrow 0$ erhalten wir $c = 1$. Es gilt also $p(0,V) = \exp(-\lambda V)$. Wir wollen nun mittels vollständiger Induktion die Differentialgleichungen (3.1) lösen. Die Induktionsannahme lautet

$$p(k,V) = \frac{(\lambda V)^k \exp(-\lambda V)}{k!}.$$

Die Annahme stimmt für $k = 0$. Es gilt unter der Annahme

$$\frac{d}{dV}p(k+1,V) = \frac{\lambda(\lambda V)^k \exp(-\lambda V)}{k!} - \lambda p(k+1,V)$$

und somit

$$e^{\lambda V}\frac{d}{dV}p(k+1,V) + \lambda e^{\lambda V}p(k+1,V) = \lambda\frac{\lambda(\lambda V)^k}{k!}.$$

Hieraus folgt

$$\frac{d}{dV}(e^{\lambda V}p(k+1,V)) = \lambda \cdot \frac{(\lambda V)^k}{k!}.$$

Eine direkte Integration ergibt nun

$$e^{\lambda V}p(k+1,V) = \frac{(\lambda V)^{k+1}}{(k+1)!} + c.$$

Es bleibt, die Konstante c zu bestimmen. Aus (ii) folgt $\lim_{V\to 0} p(k+1,V) = 0$, und dies ist nur möglich, wenn $c = 0$. Damit ist der Induktionsschritt bewiesen, und wir haben gezeigt

$$p(k,V) = \frac{(\lambda V)^k \exp(-\lambda V)}{k!}, \quad k \geq 0.$$

Beispiel 3.15.
Wir betrachten nun echte Daten, um zu sehen, wie gut stochastische Modelle in der Praxis sein können. In diesem Beispiel ist V = Zeit, und ein Ereignis ist die Emission eines $\alpha-$Teilchens:

$$P(k \quad \alpha\text{–Teilchen in einem Zeitintervall der Länge } T) = p(k,T).$$

Insgesamt wurden $n = 12169$ Beobachtungen von Intervallen der Länge 1 Sekunde erhoben. Die Ergebnisse sind wie folgt.

Anzahl der Teilchen	Häufigkeit	Theoretisch
0	5 267	5 268.7
1	4 436	4 410.5
2	1 800	1 846,0
3	534	515.1
4	111	107.8
≥ 5	21	20.9
	12 169	12 168.9

Die theoretischen Häufigkeiten ergeben sich aus der Annahme einer Poisson–Verteilung mit Parameter λ. Der Wert von λ wird aus den Daten bestimmt. Es gibt einige Gründe, hierfür den Mittelwert der Daten zu benutzen. Der Mittelwert der Daten beträgt $\tilde{\lambda} := 0.8371$ /Sekunde. Die theoretischen Häufigkeiten sind dann : $n\tilde{\lambda}^k \exp(-\tilde{\lambda})/k!, k = 0, 1, 2, \cdots, 5$ mit $n = 12.169$.

Zweite Herleitung der Poisson–Verteilung.
Wir betrachten eine Folge von Binomialverteilungen $b(n, p_n)$, wobei

$$\lim_{n\to\infty} np_n = \lambda \text{ mit } 0 < \lambda < \infty.$$

Es sei $X_n \sim b(n, p_n)$. Es gilt

$$\begin{aligned} \mathbb{P}(X_n = k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} p_n^k (1-p_n)^{n-k}. \end{aligned}$$

Es gilt wegen $\lim_{n\to\infty} np_n = \lambda$

$$\lim_{n\to\infty} n(n-1)\dots(n-k+1)p_n^k = \lim_{n\to\infty} (np_n)((n-1)p_n)\dots((n-k+1)p_n) = \lambda^k.$$

Aus $\log(1-p_n) = -p_n + o(p_n)$ folgt

$$(n-k)\log(1-p_n) = -(n-k)p_n + o((n-k)p_n)$$

und somit $\lim_{n\to\infty}(n-k)\log(1-p_n) = -\lambda$. Dies impliziert

$$\lim_{n\to\infty} (1-p_n)^{n-k} = \exp(-\lambda).$$

Wir erhalten insgesamt

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \lambda^k \exp(-\lambda)/k!,$$

d.h. fr n groß, p_n klein, $np_n \sim \lambda$, ist X_n ungefähr Poisson–verteilt mit Parameter λ.

Definition 3.16.
Die Zufallsvariable X heißt geometrisch verteilt mit Parameter p, falls

$$f_X(j) = \begin{cases} (1-p)^{j-1}p, & 1 \le j < \infty \\ 0 & , \ sonst \end{cases}.$$

Es gilt

$$\sum_{j=1}^{\infty} f_X(j) = p \cdot \sum_{j=1}^{\infty} (1-p)^{j-1} = \frac{p}{1-(1-p)} = 1.$$

Wir schreiben $X \sim g(p)$ oder $X \sim nb(1,p)$.

Beispiel 3.17.
Wir betrachten das Experiment von Beispiel 3.13. Wir führen es aber nur solange durch, bis wir zum ersten Mal einen Erfolg erzielen. Es sei X die hierzu notwendige Anzahl von Versuchen. Es ist klar, dass $X = k$ nur für die Reihenfolge

$$\underbrace{M \cdots . \cdots M}_{k-1} \quad \cdot \quad E$$

gilt und somit

$$\mathbb{P}(X = k) = (1-p)^{k-1} p.$$

3.2.2 Einige stetige Verteilungen

Eine andere wichtige Klasse von Verteilungen hat Dichtefunktionen bezüglich des Lebesgue–Maßes λ auf $\mathcal{B}(\mathbb{R})$. In diesem Falle gilt

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \mathbb{P}^X(B) = \int_B f_X d\lambda.$$

Insbesondere gilt

$$F_X(u) = \int_{-\infty}^{u} f_X(x)dx = \int_B f_X d\lambda, \text{ wobei } B = (-\infty, u].$$

Eine solche Verteilung bzw. Zufallsvariable heißt stetig. Unter stetigen Verteilungen oder Zufallsvariablen verstehen wir Verteilungen $\mathbb{P}^X$, die absolut stetig bezüglich des Lebesgue–Maßes sind. Die Randon–Nikodym'sche Ableitung heißt die Dichtefunktion.

Definition 3.18.
Die Zufallsvariable X heißt gleichverteilt über dem Intervall $[a,b]$, falls gilt

$$f_X(u) = \frac{1}{b-a}\{a \le u \le b\}.$$

Es gilt

$$\mathbb{P}(X \le u) = \begin{cases} 0, & u < a, \\ \frac{u-a}{b-a}, & a \le u \le b, \\ 1, & u > b. \end{cases}$$

Wir schreiben $X \sim U(a,b)$.

Die Gleichverteilung über $[a, b]$ ist ein stochastisches Modell für das Experiment, eine Zahl zwischen a und b zufällig auszusuchen.

Definition 3.19.
Die Zufallsvariable X heißt exponentialverteilt mit Parameter $\lambda > 0$, falls

$$f_X(u) = \begin{cases} \lambda e^{-\lambda u}, & 0 \leq u < \infty \\ 0, & sonst \end{cases}.$$

Es gilt

$$F_X(u) = \begin{cases} 0, & u \leq 0 \\ 1 - \exp(-\lambda u) & u > 0 \end{cases}.$$

Wir schreiben $X \sim E(\lambda)$.

Die Exponentialverteilung beschreibt Gedächtnislosigkeit oder die Lebensdauer eines Objektes, das nicht altert. Nicht altern heißt, dass die Verteilung der Lebensdauer eines Objektes, das mindestens s Einheiten alt ist, dieselbe ist wie die eines neuen Objekts. Dies können wir schreiben in der Form

$$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t).$$

Für die Exponentialverteilung gilt $\mathbb{P}(X > s) = e^{-\lambda s}$. Hieraus folgt

$$\begin{aligned} \mathbb{P}(X > s + t | X > s) &= \frac{\mathbb{P}(\{X > s + t\} \cap \{X > s\})}{\mathbb{P}(X > s)} \\ &= \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} \\ &= \frac{\exp(-\lambda(s + t))}{\exp(-\lambda s)} \\ &= \exp(-\lambda t) \\ &= \mathbb{P}(X > t). \end{aligned}$$

Die Exponentialverteilung ist die einzige Verteilung mit dieser Eigenschaft.

Definition 3.20.
Die Zufallsvariable X heißt Gauß–verteilt (normalverteilt) mit Parametern μ und σ, falls

$$f_X(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{u - \mu}{\sigma}\right)^2\right).$$

Wir schreiben $X \sim N(\mu, \sigma^2)$.

Es gibt keinen geschlossenen Ausdruck für die Verteilungsfunktion, dafür aber Tabellen oder Computerprogramme. μ mißt die Lage, σ die Breite der Verteilung. Die Normalverteilung ist die wichtigste aller Verteilungen, einen einleuchtenden Grund hierfür gibt es jedoch nicht.

3.3 Transformationen von Zufallsvariablen

Ist $X : \Omega \to \mathbb{R}^k$ eine Zufallsvariable und $g : \mathbb{R}^k \to \mathbb{R}^l$ Borel–meßbar, so ist $g \circ X = g(X)$ auch eine Zufallsvariable. Wir zeigen nun zuerst für den Fall $k = 1$, wie man die Verteilung von Y aus der Verteilung von X gewinnen kann. Statt einen allgemeinen Satz anzugeben, betrachten wir zwei Beispiele.

Bemerkung 3.21.
Es ist schwierig, eine nicht Borel–meßbare Abbildung $g : \mathbb{R}^k \to \mathbb{R}^l$ anzugeben. Alle Beispiele benutzen u.a. das Auswahlaxiom. Hieraus folgt, dass jede konstruktive Funktion g Borel–meßbar ist. Insbesondere sind alle stetigen Funktionen meßbar, sogar alle stetigen Funktionen mit höchstens abzählbar vielen Unstetigkeitspunkten.

Beispiel 3.22.
Es sei $Y = X^2$ und $X \sim N(0, 1)$. Es gilt

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y})$$

und somit

$$\begin{aligned} F_Y(y) &= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} \exp\left(-\frac{1}{2}x^2\right) dx \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{y}} \exp\left(-\frac{1}{2}x^2\right) dx. \end{aligned}$$

In dem Integral setzen wir $u = x^2$. Hieraus folgt $x = \sqrt{u}$ und $dx = \frac{1}{2} \cdot \frac{du}{\sqrt{u}}$. Wir erhalten also

$$F_Y(y) = \sqrt{\frac{2}{\pi}} \int_0^y \frac{1}{\sqrt{u}} \exp\left(-\frac{1}{2}u\right) du$$

Damit hat Y die Dichtefunktion

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \cdot y^{-1/2} \exp\left(-\frac{1}{2}y\right) \{y > 0\}.$$

Definition 3.23.
Die Zufallsvariable Y ist chi–quadrat–verteilt mit k Freiheitsgraden, falls Y die Dichtefunktion

$$f_Y(u) = \frac{2^{-k/2}}{\Gamma(k/2)} u^{k/2-1} \exp\left(-\frac{1}{2}u\right) \{u > 0\}$$

hat. Wir schreiben $Y \sim \chi^2_k$.

Diese Definition beinhaltet die sogenannte Γ –Funktion, die für $\lambda > 0$ durch

$$\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} \exp(-x) dx$$

definiert wird. Für $\lambda \in \mathbb{N} = \{1, 2, \dots\}$ gilt

$$\Gamma(\lambda) = (\lambda - 1)!.$$

Für alle $\lambda > 1$ gilt $\Gamma(\lambda) = (\lambda - 1)\Gamma(\lambda - 1)$. Aus Beispiel 3.23 folgt nun, dass $X \sim N(0,1) \Rightarrow X^2$ ist chi–quadrat–verteilt mit einem Freiheitsgrad.

Beispiel 3.24.
Es sei $X \sim N(\mu, \sigma^2)$. Dann ist $Y = (X - \mu)/\sigma N(0,1)$-verteilt. Es gilt nämlich

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq y\right) = \mathbb{P}(X \leq \mu + \sigma y)$$

und somit

$$F_Y(y) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\mu+\sigma y} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx.$$

In dem Integral setzen wir $u = \frac{x-\mu}{\sigma}$. Es ergibt sich

$$F_Y(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \exp\left(-\frac{1}{2}u^2\right) du$$

und somit die Behauptung.

Von allgemeinerer Bedeutung ist der nächste Satz. Wir brauchen folgende Definition.

Definition 3.25.
Es sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion. Die verallgemeinerte Inverse F^{-1} wird definiert durch

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}.$$

Satz 3.26.
Es gilt immer

$$F(F^{-1}(u) - 0) \leq u \leq F(F^{-1}(u)).$$

Insbesondere gilt $F(F^{-1}(u)) = u$, falls $F^{-1}(u)$ ein Stetigkeitspunkt von F ist.

Beweis: Es sei $(x_n)_1^\infty$ eine Folge mit $x_n \downarrow F^{-1}(u)$. Da F rechtsseitig stetig ist, folgt

$$u \leq \lim_{n\to\infty} F(x_n) = F(\lim x_n) = F(F^{-1}(u)).$$

In der anderen Richtung gilt für jede Folge $(x_n)_1^\infty$ mit $x_n \uparrow F^{-1}(u)$

$$u \geq \lim_{n\to\infty} F(x_n) = F(F^{-1}(u) - 0).$$

□

Satz 3.27.
Es sei X eine Zufallsvariable mit einer stetigen Verteilungsfunktion F_X. Dann ist die Zufallsvariable $U = F_X(X)$ gleichverteilt über das Intervall $[0, 1]$.

Beweis: Es gilt $F_X(x) \geq u$ genau dann, wenn $x \geq F_X^{-1}(u)$. Somit gilt

$$\mathbb{P}(U \geq u) = \mathbb{P}(F_X(X) \geq u) = \mathbb{P}(X \geq F_X^{-1}(u)) = 1 - F_X(F_X^{-1}(u)) = 1 - u$$

nach Satz 3.26. Hieraus folgt die Behauptung. □

Satz 3.28.
Es seien F eine stetige Verteilungsfunktion und U eine über $[0, 1]$ *gleichverteilte Zufallsvariable. Dann hat die Zufallsvariable $X := F^{-1}(U)$ die Verteilungsfunktion F.*

Beweis: Es gilt $F(x) \geq u$ genau dann, wenn $x \geq F^{-1}(u)$. Hieraus folgt

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

Bemerkung 3.29.
Ein Pseudozufallszahlenerzeuger liefert Zufallszahlen, die annähernd gleichverteilt über $[0, 1]$ sind. Mit Hilfe von Satz 3.28 kann man jede beliebige Verteilung erzeugen. Dies heißt nicht, dass die Methode von Satz 3.28 immer die Beste ist.

3.4 k–dimensionale Zufallsvariablen

Ist $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}^k$ meßbar, so heißt X eine k–dimensionale Zufallsvariable. Wir können X schreiben als

$$X = (X_1, \ldots, X_k)^T$$

mit $X_j = \Pi_j \circ X$ wobei $\Pi_j : \mathbb{R}^k \to \mathbb{R}$ die j-te Projektion ist. Es läßt sich einfach nachweisen, dass die Komponenten $(X_j)_1^k$ von X eindimensionale Zufallsvariablen sind, d.h.

$$X_j : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

ist meßbar. In der anderen Richtung sind $(X_j)_1^k : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eindimensionale Zufallsvariable, so ist $X := (X_1, \ldots, X_k)^T$ eine k–dimensionale Zufallsvariable. Analog zu dem Fall $k = 1$ kann man auch eine Verteilungsfunktion für eine k–dimensionale Zufallsvariable X wie folgt definieren: die Verteilungsfunktion $F_X : \mathbb{R}^k \to [0, 1]$ wird durch

$$\begin{aligned} F_X(x) &= \mathbb{P}^X((-\infty, x]) \\ &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(X_1 \leq x_1, \ldots, X_k \leq x_k) \end{aligned}$$

definiert, wobei $x = (x_1, \ldots, x_k)^T$.

Satz 3.30.
Es sei X eine k–dimensionale Zufallsvariable mit Verteilungsfunktion F_X. Dann gilt

(a)

$$\lim_{x_j \to -\infty} F_X(x_1, \ldots, x_k) = 0, \text{ für } 1 \leq j \leq k$$

(b)

$$\lim_{x_1, \ldots, x_k \to \infty} F_X(x_1, \ldots, x_k) = 1$$

(c) F_X ist rechtsseitig stetig in x, d.h.

$$\lim_{h \downarrow 0} F_X(x + h) = \lim_{h_1, \ldots, h_k \downarrow 0} F_X(x_1 + h_1, \ldots, x_k + h_k) = F_X(x)$$

(d) Für alle $a \leq b$, d.h. $a_j \leq b_j, 1 \leq j \leq k$, gilt

$$\sum_c (-1)^{\epsilon(c)} F_X(c) \geq 0,$$

wobei die Summen über allen $c \in \mathbb{R}^k$ von der Form

$$c = (c_1, \ldots, c_k)^T$$

mit $c_j \in \{a_j, b_j\}, 1 \leq j \leq k$, genommen wird und

$$\epsilon(c) = |\{j : c_j = a_j, 1 \leq j \leq k\}|.$$

Beweis: (a), (b) und (c) werden bewiesen wie für den Fall $k = 1$. (d) entspricht der Tatsache, dass im Falle $k = 1$ F_X isoton ist. Leider ist (d) etwas komplizierter, und es ist nicht möglich, (d) durch die Forderung zu ersetzen, F_X ist isoton in jeder Komponente. Man kann zeigen, dass gilt

$$\sum_c (-1)^{\epsilon(c)} F_X(c) = \mathbb{P}(a < X \leq b) \geq 0.$$

Wir beschränken uns auf den Fall $k = 2$. Es gilt

$$\begin{aligned}
\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = \\
&\mathbb{P}(X_1 \leq b_1, X_2 \leq b_2) - \mathbb{P}(X_1 \leq a_1, X_2 \leq b_2) \\
&-\mathbb{P}(X_1 \leq b_1, X_2 \leq a_2) + \mathbb{P}(X_1 \leq a_1, X_2 \leq a_2) \\
&= \sum_c (-1)^{\epsilon(c)} F_X(c).
\end{aligned}$$

□

Analog zu Satz 3.9 haben wir

Satz 3.31.
Es sei $F : \mathbb{R}^k \to [0, 1]$ eine Funktion, die (a), (b), (c) und (d) von Satz 3.30 erfüllt. Dann gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine k–dimensionale Zufallsvariable X mit $F_X = F$.

Beweis: Wir setzen $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{Q}_F)$ wobei

$$\mathbb{Q}_F((a, b]) := \sum_c (-1)^{\epsilon(c)} F(c).$$

Es lässt sich etwas mühsam zeigen, dass $\mathbb{Q}_F$ zu einem Maß auf $\mathcal{B}(\mathbb{R}^k)$ fortgesetzt werden kann. Die Hauptschwierigkeit dabei besteht darin nachzuweisen, dass $\mathbb{Q}_F$ endlich additiv ist. Aus (a) und (b) folgt, dass $\mathbb{Q}_F$ ein Wahrscheinlichkeitsmaß ist. Definieren wir nun $X : \mathbb{R}^h \to \mathbb{R}^k, X(x) = x$, so folgt $F_X = F$. □

Es sei ν^k das Zählmaß auf $\mathbb{Z}^k$,d.h. $\nu^k(\{j\}) = 1$ für alle $j = (j_1, \ldots, j_k)^T \in \mathbb{Z}^k$. Die Zufallsvariable X heißt diskret, falls die Verteilung $\mathbb{P}^X$ von X absolut stetig bezüglich ν^k ist. In diesem Fall können wir schreiben

$$\mathbb{P}^X(B) = \int_B f_X(x) d\nu^k(x) = \sum_{j \in \mathbb{Z}^k} f_X(j).$$

Die Funktion f_X heißt die Dichtefunktion von X, und es gilt für alle $j \in \mathbb{Z}^k$

$$\mathbb{P}(X = j) = f_X(j).$$

Definition 3.32. *(Multinomiale Verteilung)*
Betrachten wir ein Experiment, das k verschiedene Ergebnisse hat $E_1, \ldots, E_k$. Die Wahrscheinlichkeit für das Ergebnis E_j sei $p_j, 1 \leq j \leq k$. Es gilt $\sum_1^k p_j = 1$. Das Experiment sei nun n-mal nacheinander durchgeführt; die einzelnen Durchführungen seien voneinander unabhängig. Es sei nun $X = (X_1, \ldots, X_k)$, wobei X_j die Anzahl von Durchführungen bezeichnet, die zu dem Ergebnis E_j führten. Für $j \in \mathbb{Z}^k$ gilt wie für die Binomialverteilung

$$f_X(j) = \mathbb{P}(X = j) = \frac{n!}{j_1! \ldots j_k!} p_1^{j_1} \ldots p_k^{j_k} \left\{ \sum_1^k j_l = n \right\} \{j_1 \geq 0\} \ldots \{j_k \geq 0\}.$$

Im Falle $k = 2$ erhalten wir die Binomialverteilung.

Die Zufallsvariable X heißt stetig, falls die Verteilung $\mathbb{P}^X$ von X absolut stetig bezüglich des Lebesgue–Maßes λ^k auf $\mathbb{R}^k$ ist. Es gilt für alle $B \in \mathcal{B}(\mathbb{R}^k)$

$$\mathbb{P}(X \in B) = \mathbb{P}^X(B) = \int_B f_X d\lambda^k.$$

Die Funktion f_X heißt die Dichtefunktion von X.

Bemerkung 3.33.
Ist $X = (X_1, \ldots, X_k)$ eine stetige Zufallsvariable mit Dichtefunktion f, so ist $\tilde{X}_l = (X_{i_1}, \ldots, X_{i_l})$ für $i_1, \ldots, i_l \in \{1, \ldots, k\}$ eine stetige $\mathbb{R}^l$–wertige Zufallsvariable. Ferner erhält man die Dichtefunktion $\tilde{f}_l$ von $\tilde{X}_l$ durch Ausintegrieren der anderen Variablen. Für $l = 1$ und $\tilde{X}_l = X_1$ erhalten wir z.B.

