2 Diskrete und stetige W-Räume

2.1 Uniforme und geometrische Verteilung

Anmerkung. Beim schwachen Gesetz der großen Zahl haben wir die Grenzen der bisher als Modelle betrachteten endlichen W-Räume erreicht. Die in Bezeichnung 1.20.6 eingeführten projektiven Systeme $(\Omega_n, P_n \mathrm{pr}_n)$ endlicher W-Räume sind umständlich zu handhaben und nicht allgemein genug:

– Man müßte Produkte, Bildverteilungen, Bedingte Verteilungen . . . für projektive Systeme erklären. Wie verhalten sich projektive Systeme von projektiven Systemen?

– Interpretiert in einem projektiven System den Parameter $n \in \mathbb{N}$ als Zeitpunkte, in denen die Experimente durchgeführt werden, so könnte man auch an *kontinuierliche Zeitparameter* $t \in \mathbb{R}$ denken.

Der elegante und auch anschauliche Ausweg sind *unendliche* W-Räume.

Wir wollen hier zwei einfache Bespiele unendlicher W-Räume mit anschaulichen Hilfsmitteln vorführen. Im Beispiel 2.1.2 erhalten wir eine stetige W-Verteilung auf dem Intervall $[0, 1)$. im zweiten Beispiel 2.1.3 einen diskretes W-Maß auf $\mathbb{N}$. Für beide Beispiele geben wir jeweils zwei Herleitungen an,

– als **projektive Limes** eines projektiven Systems endlicher W-Räume,

– als Grenzwerte einer Folge endlicher Verteilungen auf $[0, 1]$ bzw. auf $\mathbb{N}$.

2.1.1 Bem. (Projektiver Limes)

Es sei $(\Omega_n, P_n, \mathrm{pr}_n$ ein projektives System endlicher W-Räume. Die Frage ist, kann man nicht ein W-Raum (Ω, P) finden, derart daß die Räume (Ω_n, P_n) Bilder von (Ω, P) mit der Bildverteilung sind. D.h., es gibt eine Folge von Zufallsvariablen

$$\pi_n : (\Omega, P) \to (\Omega_n, P_n)$$

mit den folgenden Eigenschaften:

(i) *Das Diagramm*

$$\begin{tikzcd} \Omega \arrow[d, "\pi_{n+1}"'] \arrow[dr, "\pi_n"] & \\ \Omega_{n+1} \arrow[r, "\mathrm{pr}_n"] & \Omega_n \end{tikzcd}$$

kommutiert für alle $n \in \mathbb{N}$,

(ii) $P_n = P_{\pi_n}$.

Häufig kommt man mit den beiden Eigenschaften (i) und (ii) aus. Zur Definition des projektiven Limes gehört noch die Eindeutigkeit, die wir der Vollständigkeit halber hier mit aufführen. Die Eindeutigkeit folgt aus einer *universellen Eigenschaft.* Zur Formulierung dieser universellen Eigenschaft benötigt man den Begriff der **meßbaren Abbildung**, den wir hier ohne Erklärung verwenden.

(iii) *Man nennt (Ω, P, π_n) den projektiven Limes des projektiven Systems $(\Omega_n, P_n, \mathrm{pr}_n)$, wenn (Ω, P, π_n) die folgende universelle Eigenschaft hat:*
Ist $\widetilde{\Omega}, \widetilde{P}, \widetilde{\pi}_n)$ ein weiterer W-Raum mit den Eigenschaften (i) *und* (ii)*, so gibt es eine eindeutig bestimmte meßbare Abbildung*

$$\rho : \widetilde{\Omega} \to \Omega \quad \textit{mit } \widetilde{\pi}_n = \pi_n \circ \rho.$$

D.h. Ω ist in diesem Sinne minimal.

2.1.2 Bsp. (Münzwurf-Raum)

Das immerwährende Werfen einer fairen Münze beschreiben wir durch das projektive System

$$(\Omega_n, P_n, \mathrm{pr}_n) := (\{0, 1\}^n, P^{\otimes n}, \mathrm{pr}_n).$$

Dabei ist P die Laplace-Wahrscheinlichkeit auf $\{0, 1\}$ und $\mathrm{pr}_n : \{0, 1\}^{n+1} \to \{0, 1\}^n$ die Projektion auf die ersten n Komponenten.