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_k) dx_2 \ldots dx_k.$$

Definition 3.34. *(k–dimensionale Normalverteilung)*
Es seien $\mu \in \mathbb{R}^k$ und Σ eine $k \times k$ positiv definite, symmetrische, nicht–singuläre Matrix. Die Zufallsvariable X heißt $N(\mu, \Sigma)$–verteilt, falls gilt

$$f_X(x) = (2\pi)^{-k/2}|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

Dass $f_X(x) \geq 0$ gilt, ist klar. Dass f_X tatsächlich eine Dichtefunktion ist, folgt aus $\int f_X d\lambda^k = 1$. Einen Beweis hierfür geben wir nicht an.

3.5 Transformation von k–dimensionalen Zufallsvariablen

Ist $X : \Omega \to \mathbb{R}^l$ eine l–dimensionale Zufallsvariable und $g : \mathbb{R}^l \to \mathbb{R}^k$ Borel-meßbar, so ist $Y = g \circ X$ eine k–dimensionale Zufallsvariable. Anhand eines Beispiels zeigen wir nun, wie man die Verteilung von Y aus der Verteilung von X gewinnen kann.

Beispiel 3.35.
Die Zufallsvariable X sei gleichverteilt über dem Einheitskreis, d.h. X hat die Dichtefunktion

$$f_X(x) = \frac{1}{\pi}\{||x|| \leq 1\}$$

bezüglich des Lebesgue–Maßes auf $\mathbb{R}^2$. Wir schreiben $X = (X_1, X_2)^T$ und definieren die Zufallsvariablen R und Θ durch

$$X_1 = R\cos\Theta, \quad X_2 = R\sin\Theta.$$

Es gilt

$$\mathbb{P}(R \leq r, \Theta \leq \theta) = \mathbb{P}((X_1, X_2) \in C(r,\theta)) = \int_{C(r,\theta)} \frac{1}{\pi}\{||x|| \leq 1\} dx_1 dx_2$$

wobei

$$C(r,\theta) = \left\{x : \|x\| \leq r, \arcsin\left(\frac{x_2}{\|x\|}\right) \leq \theta, \arccos\left(\frac{x_1}{\|x\|}\right) \leq \theta\right\}.$$

Wir evaluieren das Integral mit Hilfe folgender Transformation. Wir setzen $x_1 = \rho\cos\phi$ und $x_2 = \rho\sin\phi$. Es gilt

$$dx_1 dx_2 = \left|\begin{array}{cc} \frac{\partial x_1}{\partial \rho} & \frac{\partial x_2}{\partial \rho} \\ \frac{\partial x_1}{\partial \phi} & \frac{\partial x_2}{\partial \phi} \end{array}\right| d\rho d\phi = \left|\begin{array}{cc} \cos\phi & \sin\phi \\ -\rho\sin\phi & \rho\cos\phi \end{array}\right| d\rho d\phi = \rho d\rho d\phi.$$

Hieraus folgt

$$\mathbb{P}(R \leq r, \Theta \leq \theta) = \frac{1}{\pi} \int \{\rho \leq r\}\{\phi \leq \theta\} \rho d\rho d\phi = \frac{\theta}{\pi} \int_0^r \rho d\rho = \frac{r^2 \theta}{2\pi}.$$

Somit hat die Zufallsvariable (R, Θ) die Dichtefunktion

$$f_{R,\Theta}(r, \theta) = \frac{r}{\pi} \{0 \leq r \leq 1\}\{0 \leq \theta \leq 2\pi\}.$$

3.6 Unabhängig verteilte Zufallsvariablen

In §2.4 haben wir unabhängige Ereignisse bzw. Ereignissysteme definiert. Wir erweitern nun die Definition auf Zufallsvariablen.

Definition 3.36.
Es sei $(X_n)_1^\infty$ eine Folge von Zufallsvariablen, die auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ definiert sind $X_n : \Omega \to \mathbb{R}^{k_n}, n \in \mathbb{N}$. Die $(X_n)_1^\infty$ heißen unabhängig verteilt, falls die σ–Algebren $\mathcal{F}_n = X_n^{-1}(\mathcal{B}(\mathbb{R}^{k_n}))$, $n \in \mathbb{N}$, unabhängig sind.

Bemerkung 3.37.
Wir werden uns hauptsächlich, aber nicht immer, auf den Fall $k_n = k = 1$ beschränken.

Satz 3.38.
Es seien $(X_j)_1^n$ reellwertige Zufallsvariablen mit den Verteilungsfunktionen $(F_j)_1^n$ und gemeinsamer Verteilungsfunktion F. Die $(X_j)_1^n$ sind genau dann unabhängig verteilt, wenn gilt

$$F(x_1, \ldots, x_n) = \Pi_1^n F_j(x_j).$$

Beweis:

(1) $(X_j)_1^n$ unabhängig. Es gilt

$$\begin{aligned} F(x_1,\ldots,x_n) &= \mathbb{P}(\{X_1 \le x_1\} \cap \cdots \cap \{X_n \le x_n\}) \\ &= \mathbb{P}(\{X_1 \le x_1\}) \ldots \mathbb{P}(\{X_n \le x_n\}) \\ &= \Pi_1^n F_j(x_j). \end{aligned}$$

(2) Es seien $\mathcal{I} = \{(-\infty, c] : c \in \mathbb{R}\}$ und $\mathcal{C}_j = X_j^{-1}(\mathcal{I}), 1 \le j \le n$. Es ist klar, dass $\mathcal{I}$ ein π–System ist. Somit ist $\mathcal{C}_j, 1 \le j \le n$, ein π–System. Es gilt aber für $C_j \in \mathcal{C}_j, 1 \le j \le n, C_j = X_j^{-1}((-\infty, c_j])$

$$\begin{aligned} \mathbb{P}(C_1 \cap \cdots \cap C_n) &= \mathbb{P}(X_1 \le c_1, \ldots, X_n \le c_n) \\ &= F(c_1,\ldots,c_n) \\ &= \Pi_1^n F_j(c_j) \\ &= \Pi_1^n \mathbb{P}(X_j \le c_j) \\ &= \Pi_1^n \mathbb{P}(C_j). \end{aligned}$$

Somit sind die Ereignissysteme $(\mathcal{C}_j)_1^n$ unabhängig. Nach Satz 2.24 sind die erzeugten σ – Algebren $(\sigma - \mathcal{C}_j)_1^n$ unabhängig. Es gilt aber

$$\sigma - \mathcal{C}_j = \sigma - X_j^{-1}(\mathcal{I}) = X_j^{-1}(\sigma - \mathcal{I}) = X_j^{-1}(\mathcal{B}(\mathbb{R})).$$

Damit sind die Zufallsvariablen $(X_j)_1^n$ voneinander unabhängig.

□

Bemerkung 3.39.
Sind $(X_j)_1^n$ diskrete Zufallsvariablen, dann sind sie genau dann unabhängig verteilt, wenn gilt

$$f_X(x) = \mathbb{P}(X = x) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \Pi_1^n f_j(x_j),$$

wobei f_j die Dichtefunktion von $X_j, 1 \le j \le n$, bezeichnet. Dasselbe gilt für stetige Zufallsvariablen. Es seien $(X_j)_1^n$ Zufallsvariablen mit einer gemeinsamen Dichtefunktion f. Die $(X_j)_1^n$ sind genau dann unabhängig verteilt, wenn gilt

$$f_X(x_1,\ldots,x_n) = \Pi_1^n f_j(x_j)$$

wobei f_j die Dichtefunktion von $X_j, 1 \le j \le n$, bezeichnet.

Beispiel 3.40.
Die Zufallsvariablen X und Y genügen der Normalverteilung $N(\mu, \Sigma)$ mit $\mu = (\mu_X, \mu_Y)^T$ und

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

wobei $-1 < \rho < 1$. Die gemeinsame Dichtefunktion ist gegeben durch

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{(1-\rho^2)}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right.\right. \\ &\quad \left.\left. -\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right). \end{aligned}$$

Dann sind X und Y genau dann unabhängig, wenn $\rho = 0$.

Sind X und Y unabhängige reellwertige Zufallsvariablen, so ist es plausibel, dass $U = X^2$ und $W = \exp(Y)$ auch unabhängige Zufallsvariablen sind. Ferner, sind X, Y und Z unabhängig, so ist zu erwarten, dass $U = X^2 + X\exp(YX - Y^2)$ und $W = Z^2 - 3$ auch unabhängig sind. Dies ist in der Tat der Fall und ist Inhalt des nächsten Satzes.

Satz 3.41.
Es seien $(X_n)_1^\infty$ unabhängig verteilte Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Es gelte $X_n : \Omega \to \mathbb{R}^{k_n}$. Es sei noch $(I_k)_1^\infty$ eine Zerlegung von $\mathbb{N} = \{1, 2, \dots\}$, d. h.

$$\begin{aligned} I_i \cap I_j &= \emptyset, i \neq j, \\ \sum_k I_k &= \mathbb{N}, \end{aligned}$$

mit $|I_k| < \infty$. Ferner seien

$$h_k : \Pi_{j\in I_k}\mathbb{R}^{k_j} =: \mathbb{R}^{N_k} \to \mathbb{R}^{m_k}, k \in \mathbb{N},$$

meßbare Funktionen, wobei $N_k = \sum_{j\in I_k} k_j$. Dann sind die Zufallsvariablen

$$Y_k = h_k \circ (\Pi_{j\in I_k} X_j), k \in \mathbb{N},$$

auch unabhängig verteilt.

Beweis: Es sei $\tilde{X}_k = \Pi_{j\in I_k} X_j$. Es gilt $\tilde{X}_k : \Omega \to \Pi_{j\in I_k}\mathbb{R}^{k_j} = \mathbb{R}^{N_k}$ und

$$\tilde{X}_k^{-1}(\mathcal{B}(\mathbb{R}^{N_k})) = \sigma - \{\Pi_{j\in I_k} X_j^{-1}(\mathcal{B}(\mathbb{R}^{k_j}))\}.$$

Nach der Definition der Unabhängigkeit sind die σ–Algebren $X_j^{-1}(\mathcal{B}(\mathbb{R}^j))$, $j \in \mathbb{N}$, unabhängig. Aus Satz 2.25 folgt, dass die σ–Algebren

$$\tilde{X}_k^{-1}(\mathcal{B}(\Pi_{j\in I_k}\mathbb{R}^{k_j})), k \in \mathbb{N}$$

unabhängig sind. Da aber

$$h_k : \Pi_{j\in I_k}\mathbb{R}^{k_j} \to \mathbb{R}^{m_k}$$

meßbar ist, gilt

$$h_k^{-1}(\mathcal{B}(\mathbb{R}^{m_k})) \subset \mathcal{B}(\Pi_{j\in I_k}\mathbb{R}^{k_j})$$

und somit

$$Y_k^{-1}(\mathcal{B}(\mathbb{R}^{m_k})) \subset \tilde{X}_k^{-1}(\mathcal{B}(\Pi_{j\in I_k}\mathbb{R}^{k_j})).$$

Somit sind die σ–Algebren $Y_k^{-1}(\mathcal{B}((\mathbb{R}^{m_k})), k \in \mathbb{N}$, unabhängig. □

Bemerkung 3.42.
Der Beweis des Satzes ist auch für den Fall $|I_k| = \infty$ korrekt, vorausgesetzt, die Meßbarkeit h von einer Funktion $\mathbb{R}^{\mathbb{N}} \to \mathbb{R}^m$ wird richtig definiert.

4 Erwartungswerte

4.1 Erwartungswert und Varianz

Die Verteilung $\mathbb{P}^X$ einer Zufallsvariablen X enthält die ganze Information über X. In vielen Situationen ist es aber nicht möglich oder zu umständlich, die genaue Verteilung zu bekommen. Man muß sich mit weniger Information oder mit einer Approximation zufrieden geben. In beiden Fällen spielt der Erwartungswert einer Zufallsvariablen eine bedeutende Rolle.

Definition 4.1.
Es sei X eine reellwertige Zufallsvariable. Falls $\int |X| \, d\mathbb{P} < \infty$, wird der Erwartungswert $\mathbb{E}(X)$ von X durch

$$\mathbb{E}(X) := \int X d\mathbb{P}$$

definiert.

Der folgende Satz zeigt, wie man den Erwartungswert $\mathbb{E}(X)$ von X von der Verteilung $\mathbb{P}^X$ gewinnen kann.

Satz 4.2.
Es sei X eine reellwertige Zufallsvariable mit Erwartungswert $\mathbb{E}(X)$ und Verteilung $\mathbb{P}^X$. Dann gilt

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} u \; d\mathbb{P}^X(u).$$

Beweis: Nach Definition gilt

$$\mathbb{E}(X) = \int_\Omega X \; d\mathbb{P} = \int_\Omega g \circ X \; d\mathbb{P}$$

mit $g \equiv$ Identität; $g(x) = x$. Aus dem Transformationssatz für Integrale folgt

$$\mathbb{E}(X) = \int_\Omega g \circ X \; d\mathbb{P} = \int_{X^{-1}(\mathbb{R})} g \circ X \; d\mathbb{P} = \int_{\mathbb{R}} g(u) \; d\mathbb{P}^X(u) = \int_{\mathbb{R}} u \, d\mathbb{P}^X(u).$$

□

Bemerkung 4.3.
Ist X eine diskrete Zufallsvariable mit Dichtefunktion f_X, so gilt

$$\mathbb{E}(X) = \sum_{-\infty}^{\infty} j f_X(j),$$

insofern

$$\sum_{-\infty}^{\infty} |j| f_X(j) < \infty.$$

Ist X eine stetige Zufallsvariable mit Dichtefunktion f_X, so gilt

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} u f_X(u)\, du$$

insofern

$$\int_{-\infty}^{\infty} |u| f_X(u)\, du < \infty.$$

Beispiel 4.4.
X sei gleichverteilt über $\{1, \ldots, n\}$. Es gilt

$$\mathbb{E}(X) = \sum_{-\infty}^{\infty} j f_X(j) = \sum_{1}^{n} j \cdot \frac{1}{n} = \frac{1}{n} \cdot \frac{1}{2} n(n+1) = \frac{n+1}{2}.$$

Beispiel 4.5.
X sei binomialverteilt mit Parametern n und p, $X \sim b(n,p)$. Es gilt

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j} \\
&= \sum_{j=1}^{n} \frac{n!}{(n-j)!(j-1)!} p^j (1-p)^{n-j} \\
&= np \sum_{j=1}^{n} \frac{(n-1)!}{(n-1-(j-1)!(j-1)!} p^{j-1} (1-p)^{n-1-(j-1)} \\
&= np \sum_{j=1}^{n} \binom{n-1}{j-1} p^{j-1} (1-p)^{n-1-(j-1)} \\
&= np \sum_{0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \\
&= np(p + (1-p))^{n-1} = np.
\end{aligned}$$

Beispiel 4.6.
X sei gleichverteilt über das Intervall $[a, b]$. Es gilt

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} u f_X(u)\, du = \int_a^b u \cdot \frac{1}{b-a}\, du = \\ &= \frac{1}{b-a}\left[\frac{1}{2}u^2\right]_a^b = \frac{1}{2}\cdot\frac{1}{b-a}(b^2-a^2) = \frac{1}{2}(a+b).\end{aligned}$$

Nicht alle reellwertigen Zufallsvariablen haben einen Erwartungswert.

Definition 4.7.
Die stetige Zufallsvariable X heißt Cauchy-verteilt mit Parametern μ und $\sigma, X \sim C(\mu, \sigma)$, falls gilt

$$f_X(x) = \frac{\sigma}{\pi(\sigma^2 + (x-\mu)^2)}.$$

Bemerkung 4.8.
Die Funktion f_X ist in der Tat eine Dichtefunktion, denn es gilt

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{1}{\pi}\frac{\sigma}{\sigma^2+(x-\mu)^2}\, dx &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1}{1+u^2}\, du \\ &= \frac{1}{\pi}\left[\arctan(u)\right]_{-\infty}^{\infty} = \frac{1}{\pi}\left[\frac{\pi}{2} - (-\frac{\pi}{2})\right] = 1.\end{aligned}$$

Beispiel 4.9.
Es sei X eine Cauchy-verteilte Zufallsvariable. Dann ist $\mathbb{E}(X)$ nicht definiert. Es gilt nämlich

$$\begin{aligned}\int |X|\, d\mathbb{P} &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{|u|}{1+u^2}\, du \\ &= 2\frac{1}{\pi}\left[\frac{1}{2}\log(1+u^2)\right]_0^{\infty} \\ &= \infty.\end{aligned}$$

Satz 4.10.
Es seien X und Y reellwertige Zufallsvariablen mit Erwartungswerten $\mathbb{E}(X)$ und $\mathbb{E}(Y)$. Dann hat die Zufallsvariable $\alpha X + \beta Y$ den Erwartungswert

$$\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y).$$

Beweis: Es gilt $|\alpha X + \beta Y| \leq |\alpha||X| + |\beta||Y|$. Aus

$$\int |X| \; d\mathbb{P}, \quad \int |Y| \; d\mathbb{P} < \infty$$

folgt

$$\int |\alpha X+\beta Y| \; d\mathbb{P} \leq \int (|\alpha||X|+|\beta||Y|) \; d\mathbb{P} = |\alpha| \int |X| \; d\mathbb{P}+|\beta| \int |Y| \; d\mathbb{P} < \infty.$$

Somit ist der Erwartungswert von $\alpha X + \beta Y$ wohl definiert, und es gilt

$$\mathbb{E}(\alpha X+\beta Y) = \int (\alpha X+\beta Y) \; d\mathbb{P} = \alpha \int X \; d\mathbb{P}+\beta \int Y \; d\mathbb{P} = \alpha\mathbb{E}(X)+\beta\mathbb{E}(Y).$$

□

Satz 4.11.
Es sei X eine Zufallsvariable $X : \Omega \to \mathbb{R}^k$ und $g : \mathbb{R}^k \to \mathbb{R}$ eine Borel-meßbare Funktion. Die Zufallsvariable $Y = g \circ X$ hat genau dann einen Erwartungswert, wenn $\int_{\mathbb{R}^k} |g(u)| \; d\mathbb{P}^X(u) < \infty$, und es gilt

$$\mathbb{E}(Y) = \int_{\mathbb{R}^k} g(u) \; d\mathbb{P}^X(u).$$

Beweis: Dies ist nichts anderes als der Transformationssatz für Integrale. □

Definition 4.12.
Es sei X eine Zufallsvariable mit $\mathbb{E}(X^2) < \infty$. Die Varianz $\mathbb{V}(X)$ wird durch

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

definiert. Falls $\mathbb{V}(X) < \infty$, so wird die Standardabweichung $SA(X)$ von X durch

$$SA(X) = \sqrt{\mathbb{V}(X)}$$

definiert.

Bemerkung 4.13.
Es gilt

$$\begin{aligned} \mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2. \end{aligned}$$

Beispiel 4.14.
X sei gleichverteilt über $\{1, \ldots, n\}$. Es gilt

$$\mathbb{E}(X(X+1)) = \sum_1^n \frac{j(j+1)}{n} = \frac{1}{3}\frac{n(n+1)(n+2)}{n} = \frac{1}{3}(n+1)(n+2).$$

Hieraus folgt

$$\begin{aligned}\mathbb{E}(X^2) &= \frac{1}{3}(n+1)(n+2) - \frac{1}{2}(n+1) \\ &= \frac{n+1}{6}(2n+4-3) = \frac{(n+1)(2n+1)}{6}\end{aligned}$$

und somit

$$\begin{aligned}\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\ &= \frac{n+1}{12}(4n+2-3(n+1)) \\ &= \frac{1}{12}(n+1)(n-1) = \frac{1}{12}(n^2-1).\end{aligned}$$

Beispiel 4.15.
Es sei $X \sim b(n,p)$. Es gilt

$$\mathbb{E}(X(X-1)) = \sum_{j=0}^n j(j-1)\binom{n}{j}p^j(1-p)^{(n-j)} = n(n-1)p^2$$

wie bei Beispiel 4.5. Es gilt also $\mathbb{E}(X^2) = n(n-1)p^2 + np$ und somit

$$\mathbb{V}(X) = n(n-1)p^2 + np - n^2p^2 = np((n-1)p + 1 - np) = np(1-p).$$

Beispiel 4.16.
Es sei $X \sim \mathrm{U}(a,b)$. Es gilt

$$\mathbb{E}(X^2) = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a}\frac{(b^3-a^3)}{3} = \frac{b^2+ab+a^2}{3}.$$

Somit gilt

$$\begin{aligned}\mathbb{V}(X) &= \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} = \frac{1}{12}[4b^2+4ab+4a^2-3a^2-3ab-3b^2] \\ &= \frac{1}{12}\left[a^2+b^2-2ab\right] = \frac{(b-a)^2}{12}.\end{aligned}$$

Definition 4.17.
Es seien X und Y zwei reellwertige Zufallsvariablen mit $\mathbb{E}(X^2) < \infty$ und $\mathbb{E}(Y^2) < \infty$. Dann wird die Kovarianz zwischen X und Y durch

$$\mathrm{cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

definiert. Der Korrelationskoeffizient $\rho(X,Y)$ von X und Y wird definiert durch

$$\rho(X,Y) := \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}.$$

X und Y heißen unkorreliert, falls $\rho(X,Y) = 0$.

Bemerkung 4.18.
$|X| \leq 1 + X^2$ für alle X und somit gilt $\mathbb{E}(|X|) < \infty$, d.h. $\mathbb{E}(X)$ und $\mathbb{E}(Y)$ sind wohl definiert. Ferner gilt

$$2|XY| \leq X^2 + Y^2$$

und somit

$$|(X - \mathbb{E}(X))(Y - \mathbb{E}Y))| \leq \frac{1}{2}(X^2 + Y^2) + |Y||\mathbb{E}(X)| + |X||\mathbb{E}(Y)| + |\mathbb{E}(X)\mathbb{E}(Y)|.$$

Hieraus folgt, dass $\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$ wohl definiert ist.

Satz 4.19.
Es seien X und Y zwei reellwertige Zufallsvariablen mit $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$. Dann gilt

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathrm{cov}(X,Y).$$

Beweis:

$$\begin{aligned}
\mathbb{V}(X+Y) &= \mathbb{E}((X + Y - \mathbb{E}(X+Y))^2) = \mathbb{E}((X - \mathbb{E}(X) + Y - \mathbb{E}(Y))^2) \\
&= \mathbb{E}((X - \mathbb{E}(X))^2 + 2(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) + (Y - \mathbb{E}(Y))^2) \\
&= \mathbb{V}(X) + 2\mathrm{cov}(X,Y) + \mathbb{V}(Y).
\end{aligned}$$

□

Korollar
Sind X und Y unkorrelierte Zufallsvariablen, so gilt

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y).$$

4.2 Unabhängig verteilte Zufallsvariablen

In der Regel stellen Zufallsvariablen Meßergebnisse von Experimenten dar. Sind die einzelnen Durchführungen des Experiments voneinander unabhängig, so sind die resultierenden Meßergebnisse auch unabhängig. Die Untersuchung von unabhängig verteilten Zufallsvariablen ist deswegen von überragender Bedeutung. Der nächste Satz stellt die Basis aller weiteren Überlegungen dar.

Satz 4.20.
Es seien X und Y unabhängig verteilte reellwertige Zufallsvariablen mit $\mathbb{E}(|X|) < \infty$ und $\mathbb{E}(|Y|) < \infty$. Dann gilt

$$\mathbb{E}(|XY|) < \infty \quad \textit{und} \quad \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Beweis: Der Beweis läuft wie folgt. Wir beweisen den Satz zuerst für X und Y Indikatorfunktionen $X = F, Y = G$. Dann halten wir Y fest und zeigen, dass die Behauptung für primitive Funktionen X gilt. Durch Grenzübergang erhalten wir den Satz für $X \geq 0$ und $Y = G$. Nun kann man X festhalten und das ganze Argument für Y wiederholen. Schließlich erhalten wir den Satz für $X \geq 0, Y \geq 0$. Der allgemeine Fall erfolgt nun aus $X = X_+ - X_-, Y = Y_+ - Y_-$. Eine solche Beweismethode ist Standard in der Wahrscheinlichkeitstheorie. Es sei $X = F, Y = G$. Dann sind F und G unabhängig und es gilt

$$\mathbb{E}(XY) = \mathbb{P}(F \cap G) = \mathbb{P}(F) \cdot \mathbb{P}(G) = \mathbb{E}(X)\mathbb{E}(Y).$$

Es sei nun $Y = G$ fest und $X = \sum_1^n a_j F_j$ mit $a_j \geq 0$. Da X und Y unabhängig sind, sind F_j und G unabhängig, $1 \leq j \leq n$. Es gilt

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}(\sum_1^n a_j F_j G) = \sum_1^n a_j \mathbb{E}(F_j G) = \\
&= \sum_1^n a_j \mathbb{E}(F_j)\mathbb{E}(G) = \mathbb{E}(\sum_1^n a_j F_j)\mathbb{E}(G) \\
&= \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

Es sei nun $X \geq 0$. Dann gibt es eine Folge $(X_n)_1^\infty$ nicht-negativer primitiver Funktionen $X_n = \sum_{j=1}^{m_n} a_{n,j} F_{n,j}$ mit $X_n \uparrow X$ und $F_{n,j} \in X^{-1}(\mathcal{B}(\mathbb{R}))$. Die $(F_{n,j})$ und somit die X_n sind unabhängig von Y und $X_n Y \uparrow XY$. Es gilt

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}(\lim_{n\to\infty} X_n Y) \\
&= \lim_{n\to\infty} \mathbb{E}(X_n Y) \text{(monotone Konvergenz bzw. Definition des Integrals)} \\
&= \lim_{n\to\infty} \mathbb{E}(X_n)\mathbb{E}(Y) \\
&= \mathbb{E}(X)\mathbb{E}(Y) \text{(monotone Konvergenz bzw. Definition des Integrals).}
\end{aligned}$$

Somit gilt der Satz für $X \geq 0$ und $Y = G$. Wir halten nun X fest und wiederholen das Argument für Y. Es ergibt sich

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

falls $X, Y \geq 0$. Im Allgemeinen schreiben wir $X = X_+ - X_-, Y = Y_+ - Y_-$. Es gilt

$$\begin{aligned}
\mathbb{E}(|XY|) &= \mathbb{E}((X_+ + X_-)(Y_+ + Y_-)) \\
&= \mathbb{E}(X_+Y_+ + X_+Y_- + X_-Y_+ + X_-Y_-) \\
&= \mathbb{E}(X_+Y_+) + \cdots + \mathbb{E}(X_-Y_-) \\
&= \mathbb{E}(X_+)\mathbb{E}(Y_+) + \cdots + \mathbb{E}(X_-)\mathbb{E}(Y_-) \\
&= \mathbb{E}(X_+ + X_-)\mathbb{E}(Y_+ + Y_-) \\
&= \mathbb{E}(|X|)\mathbb{E}(|Y|) < \infty.
\end{aligned}$$

Somit ist $\mathbb{E}(|XY|) < \infty$, und dasselbe Argument ergibt nun $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. □

Korollar.
Es seien X und Y unabhängig verteilte Zufallsvariablen mit endlichen Varianzen. Dann gilt

$$\operatorname{cov}(X, Y) = 0,$$

d.h. X und Y sind unkorreliert.
Beweis: Es gilt

$$\begin{aligned}
\operatorname{cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(X - \mathbb{E}(X))\mathbb{E}(Y - \mathbb{E}(Y)) \\
&= 0 \cdot 0.
\end{aligned}$$

□

4.3 Momente und Ungleichungen

Viele Überlegungen in der Wahrscheinlichkeitstheorie basieren auf der Abschätzung von Wahrscheinlichkeiten. Dabei spielen die Momente einer Zufallsvariablen eine bedeutende Rolle.

Definition 4.21.
Es sei X eine reellwertige Zufallsvariable. Das Absolutmoment von X von Ordnung r wird durch $\mathbb{E}(|X|^r)$ definiert. Falls für $n \in \mathbb{N}$, $\mathbb{E}(|X|^n) < \infty$, so wird das n–Moment von X durch

$$m_n = \mathbb{E}(X^n)$$

definiert.

Beispiel 4.22.
Es sei $X \sim \mathrm{E}(\lambda)$. Es gilt

$$\mathbb{E}(X^n) = \int_0^\infty x^n \lambda \exp(-\lambda x)\, dx.$$

Die Substitution $u = \lambda x$ ergibt

$$\begin{aligned} \mathbb{E}(X^n) &= \int_0^\infty \left(\frac{u}{\lambda}\right)^n \exp(-u)\, du \\ &= \lambda^{-n} \cdot \int_0^\infty u^n \exp(-u)\, du \\ &= \Gamma(n+1)\lambda^{-n} = n!\lambda^{-n}. \end{aligned}$$

Satz 4.23. *(Hölder)*
Es seien X und Y reellwertige Zufallsvariablen und p und q positive Zahlen mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt

$$\mathbb{E}(|XY|) \leq \mathbb{E}(|X|^p)^{1/p} \cdot \mathbb{E}(|Y|^q)^{1/q}.$$

Korollar 1
Es sei X eine reellwertige Zufallsvariable mit endlichem absoluten Moment von Ordnung s. Dann sind die absoluten Momente von Ordnung $r \leq s$ auch endlich, und es gilt

$$\mathbb{E}(|X|^r)^{\frac{1}{r}} \leq \mathbb{E}(|X|^s)^{\frac{1}{s}}$$

Beweis: Wir setzen in Satz 4.23

$$Y \equiv 1, p = s/r \text{ und } q = s/(s-r)$$

Es gilt

$$\begin{aligned} \mathbb{E}(|X|^r \cdot 1) &\leq \mathbb{E}((|X|^r)^{s/r})^{r/s} \mathbb{E}(1^q)^{1/q} \\ &= \mathbb{E}(|X|^s)^{r/s} < \infty. \end{aligned}$$

□

Korollar 2
Es seien X und Y reelle Zufallsvariablen mit endlichen zweiten Momenten. Dann gilt $-1 \leq \rho_{XY} \leq 1$.
Beweis: Es gilt

$$\begin{aligned}\left|\int (X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\ d\mathbb{P})\right| &\leq \int |X-\mathbb{E}(X)| \cdot |Y-\mathbb{E}(Y)|\ d\mathbb{P} \\ &\leq \left(\int (X-\mathbb{E}(X))^2\ d\mathbb{P}\right)^{1/2} \left(\int (Y-\mathbb{E}(Y))^2\ d\mathbb{P}\right)^{1/2}\end{aligned}$$

woraus folgt: $\rho_{XY}^2 \leq 1$. □

Satz 4.24. *(Minkowski)*
Es seien X und Y reellwertige Zufallsvariablen mit endlichem absoluten s–Moment. Dann gilt

$$\mathbb{E}(|X+Y|^s)^{1/s} \leq \mathbb{E}(|X|^s)^{1/s} + \mathbb{E}(|Y|^s)^{1/s}.$$

Satz 4.25. *(Markov)*
Es sei X eine Zufallsvariable mit endlichem absoluten s–Moment. Dann gilt

$$\mathbb{P}(|X| \geq \lambda) \leq \mathbb{E}(|X|^s)\lambda^{-s}$$

für alle $\lambda > 0$.

Beweis: Es gilt

$$\begin{aligned}\mathbb{E}(|X|^s) &= \int |X|^s\ d\mathbb{P} = \int_{\{|X|\geq\lambda\}} |X|^s\ d\mathbb{P} + \int_{\{|X|<\lambda\}} |X|^s\ d\mathbb{P} \\ &\geq \int_{\{|X|\geq\lambda\}} |X|^s d\mathbb{P} \geq \lambda^s \int_{\{|X|\geq\lambda\}} d\mathbb{P} \\ &= \lambda^s \mathbb{P}(|X| \geq \lambda)\end{aligned}$$

□

Korollar(Tschebychev)
Es sei X eine reellwertige Zufallsvariable mit endlicher Varianz. Dann gilt

$$\mathbb{P}(|X-\mathbb{E}(X)| \geq \lambda) \leq \mathbb{V}(X)/\lambda^2.$$

Beweis: Wir setzen $Y = X - \mathbb{E}(X)$. Nach dem Satz 4.25 gilt

$$\begin{aligned}\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) &= \mathbb{P}(|Y| \geq \lambda) \leq \mathbb{E}(|Y|^2)/\lambda^2 \\ &= \mathbb{E}((X - \mathbb{E}(X))^2)/\lambda^2 = \mathbb{V}(X)/\lambda^2.\end{aligned}$$

□

Beispiel 4.26.
Wir betrachten eine Zufallsvariable X, die $\mathrm{N}(0,1)-$verteilt ist. Es gilt

$$\mathbb{E}(X) = 0 \quad \mathbb{V}(X) = 1.$$

Eine Tabelle für die $\mathrm{N}(0,1)-$Verteilung ergibt

$$\mathbb{P}(|X| \geq 1.96) = 0.05.$$

Nach Tschebychev erhält man die Ungleichung

$$\mathbb{P}(|X| \geq 1.96) \leq \frac{1}{1.96^2} = 0.2603.$$

Die Tschebychev'sche Abschätzung ist zwar richtig, aber verhältnismäßig grob.

Satz 4.27.
Es seien X und Y reelle Zufallsvariablen. Dann gilt $\mathbb{P}^X = \mathbb{P}^Y$ genau dann, wenn

$$\mathbb{E}(u(X)) = \mathbb{E}(u(Y))$$

für alle unendlich oft differenzierbaren Funktionen u mit kompaktem Träger.

Beweis: Aus $\mathbb{P}^X = \mathbb{P}^Y$ folgt sofort $\mathbb{E}(u(X)) = \mathbb{E}(u(Y))$ für alle u, für die die Erwartungswerte definiert sind. Es genügt also, die andere Richtung zu beweisen. Es sei nun $\epsilon > 0$ gegeben. Wir wählen K so, dass gilt

$$\mathbb{P}(X \leq K) < \epsilon.$$

Für x und $\delta > 0$ sei nun u eine unendlich oft differenzierbare Funktion mit $0 \leq u \leq 1, u(y) = 1$ für $K \leq y \leq x, u(y) = 0$ für $y \geq x + \delta$. Es gilt

$$\begin{aligned}\mathbb{P}(X \leq x) &= \mathbb{P}(X \leq K) + \mathbb{P}(K < X \leq x) \\ &= \mathbb{P}(X \leq K) + \int_{K<X\leq x} u(X)\, d\mathbb{P} \\ &\leq \epsilon + \int u(X)\, d\mathbb{P} \\ &= \epsilon + \int u(Y)\, d\mathbb{P} \\ &\leq \epsilon + \mathbb{P}(Y \leq x + \delta).\end{aligned}$$

Da aber ϵ und δ beliebig sind, folgt nun $\mathbb{P}(X \leq x) \leq \mathbb{P}(Y \leq x)$. Analog gilt $\mathbb{P}(Y \leq x) \leq \mathbb{P}(X \leq x)$ und somit $F_X = F_Y$, d.h. $\mathbb{P}^X = \mathbb{P}^Y$. □

Satz 4.28.
Es seien X und Y beschränkte Zufallsvariablen mit $\mathbb{E}(X^n) = \mathbb{E}(Y^n)$ für alle n. Dann gilt

$$\mathbb{P}^X = \mathbb{P}^Y.$$

Beweis: Es sei K so, dass $-K \leq X, Y \leq K$. Die Zufallsvariablen

$$X_1 = \frac{K - X}{2K}, \quad Y_1 = \frac{K - Y}{2K}$$

liegen zwischen 0 und 1, und aus $\mathbb{E}(X^n) = \mathbb{E}(Y^n)$ für alle n folgt $\mathbb{E}(X_1^n) = \mathbb{E}(Y_1^n)$ für alle n. Ferner gilt $\mathbb{P}^X = \mathbb{P}^Y$ genau dann, wenn $\mathbb{P}^{X_1} = \mathbb{P}^{Y_1}$. Es genügt also, den Fall zu betrachten, dass X und Y in $[0, 1]$ liegen. Nach Satz 4.27 genügt es zu zeigen, dass $\mathbb{E}(u(X)) = \mathbb{E}(u(Y))$ für alle stetigen Funktionen $u : [0, 1] \to \mathbb{R}$. Ist u eine solche Funktion, dann gibt es zu jedem $\epsilon > 0$ ein Polynom p derart, dass gilt

$$\sup_{0 \leq x \leq 1} |u(x) - p(x)| < \epsilon.$$

Aus $\mathbb{E}(X^n) = \mathbb{E}(Y^n)$ für alle n folgt $\mathbb{E}(p(X)) = \mathbb{E}(p(Y))$. Wir erhalten also

$$\begin{aligned} |\mathbb{E}(u(X)) - \mathbb{E}(u(Y))| \quad &\leq \quad \mathbb{E}(|u(X) - p(X)|) + \mathbb{E}(|u(Y - p(Y)|) \\ &\quad + |\mathbb{E}(p(X) - \mathbb{E}(p(Y))| \leq 2\epsilon. \end{aligned}$$

Da ϵ aber beliebig klein gewählt werden kann, folgt $\mathbb{E}(u(X)) = \mathbb{E}(u(Y))$ und somit $\mathbb{P}^X = \mathbb{P}^Y$. □

4.4 Charakteristische Funktionen

Definition 4.29. *Es sei X eine reellwertige Zufallsvariable. Die charakteristische Funktion ϕ_X von X wird durch*

$$\phi_X(t) = \mathbb{E}(\exp(itX)) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX)), t \in \mathbb{R}$$

definiert.

Bemerkung 4.30. Wegen $\mathbb{E}(|\exp(itX)|) = 1$ folgt, dass ϕ_X für alle reellwertige Zufallsvariablen wohl definiert ist.

Beispiel 4.31.
Es sei $X \sim \mathrm{C}(0,1)$ Cauchy-verteilt mit den Parametern 0 und 1. Es gilt

$$\phi_X(t) = \mathbb{E}(\exp(itX)) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\exp(itx)}{1+x^2}\,dx.$$

Für $t > 0$ gilt nach dem Satz von Cauchy

$$\begin{aligned}\frac{1}{\pi}\int_{\underline{\cap}} \frac{\exp(itz)}{1+z^2}\,dz &= 2\pi i \quad \sum \text{ Residuen} \\ &= 2\pi i \cdot \frac{\exp(-t)}{\pi \cdot 2i},\end{aligned}$$

wobei $\underline{\cap}$ den Halbkreis $|z| = R, \mathcal{I}m(z) \geq 0$ sowie

$$\mathcal{I}m(z) = 0, -R \leq \mathcal{R}e(z) \leq R,$$

bezeichnet. Hieraus folgt

$$\begin{aligned}\exp(-t) &= \frac{1}{\pi}\int_{\underline{\cap}} \frac{\exp(itz)}{1+z^2}\,dz \\ &= \frac{1}{\pi}\int_{-R}^{R} \frac{\exp(itx)}{1+x^2}\,dx + \frac{1}{\pi}\int_{\cap} \frac{\exp(itz)}{1+z^2}\,dz \qquad (4.1)\end{aligned}$$

wobei $\cap = \{|z| = R, \mathcal{I}m(z) > 0\}$. Es gilt $Z = Re^{i\theta} = R\cos\theta + iR\sin\theta$ und somit

$$\begin{aligned}|\exp(itz)| &= |\exp(itR\cos\theta - tR\sin\theta)| \\ &= |\exp(-tR\sin\theta)|.\end{aligned}$$

Wegen $\sin\theta > 0$ für $0 < \theta < \pi$ erhalten wir $|\exp(itz)| \leq 1$ auf $\cap$. Hieraus folgt

$$\frac{1}{\pi}\left|\int_{\cap} \frac{\exp(itz)}{1+z^2}\,dz\right| \leq \frac{1}{\pi}\cdot\frac{\pi R}{R^2-1}, \quad R > 1.$$

Unter Berücksichtigung von (4.1) erhalten wir also

$$\phi_X(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\exp(itx)}{1+x^2}\,dx = \lim_{R\to\infty}\frac{1}{\pi}\int_{-R}^{R} \frac{\exp(itx)}{1+x^2}\,dx = \exp(-t)$$

für $t > 0$. Für $t < 0$ gilt analog $\phi_X(t) = \exp(t)$ und somit

$$\phi_X(t) = \exp(-|t|).$$

Satz 4.32.
Es sei ϕ_X eine charakteristische Funktion. Dann gilt

(a) $|\phi_X(t)| \leq \phi_X(0) = 1$

(b) $\phi_X : \mathbb{R} \to \mathbb{C}$ *ist gleichmäßig stetig*

(c) $\phi_{aX+b}(t) = \exp(ibt)\phi_X(at)$.

Beweis:

(a) Es gilt

$$|\phi_X(t)| = |\mathbb{E}(\exp(itX))| \leq \mathbb{E}(|\exp(itX)|) = 1.$$

Wegen $\phi_X(0) = \mathbb{E}(\exp(0)) = 1$ erhalten wir die erste Behauptung.

(b) Es gilt

$$\begin{aligned} |\phi_X(t+h) - \phi_X(t)| &= |\mathbb{E}(\exp(i(t+h)X) - \exp(itX))| \\ &= |\mathbb{E}(\exp(itX)(\exp(ihX) - 1))| \\ &\leq \mathbb{E}(|\exp itX)(\exp(ihX) - 1)|) \\ &= \mathbb{E}(|\exp(ihX) - 1|). \end{aligned}$$

Aus $|\exp(ihX) - 1| \leq 2$ folgt mit dominierter Konvergenz (2 ist eine Majorante)

$$\lim_{h \to 0} \mathbb{E}(|\exp(ihX) - 1|) = 0.$$

Da die rechte Seite unabhängig von t ist, ist ϕ_X gleichmäßig stetig.

c) Schließlich gilt

$\phi_{aX+b}(t) = \mathbb{E}(\exp(it(aX+b)) = \exp(itb)\mathbb{E}(\exp(itaX)) = \exp(itb)\phi_X(at)$

und somit die letzte Behauptung.