Als projektiver Limes (Ω, P) bietet sich hier ein guter Bekannter aus der Analysis an, das Intervall $[0, 1) \subset \mathbb{R}$. Wir beschränken uns auf die Eigenschaften Bemerkung 2.1.1 (i) und (ii).

Die universelle Eigenschaft (iii) gilt auch. Wir zeigen sie aber nicht, da man für den Nachweis etwas Maßtheorie benötigt.

Damit die folgenden Formeln schön symmetrisch sind, lassen wir bei allen Intervalle den rechten Endpunkt weg, was aber nicht weiter wichtig ist. Die Wahrscheinlichkeit eines Teilintervalls $[a, b) \subset [0, 1)$ ist seine Länge:

$$\mathcal{U}([a, b]) := b - a.$$

Für endliche Vereingungen disjunkter Intervalle addiere man die Längen. Man bezeichnet diese Wahrscheinlichkeitsverteilung mit $\mathcal{U} := \mathcal{U}_{[0,1)}$ und nennt sie die Gleichverteilung oder *uniforme* Verteilung auf $[0, 1)$.

Die Teilmengen

$$A_n := \Big[\frac{1}{2^n}, \frac{2}{2^n}\Big) \cup \Big[\frac{3}{2^n}, \frac{4}{2^n}\Big) \cup \cdots \cup \Big[\frac{2^n - 1}{2^n}, \frac{2^n}{2^n}\Big)$$

sind unabhängig, da für alle $\{\nu_1, \ldots, \nu_k\} \subset \{1, \ldots, n\}$ gilt:

$$\mathcal{U}\big(\bigcap_{i=1}^{k} A_{\nu_i}\big) = 2^{-k} = \prod_{i=1}^{k} \mathcal{U}(A_{\nu_i}).$$

Die Indikatorfunktionen $\mathbb{1}_{A_n}$, $(n \in \mathbb{N})$ sind n-fach Bernoulli-verteilt mit Parameter $p = \frac{1}{2}$. Es gilt also:

(i) Die Bildverteilung von

$$\pi_n := (\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n})$$

ist die Produktverteilung $P^{\otimes n}$ auf $\{0, 1\}^n$.

(ii) Nach Konstruktion ist $\pi_n = \mathrm{pr}_n \circ \pi_{n+1}$.

Auf $([0, 1), \mathcal{U})$ gibt es eine abzählbare Familie $(\mathbb{1}_{A_n})_n$ unabhängiger und identisch verteilter reeller Zufallsvariabler. Diese beschreiben den unendlichen Münzwurf.

Anmerkung. Man fragt sich, ob man nicht statt $([0, 1), \mathcal{U})$ den unendlichen Produktraum $\{0, 1\}^{\mathbb{N}}$ versehen mit einem *unendlichen Produktmaß* $P^{\otimes \mathbb{N}}$ nehmen kann? In der Maßtheorie zeigt man, daß das geht und daß die W-Räume $([0, 1), \mathcal{U})$ und $(\{0, 1\}^{\mathbb{N}}, P^{\otimes \mathbb{N}})$ in einem noch zu präzisierenden Sinne isomorph sind. Beide Räume repräsentieren den sogenannten *projektiven Limes* des projektiven Systems:

$$\lim_{n\to\infty} \mathrm{proj}\,(\{0, 1\}^n, P^{\otimes n}, \mathrm{pr}_n),$$

was wir hier aber mangels Maßtheorie nicht beweisen können.

2.1.3 Bsp. (geometrische Verteilung)
In einem n-fachen Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p > 0$ ist die Wahrscheinlichkeit, den ersten Erfolg genau im $k+1$-ten Teilversuch zu haben, d.h. man hat zuvor k Mißerfolge

$$\mathcal{G}_{n;p}\{k\} := \begin{cases} p(1-p)^k & \text{für } k = 0, \dots, n-1, \\ (1-p)^n & \text{wenn gar kein Erfolg eintritt.} \end{cases}$$