□

Bemerkung 4.33.
Wir benutzen unten den Satz von Fubini. Er sagt Folgendes aus: Es sei $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ ein Produktraum und $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ eine meßbare Funktion. Dann gilt

$$\begin{aligned} \int f \; d\mathbb{P}_1 \otimes \mathbb{P}_2 &= \int \left(\int f(\omega_1, \omega_2) \; d\mathbb{P}_1(\omega_1) \right) \; d\mathbb{P}_2(\omega_2) \\ &= \int \left(\int f(\omega_1, \omega_2) \, d\mathbb{P}_2(\omega_2) \right) \; d\mathbb{P}_1(\omega_1) \end{aligned}$$

insofern eine der folgenden äquivalenten Bedingungen gilt:

(i) $\int |f| \, d\mathbb{P}_1 \otimes \mathbb{P}_2 < \infty$

(ii) $\int \left(\int |f(\omega_1, \omega_2)| \, d\mathbb{P}_1(\omega_1) \right) \, d\mathbb{P}_2(\omega_2) < \infty$

(iii) $\int \left(\int |f(\omega_1, \omega_2)| \, d\mathbb{P}_2(\omega_2) \right) \, d\mathbb{P}_1(\omega_1) < \infty.$

Satz 4.34.
Es seien X und Y Zufallsvariablen mit Verteilungsfunktionen F_X und F_Y und charakteristischen Funktionen ϕ_X und ϕ_Y. Dann gilt

$$\int \exp(-ist)\phi_X(s) dF_Y(s) = \int \phi_Y(x-t) dF_X(x).$$

Beweis: Es gilt $\phi_X(s) = \int \exp(isx) dF_X(x)$ und somit

$$\exp(-ist)\phi_X(s) = \int \exp(is(x-t) dF_X(x).$$

Hieraus folgt

$$\int \exp(-ist)\phi_X(s) dF_Y(s) = \int \left(\int \exp(is(x-t)) dF_X(x) \right) dF_Y(s).$$

Wegen

$$\int \int |\exp(is(x-t))| dF_X dF_Y = 1 < \infty$$

können wir nach Fubini die Integrationsreihenfolge vertauschen. Wir erhalten

$$\int \exp(-ist)\phi_X(s) \, ds = \int \phi_Y(x-t) dF_X(x).$$

□

Satz 4.35.
Es sei X eine Zufallsvariable mit Verteilungsfunktion F_X und charakteristischer Funktion ϕ_X. Dann gilt für jeden Stetigkeitspunkt y von F_X:

$$F_X(y) = \lim_{\sigma \to \infty} \frac{1}{2\pi} \int_{-\infty}^{y} \left(\int_{-\infty}^{\infty} \exp(-ist)\phi_X(s) \exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2} \right) ds \right) dt.$$

Beweis: Im Satz 4.34 sei $Y \sim \mathrm{N}(0, \sigma^2)$. Dann gilt $\phi_Y(s) = \exp\left(-\frac{1}{2}s^2\sigma^2\right)$, und aus Satz 4.34 folgt nun

$$\frac{1}{\sqrt{2\pi}\sigma} \int \exp(-ist)\phi_X(s) \exp\left(-\frac{1}{2}\frac{s}{\sigma^2} \right) ds = \int \exp\left(-\frac{1}{2}\sigma^2(x-t)^2 \right) dF_X(x).$$

Dies impliziert

$$\frac{1}{2\pi}\int \exp(-ist)\phi_X(s)\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right)\,ds = \frac{1}{\sqrt{2\pi}}\sigma\cdot\int \exp\left(-\frac{1}{2}\sigma^2(x-t)^2\right)dF_X(x)$$

und somit

$$\begin{aligned} & \frac{1}{2\pi}\int_{-\infty}^{y}\left(\int \exp(-ist)\phi_X(s)\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right)\,ds\right)dt \\ = & \int_{-\infty}^{\infty}\left(\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{y}\exp\left(-\frac{1}{2}\sigma^2(x-t)^2\right)\,dt\right)dF_X(x), \end{aligned}$$

wobei wir wieder den Satz von Fubini benutzt haben. Die Subsitution $u = \sigma(t-x)$ ergibt

$$\begin{aligned} & \frac{1}{2\pi}\int_{-\infty}^{y}\left(\int \exp(-ist)\phi_X(s)\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right)\,ds\right)dt \\ = & \int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\sigma(y-x)}\exp\left(-\frac{1}{2}u^2\right)\,du\right)dF_X(x). \end{aligned}$$

Es gilt aber

$$\lim_{\sigma\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\sigma(y-x)}\exp\left(-\frac{1}{2}u^2\right)\,du = \{x<y\}+\frac{1}{2}\{x=y\}$$

und somit

$$\begin{aligned} & F_X(y-0)+\frac{1}{2}F_X(\{y\}) \\ = & \lim_{\sigma\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}\left(\int \exp(-ist)\phi_X(s)\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right))\,ds\right)\,dt. \end{aligned}$$

Nun ist y ein Stetigkeitspunkt von F_X genau dann, wenn

$$F_X(y-0) = F_X(y) \Leftrightarrow F_X(\{y\}) = 0.$$

Es folgt also

$$F_X(y) = \lim_{\sigma\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}\left(\int \exp(-ist)\phi_X(s)\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right)\,ds\right)\,dt.$$

□

Korollar.
Es seien X und Y reelle Zufallsvariablen, und es gelte $\phi_X(t) = \phi_Y(t)$ für alle $t \in \mathbb{R}$. Dann folgt $\mathbb{P}^X = \mathbb{P}^Y$.
Beweis: Aus Satz 4.35 folgt $F_X(t) = F_Y(t)$ in allen gemeinsamen Stetigkeitspunkten t von F_X und F_Y. Da diese Punkte dicht in $\mathbb{R}$ liegen, folgt mit der rechtsseitigen Stetigkeit der Verteilungsfunktionen, dass gilt $F_X = F_Y$. Die Behauptung folgt nun aus Satz 3.8. □

Satz 4.36.
Es sei X eine Zufallsvariable mit charakteristischer Funktion ϕ_X. Gilt $\int |\phi_X(t)|\, dt < \infty$, so folgt, dass X eine stetige Zufallsvariable mit Dichtefunktion

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ixt)\phi_X(t)\, dt$$

ist.

Beweis: Es seien x und $y, x < y$, Stetigkeitspunkte von F_X. Nach Satz 4.34 gilt

$$F_X(y) - F_X(x) = \lim_{\sigma\to\infty} \frac{1}{\sqrt{2\pi}} \int_x^y \left(\int \exp(-ist)\phi_X(s) \exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) ds \right) dt.$$

Es gilt

$$\begin{aligned} & \int_x^y \left(\int_{-\infty}^{\infty} \left| \exp(-ist)\phi_X(s) \exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) \right| ds \right) dt \\ \leq\ & \int_x^y \left(\int_{-\infty}^{\infty} |\phi_X(s)|\, ds \right) dt \\ =\ & (y-x) \cdot \int_{-\infty}^{\infty} |\phi_X(s)|\, ds < \infty. \end{aligned}$$

Somit folgt mit dominierter Konvergenz

$$\begin{aligned} F_X(y) - F_X(x) &= \lim_{\sigma\to\infty} \frac{1}{\sqrt{2\pi}} \int_x^y \left(\int_{-\infty}^{\infty} \exp(-ist)\phi_X(s) \exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) ds \right) dt \\ &= \frac{1}{2\pi} \int_x^y \left(\int_{-\infty}^{\infty} \lim_{\sigma\to\infty} \exp(-ist)\phi_X(s) \exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right) ds \right) dt \\ &= \frac{1}{2\pi} \int_x^y \left(\int_{-\infty}^{\infty} \exp(-ist)\phi_X(s)\, ds \right) dt. \end{aligned}$$

Damit ist

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-ist)\phi_X(s)\,ds$$

eine Dichtefunktion für F_X. □

Satz 4.37.
Es sei X eine reellwertige Zufallsvariable mit endlichem absoluten Moment von Ordnung $n \in \mathbb{N}$. Dann ist ϕ_X n-mal stetig differenzierbar, und es gilt für $0 \leq m \leq n$

$$\phi_X^{(m)}(t) = i^m \int_{-\infty}^{\infty} x^m \exp(ixt)\,dF_X(x) = i^m \mathbb{E}(X^m \exp(iXt)).$$

Beweis: Wir führen den Beweis über Induktion. Der Satz gilt für $m = 0$. Es sei nun $0 \leq m < n$. Aus der Induktionsannahme folgt

$$\frac{\phi_X^{(m)}(t+h) - \phi_X^{(m)}(t)}{h} = i^m \mathbb{E}\left(X^m \frac{(\exp(iX(t+h)) - \exp(iXt))}{h}\right).$$

Es gilt

$$\left|X^m \exp(iXt)\frac{(\exp(iXh)-1)}{h}\right| \leq |X|^m \cdot |X| \leq |X|^{m+1} \leq 1 + |X|^n.$$

Wegen $\mathbb{E}(|X^n|) < \infty$ ist $1 + |X|^n$ eine Majorante. Es folgt mit dominierter Konvergenz

$$\begin{aligned}\lim_{h\to 0}\frac{\phi_X^{(m)}(t+h) - \phi_X^{(m)}(t)}{h} &= i^m \mathbb{E}\left(X^m \exp(itX)\lim_{h\to 0}\frac{(\exp(iXh)-1)}{h}\right)\\ &= i^{m+1}\mathbb{E}(X^{m+1}\exp(itX)).\end{aligned}$$

Da $\phi_X^{(m)}$ für $0 \leq m < n$ differenzierbar ist, ist $\phi_X^{(m)}$ auch stetig. Im Falle $m = n$ gilt

$$\phi_X^{(n)}(t+h) - \phi_X^{(n)}(t) = i^n \mathbb{E}(X^n \exp(iXt)(\exp(iXh) - 1)).$$

Es gilt

$$|X^n \exp(iXt)(\exp(iXh) - 1) \leq 2|X|^n$$

und somit folgt mit dominierter Konvergenz

$$\begin{aligned}\lim_{h\to 0}(\phi_X^{(n)}(t+h)) - \phi_X^{(n)}(t)) &= i^n \mathbb{E}(X^n \exp(iXt)\lim_{h\to 0}(\exp(iXh)-1))\\ &= 0.\end{aligned}$$

□

Bemerkung 4.38.
Der Beweis zeigt, dass $\phi_X^{(m)}$ gleichmäßig stetig ist für $0 \leq m \leq n$. Die Umkehrung von Satz 4.37 gilt nur für gerade Momente. Wir beschränken uns auf den Fall $k = 2$.

Satz 4.39.
Es sei X eine reellwertige Zufallsvariable mit charakteristischer Funktion ϕ_X. Ist ϕ_X zweimal differenzierbar im Nullpunkt, dann ist $\mathbb{E}(X^2) < \infty$.

Beweis: Nach Voraussetzung gibt es ein $\epsilon > 0$ derart, dass ϕ_X einmal differenzierbar im Intervall $[-\epsilon, \epsilon]$ ist. Ferner gibt es eine komplexe Zahl z so, dass

$$\lim_{t \to 0} \frac{\phi_X^{(1)}(t) - \phi_X^{(1)}(0)}{t} = z.$$

Es gilt

$$\mathcal{R}e(\phi_X(t)) = \mathbb{E}(\cos(tX))$$

und somit ist $\mathcal{R}e(\phi_X(t)) = u(t)$ eine gerade Funktion. Nach Voraussetzung existiert $u^{(1)}(t)$ für $|t| < \epsilon$ und $u^{(1)}$ ist eine ungerade Funktion von t. Es gilt also $u^{(1)}(0) = 0$. Nach dem Mittelwertsatz gilt

$$\left| \frac{u(t) - 1}{t^2} \right| = \left| \frac{u^{(1)}(\theta t)}{\theta t} \right|$$

mit $0 < \theta < 1$ und somit

$$\overline{\lim_{t \to 0}} \left| \frac{u(t) - 1}{t^2} \right| = \overline{\lim_{t \to 0}} \left| \frac{u^{(1)}(\theta t)}{\theta t} \right| = |u^{(2)}(0)| < \infty.$$

Es gilt also nach dem Lemma von Fatou

$$\begin{aligned}
\frac{1}{2} \int x^2 F_X(x) &= \int \lim_{t \to 0} \frac{1 - \cos(tx)}{t^2} dF_X \\
&\leq \overline{\lim_{t \to 0}} \int \frac{1 - \cos(tx)}{t^2} dF_X \\
&= \overline{\lim_{t \to 0}} \frac{1 - u(t)}{t^2} \\
&\leq |u^{(2)}(0)| \\
&< \infty.
\end{aligned}$$

□

5 Summen von unabhängig verteilten Zufallsvariablen

5.1 Einführung

Viele wichtige Statistiken lassen sich als Summen von Zufallsvariablen entweder direkt darstellen oder approximinieren.

Beispiel 5.1.

(a) Das arithmetische Mittel

$$\overline{X}_n = \frac{1}{n}\sum_1^n X_j.$$

(b) Die empirische Varianz

$$S_n^2 = \frac{1}{n}\sum_1^n (X_j - \overline{X}_n)^2 = \frac{1}{n}\sum_1^n X_j^2 - (\frac{1}{n}\sum_1^n X_j)^2.$$

Satz 5.2. *Es seien X und Y unabhängig verteilte Zufallsvariablen mit Verteilungsfunktionen F_X und F_Y. Dann ist die Verteilungsfunktion F_{X+Y} der Summe $X+Y$ gegeben durch*

$$F_{X+Y}(t) = \int F_X(t-y)dF_Y(y) = \int F_Y(t-x)dF_X(x).$$

Beweis: Da Xund Y unabhängig verteilt sind, gilt für jede Borel-Menge B von $\mathbb{R}^2$

$$\mathbb{P}(\{\omega : (X(\omega), Y(\omega)) \in B\}) = \int \{(x,y) \in B\} dF_X(x) dF_Y(y)$$

Für $B = \{(x,y) : x + y \leq t\}$ erhalten wir

$$\mathbb{P}(X+Y \leq t) = \mathbb{P}((X,Y) \in B) = \int \{(x,y) \in B\} dF_X(x) dF_Y(y)$$

$$= \int \{x+y \leq t\} dF_X(x) dF_Y(y) = \int \left(\int \{x+y \leq t\} dF_Y(y) \right) dF_X(x)$$

nach dem Satz von Fubini. Hieraus folgt

$$\mathbb{P}(X+Y\leq t)=\int\left(\int\{y\leq t-x\}dF_y(y)\right)dF_X(x)=\int F_Y(t-x)dF_X(x).$$

Analog gilt

$$\mathbb{P}(X+Y\leq t)=\int F_X(t-y)dF_Y(y)$$

und somit die Behauptung des Satzes. □

Korollar
Sind X und Y beide diskret bzw. stetig, dann ist $X+Y$ diskret bzw. stetig mit Dichtefunktion

$$f_{X+Y}(n)=\sum_m f_X(n-m)f_Y(m)=\sum_m f_Y(n-m)f_X(m)$$

bzw.

$$f_{X+Y}(t)=\int f_Y(t-x)f_X(x)dx=\int f_X(t-y)f_Y(y)dy.$$

Beweis:

(i) Sind X und Y beide diskret, so gilt $X\in\mathbb{Z}$ und $Y\in\mathbb{Z}$. Hieraus folgt $X+Y\in\mathbb{Z}$, d.h. $X+Y$ ist diskret. Es gilt

$$f_{X+Y}(n)=\mathbb{P}(X+Y=n)=\sum_{i+j=n}\mathbb{P}(X=i,Y=j).$$

Da X und Y unabhängig sind, gilt

$$\mathbb{P}(X=i,Y=j)=\mathbb{P}(X=i)\mathbb{P}(Y=j)=f_X(i)f_Y(j).$$

Wir erhalten also

$$f_{X+Y}(n)=\sum_{i+j=n}f_X(i)f_Y(j)=\sum_j f_X(n-j)f_Y(j)=\sum_i f_Y(n-i)f_X(i).$$

(ii) Aus dem Satz folgt

$$\begin{aligned}F_{X+Y}&=\int F_Y(t-x)dF_X(x)\\&=\int\left(\int_{-\infty}^{t-x}f_Y(y)dy\right)f_X(x)dx\\&=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{t}f_Y(y-x)dy\right)f_X(x)dx\\&=\int_{-\infty}^{t}\int_{-\infty}^{\infty}f_Y(y-x)f_X(x)dx\,dy\end{aligned}$$

nach dem Satz von Fubini. Dies impliziert, dass

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_Y(t-x) f_X(x) dx$$

eine Dichtefunktion für $X+Y$ ist. Analog ist $\int_{-\infty}^{\infty} f_X(t-y) f_Y(y) dy$ eine Dichtefunktion für $X+Y$. □

Beispiel 5.3.
Es seien $X \sim b(m,p), Y \sim b(n,p)$ und X und Y unabhängig verteilt. Es gilt

$$\begin{aligned}
f_{X+Y}(k) &= \sum_{j=0}^{k} \mathbb{P}(X=j, Y=k-j) \\
&= \sum_{j=0}^{k} \mathbb{P}(X=j)\mathbb{P}(Y=k-j) \\
&= \sum_{j=0}^{k} \binom{m}{j} p^j (1-p)^{m-j} \binom{n}{k-j} p^{k-j} (1-p)^{n-(k-j)} \\
&= p^k (1-p)^{n+m-k} \sum_{j=0}^{k} \binom{m}{j} \binom{n}{k-j} \\
&= p^k (1-p)^{n+m-k} \binom{m+n}{k}.
\end{aligned}$$

Damit ist $X+Y \sim b(n+m,p)$. Die Identität

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j} \binom{n}{k-j}$$

kann man durch vollständige Induktion beweisen. Wenn $X_1, \cdots, X_n$ unabhängig und $b(1,p)$ verteilt sind, so folgt insbesondere

$$\sum_{1}^{n} X_j \sim b(n,p).$$

Beispiel 5.4.
Es seien $X \sim U(-1,1), Y \sim U(-1,1)$ und X und Y unabhängig verteilt. Nach dem Korollar zu Satz 5.2 ist $X+Y$ stetig mit Dichtefunktion

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(t-u) f_Y(u) du = \frac{1}{2} \int_{-1}^{1} f_X(t-u) du.$$

Das letzte Integral berechnen wir wie folgt.
Fall 1: $-2 \leq t \leq 0$.
Es gilt

$$\begin{aligned} f_{X+Y}(t) &= \frac{1}{4}\int_{-1}^{1}\{-1 \leq t-u \leq 1\}du \\ &= \frac{1}{4}\int_{-1}^{1}\{t-1 \leq u \leq t+1\}du \\ &= \frac{1}{4}\int_{-1}^{t+1} du, \end{aligned}$$

da $-1 \leq t+1 \leq 1$. Hieraus folgt

$$f_{X+Y}(t) = \frac{1}{4}(t+2)\{-2 \leq t \leq 0\}.$$

Fall 2: $0 \leq t \leq 2$.
Es gilt

$$\begin{aligned} f_{X+Y}(t) &= \frac{1}{4}\int_{-1}^{1}\{t-1 \leq u \leq t+1\}du \\ &= \frac{1}{4}\int_{t-1}^{1} du \end{aligned}$$

da $-1 \leq t-1 \leq 1$. Hieraus folgt

$$f_{X+Y}(t) = \frac{1}{4}(2-t)\{0 \leq t \leq 2\}.$$

5.2 Charakteristische Funktionen

Satz 5.5.
Es seien X und Y reellwertige, unabhängig verteilte Zufallsvariablen mit charakteristischen Funktionen φ_X und φ_Y. Dann hat $X+Y$ die charakteristische Funktion

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

Beweis: Die Zufallsvariablen $\exp(itx)$ und $\exp(ity)$ sind unabhängig verteilt. Hieraus folgt

$$\begin{aligned} \varphi_{X+Y}(t) &= \mathbb{E}(\exp(it(X+Y))) \\ &= \mathbb{E}(\exp(itX)\exp(itY)) \\ &= \mathbb{E}(\exp(itX)\mathbb{E}(\exp(itY)) \\ &= \varphi_X(t)\varphi_Y(t) \end{aligned}$$

□

Beispiel 5.6.
$X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2)$ und X und Y unabhängig verteilt.
Es gilt

$$\varphi_X(t) = \exp(it\mu_X - \frac{1}{2}t^2\sigma_X^2)$$

$$\varphi_Y(t) = \exp(it\mu_y - \frac{1}{2}t^2\sigma_Y^2)$$

und somit

$$\varphi_{X+Y}(t) = \exp(it(\mu_X + \mu_Y) - \frac{1}{2}t^2(\sigma_X^2 + \sigma_Y^2).$$

Hieraus folgt

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

Beispiel 5.7.
Es seien X und Y unabhängig verteilt mit $X \sim \mathrm{C}(0,1)$ und $Y \sim \mathrm{C}(0,1)$.
Es gilt

$$\varphi_X(t) = \exp(-|t|) = \varphi_Y(t)$$

und somit

$$\varphi_{X+Y}(t) = \exp(-2|t|).$$

Hieraus folgt

$$\varphi_{\frac{1}{2}(X+Y)}(t) = \exp(-|t|)$$

d.h.

$$\frac{1}{2}(X + Y) \sim \mathrm{C}(0,1).$$

5.3 Gesetze der großen Zahlen

Im Allgemeinen ist es schwierig, die Verteilung von $\frac{1}{n}\sum_1^n X_j$ zu bestimmen. Asymptotische Aussagen können aber gemacht werden.

Definition 5.8.
Es sei $(X_n)_1^\infty$ eine Folge reellwertiger Zufallsvariablen, die auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ definiert sind.

(a) Wir schreiben plim$X_n = X$, falls für alle $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

($(X_n)_1^\infty$ konvergiert der Wahrscheinlichkeit nach gegen X.)

(b) Wir schreiben $\lim X_n = X$ f.s., falls

$$\mathbb{P}(\{\omega : \lim X_n(\omega) = X(\omega)\}) = 1.$$

($(X_n)_1^\infty$ konvergiert fast sicher gegen X.)

Definition 5.9. *Es sei $(F_n)_1^\infty$ eine Folge von Ereignissen. Wir definieren*

(a)

$$\underline{\lim} F_n \quad = \quad \liminf F_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} F_m.$$

(b)

$$\overline{\lim} F_n \quad = \quad \limsup F_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} F_m.$$

(c) Falls $\underline{\lim} F_n = \overline{\lim} F_n$, so definieren wir

$$\lim F_n = \underline{\lim} F_n = \overline{\lim} F_n.$$

Bemerkung 5.10.

(a) Es gilt $\omega \in \underline{\lim} F_n$ genau dann, wenn $\omega \in F_n$ für alle bis auf höchstens endlich viele n-Werte.

(b) Es gilt $\omega \in \overline{\lim} F_n$ genau dann, wenn ω zu unendlichen vielen F_n gehört.

(c) Es gilt immer $\underline{\lim} F_n \subset \overline{\lim} F_n$.

Satz 5.11.
Es sei $(X_n)_1^\infty$ eine Folge reellwertiger Zufallsvariablen mit $\lim_{n\to\infty} X_n = X$ f.s. Dann gilt

$$plim X_n = X.$$

Beweis: Es sei $B_{n,j} = \{\omega : |X_n(\omega) - X(\omega)| > \frac{1}{j}\}$. Es gilt

$$\{\omega : \lim X_n(\omega) \neq X(\omega)\} = \cup_{j=1}^\infty (\cap_{n=1}^\infty (\cup_{m=n}^\infty B_{m,j})) = \cup_{j=1}^\infty (\overline{\lim} B_{n,j}). \quad (5.1)$$

Denn

(i) $\lim_{n\to\infty} X_n(\omega) \neq X(\omega)$ impliziert, es existiert ein $j = j(\omega)$ mit

$$\overline{\lim_{n\to\infty}} |X_n(\omega) - X(\omega)| > \frac{1}{j}.$$

Hieraus folgt, dass für alle n gilt

$$\sup_{m\geq n} |X_m(\omega) - X(\omega)| > \frac{1}{j}$$

und somit $\omega \in \cup_{m=1}^\infty (\cup_{m=n}^\infty B_{m,j})$ und insbesondere $\omega \in \cup_{j=1}^\infty (\overline{\lim} B_{n,j})$.

(ii) In der anderen Richtung gilt für $\omega \in \cup_{j=1}^\infty (\overline{\lim} B_{n,j})$, dass ein j existiert mit $\omega \in \cap_{n=1}^\infty (\cup_{m=n}^\infty B_{m,j})$. Dies impliziert $\omega \in \cup_{m=n}^\infty B_{m,j}$ für alle n, woraus folgt $\sup_{m\geq n} |X_m(\omega)| > \frac{1}{j}$. Somit konvergiert die Folge $(X_n(\omega))_1^\infty$ nicht gegen $X(\omega)$.

Wir haben also (5.1) gezeigt. Hieraus folgt

$$\mathbb{P}(\{\omega : \lim X_n(\omega) \neq X(\omega)\}) = 0$$

genau dann, wenn $\mathbb{P}(\cup_{j=1}^\infty (\overline{\lim} B_{n,j})) = 0$. Insbesondere gilt $\mathbb{P}(\overline{\lim} B_{n,j}) = 0$ für alle j. Da ein Wahrscheinlichkeitsmaß stetig ist, gilt

$$0 = \lim_{N\to\infty} \mathbb{P}(\cap_{n=1}^N (\cup_{m=n}^\infty B_{m,j})) = \mathbb{P}(\cup_{m=n}^\infty B_{m,j})$$

für alle j und somit $\lim_{N\to\infty} \mathbb{P}(B_{N,j}) = 0$ für alle j. Wir haben also gezeigt

$$\lim_{N\to\infty} \mathbb{P}(|X_N(\omega) - X(\omega)| > \frac{1}{j}) = 0$$

für alle j. Ist $\epsilon > 0$ gegeben, wählen wir $\frac{1}{j} < \epsilon$. Es folgt

$$\lim_{n\to\infty} \mathbb{P}(|X_n(\omega) - X(\omega)| > \epsilon) \leq \lim_{n\to\infty} \mathbb{P}(|X_n(\omega) - X_n| > \frac{1}{j}) = 0$$

und somit plim$X_n = X$. □

Dass die Umkehrung falsch ist, zeigt folgendes Beispiel.

Beispiel 5.12.
Wir setzen $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda_{[0,1]})$. Für $2^n + 1 \leq j \leq 2^{n+1}$ sei $X_j(\omega) = \left\{\frac{j-2^n-1}{2^n} \leq \omega \leq \frac{j-2^n}{2^n}\right\}$. Es gilt für $0 < \epsilon < 1$

$$\begin{aligned}
\mathbb{P}(|X_j(\omega) - 0| > \epsilon) &= \mathbb{P}(X_j(\omega) = 1) \\
&= \mathbb{P}\left(\left\{\omega : \frac{j-2^n-1}{2^n} \leq \omega \leq \frac{j-2^n}{2^n}\right\}\right) \\
&= \frac{1}{2^n}.
\end{aligned}$$

Somit folgt $\lim_{j\to\infty} \mathbb{P}(|X_j(\omega) - 0| > \epsilon) = 0$ für alle $\epsilon > 0$. In anderen Worten, es gilt plim$X_j = 0$. Für jedes ω und für jedes n gibt es aber ein j mit $\left\{\frac{j-2^n-1}{2^n} \leq \omega \leq \frac{j-2^n}{2^n}\right\}$. Somit gilt $\overline{\lim} X_n = 1$. Es gilt also plim$X_j = 0$, aber $\lim X_j \neq 0$ f.s.

Satz 5.13.
Es sei $(X_n)_1^\infty$ eine Folge unkorrelierter Zufallsvariablen mit endlichen Varianzen $(\mathbb{V}(X_n))_1^\infty$. Gilt

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{j=1}^{n} \mathbb{V}(X_j) = 0,$$

so folgt

$$plim\ \frac{1}{n} \sum_{j=1}^{n} (X_j - \mathbb{E}(X_j)) = 0.$$

Beweis: Nach der Ungleichung von Tchebychev gilt

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n}(X_j - \mathbb{E}X_j)\right| \geq \epsilon\right) \leq \frac{\mathbb{V}\left(\frac{1}{n}\sum_{j=1}^{n}(X_j - \mathbb{E}X_j)\right)}{\epsilon^2}.$$

Da aber die Zufallsvariablen unkorreliert sind, gilt

$$\mathbb{V}\left(\frac{1}{n}\sum_{j=1}^{n}(X_j - \mathbb{E}(X_j))\right) = \frac{1}{n^2}\mathbb{V}(\sum_{j=1}^{n} X_j) = \frac{1}{n^2}\sum_{j=1}^{n}\mathbb{V}(X_j).$$

Wir erhalten also

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{1}^{n}(X_j - \mathbb{E}(X_j))\right| \geq \epsilon\right) \leq \frac{1}{n^2\epsilon^2}\sum_{j=1}^{n}\mathbb{V}(X_j)$$

und somit

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{n}\sum_{1}^{n}(X_j - \mathbb{E}(X_j))\right| \geq \epsilon\right) = 0$$

für alle $\epsilon > 0$. □

Satz 5.14. *(Borel-Cantelli)*
Es sei $(F_n)_1^\infty$ eine Folge von Ereignissen. Dann:

(a) Gilt $\sum_1^\infty \mathbb{P}(F_n) < \infty$, so folgt $\mathbb{P}(\overline{\lim} F_n) = 0$.

(b) Sind die $(F_n)_1^\infty$ unabhängig und $\sum_1^\infty \mathbb{P}(F_n) = \infty$, so folgt $\mathbb{P}(\overline{\lim} F_n) = 1$.

Beweis:

(a) Es gilt $\overline{\lim} F_n = \bigcap_{n=1}^\infty (\bigcup_{m=n}^\infty F_m)$ und somit $\overline{\lim} F_n \subset \bigcup_{m=n}^\infty F_m$ für alle m. Hieraus folgt $\mathbb{P}(\overline{\lim} F_n) \leq \sum_{m=n}^\infty \mathbb{P}(F_m)$ für alle m und wegen $\sum_1^\infty \mathbb{P}(F_m) < \infty$ gilt $\lim_{n\to\infty}\sum_{m=n}^\infty \mathbb{P}(F_m) = 0$ und somit die Behauptung.

(b) Es gilt
$\Omega\backslash\overline{\lim} F_n = \Omega\backslash\left(\bigcap_{n=1}^\infty \bigcup_{m=n}^\infty F_m\right) = \bigcup_{n=1}^\infty \left(\Omega\backslash \bigcup_{m=n}^\infty F_m\right)$.
Wir schreiben $G_n = \Omega\backslash\bigcup_{m=n}^\infty F_m$. Es gilt also $\Omega\backslash\overline{\lim} F_n = \bigcup_{n=1}^\infty G_n$.
Da aber die Folge $(G_n)_1^\infty$ isoton ist ($G_n \subset G_{n+1}$), erhalten wir $\bigcup_{n=1}^\infty G_n = \lim_{n\to\infty} G_n$ und somit $\mathbb{P}(\Omega\backslash\overline{\lim} F_n) = \mathbb{P}(\lim_{n\to\infty} G_n) = \lim_{n\to\infty}\mathbb{P}(G_n)$.
Für alle $k \geq n$ gilt $G_n \subset \Omega\backslash\bigcup_{m=n}^k F_m = \bigcap_{m=n}^k(\Omega\backslash F_m)$
und somit $\mathbb{P}(G_n) \leq \mathbb{P}(\bigcap_{m=n}^k(\Omega\backslash F_m))$.
Da aber die $(F_n)_1^\infty$ unabhängig sind, erhalten wir

$$\mathbb{P}(\bigcap_{m=n}^{k}(\Omega\backslash F_n)) = \Pi_{m=n}^k(1 - \mathbb{P}(F_m)) \leq \exp(-\sum_{m=n}^{k}\mathbb{P}(F_m)).$$

Dies impliziert

$$\mathbb{P}(G_n) \leq \exp(-\sum_{m=n}^{k} \mathbb{P}(F_m))$$

für alle $k \geq n$ und wegen $\sum_{n=1}^{\infty} \mathbb{P}(F_n) = \infty$ folgt $\mathbb{P}(G_n) = 0$.
Hieraus folgt $\mathbb{P}(\Omega \backslash \overline{\lim} F_n) = 0$ und somit $\mathbb{P}(\overline{\lim} F_n) = 1$.

Wenn man die Bedingung an die Varianzen etwas verstärkt, kann man sogar fast sichere Konvergenz beweisen.

Satz 5.15.
Es seien $(X_j)_1^{\infty}$ eine Folge unkorrelierter Zufallsvariablen mit

$$\sup_j \mathbb{V}(X_j) < \infty.$$

Dann gilt

$$\frac{1}{n} \sum_{j=1}^{n} (X_j - \mathbb{E}(X_j)) \to 0 \quad \textit{f.s.}$$

Beweis: Wir betrachten zuerst die Folge

$$\frac{1}{n^2} \sum_{j=1}^{n^2} (X_j - \mathbb{E}(X_j)) \quad n = 1, 2, \ldots$$

Es gilt

$$\mathbb{P}\left(\left| \frac{1}{n^2} \sum_{j=1}^{n^2} (X_j - \mathbb{E}(X_j)) \right| \geq \epsilon \right) \leq \frac{1}{n^4 \epsilon^2} \sum_{j=1}^{n^2} \mathbb{V}(X_j).$$

Nach Voraussetzung gibt es ein K mit $\mathbb{V}(X_j) \leq K$ für alle j. Hieraus folgt

$$\mathbb{P}\left(\left| \frac{1}{n^2} \sum_{j=1}^{n^2} (X_j - \mathbb{E}(X_j)) \right| \geq \epsilon \right) \leq \frac{K n^2}{n^4 \epsilon^2} = \frac{K}{n^2 \epsilon^2}.$$

Es sei

$$B_{n^2}(\epsilon) = \left\{ \omega : \left| \frac{1}{n^2} \sum_{j=1}^{n^2} (X_j - \mathbb{E}(X_j)) \right| \geq \epsilon \right\}.$$

Es gilt $\mathbb{P}(B_{n^2}(\epsilon)) < \frac{K}{n^2 \epsilon^2}$ und somit $\sum_{n=1}^{\infty} \mathbb{P}(B_{n^2}(\epsilon)) < \infty$. Aus Borel-Cantelli folgt

$$\mathbb{P}(\overline{\lim} B_{n^2}(\epsilon)) = 0,$$

d.h.

$$\overline{\lim}\left|\frac{1}{n^2}\sum_{j=1}^{n^2}(X_j-\mathbb{E}(X_j))\right|\leq\epsilon \text{ f.s.}$$

Da aber dies für jedes $\epsilon>0$ gilt,folgt

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{j=1}^{n^2}(X_j-\mathbb{E}(X_j))=0 \quad \text{f. s.}$$

Wir müssen nun die Lücken zwischen den quadratischen Zahlen ausfüllen. Um die Schreibweise zu erleichtern, setzen wir $\mathbb{E}(X_j)=0$ für alle j. Dies kann erreicht werden, indem wir die Zufallsvariablen $X_j'=X_j-\mathbb{E}(X_j)$ betrachten. Es sei nun $n^2\leq m<(n+1)^2$. Wir vergleichen $\sum_1^m X_j'$ mit $\sum_1^{n^2} X_j'$. Es gilt

$$\mathbb{V}\left(\sum_{j=1}^{m}X_j'-\sum_{j=1}^{n^2}X_j'\right)=\mathbb{V}\left(\sum_{j=n^2+1}^{m}X_j'\right)=\sum_{j=n^2+1}^{m}\mathbb{V}(X_j')\leq(m-n^2)K.$$

Mit Tschebyschev erhalten wir

$$\mathbb{P}\left(\left|\sum_{j=1}^{m}X_j'-\sum_{j=1}^{n^2}X_j'\right|\geq\epsilon n^2\right)\leq\frac{(m-n^2)K}{n^4\epsilon^2}<\frac{(2n+1)K}{n^4\epsilon^2}\leq\frac{3nK}{n^4\epsilon^2}\leq\frac{3K}{n^3\epsilon^2}.$$

Es gilt

$$\begin{aligned}
&\mathbb{P}\left(\max_{n^2\leq m<(n+1)^2}\left|\sum_{j=1}^{m}X_j'-\sum_{j=1}^{n^2}X_j'\right|\geq\epsilon n^2\right)\\
=\;&\mathbb{P}\left(\bigcup_{m=n^2}^{(n+1)^2-1}\left\{\left|\sum_{j=1}^{m}X_j'-\sum_{j=1}^{n^2}X_j'\right|\leq\epsilon n^2\right\}\right)\\
\leq\;&\sum_{m=n^2}^{(n+1)^2-1}\mathbb{P}\left(\left|\sum_{j=1}^{m}X_j'-\sum_{j=1}^{n^2}X_j'\right|\geq\epsilon^2 n^2\right)\\
\leq\;&\sum_{m=n^2}^{(n+1)^2-1}\frac{3K}{n^3\epsilon^2}\leq\frac{3n\cdot 3K}{n^3\epsilon^2}\leq\frac{9K}{n^2\epsilon^2}.
\end{aligned}$$

Es sei nun

$$C_{n^2}=\{\omega:\max_{n^2\leq m<(n+1)^2}\left|\sum_{j=1}^{m}X_j'-\sum_{j=1}^{n^2}X_j'\right|\geq\epsilon n^2\}.$$

Es gilt

$$\mathbb{P}(C_{n^2}) \leq \frac{9K}{n^2\epsilon^2}$$

und somit $\sum_{n=1}^{\infty} \mathbb{P}(C_{n^2}) < \infty$. Mit Borel-Cantelli folgern wir, dass es für fast alle ω ein $n_\circ(\omega)$ derart gibt, dass für alle $n \geq n_\circ$ und m mit $n^2 \leq m < (n+1)^2$

$$\left| \sum_{j=1}^{m} X'_j - \sum_{j=1}^{n^2} X'_j \right| < \epsilon n^2.$$

Hieraus folgt

$$\left| \frac{1}{m} \sum_{j=1}^{m} X'_j \right| \leq \left| \frac{1}{m} \sum_{j=1}^{n^2} X'_j \right| + \frac{\epsilon n^2}{m} \leq \frac{1}{n^2} \left| \sum_{j=1}^{n^2} X'_j \right| + \epsilon$$

für alle $n \geq n_\circ(\omega)$, und m mit $n^2 \leq m < (n+1)^2$. Da aber $\frac{1}{n^2} \sum_{j=1}^{n^2} X'_j \to 0$ f.s. erhalten wir

$$\overline{\lim} \left| \frac{1}{m} \sum_{j=1}^{m} X'_j \right| \leq \epsilon \text{ f.s.}$$

Da dies aber für jedes $\epsilon > 0$ gilt, folgt

$$\lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^{m} X'_j = 0 \quad \text{f.s.}$$

□

Wenn man sich etwas mehr anstrengt, kann man Folgendes beweisen.

Satz 5.16.
Es sei $(X_j)\infty$ eine Folge unabhängig und identisch verteilter Zufallsvariablen mit $\mathbb{E}(|X_1|) < \infty$. Dann gilt

$$\lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} X_j = \mathbb{E}(X_1) \text{ f. s.}$$

Beweis: Dudley, S. 206-207. Der Beweis verläuft wie der Beweis von Satz 5.15, ist etwas langwieriger, aber nicht wesentlich komplizierter. □

Beispiel 5.17.
Es seien $(X_j)_1^\infty$ unabhängig und identisch verteilt mit einer gemeinsamen $\mathrm{C}(0,1)$-Verteilung. Es gilt

$$\frac{1}{n}\sum_{j=1}^{n} X_j \sim \mathrm{C}(0,1) \tag{5.2}$$

Es handelt sich um eine Verallgemeinerung von Beispiel 5.7.
Das Ereignis $\{\omega : \lim_{n\to\infty} \frac{1}{n}\sum_1^n X_j(\omega) \text{ existiert}\}$ ist ein Terminalereignis, da es ja nicht von endlich vielen X_n abhängt. Nach dem $0-1-$Gesetz von Kolmogoroff gilt also

$$\mathbb{P}\left(\lim \frac{1}{n}\sum_{j=1}^{n} X_j \text{ existiert } \right) = 0,1.$$

Wenn der Limes existiert, so folgt

$$\mathbb{P}\left(\lim_{n\to\infty}\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \frac{1}{2n}\sum_{j=1}^{2n} X_j\right| = 0\right) = 1.$$

Es gilt aber (wieder Beispiel 5.7)

$$\frac{1}{2n}\sum_{j=1}^{n} X_j - \frac{1}{2n}\sum_{j=n+1}^{2n} X_j \sim \mathrm{C}(0,1)$$

und somit

$$\mathbb{P}\left(\left|\frac{1}{2n}\sum_{j=1}^{n} X_j - \frac{1}{2n}\sum_{j=n+1}^{2n} X_j\right| \geq 1\right) = 1 - \frac{2}{\pi}\arg\tan(1) = \frac{1}{2}$$

für alle n. Dies widerspricht (5.2). Es gilt also

$$\mathbb{P}\left(\lim \frac{1}{n}\sum_{j=1}^{n} X_j \text{ existiert } \right) = 0.$$

Dies zeigt, dass die Annahme $\mathbb{E}(|X_1|) < \infty$ in Satz 5.16 nicht ersatzlos gestrichen werden kann.

6 Schwache Konvergenz

6.1 Einführung

Definition 6.1.
Es seien $(X_n)_1^\infty$ und X Zufallsvariablen. Wir schreiben $X_n \Rightarrow X$ (X_n konvergiert schwach gegen X), wenn gilt

$$\lim_{n\to\infty} \mathbb{E}(u(X_n)) = \mathbb{E}(u(X))$$

für alle $u : \mathbb{R} \to \mathbb{R}$ stetig und beschränkt.

Bemerkung 6.2.
Sind $(F_n)_1^\infty$ und F die entsprechenden Verteilungsfunktionen, so ist $X_n \Rightarrow X$ damit äquivalent, dass

$$\lim_{n\to\infty} \int u(x) dF_n(x) = \int u(x) dF(x).$$

Wir schreiben $F_n \Rightarrow F$. Es ist nicht notwendig, dass die Zufallsvariablen $(X_n)_1^\infty$ und X auf demselben Wahrscheinlichkeitsraum definiert sind. Entscheidend ist nur das Verhalten der Verteilungen.

Bemerkung 6.3.
Gilt $X_n \Rightarrow X$ und $X_n \Rightarrow Y$, so folgt $\mathbb{E}(u(X)) = \mathbb{E}(u(Y))$ für alle $u : \mathbb{R} \to \mathbb{R}$ stetig und beschränkt. Aus Satz 4.27 folgt nun $\mathbb{P}^X = \mathbb{P}^Y$.

Satz 6.4.
Es seien $(X_n)_1^\infty$ und X Zufallsvariablen. Dann gilt $X_n \Rightarrow X$ genau dann, wenn

$$\lim_{n\to\infty} \mathbb{E}(u(X_n)) = \mathbb{E}(u(X))$$

für alle $u : \mathbb{R} \to \mathbb{R}$ $\infty-$oft differenzierbar mit kompaktem Träger.

Beweis: In der einen Richtung ist die Behauptung trivial. Es gelte also

$$\lim_{n\to\infty} \mathbb{E}(u(X_n)) = \mathbb{E}(u(X))$$

für alle $u : \mathbb{R} \to \mathbb{R}$ $\infty-$unendlich oft differenzierbar mit kompaktem Träger. Indem wir $(u-\inf u)/(\sup u - \inf u)$ betrachten, können wir o.B.d.A.annehmen, dass $0 \le u \le 1$ gilt. Wir zeigen zuerst, dass gilt

$$\lim_{K\to\infty} \overline{\lim}_{n\to\infty} \mathbb{P}(|X_n| > K) = 0. \tag{6.1}$$

Es sei $K > 0$ gegeben und $v : \mathbb{R} \to [0,1]$ stetig mit $v(x) = 1$, falls $|x| < \frac{1}{2}K$ und $v(x) = 0$ falls $|x| \geq K$. Es gilt

$$\mathbb{P}(|X_n| > K) = 1 - \mathbb{P}(|X_n| \leq K) \leq 1 - \mathbb{E}(v(X_n))$$

und somit

$$\overline{\lim}_{n\to\infty}\mathbb{P}(|X_n| > K) \leq 1 - \overline{\lim}_{n\to\infty}\mathbb{E}(v(X_n)) \leq 1 - \overline{\lim}_{n\to\infty}\mathbb{P}(|X_n| \leq \frac{1}{2}K)$$

woraus (6.1) folgt. Es seien nun $v : \mathbb{R} \to [0,1]$ und $K > 0$ und $\epsilon > 0$ gegeben. Es sei $u : \mathbb{R} \to [0,1]$ $\infty-$oft differenzierbar mit kompaktem Träger und $\sup_{|x|\leq K} |v(x)-u(x)| < \epsilon$ (zur Existenz von u siehe nachfolgende Bemerkung). Es gilt

$$\begin{aligned}
\mathbb{E}(v(X_n)) &\leq \mathbb{E}(v(X_n)\{|X_n| \leq K\}) + \mathbb{P}(|X_n| > K) \\
&\leq \mathbb{E}(u(X_n)\{|X_n| \leq K\}) + \epsilon + \mathbb{P}(|X_n| > K) \\
&\leq \mathbb{E}(u(X_n)) + \epsilon + \mathbb{P}(|X_n| > K).
\end{aligned}$$

Hieraus folgt

$$\begin{aligned}
\overline{\lim}_{n\to\infty}\mathbb{E}(v(X_n)) &\leq \mathbb{E}(u(X)) + \epsilon + \overline{\lim}_{n\to\infty}\mathbb{P}(|X_n| > K) \\
&\leq \mathbb{E}(u(X)\{|X| \leq K\} + \mathbb{P}(|X| > K) + \epsilon + \overline{\lim}_{n\to\infty}\mathbb{P}(|X_n| > K) \\
&\leq \mathbb{E}(v(X)\{|X| \leq K\}) + \epsilon + \mathbb{P}(|X| > K) + \epsilon + \overline{\lim}_{n\to\infty}\mathbb{P}(|X_n| > K) \\
&\leq \mathbb{E}(v(X)) + 2\epsilon + \mathbb{P}(|X| > K) + \overline{\lim}_{n\to\infty}\mathbb{P}(|X_n| > K).
\end{aligned}$$

Indem wir mit $K \to \infty$ gehen, folgt aus (6.1)

$$\overline{\lim}_{n\to\infty}\mathbb{E}(v(X_n)) \leq \mathbb{E}(v(X)) + 2\epsilon$$

und somit

$$\overline{\lim}_{n\to\infty}\mathbb{E}(v(X_n)) \leq \mathbb{E}(v(X)).$$

Indem man nun v durch $1 - v$ ersetzt, folgt

$$\underline{\lim}\,\mathbb{E}_{n\to\infty}(v(X_n)) \geq \mathbb{E}(v(X))$$

und somit

$$\lim_{n\to\infty}\mathbb{E}(v(X_n)) = \mathbb{E}(v(X)).$$

□

Bemerkung 6.5.
Wir zeigen nun, dass es eine Funktion u mit den behaupteten Eigenschaften gibt. Es sei zuerst $K > 0$. Wir definieren $v_K : \mathbb{R} \to [0,1]$ durch

$$v_K(x) = \begin{cases} v(x) & |x| < K \\ 0 & |x| > 2K \\ \text{linear} & \text{dazwischen} \end{cases}$$

Dann hat v_K einen kompakten Träger, ist stetig, und es gilt $0 \leq v_K \leq 1$. Es sei nun $f : \mathbb{R} \to \mathbb{R}_+$ mit Träger $[-1,1]$, unendlich oft differenzierbar und $\int f(x)dx = 1$, z.B.

$$f(x) = \begin{cases} c \exp\left(-\frac{1}{(1-x^2)}\right) & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}$$

Wir setzen

$$u_\sigma(x) = \frac{1}{\sigma} \int f\left(\frac{x-y}{\sigma}\right) v_K(y) dy.$$

Dann gilt

(i) $0 \leq u_\sigma \leq 1$

(ii) u_σ hat einen kompakten Träger

(iii) u_σ ist unendlich oft differenzierbar.

Ferner gilt

$$\begin{aligned} u_\sigma(x) - v_K(x) &= \frac{1}{\sigma} \int f\left(\frac{x-y}{\sigma}\right) (v_K(y) - v_K(x)) dy \\ &= \int f(z)(v_K(x) - v_K(x - \sigma z)) dz. \end{aligned}$$

Hieraus folgt

$$\sup_x |u_\sigma(x) - v_K(x)| \leq \sup_{|z| \leq \sigma} |v_K(x) - v_K(x-z)|.$$

Da aber v_K stetig mit kompaktem Träger ist, ist v_K gleichmäßig stetig. Hieraus folgt

$$\sup_x \sup_{|z| \leq \sigma} |v_K(x) - v_K(x-z)| < \epsilon$$

falls $\sigma \leq \sigma(\epsilon)$. Hieraus folgt

$$\sup_{|x| \leq K} |u_{\sigma(\epsilon)}(x) - v(x)| = \sup_{|x| \leq K} |u_{\sigma(\epsilon)}(x) - v_K(x)| < \epsilon.$$

Beispiel 6.6.
Es sei $X_n \sim b(n, p_n)$, und es gelte $\lim_{n\to\infty} np_n = \lambda, 0 < \lambda < \infty$. Wir zeigen, dass $X_n \Rightarrow X \sim P_0(\lambda)$. Nach Satz 6.4 genügt es zu zeigen, dass gilt

$$\lim_{n\to\infty} \mathbb{E}(u(X_n)) = \mathbb{E}(u(X))$$

für alle $u : \mathbb{R} \to \mathbb{R}$ stetig mit kompaktem Träger. Es sei N so, dass $u(x) = 0$ für $|x| > N$ gilt. Es folgt

$$\mathbb{E}(u(X_n)) = \sum_{j=0}^{\min(N,n)} u(j)\binom{n}{j} p_n^j (1-p_n)^{n-j}$$

und somit

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}(u(X_n)) &= \sum_{j=0}^{N} u(j) \lim_{n\to\infty} \binom{n}{j} p_n^j (1-p_n)^{n-j} \\
&= \sum_{j=0}^{N} u(j) \frac{\lambda^j \exp(-\lambda)}{j!} \\
&= \sum_{0}^{\infty} u(j) \frac{\lambda^j \exp(-\lambda)}{j!} \\
&= \mathbb{E}(u(X)).
\end{aligned}$$

Satz 6.7.
Es seien $(X_n)_1^\infty$ und X reelle Zufallsvariablen mit Verteilungsfunktionen $(F_n)_1^\infty$ und F. Dann gilt

$$X_n \Rightarrow X \text{ genau dann, wenn } \lim_{n\to\infty} F_n(x) = F(x)$$

für alle Stetigkeitspunkte x von F.

Beweis: Es gelte $X_n \Rightarrow X$. Es sei $u_{x,\epsilon}$ eine stetige Funktion mit

(i) $0 \le u_{x,\epsilon} \le 1$,

(ii) $u_{x,\epsilon}(y) = 1$, falls $y \le x$,

(iii) $u_{x,\epsilon}(y) = 0$, falls $y > x + \epsilon$.

Es gilt

$$\begin{aligned}
\overline{\lim}_{n\to\infty} F_n(x) &= \overline{\lim}_{n\to\infty} \mathbb{E}(\{X_n \leq x\}) \\
&\leq \lim_{n\to\infty} \mathbb{E}(u_{x,\epsilon}(X_n)) \\
&= \mathbb{E}(u_{x,\epsilon}(X)) \\
&\leq \mathbb{E}(\{X \leq x+\epsilon\}) \\
&= F(x+\epsilon).
\end{aligned}$$

Hieraus folgt mit $\epsilon \downarrow 0$, dass gilt $\overline{\lim}_{n\to\infty} F_n(x) \leq F(x)$ für alle x. Ferner gilt

$$\begin{aligned}
\underline{\lim}_{n\to\infty} F_n(x+\epsilon) &\geq \underline{\lim}_{n\to\infty} \mathbb{E}(u_{x,\epsilon}(X_n)) \\
&= \mathbb{E}(u_{x,\epsilon}(X)) \\
&\geq \mathbb{E}(\{X \leq x\}) \\
&= F(x)
\end{aligned}$$

für alle x, ϵ. Ersetzen wir x durch $x-\epsilon$, so folgt

$$F(x-\epsilon) \leq \underline{\lim}_{n\to\infty} F_n(x)$$

und, mit $\epsilon \downarrow 0$,

$$F(x-0) \leq \underline{\lim}_{n\to\infty} F_n(x) \leq \overline{\lim}_{n\to\infty} F_n(x) \leq F(x).$$

Ist x ein Stetigkeitspunkt von F , so folgt $F(x-0) = F(x)$ und somit

$$\lim_{n\to\infty} F_n(x) = F(x).$$

Es gelte nun $F_n(x) \to F(x)$ für alle x Stetigkeitspunkte von F. Es genügt zu zeigen, dass gilt

$$\lim_{n\to\infty} \int u dF_n = \int u dF$$

für alle $u : \mathbb{R} \to \mathbb{R}$ unendlich oft differenzierbar mit kompaktem Träger. Für ein solches u gilt (partielle Integration)

$$\int u dF_n = -\int u^{(1)}(x) F_n(x) dx.$$

Da $u^{(1)} = 0$ außerhalb einer kompakten Menge und $\lim_{n\to\infty} F_n(x) = F(x)$ für Lebesgue-fast alle x, folgt mit dominierter Konvergenz

$$\begin{aligned}\lim_{n\to\infty}\int u dF_n &= -\lim_{n\to\infty}\int u^{(1)}(x)F_n(x)dx \\ &= -\int u^{(1)}(x)\lim_{n\to\infty} F_n(x)dx \\ &= -\int u^{(1)}(x)F(x)dx \\ &= \int u dF.\end{aligned}$$

□

Beispiel 6.8.
Es seien X_n und X wie folgt definiert:

$$\mathbb{P}\left(X_n = \frac{1}{n}\right) = \frac{1}{2} = \mathbb{P}(X_n = 1) \text{ und } \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

Es gilt

$$\mathbb{E}(u(X_n)) = \frac{1}{2}u\left(\frac{1}{n}\right) + \frac{1}{2}u(1) \to \frac{1}{2}(u(0) + u(1)) = \mathbb{E}(u(X)).$$

Somit gilt $X_n \Rightarrow X$. Es gilt aber $F_n(0) = \mathbb{P}(X_n \leq 0) = 0$ und $F(0) = \mathbb{P}(X \leq 0) = \frac{1}{2}$.

6.2 Straffheit

In der Definition der schwachen Konvergenz wird vorausgesetzt, dass es eine Grenzverteilung gibt. Wir wollen nun Bedingungen herleiten, die dies garantieren.

Beispiel 6.9.
Die Zufallsvariablen X_n seien wie folgt definiert: $\mathbb{P}(X_n = 0) = \frac{1}{2}$, $\mathbb{P}(X_n = n) = \frac{1}{2}$. Es gilt $F_n(x) = \frac{1}{2}\{0 \leq x < n\} + \{x \geq n\}$ und somit $\lim_{n\to\infty} F_n(x) = \frac{1}{2}\{0 \leq x < \infty\} = G(x)$. Die Funktion $G(x)$ ist isoton und rechtsseitig stetig, aber keine Verteilungsfunktion. Es gilt aber für alle u stetig mit kompaktem Träger

$$\lim_{n\to\infty}\int u dF_n = \frac{1}{2}u(0) = \int u dG.$$

Dies gilt aber nicht, falls man nur verlangt, dass u stetig und beschränkt ist. Insbesondere gilt für $u(x) \equiv 1$

$$\int u dF_n = 1 \neq \frac{1}{2} = \int u dG.$$

Wir brauchen Bedingungen, die garantieren, dass, falls $\lim_{n\to\infty} F_n(x) = F(x)$ für alle Stetigkeitspunkte von x von F gilt, F eine Verteilungsfunktion ist.

Definition 6.10.
Eine Folge von Verteilungsfunktionen $(F_n)_1^\infty$ heißt straff, falls es zu jedem $\epsilon > 0$ ein $K > 0$ derart gibt, dass gilt

$$F_n(K) - F_n(-K) \geq 1 - \epsilon$$

für alle n.

Satz 6.11. *(Helly)*
Es sei $(F_n)_1^\infty$ eine straffe Folge von Verteilungsfunktionen. Dann gibt es eine Teilfolge $(n_j)_1^\infty$ und eine Verteilungsfunktion F mit

$$F_{n_j} \Rightarrow F.$$

Beweis: Wir benutzen das Diagonalverfahren von Cantor. Es seien $(r_j)_1^\infty$ die rationalen Zahlen. Die Folge $(F_n(r_1))_{n=1}^\infty$ ist beschränkt, und somit existiert eine Teilfolge, die wir mit $(F_{1,n})_{n=1}^\infty$ bezeichen, so dass gilt

$$\lim_{n\to\infty} F_{1,n}(r_1) = s_1.$$

Die Folge $(F_{1,n}(r_2))_{n=1}^\infty$ ist beschränkt, und somit existiert hiervon eine Tielfolge $(F_{2,n})_{n=1}^\infty$ mit

$$\lim_{n\to\infty} F_{2,n}(r_2) = s_2.$$

Auf diese Art und Weise erhalten wir Teilfolgen $(F_{k,n})_{n=1}^\infty$ für $k = 1, \ldots$ mit folgenden Eigenschaften:

(i) $(F_{k+1,n})_{n=1}^\infty$ ist eine Teilfolge von $(F_{k,n})_{n=1}^\infty$

(ii) $\lim_{n\to\infty} F_{k,n}(r_k) = s_k$.

Betrachten wir nun die Diagonalfolge $(F_{j,j})_{j=1}^\infty$.

$$\begin{array}{cccc} F_{1,1} & F_{1,2} & F_{1,3} & \ldots \\ F_{2,1} & F_{2,2} & F_{2,3} & \ldots \\ F_{3,1} & F_{3,2} & F_{3,3} & \ldots \end{array}$$

$(F_{j,j})_{j=k}^{\infty}$ ist eine Teilfolge von $(F_{k,n})_{n=1}^{\infty}$, und somit existiert

$$\lim_{j\to\infty} F_{j,j}(r_k) = s_k, \quad k = 1, \ldots$$

Da jede Verteilungsfunktion isoton ist, folgt aus $r_i < r_j$, dass gilt $s_i \leq s_j$. Wir definieren nun eine Funktion $F : \mathbb{R} \to [0, 1]$ durch

$$F(x) = \inf_{r_i > x} s_i.$$

Die Funktion F ist eine Verteilungsfunktion. Denn

(i) F ist isoton. Ist $x < x'$, so gilt

$$F(x) = \inf_{r_i > x} s_i \leq \inf_{r_i > x'} s_i = F(x').$$

(ii) F ist rechtsseitig stetig. Denn es gibt zu jedem x und zu jedem $\epsilon > 0$ ein $r_i > x$ mit $s_i \leq F(x) + \epsilon$. Für alle y mit $x < y < r_i$ gilt $F(x) \leq F(y) \leq F(x) + \epsilon$ und somit

$$F(x) \leq \lim_{y\downarrow x} F(y) \leq F(x) + \epsilon.$$

Da ϵ beliebig ist, folgt $\lim_{y\downarrow x} F(y) = F(x)$.

(iii) $\lim_{x\to-\infty} F(x) = 0$. Hierzu benutzen wir die Annahme, die Folge $(F_n)_1^{\infty}$ sei straff. Zu jedem $\epsilon > 0$ gibt es ein $K = K(\epsilon)$ derart, dass gilt $F_n(K) - F_n(-K) \geq 1 - \epsilon$ für alle n. Insbesonders gilt

$$F_n(-K) \leq F_n(K) - 1 + \epsilon \leq \epsilon.$$

Ist nun $x < r_i < -K$, so folgt

$$F(x) \leq \lim_{k\to\infty} F_{k,k}(r_i) \leq \epsilon$$

und somit $\overline{\lim}_{x\to-\infty} F(x) \leq \epsilon$. Da aber ϵ beliebig ist, folgt die Behauptung.

(iv) $\lim_{x\to\infty} F(x) = 1$. Folgt wie bei (iii).

Wir zeigen nun, dass gilt $F_{j,j} \Rightarrow F$. Nach Satz 6.7 genügt es zu zeigen, dass gilt

$$\lim_{j\to\infty} F_{j,j}(x) = F(x)$$

in allen Stetigkeitspunkten x von F. Es sei $y > x$. Es gilt für $x < r_i < y$

$$\overline{\lim}_{j\to\infty} F_{j,j}(x) \leq \lim_{j\to\infty} F_{j,j}(r_i) \leq F(y),$$

woraus mit $y \downarrow x$ und der Stetigkeit von F in x folgt

$$\overline{\lim}_{j\to\infty} F_{j,j}(x) \leq F(x). \tag{6.2}$$

In der anderen Richtung mit $y < r_i < x$ gilt

$$F(y) \leq \underline{\lim}_{j\to\infty} F_{j,j}(r_i) \leq \underline{\lim}_{j\to\infty} F_{j,j}(x)$$

woraus mit $y \uparrow x$ und der Stetigkeit von F in x folgt

$$F(x) \leq \underline{\lim}_{j\to\infty} F_{j,j}(x).$$

Dies zusammen mit (6.2) ergibt die Behauptung. □.

Satz 6.12.
Es sei $(F_n)_1^\infty$ eine Folge von Verteilungsfunktionen mit zugehörigen charakteristischen Funktionen $(\phi_n)_1^\infty$. Gilt

$$\lim_{n\to\infty} \phi_n(t) = \phi(t),$$

wobei $\phi : \mathbb{C} \to \mathbb{R}$ stetig in 0 ist, dann ist ϕ eine charakteristische Funktion mit Verteilungsfunktion F, und es gilt

$$F_n \Rightarrow F.$$

Beweis: Das Hauptproblem besteht darin, zu zeigen, dass die Folge $(F_n)_1^\infty$ straff ist. Es gilt

$$\begin{aligned}
\frac{1}{t}\int_0^t (1 - \mathrm{Re}(\phi(u)))du &= \frac{1}{t}\int_0^t (1 - \mathrm{Re}(\lim_{n\to\infty} \phi_n(u)))du \\
&= \lim_{n\to\infty} \frac{1}{t}\int_0^t (1 - \mathrm{Re}(\phi_n(u)))du \\
&= \lim_{n\to\infty} \frac{1}{t}\int_0^t \left(1 - \int_{-\infty}^{\infty} \cos(ux) dF_n(x)\right) du \\
&= \lim_{n\to\infty} \int_{-\infty}^{\infty} \left(\frac{1}{t}\int_0^t (1 - \cos(ux))du\right) dF_n(x)
\end{aligned}$$

nach den Sätzen von Lebesgue und Fubini. Aus $0 \leq 1 - \cos(ux)$ folgt

$$\begin{aligned}
\frac{1}{t}\int_0^t (1 - \mathrm{Re}(\phi(u)))du &\geq \overline{\lim}_{n\to\infty} \int_{|x|\geq 2/t} \left(\frac{1}{t}\int_0^t (1-\cos(ux))du\right) dF_n(x) \\
&= \overline{\lim}_{n\to\infty} \int_{|x|\geq 2/t} \left(1 - \frac{\sin(tx)}{tx}\right) dF_n(x) \\
&\geq \overline{\lim}_{n\to\infty} \int_{|x|\geq 2/t} \frac{1}{2} dF_n(x) \\
&= \frac{1}{2}(1 - \underline{\lim}_{n\to\infty} \int_{|x|<2/t} dF_n(x) \\
&\geq \frac{1}{2}(1 - \underline{\lim}_{n\to\infty} \left(F_n(4/t) - F_n(-4/t)\right).
\end{aligned}$$

Da ϕ im Nullpunkt stetig ist, gibt es zu jedem $\epsilon > 0$ ein t_0 so, dass $0 \leq 1 - \mathrm{Re}(\cos(u)) \leq \epsilon$ für alle $u, 0 \leq u \leq t_0$. Hieraus folgt

$$\lim_{n\to\infty} (F_n(4/t_0) - F_n(-4/t_0)) \geq 1 - 2\epsilon.$$

Somit gibt es n_0 so, dass

$$F_n(4/t_0) - F_n(-4/t_0) \geq 1 - 3\epsilon$$

für alle $n \geq n_0$. Es gibt ferner ein t_1 so, dass

$$F_n(4/t_1) - F_n(-4/t_1) \geq 1 - 3\epsilon$$

für alle $n, 1 \leq n \leq n_0$. Hieraus folgt

$$F_n(4/t_2) - F_n(-4/t_2) \geq 1 - 3\epsilon.$$

für alle n mit $t_2 = \min(t_0, t_1)$. Da ϵ beliebig klein gewählt werden kann, ist die Folge $(F_n)_1^\infty$ straff. Nach dem Satz von Helly (Satz 6.11) gibt es eine schwach konvergente Teilfolge $(F_{n_j})_1^\infty$ und eine Verteilungsfunktion F^* mit $F_{n_j} \Rightarrow F^*$. Da die Funktion $\exp(i \cdot t) : \mathbb{R} \to \mathbb{C}, x \to \exp(ixt)$ beschränkt und stetig ist, folgt

$$\phi^*(t) = \int_{-\infty}^{\infty} \exp(ixt) dF^*(x) = \lim_{n\to\infty} \int_{-\infty}^{\infty} \exp(ixt) dF_{n_j}(x) = \phi(x).$$

Aus $\phi^* = \phi$ folgt mit dem Korollar 4.4, dass gilt $F^* = F$ und somit $F_{n_j} \Rightarrow F$. Es bleibt zu zeigen, dass die Folge $(F_n)_1^\infty$ selbst schwach gegen F konvergiert.

Ist dies nicht der Fall, dann gibt es eine stetige und beschränkte Funktion $u : \mathbb{R} \to \mathbb{R}$ mit

$$\overline{\lim}_{n\to\infty} \left| \int u(x)dF_n(x) - \int u(x)dF(x) \right| > 0,$$

und somit gibt es eine Teilfolge F_{n_j} mit

$$\underline{\lim}_{n\to\infty} \left| \int u(x)dF_n(x) - \int u(x)dF(x) \right| > 0. \tag{6.3}$$

Diese Folge ist straff und hat somit eine schwach konvergente Teilfolge $(F_{n_{j_k}})_1^\infty$. Diese Teilfolge konvergiert schwach gegen F und somit

$$\lim_{k\to\infty} \left| \int u(x)dF_{n_{j_k}}(x) - \int u(x)dF(x) \right| = 0$$

im Widerspruch zu (6.3). □

Beispiel 6.13.
Es sei F_n die Verteilungsfunktion einer $U(0, n)$-verteilten Zufallsvariablen. Es gilt

$$\begin{aligned} \phi_n(t) &= \int \exp(itx)dF_n(x) = \frac{1}{n}\int_0^n \exp(itx)dx \\ &= \left(\frac{\exp(itn) - 1}{itn}\right)\{t \neq 0\} + \{t = 0\} \end{aligned}$$

und somit

$$\lim_{n\to\infty} \phi_n(t) = \{t = 0\} = \phi(t).$$

Die Funktion $\{t = 0\}$ ist aber im Nullpunkt nicht stetig. Damit sind die Voraussetzungen des Satzes nicht erfüllt.

Satz 6.14. *(Zentraler Grenzwertsatz)*
Es sei $(X_n)_1^\infty$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit Erwartungswert μ und endlicher Varianz $\sigma^2 > 0$. Dann gilt

$$\frac{1}{\sqrt{n}\sigma}\sum_1^n (X_i - \mu) \Rightarrow N(0, 1).$$

Beweis: Es sei ϕ die charakteristische Funktion von $(X_1 - \mu)/\sigma$ und ϕ_n die charakteristische Funktion von

$$\frac{1}{\sqrt{n}\sigma} \sum_1^n (X_i - \mu).$$

Es gilt

$$\phi_n(t) = \phi\left(\frac{t}{\sqrt{n}}\right)^n.$$

Nach Satz 4.37 gilt

$$\phi(t) = 1 + i\mu_1 t - \frac{1}{2}\mu_2 t^2 + o(|t|^2)$$

mit $\mu_1 = \mathbb{E}((X_1 - \mu)/\sigma) = 0$ und $\mu_2 = \mathbb{E}((X_1 - \mu)^2/\sigma^2) = 1$. Es gilt also

$$\phi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)$$

und somit

$$\phi_n(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n.$$

Hieraus folgt

$$\lim_{n\to\infty} \phi_n(t) = \lim_{n\to\infty} \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n = \exp\left(-\frac{1}{2}t^2\right)$$

für alle t. Die Funktion $\exp\left(-\frac{1}{2}t^2\right) : \mathbb{R} \to \mathbb{C}$ ist aber stetig in 0, und wir können Satz 6.12 anwenden. Danach konvergiert $\frac{1}{\sqrt{n}\sigma}\sum_1^n (X_i - \mu)$ schwach gegen die Verteilung mit charakteristischer Funktion $\exp\left(-\frac{1}{2}t^2\right)$. Dies ist aber die chrakteristsiche Funktion der $\mathrm{N}(0,1)$-Verteilung. □

Beispiel 6.15.
Es sei $X \sim \mathrm{b}(n,p)$. Wir können schreiben $X = \sum_1^n X_i$ mit $X_i \sim \mathrm{b}(1,p)$. Hieraus folgt

$$\frac{1}{\sqrt{n}} \frac{X - np}{\sqrt{p(1-p)}} \Rightarrow \mathrm{N}(0,1).$$

Für $n = 10$ und $p = 0,4$ liefert die Approximation durch die Normalverteilung

$$\begin{aligned}
\mathbb{P}(3 \leq X \leq 6) &= \mathbb{P}\left(\frac{3-np}{\sqrt{np(1-p)}} \leq \frac{X-np}{\sqrt{np(1-p)}} \leq \frac{6-np}{\sqrt{np(1-p)}}\right) \\
&\approx \Phi\left(\frac{6-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{3-np}{\sqrt{np(1-p)}}\right) \\
&= \Phi(1,291) - \Phi(-0,6455) \\
&= 0.6423.
\end{aligned}$$

In diesem Fall kann man den exakten Wert berechnen. Es gilt

$$\mathbb{P}(3 \leq X \leq 6) = \sum_{3}^{6} \binom{10}{j} 0,4^j 0,6^{10-j} = 0,778.$$

Die $\mathrm{b}(n,p)$-Verteilung ist diskret und die $\mathrm{N}(0,1)$-Verteilung stetig. Um die Güte der Approximation zu verbessern, gibt es eine sogenannte Stetigkeitskorrektur. Sie lautet

$$\mathbb{P}(a \leq X \leq b) \approx \Phi\left(\frac{b-np+0,5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np-0,5}{\sqrt{np(1-p)}}\right).$$

Mit den Werten $a = 3, b = 6, n = 10$ und $p = 0,4$ erhalten wir

$$\mathbb{P}(a \leq X \leq b) \approx \Phi(1,614) - \Phi(-0,9682) = 0,7803.$$

Der Fehler beträgt ungefähr $0,24\%$.

7 Bedingte Erwartungswerte

7.1 Definition

Wir bezeichnen mit $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{Y}$ eine diskrete Zufallsvariable, d.h. der Raum $\mathcal{Y}$ ist abzählbar mit Elementen $y_j, j \in I \subset \mathbb{N}$. Wir setzen ferner voraus, dass

$$\mathbb{P}(Y = y) > 0 \quad \text{für alle } y \in \mathcal{Y}.$$

Für alle F in $\mathcal{F}$ und $y \in \mathcal{Y}$ definieren wir die bedingte Wahrscheinlichkeit für F gegeben $Y = y$ durch

$$\mathbb{P}(F \mid Y = y) := \quad \frac{\mathbb{P}(F \cap \{Y = y\})}{\mathbb{P}(Y = y)}.$$

Wegen $\mathbb{P}(Y = y) > 0$ ist $\mathbb{P}(F \mid Y = y)$ wohl definiert und

$$\mathbb{P}(\cdot \mid Y = y) : \mathcal{F} \to [0, 1]$$

ist für alle y ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$.

Es sei nun $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ eine reellwertige Zufallsvariable mit

$$\mathbb{E}(|X|) < \infty.$$

Es gilt mit Fubini

$$\begin{aligned}
\mathbb{E}(|X|) &= \mathbb{E}\left(\int_0^{|X|} du\right) \\
&= \mathbb{E}\left(\int_0^{\infty} \{0 \le u \le |X|\}\, du\right) \\
&= \int_0^{\infty} \mathbb{E}(\{0 \le u \le |X|\})\, du \\
&= \int_0^{\infty} \mathbb{P}(|X| \ge u)\, du.
\end{aligned}$$

Hieraus folgt

$$\int |X(\omega)|\, d\mathbb{P}(\omega\ |Y=y)$$

$$\begin{aligned}
&= \int \mathbb{P}(|X| \geq u\,|Y=y)\, du \\
&= \int \frac{\mathbb{P}(\{|X| \geq u\}\{Y=y\})\, du}{\mathbb{P}(Y=y)} \\
&\leq \int \frac{\mathbb{P}(|X| \geq u)\, du}{\mathbb{P}(Y=y)} \\
&= \frac{\mathbb{E}(|X|)}{\mathbb{P}(Y=y)} \\
&\leq \mathbb{E}(|X|) < \infty.
\end{aligned}$$

Somit ist $\int X(\omega)\, d\mathbb{P}(\omega\ |Y=y)$ wohl definiert für alle y.

Definition 7.1. *Der bedingte Erwartungswert von X gegeben $Y=y$ wird durch*

$$\mathbb{E}(X|Y=y) := \int X(\omega)\, d\mathbb{P}(\omega\ |Y=y)$$

definiert.

Beispiel 7.2. Zwei Würfeln werden geworfen, und die Augenzahlen bezeichnen wir mit Z_1 und Z_2. Wie setzen $X = Z_1 + Z_2$ und $Y = \min\{Z_1, Z_2\}$. Es gilt

$$\begin{aligned}
\mathbb{P}(Y=1) &= \mathbb{P}(Z_1=1, Z_2>1) + \mathbb{P}(Z_2=1, Z_1>1) + \mathbb{P}(Z_1=Z_2=1) \\
&= \frac{1}{6}\cdot\frac{5}{6} + \frac{1}{6}\cdot\frac{5}{6} + \frac{1}{6}\cdot\frac{1}{6} \\
&= \frac{11}{36}
\end{aligned}$$

Im Allgemeinen erhalten wir $\mathbb{P}(Y=k) = \frac{13-2k}{36}$, $k=1,\ldots,6$. Es gilt

$$\mathbb{P}(\{X=j\}\cap\{Y=k\}) = \begin{cases} 0, & j \leq 2k-1 \\ \frac{1}{36}, & j=2k \\ \frac{2}{36}, & 2k+1 \leq j \leq k+6 \\ 0, & j \geq k+7 \end{cases}$$

Hieraus folgt

$$\begin{aligned}\mathbb{E}(X\,|Y=k) &= (2k+2(2k+1+\ldots+k+6))/(36\mathbb{P}(Y=k)) \\ &= \frac{42+13k-3k^2}{(13-2k)}\end{aligned}$$

Satz 7.3. *Es seien* $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{Y}$ *eine diskrete Zufallsvariable und* $X, (X_n)_1^\infty : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ *reellwertige Zufallsvariablen mit* $\mathbb{E}(|X|) < \infty$, $\mathbb{E}(|X_n|) < \infty$. *Dann gilt*

(a) $\mathbb{E}(\alpha_1 X_1 + \alpha_2 X_2|Y=y) = \alpha_1\mathbb{E}(X_1|Y=y) + \alpha_2\mathbb{E}(X_2|Y=y)$.

(b) $0 \le X \implies 0 \le \mathbb{E}(X|Y=y)$.

(c) $0 \le X_n \uparrow X \implies 0 \le \mathbb{E}(X_n|Y=y) \uparrow \mathbb{E}(X|Y=y)$.

(d) $0 \le |X_n| \le X \implies \mathbb{E}(\lim X_n|Y=y) = \lim \mathbb{E}(X_n|Y=y)$.

Beweis:
Alle Behauptungen folgen aus den entsprechenden Eigenschaften von Erwarungswerten, weil für jedes $y \quad \mathbb{P}(\cdot\,|Y=y)$ ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$ ist.

Definition 7.4. *Es seien* $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ *eine Zufallsvariable mit* $\mathbb{E}(|X|) < \infty$ *und* $Y : \mathcal{Y} \to \mathbb{R}$ *eine diskrete Zufallsvariable. Wir definieren die Funktion* $h : \mathcal{Y} \to \mathbb{R}$ *durch*

$$h(y) = \mathbb{E}(X|Y=y), \quad y \in \mathcal{Y}.$$

Der bedingte Erwartungswert von X *gegeben* Y *wird nun durch*

$$\mathbb{E}(X|Y) := h(Y)$$

definiert.

Beispiel 7.5. Es seine X und Y wie in Beispiel 7.2. Es gilt

$$\mathbb{E}(X\,|Y=k) = \frac{42+13k-3k^2}{(13-2k)}$$

und somit

$$\mathbb{E}(X\,|Y) = \frac{42+13Y-3Y^2}{(13-2Y)}$$

Satz 7.6. *Die Behauptungen von Satz 7.3 gelten entsprechend für die bedingten Erwartungswerte* $\mathbb{E}(X|Y)$.

Beweis:
(c) Es sei $h_n(y) = \mathbb{E}(X_n\,|Y = y)$. Aus Satz 8.2 (c) folgt

$$h(y) = \mathbb{E}(X|Y = y) = \lim_{n\to\infty} \mathbb{E}(X_n|Y = y) = \lim_{n\to\infty} h_n(y).$$

Hieraus folgt

$$\mathbb{E}(X|Y) = h(Y) = \lim_{n\to\infty} h_n(Y) = \mathbb{E}(X_n|Y)$$

und somit die Behauptung.

Satz 7.7. *Ist* $X \geq 0$, *so gilt*

$$\mathbb{E}(X\,|Y = y) = \int_0^\infty \mathbb{P}(X \geq u\,|Y = y)\,du$$

Beweis:
Dies folgt direkt aus der entsprechenden Aussage für normale Erwartungswerte.

7.2 Eigenschaften des bedingten Erwartungswertes

Wir leiten nun die wichtigsten Eigenschaften des bedingten Erwartungswertes her.

Definition 7.8. *Die Zufallsvariable* X *heißt* Y*–messbar, falls es eine Funktion* g *mit* $X = g(Y)$ *gibt.*

Satz 7.9. *Es seien* $X, Z : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ *reellwertige Zufallsvariablen mit* $\mathbb{E}(|X|)$, $\mathbb{E}(|Z|) < \infty$. *Dann gilt*

(a) $\mathbb{E}(|\mathbb{E}(X|Y)|) < \infty$ *und* $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$.

(b) Ist X Y*–messbar, so gilt* $\mathbb{E}(X|Y) = X$.

(c) Ist X Y*–messbar, so gilt* $\mathbb{E}(XZ|Y) = X\mathbb{E}(Z|Y)$.

(d) Ist X *unabhängig von* Y, *so gilt* $\mathbb{E}(X|Y) = \mathbb{E}(X)$.

(e) Für alle $B \subset \mathcal{Y}$ *gilt* $\int_{\{Y\in B\}} X\,d\mathbb{P} = \int_{\{Y\in B\}} \mathbb{E}(X|Y)\,d\mathbb{P}$.

Beweis:

(a) Es gilt $\mathbb{E}(X|Y) = h(Y)$ mit $h(y) = \int X(\omega)\, d\mathbb{P}(\omega\,|Y = y)$ und somit

$$|h(y)| \leq \int |X(\omega)|\, d\mathbb{P}(\omega\,|Y = y) = \int_0^\infty \mathbb{P}(|X(\omega)| \geq u\,|Y = y)du$$

Hieraus folgt

$$\begin{aligned}
\mathbb{E}(|h(Y)|) &= \sum |h(y)| \cdot \mathbb{P}(Y = y) \\
&\leq \sum \int_0^\infty \mathbb{P}(|X(\omega)| \geq u|Y = y)\mathbb{P}(Y = y)\, du \\
&= \int_0^\infty \mathbb{P}(|X(\omega)| \geq u)\, du \\
&= \mathbb{E}(|X(\omega)|) < \infty.
\end{aligned}$$

Somit ist $\mathbb{E}(X|Y)$ integrierbar und $\mathbb{E}(\mathbb{E}(X|Y))$ wohl definiert. Für $a \in \mathbb{R}$ schreiben wir $a^+ = \max(a, 0)$, $a^- = \max(-a, 0)$ sowie

$$h_+(y) = \mathbb{E}(X^+\,|Y = y).$$

Es gilt $a = a^+ - a^-$. Das obige Argument ergibt

$$\mathbb{E}(\mathbb{E}(X^+|Y)) = \sum h_+(y)\mathbb{P}(Y = y) = \int_0^\infty \mathbb{P}(X^+ \geq u)\, du$$

und somit

$$\mathbb{E}(\mathbb{E}(X^+|Y)) = \mathbb{E}(X^+), \quad \mathbb{E}(\mathbb{E}(X^-|Y)) = \mathbb{E}(X^-).$$

Hieraus folgt

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(X|Y)) &= \mathbb{E}(\mathbb{E}(X^+ - X^-|Y)) \\
&= \mathbb{E}(\mathbb{E}(X^+\,|Y)) - \mathbb{E}(\mathbb{E}(X^-\,|Y)) \\
&= \mathbb{E}(X^+) - \mathbb{E}(X^-) \\
&= \mathbb{E}(X).
\end{aligned}$$

(b) Da X Y–messbar ist, gibt es eine Funktion $g : \mathcal{Y} \to \mathbb{R}$ mit $X = g(Y)$. Hieraus folgt mit $X^+ = g^+(Y)$

$$\begin{aligned}
\mathbb{E}(X^+|Y=y) &= \int \mathbb{P}(X^+ \geq u|Y=y)\,du \\
&= \int \mathbb{P}(g^+(Y) \geq u|Y=y)\,du \\
&= \int \frac{\mathbb{P}(g^+(Y) \geq u|Y=y)\,du}{\mathbb{P}(Y=y)} \\
&= \int \{g^+(y) \geq u\}\,du \\
&= g^+(y)
\end{aligned}$$

Somit gilt $\mathbb{E}(X^+|Y) = g^+(Y) = X^+$. Analog ist $\mathbb{E}(X^-|Y) = X^-$, woraus die Behauptung folgt.

(c) X ist Y–messbar, also $X = g(Y)$. Es gilt für $X \geq 0, Z \geq 0$.

$$\begin{aligned}
\mathbb{E}(XZ|Y=y) &= \int \mathbb{P}(XZ \geq u|Y=y)\,du. \\
&= \int \frac{\mathbb{P}(\{XZ \geq u\} \cap \{Y=y\})\,du}{\mathbb{P}(Y=y)} \\
&= \int \frac{\mathbb{P}(\{g(Y)Z \geq u\} \cap \{Y=y\})\,du}{\mathbb{P}(Y=y)} \\
&= \int \frac{\mathbb{P}(\{g(y)Z \geq u\} \cap \{Y=y\})\,du}{\mathbb{P}(Y=y)}.
\end{aligned}$$

Ist nun $g(y) = 0$, so gilt $\mathbb{P}(\{g(y)Z \geq u\}) = 0$ für alle $u \neq 0$ und somit

$$\int \frac{\mathbb{P}(\{g(y)Z \geq u\} \cap \{Y=y\})\,du}{\mathbb{P}(Y=y)} = 0 = g(y)\mathbb{E}(Z\,|Y=y)$$

Ist $g(y) > 0$ so erhalten wir

$$\begin{aligned}
\mathbb{E}(XZ|Y=y) &= \int_0^\infty \frac{\mathbb{P}(\{g(y)Z \geq u\} \cap \{Y=y\})\,du}{\mathbb{P}(Y=y)} \\
&= \int_0^\infty \frac{\mathbb{P}(\{Z \geq u/g(y)\} \cap \{Y=y\})\,du}{\mathbb{P}(Y=y)} \\
&= g(y)\int_0^\infty \frac{\mathbb{P}(\{Z \geq v)\} \cap \{Y=y\})\,dv}{\mathbb{P}(Y=y)} \\
&= g(y)\mathbb{E}(Z\,|Y=y).
\end{aligned}$$

Hieraus folgt $\mathbb{E}(XZ|Y) = g(Y)\mathbb{E}(Z|Y)$ falls $X \geq 0$ und $Z \geq 0$. Die anderen drei Fälle zeigt man analog, und wir erhalten dann die gewünschte Behauptung.

(d) Es gilt für $X \geq 0$ und unabhängig von Y

$$\begin{aligned}
\mathbb{E}(X\,|Y=y) &= \int_0^\infty \frac{\mathbb{P}(\{X \geq u\} \cap \{Y=y\})}{\mathbb{P}(Y=y)}\,du \\
&= \int_0^\infty \frac{\mathbb{P}(X \geq u)\mathbb{P}(Y=y)}{\mathbb{P}(Y=y)}\,du \\
&= \int_0^\infty \mathbb{P}(X \geq u)\,du \\
&= \mathbb{E}(X).
\end{aligned}$$

(e) Es gilt für $X \geq 0$

$$\begin{aligned}
\int_{\{Y\in B\}} X\,d\mathbb{P} &= \sum_{y\in B} \int X\{Y=y\}\,d\mathbb{P} \\
&= \sum_{y\in B} \mathbb{E}(X\{Y=y\}) \\
&= \sum_{y\in B} \mathbb{E}(\mathbb{E}(X\{Y=y\}\,|Y)) \text{ nach (a)} \\
&= \sum_{y\in B} \mathbb{E}(\{Y=y\}\mathbb{E}(X\,|Y)) \text{ nach (c)} \\
&= \sum_{y\in B} \int \mathbb{E}(X\,|Y)\{Y=y\}\,d\mathbb{P} \\
&= \int_{\{Y\in B\}} \mathbb{E}(X\,|Y)\,d\mathbb{P}
\end{aligned}$$

Satz 7.10. *Es seien $X, Z : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ reellwertige Zufallsvariablen. Dann gilt*

(a) (Hölder) Es seien $0 < p, q$ relle Zahlen mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt

$$|\mathbb{E}(XZ\,|Y)| \leq \mathbb{E}(|X|^p\,|Y)^{1/p}\mathbb{E}(|Z|^q\,|Y)^{1/q}$$

(b) (Minkowski) Für alle $p > 0$ gilt

$$\mathbb{E}(|X+Z|^p\,|Y)^{1/p} \leq \mathbb{E}(|X|^p\,|Y)^{1/p} + \mathbb{E}(|Z|^p\,|Y)^{1/p}$$

(c) (Jensen) Es sei $\phi : I \subset \mathbb{R} \to \mathbb{R}$ konvex, I ein Intervall, und es gelte $\mathbb{P}(X \in I) = 1$ sowie $\mathbb{E}(|\phi(X)|) < \infty$. Dann gilt

$$\phi(\mathbb{E}(X\,|Y)) \leq \mathbb{E}(\phi(X)\,|Y).$$

Beweis:
(a) und (b) folgen direkt aus den entsprechenden Sätzen, weil $\mathbb{P}(\cdot\,|Y = y)$ für alle $y \in \mathcal{Y}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$ ist.

(c) Da $\phi : I \to \mathbb{R}$ konvex, gibt es zu jedem t im Inneren von I ein α so, dass

$$\phi(s) \geq \phi(t) + \alpha(s - t), \text{ für alle } s \in I$$

Es folgt mit $t = \mathbb{E}(X\,|Y)$ und $s = X$

$$\phi(X) \geq \phi(\mathbb{E}(X\,|Y)) + \alpha(X - \mathbb{E}(X\,|Y)),$$

wobei α eine Funktion von $\mathbb{E}(X\,|Y)$ und somit Y–messbar ist. Es folgt

$$\begin{aligned}
\mathbb{E}(\phi(X)\,|Y) &\geq \mathbb{E}(\phi(\mathbb{E}(X\,|Y))\,|Y) + \mathbb{E}(\alpha(X - \mathbb{E}(X\,|Y))\,|Y) \\
&= \phi(\mathbb{E}(X\,|Y)) + \alpha\mathbb{E}((X - \mathbb{E}(X\,|Y))\,|Y) \text{ (nach (c))} \\
&= \phi(\mathbb{E}(X\,|Y)).
\end{aligned}$$

8 Martingale

8.1 Filtrationen und Stoppzeiten

Definition 8.1. *Eine Folge $(Y_n)_0^\infty$ von Zufallsvariablen heißt eine Filtration, falls Y_n für alle n Y_{n+1}–messbar ist.*

Beispiel 8.2. Es seien $(Z_n)_0^\infty$ eine Folge von Zufallsvariablen. Wir setzen $Y_n = (Z_0, \ldots, Z_n)$. Dann ist $(Y_n)_0^\infty$ eine Filtration. Denn $Y_n = \Pi_0^n(Y_{n+1}) = \Pi_0^n((Z_0, \ldots, Z_{n+1}))$ wobei Π_m^n die Projektion auf die m-te bis n-te Koordinaten bezeichnet.

Definition 8.3. *Es seien $(X_n)_0^\infty$ Zufallsvariablen und $(Y_n)_0^\infty$ eine Filtration. Ist X_n Y_n–meßbar für alle n, so heißt $((X_n, Y_n))_0^\infty$ adaptiert.*

Beispiel 8.4. Es seien $(Z_n)_0^\infty$ und $(Y_n)_0^\infty$ wie im Beispiel 8.2. Für rellwertige Z_n setzen wir $X_n = \sum_{j=0}^n Z_j$. Dann ist $((X_n, Y_n))_0^\infty$ adaptiert.

Definition 8.5. *Es sei $(Y_n)_0^\infty$ eine Filtration. Eine Abbildung $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ heißt eine Stoppzeit, falls $\{\tau = n\}$ Y_n-messbar.*

Beispiel 8.6. Es sei $((X_n, Y_n))_0^\infty$ wie im Beispiel 8.4. Für $a \in \mathbb{R}$ sei

$$\tau = \begin{cases} \min\{n : X_n > a\} & \text{falls } \sup_n X_n > a \\ \infty & \text{falls } \sup_n X_n \leq a \end{cases}$$

Dann ist τ eine Stoppzeit: Denn

$$\{\tau = n\} = \{X_1 \leq a, \ldots X_{n-1} \leq a, X_n > a\}.$$

Satz 8.7. *Es seien $(Y_n)_0^\infty$ eine Filtration und τ, τ_1 und τ_2 Stoppzeiten. Dann gilt*

(a) τ eine Stoppzeit $\Leftrightarrow \{\tau \leq n\}$ Y_n-meßbar.

(b) $\tau_1 + \tau_2$ ist eine Stoppzeit.

(c) $\max(\tau_1, \tau_2)$ ist eine Stoppzeit.

Beweis:

(a) Es sei τ eine Stoppzeit. Es gilt $\{\tau \leq n\} = \sum_{j=0}^n \{\tau = j\}$, und somit ist $\{\tau \leq n\}$ Z_n–messbar. In der anderen Richtung, ist $\{\tau \leq n\}$ Y_n–messbar, dann ist

$$\{\tau = n\} = \{\tau \leq n\} - \{\tau \leq n-1\}$$

Y_n–messbar.

(b) Es gilt

$$\{\tau_1 + \tau_2 = n\} = \sum_{j=0}^{n} \{\tau_1 = j\}\{\tau_2 = n - j\},$$

und somit ist $\{\tau_1 + \tau_2 = n\}$ Y_n–messbar.

(c) Es gilt

$$\{\max\{\tau_1, \tau_2\} = n\} = \{\tau_1 \leq n\}\{\tau_2 = n\} + \{\tau_1 = n\}\{\tau_2 < n\}$$

und somit ist $\{\max\{\tau_1, \tau_2\} = n\}$ Y_n–messbar.

Definition 8.8. *Es seien $(Y_n)_0^\infty$ eine Filtration, τ eine Stoppzeit und X eine auf $\{\tau < \infty\}$ definierte Zufallsvariable. X heißt Y_τ–messbar, falls $X\{\tau = n\}$ Y_n–messbar ist.*

Satz 8.9. *Es seien $(X_n, Y_n)_0^\infty$ adaptiert und τ eine Stoppzeit. Dann ist*

$$X_\tau = \sum_{j=0}^{\infty} X_j\{\tau = j\}$$

Y_τ–messbar.

Beweis:
Es gilt $X_\tau\{\tau = n\} = X_n\{\tau = n\}$, und somit ist $X_\tau\{\tau = n\}$ Y_n–messbar. Damit ist X_τ Y_τ–messbar.

8.2 Sub-, Super- und Martingale

Definition 8.10. *Es seien $(X_n, Y_n)_0^\infty$ adaptiert und Y_n diskrete Zufallsvariablen, $n \in \mathbb{N}_0$. Dann heißt $(X_n, Y_n)_0^\infty$ ein*

(a) Martingal, falls

$$\mathbb{E}(X_{n+1}|Y_n) = X_n,\ n \in \mathbb{N}_0.$$

(b) Supermartingal, falls

$$\mathbb{E}(X_{n+1}|Y_n) \leq X_n,\ n \in \mathbb{N}_0.$$

(c) Submartingal, falls

$$\mathbb{E}(X_{n+1}|Y_n) \geq X_n,\ n \in \mathbb{N}_0.$$

Beispiel 8.11. Es seien $(Z_n)_0^\infty$ u.v. diskrete, rellwertige Zufallsvariablen mit $\mathbb{E}(Z_n) = 0$ für alle n. Wir setzen $Y_n = (Z_0, \ldots, Z_n)$ sowie $X_n := \sum_{j=0}^n Z_j$. Dann ist $(X_n, Y_n)_0^\infty$ ein Martingal. Denn $(X_n, Y_n)_0^\infty$ ist adaptiert und

$$\begin{aligned}
\mathbb{E}(X_{n+1}\,|Y_n) &= \mathbb{E}(X_n + Z_{n+1}|Y_n) \\
&= \mathbb{E}(X_n|Y_n) + \mathbb{E}(Z_{n+1}|Y_n) \\
&= X_n + \mathbb{E}(Z_{n+1}) \text{ (Satz 7.9 (b), (d))} \\
&= X_n.
\end{aligned}$$

Beispiel 8.12. Es seien $(Z_n)_0^\infty$ u.v. diskrete, reellwertige Zufallsvariablen mit $\mathbb{E}(Z_n) = 1$. Wir setzen $Y_n = (Z_0, \ldots, Z_n)$ sowie $X_n := \prod_{j=0}^n Z_j$. Dann ist $(X_n, Y_n)_0^\infty$ ein Martingal. Denn $(X_n, Y_n)_0^\infty$ ist adaptiert und

$$\begin{aligned}
\mathbb{E}(X_{n+1}|Y_n) &= \mathbb{E}(X_n Z_{n+1}|Y_n) \\
&= X_n \mathbb{E}(Z_{n+1}|Y_n) \text{ (Satz 7.9 (c))} \\
&= X_n \mathbb{E}(Z_{n+1}) \text{ (Satz 7.9 (d))} \\
&= X_n.
\end{aligned}$$

Beispiel 8.13. Es seien $(Z_{n,j})_{n,j=1}^\infty$ u.i.v. mit Werten in $\mathbb{N}_0$. Wir setzen $X_0 \equiv 1$ und definieren rekursiv $X_{n+1} = \sum_{j=1}^{X_n} Z_{n+1,j}$ wobei $\sum_{j=1}^0 Z_{n+1,j} := 0$. Wir setzen $Y_n = (X_0, \ldots, X_n)$. Ferner sei $m = \mathbb{E}(Y_{1,1}) < \infty$. Es gilt

$$\begin{aligned}
\mathbb{E}(X_{n+1}|Y_n) &= \mathbb{E}\left(\sum_{j=1}^{X_n} Z_{n+1,j}|Y_n\right) \\
&= \mathbb{E}\left(\sum_{j=1}^{\infty} Z_{n+1,j}\{j \le X_n\}|Y_n\right) \\
&= \sum_{j=1}^{\infty} \mathbb{E}(Z_{n+1,j}\{j \le X_n\}|Y_n) \\
&= \sum_{j=1}^{\infty} \{j \le X_n\}\mathbb{E}(Z_{n+1,j}|Y_n) \\
&= \sum_{j=1}^{\infty} \{j \le X_n\} \cdot m \\
&= mX_n.
\end{aligned}$$

Wir definieren $W_n := X_n/m^n$. Es folgt nun unmittelbar, dass $(W_n, Y_n)_0^\infty$ ein Martingal ist.

Satz 8.14. (Switching-Prinzip) *Es seien $(X_n, Y_n)_0^\infty$ und $(Z_n, Y_n)_0^\infty$ Supermartingale und τ eine Stoppzeit mit $X_\tau \leq Z_\tau$ auf $\{\tau < \infty\}$, d.h. $X_n \leq Z_n$ auf $\{\tau = n\}$. Dann ist $(W_n, Y_n)_0^\infty$ mit*

$$W_n = \begin{cases} Z_n & \{n < \tau\} \\ X_n & \{n \geq \tau\} \end{cases}$$

ein Supermartingal.

Beweis:
Es gilt

$$\begin{aligned}
&\mathbb{E}(W_{n+1}|Y_n) \\
&= \mathbb{E}(W_{n+1}\{\tau \leq n\}|Y_n) + \mathbb{E}(W_{n+1}\{\tau = n+1\}|Y_n) + \mathbb{E}(W_{n+1}\{\tau > n+1\}|Y_n) \\
&= \mathbb{E}(X_{n+1}\{\tau \leq n\}|Y_n) + \mathbb{E}(X_{n+1}\{\tau = n+1\}|Y_n) + \mathbb{E}(Z_{n+1}\{\tau > n+1\}|Y_n) \\
&\leq \mathbb{E}(X_{n+1}\{\tau \leq n\}|Y_n) + \mathbb{E}(Z_{n+1}\{\tau = n+1\}|Y_n) + \mathbb{E}(Z_{n+1}\{\tau > n+1\}|Y_n) \\
&= \{\tau \leq n\}\mathbb{E}(X_{n+1}|Y_n) + \mathbb{E}(Z_{n+1}\{\tau \geq n+1\}|Y_n) \\
&= \{\tau \leq n\}\mathbb{E}(X_{n+1}|Y_n) + \{\tau \geq n+1\}\mathbb{E}(Z_{n+1}|Y_n) \\
&\leq \{\tau \leq n\}X_n + \{\tau \geq n+1\}Z_n \\
&= W_n.
\end{aligned}$$

Satz 8.15. *Es sei $(X_n, Y_n)_0^\infty$ ein nicht-negatives Supermartingal. Dann gilt für alle $\lambda > 0$:*

$$\mathbb{P}(\max_{0 \leq j \leq n} X_j \geq \lambda) \leq \mathbb{E}\left(\min\left(\frac{X_0}{\lambda}, 1\right)\right).$$

Beweis:
$(Z_n, Y_n)_0^\infty$ mit $Z_n \equiv \lambda$ ist ein Supermartingal. Es sei $\tau = \min\{n : X_n \geq \lambda\}$. Dann gilt auf $\{\tau < \infty\}$ $Z_\tau \leq X_\tau$. So ist $W_n = X_n\{n < \tau\} + \lambda\{n \geq \tau\}$ ein Supermartingal. Hieraus folgt $\mathbb{E}(W_{n+1}|Y_n) \leq W_n$ und somit

$$\begin{aligned}
\mathbb{E}(W_{n+1}\{\tau \leq n\}|Y_n) &= \{\tau \leq n\}\mathbb{E}(W_{n+1}|Y_n) \\
&\leq \{\tau \leq n\}W_n
\end{aligned}$$

Auf $\{\tau \leq n\}$ ist $W_n = \lambda$ und somit

$$\lambda\{\tau \leq n\} = \{\tau \leq n\}W_n \leq W_n$$

woraus folgt

$$\lambda \mathbb{P}(\tau \leq n) \leq \mathbb{E}(W_n) \leq \cdots \leq \mathbb{E}(W_0) = \mathbb{E}(\min(X_0, \lambda)).$$

Wegen $\{\tau \leq n\} = \{\max_{0 \leq j \leq n} X_j \geq \lambda\}$ erhalten wir

$$\mathbb{P}(\max_{0 \leq j \leq n} X_j \geq \lambda) \leq \mathbb{E}\left(\min\left(\frac{X_0}{\lambda}, 1\right)\right).$$

Korollar 8.16. *Es sei $(X_n, Y_n)_0^\infty$ ein nicht-negatives Supermartingal. Dann gilt $\mathbb{P}(\sup_n X_n < \infty) = 1$.*

Beweis:
Es gilt

$$\mathbb{P}(\sup_n X_n \geq \lambda) = \lim_{n \to \infty} \mathbb{P}(\max_{0 \leq j \leq n} X_j \geq \lambda) \leq \mathbb{E}\left(\min\left(\frac{X_0}{\lambda}, 1\right)\right)$$

und mit $\lambda = m^2$ erhalten wir

$$\mathbb{P}(\sup_n X_n \geq m^2) \leq \mathbb{E}\left(\min\left(\frac{X_0}{m^2}, 1\right)\right) \leq \frac{1}{m^2}\mathbb{E}(X_0)$$

Hieraus folgt $\sum_{m=1}^\infty \mathbb{P}(\sup_n X_n \geq m^2) < \infty$, und mit Borel-Cantelli erhalten wir $\mathbb{P}(\sup_n X_n < \infty) = 1$.

Definition 8.17. *Es seien $\alpha < \beta$ und $(x_n)_0^\infty$ reelle Zahlen. Wir definieren*

$$\nu_1 = \begin{cases} \min\{n : x_n < \alpha\} & falls\ \inf x_n < \alpha \\ \infty & sonst \end{cases}$$

$$\nu_2 = \begin{cases} \min\{n : n > \nu_1,\, x_n > \beta\} & falls\ \sup_{n > \nu_1} x_n > \beta \\ \infty & sonst \end{cases}$$

$$\nu_{2k+1} = \begin{cases} \min\{n : n > \nu_{2k},\, x_n < \alpha\} & falls\ \inf_{n > \nu_{2k}} x_n < \alpha \\ \infty & sonst \end{cases}$$

$$\nu_{2k+2} = \begin{cases} \min\{n : n > \nu_{2k+1},\, x_n > \beta\} & falls\ \sup_{n > \nu_{2k+1}} x_n > \beta \\ \infty & sonst. \end{cases}$$

u.s.w. Die Anzahl der Überquerungen von unten nach oben des Intervalls (α, β) durch die Folge $(x_n)_0^\infty$ bezeichnen wir mit

$$\mathbb{N}(\alpha, \beta) = \max\{2j : \nu_{2j} < \infty\}.$$

Satz 8.18. *Es sei $(X_n, Y_n)_0^\infty$ ein nicht-negatives Supermartingal. Dann gilt*

$$\mathbb{P}(\mathbb{N}(\alpha,\beta) \geq k) \leq \left(\frac{\alpha}{\beta}\right)^k \mathbb{E}\left(\min\left(\frac{X_0}{\alpha}, 1\right)\right)$$

für alle α, β, $0 < \alpha < \beta < \infty$.

Beweis:
Es sei $(\nu_j)_1^\infty$ die Folge von Definition 8.17. Die $(\nu_j)_1^\infty$ sind Stoppzeiten.
Wir definieren nun eine Folge $(Z_n)_0^\infty$ wie folgt

$$Z_n = \begin{cases} 1 & 0 \leq n < \nu_1 \\ \frac{X_n}{\alpha} & \nu_1 \leq n < \nu_2 \\ \frac{\beta}{\alpha} \cdot 1 & \nu_2 \leq n < \nu_3 \\ \frac{\beta}{\alpha} \cdot \frac{X_n}{\alpha} & \nu_3 \leq n < \nu_4 \\ \vdots & \\ \left(\frac{\beta}{\alpha}\right)^{k-1} \frac{X_n}{\alpha} & \nu_{2k-1} \leq n < \nu_{2k} \\ \left(\frac{\beta}{\alpha}\right)^k & \nu_{2k} \leq n. \end{cases}$$

Es gilt $X_{\nu_1} < \alpha$, und somit ist $\frac{X_{\nu_1}}{\alpha} < 1 = Z_{\nu_1}$. Nach dem Switching-Prinzip ist $(U_{1,n}, Y_n)_0^\infty$ mit

$$U_{1,n} = \begin{cases} Z_1 & 0 \leq n < \nu_1 \\ \frac{X_n}{\alpha} & \nu_1 \leq n \end{cases}$$

ein Supermartingal. Wir definieren

$$U_{2,n} = \begin{cases} U_{1,n} & 0 \leq n < \nu_2 \\ \frac{\beta}{\alpha} & \nu_2 \leq n \end{cases}$$

und wegen $X_{\nu_2} > \beta$ ist $U_{2,\nu_2} < U_{1,\nu_2}$. Wieder nach dem Switching-Prinzip ist $(U_{2,n}, Y_n)_0^\infty$ ein Supermartingal. Im Allgemeinen ist $(U_{2k,n}, Z_n)_0^\infty$ ein Supermartingal und

$$U_{2k,n} \geq \left(\frac{\beta}{\alpha}\right)^k \{\nu_{2k} \leq n\}.$$

Hieraus folgt

$$\left(\frac{\beta}{\alpha}\right)^k \mathbb{P}(\nu_{2k} \leq n) \leq \mathbb{E}(U_{2k,n}) \leq \ldots \leq \mathbb{E}(U_{1,n}) = \mathbb{E}\left(\min\left(\frac{X_0}{\alpha}, 1\right)\right)$$

und somit

$$\mathbb{P}(\nu_{2k} \leq n) \leq \left(\frac{\alpha}{\beta}\right)^k \mathbb{E}\left(\min\left(\frac{X_0}{\alpha}, 1\right)\right).$$

Mit $n \to \infty$ erhalten wir

$$\mathbb{P}(\nu_{2k} < \infty) \leq \left(\frac{\alpha}{\beta}\right)^k \mathbb{E}\left(\min\left(\frac{X_0}{\alpha}, 1\right)\right).$$

Es gilt $\nu_{2k} < \infty$ genau dann, wenn $\mathbb{N}(\alpha, \beta) \geq k$, und wir erhalten

$$\mathbb{P}(\mathbb{N}(\alpha, \beta) \geq k) \leq \left(\frac{\alpha}{\beta}\right)^k \mathbb{E}\left(\min\left(\frac{X_0}{\alpha}, 1\right)\right).$$

Satz 8.19. *Es sei $(X_n, Y_n)_0^\infty$ ein nicht-negatives Supermartingal. Dann gilt*

$$\lim_{n\to\infty} X_n =: X_\infty \quad \textit{f.s. und } X_n \geq \mathbb{E}(X_\infty | Y_n).$$

Beweis:
Eine beschränkte Folge $(X_n)_0^\infty$ nicht-negativer Zahlen konvergiert genau dann, wenn $\mathbb{N}(\alpha, \beta) < \infty$ für alle $0 < \alpha < \beta < \infty, \alpha, \beta \in \mathbb{Q}$. Aus Satz 8.18 folgt

$$\sum_{k=0}^{\infty} \mathbb{P}(\mathbb{N}(\alpha, \beta) \geq k) \leq \sum_{k=0}^{\infty} \left(\frac{\alpha}{\beta}\right)^k < \infty$$

und mit dem Borel-Cantelli Lemma können wir nun schließen, dass gilt $\mathbb{P}(\mathbb{N}(\alpha, \beta) < \infty) = 1$. Hieraus folgt

$$\mathbb{P}(\mathbb{N}(\alpha, \beta) < \infty \quad \forall\, \alpha, \beta \in \mathbb{Q},\, 0 < \alpha < \beta < \infty) = 1$$

und somit konvergiert die Folge $(X_n)_0^\infty$ nach der Vorbemerkung f.s. Da $(X_n, Y_n)_0^\infty$ ein Supermartingal ist, gilt $X_n \geq \mathbb{E}(X_m | Y_n)$ für $m \geq n$ und mit Fatou folgt

$$X_n \geq \lim_{m\to\infty} \mathbb{E}(X_m | Y_n) \geq \mathbb{E}(\lim_{m\to\infty} X_m | Y_n) = \mathbb{E}(X_\infty | Y_n).$$

Beispiel 8.20. Es sei $(X_n, Y_n)_0^\infty$ wie in Beispiel 8.12, $X_n = \prod_0^n Z_j$ mit $Z_j \geq 0$ sowie $\mathbb{E}(Z_j) = 1$. Nach Satz 8.19 konvergiert die Folge $(X_n)_0^\infty$ f.s. gegen eine Zufallsvariable X_∞. Sind die Z_n sind u.i.v. und nicht degeneriert, so folgt aus der Ungleichung von Jensen $\mathbb{E}(\log Z_n) < \log(\mathbb{E}(Z_n)) = 0$. Aus dem starken Gesetz der großen Zahlen folgt

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_0^n \log Z_j = \mathbb{E}(\log Z_j) < 0.$$

Dies impliziert $\sum_0^n \log Z_j \to -\infty$ und somit ist $X_\infty = 0$ f.s. Insbesondere gilt $X_n > 0 = \mathbb{E}(X_\infty \,|Y_n)$ falls $Z_n > 0$ f.s.

Beispiel 8.21. Es sei $(X_n, Y_n)_0^\infty$ wie in Beispiel 8.13. Dann ist $(W_n, Y_n)_0^\infty$ ein nicht-negatives Martingal. Aus Satz 8.19 folgt

$$\lim_{n\to\infty} \frac{X_n}{m^n} = \lim_{n\to\infty} W_n = W_\infty \; f.s.$$

Wegen $W_n \geq \mathbb{E}(W_\infty \,|Y_n)$ folgt $X_n \geq m^n \mathbb{E}(W_\infty \,|Y_n)$.

Beispiel 8.22. Zum Zeitpunkt $n = 0$ enthält eine Urne $a > 0$ rote und $b > 0$ weiße Kugeln. Eine Kugel wird der Urne entnommen und zusammen mit $c > 0$ weiteren Kugeln derselben Farbe wieder in die Urne gelegt. Es sei X_n die Proportion von weißen Kugeln zum Zeitpunkt n. Wir setzen $Y_n = (X_0, \ldots, X_n)$. Zum Zeitpunkt n enthält die Urne $a + b + nc$ Kugeln, wovon $X_n(a+b+nc)$ weisse Kugeln sind. Gegeben X_n wird zum Zeitpunkt n eine weiße bzw. rote Kugel mit Wahrscheinlichkeit X_n bzw. $1 - X_n$ gezogen. Hieraus folgt

$$\begin{aligned}
\mathbb{E}(X_{n+1} \,|Y_n) &= \frac{X_n(a+b+nc)+c}{a+b+(n+1)c} X_n + \frac{X_n(a+b+nc)}{a+b+(n+1)c}(1-X_n) \\
&= \frac{X_n^2(a+b+nc) + cX_n + X_n(a+b+nc) - X_n^2(a+b+nc)}{a+b+(n+1)c} \\
&= \frac{X_n(a+b+(n+1)c)}{a+b+(n+1)c} \\
&= X_n
\end{aligned}$$

Somit ist $(X_n, Y_n)_0^\infty$ ein nicht-negatives Martingal. Nach Satz 8.19 konvergiert $(X_n)_0^\infty$ f.s. gegen eine Zufallsvariable X_∞. Da die X_n beschränkt sind, $-1 < X_n < 1$, folgt aus dominierter Konvergenz für $m \geq n$

$$X_n = \mathbb{E}(X_m \,|Y_n) = \lim_{m\to\infty} \mathbb{E}(X_m \,|Y_n) = \mathbb{E}(\lim_{m\to\infty} X_m \,|Y_n) = \mathbb{E}(X_\infty \,|Y_n).$$