Wir werden den letzteren Fall, daß kein Erfolg eintritt, mit $k = n$ kodieren. Da nach der geometrischen Summenformel

$$\sum_{k=0}^{n-1} p(1-p)^k = (1-p)^n.$$

gilt, ist $\mathcal{G}_{n;p}$ ein W-Maß auf $\Omega_n := (\{1, \dots, n, n+1\}, \mathcal{G}_{n;p})$. Man nennt daher $\mathcal{G}_{n;p}$ die **gestoppte n-te geometrische Verteilung**. *„gestoppt"*, da man ja weiter experimentieren kann, bis ein Erfolg eintritt. Mit der Projektion $\mathrm{pr}_n : \Omega_{n+1} \to \Omega_n$,

$$\mathrm{pr}_n : k \mapsto \begin{cases} k & \text{für } k = 1, \dots, n, \\ n & \text{für } k = n+1, n+2. \end{cases}$$

erhält man ein projektives System $(\Omega_n, \mathcal{G}_{n;p}, \mathrm{pr}_n)$ von endlichen W-Räumen (→ Bezeichnung 1.20.6).

Den projektiven Limes kann man in diesem Fall sofort erraten, es ist der Raum $\mathbb{N}$ der natürlichen Zahlen mit der **geometrischen Verteilung**

$$\mathcal{G}_p\{k\} = p(1-p)^k \quad \text{für } k \in \mathbb{N}_0.$$

Mit der geometrischen Reihe überprüft man

$$\sum_{k=0}^{\infty} P\{k\} = p \sum_{k=0}^{\infty} (1-p)^k = p \frac{1}{1-(1-p)} = 1.$$

Die gestoppten W-Räume $(\Omega_n . \mathcal{G}_{n;p})$ sind die Bilder von $(\mathbb{N}, \mathcal{G}_p)$ unter der Projektion

$$\pi_n : \mathbb{N}_0 \to \{0, \dots, n\}$$

mit

$$\pi_n : k \mapsto \begin{cases} k & \text{für } k = 1, \dots, n, \\ n & \text{für } k = n+1, n+2, \dots. \end{cases}$$

Das System $(\mathbb{N}_0, \mathcal{G}_p, \pi_n)$ erfüllt also die Bedingen (i) und (ii) von Bemerkung 2.1.1.

Man sieht leicht, daß auch die universelle Eigenschaft → Bemerkung 2.1.1 (iii) erfüllt ist:
Ist $(\widetilde{\Omega}\widetilde{P}, \widetilde{\pi}_n)$ ein weiteres System mit (i) und (ii), so bilde man $\rho : \widetilde{\Omega} \to \mathbb{N}$ mit

$$\rho(\widetilde{\omega}) := \lim_{n\to\infty} \widetilde{\pi}_n(\widetilde{\omega}).$$

Man beachte, daß die Folge $\left(\widetilde{\pi}_n(\widetilde{\omega})\right)_n$ moton wächst und schließlich konstant ist.

Anmerkung. In der folgenden Feststellung verwenden wir bereits einige Eigenschaften diskreter W-Räume, die erst im folgenden Kapiteln sauber definiert werden.

2.1.4 Bez. (Geometrische Verteilung)
(i) Die geometrische Verteilung auf $\mathbb{N}_0$ mit dem Parameter $p \in (0,1)$ ist gegeben durch

$$\mathcal{G}_p\{k\} := p(1-p)^k \quad \text{für } k \in \mathbb{N}_0.$$

(ii) Eine Zufallsvariable X mit Werten in $\mathbb{N}$ ist geometrisch verteilt mit Parameter p, $(0 < p < 1)$, wenn ihre Bildverteilung auf $\mathbb{N}$ die geometrische Verteilung $\mathcal{G}_p$ ist:

$$P\{X = k\} = p(1-p)^{k-1} \quad \text{für } k \in \mathbb{N}_0.$$

2.2 Diskrete W-Räume

2.2.1 Def. (diskreter W-Raum)
Es sei Ω eine beliebige Menge und $\{\omega_n \mid n \in \mathbb{N}\}$ eine abzählbare Teilmenge von Ω. Jedem ω_n sei eine Wahrscheinlichkeit p_n zugeordnet, derart daß

$$0 \le p_n \le 1 \quad \textit{für } n \in \mathbb{N}, \tag{2.2.1}$$

$$\sum_{n=1}^{\infty} p_n = 1. \tag{2.2.2}$$

Man definiert dann die Wahrscheinlichkeit einer Teilmenge $A \in 2^\Omega$, indem man die Wahrscheinlichkeiten der $\omega_n \in A$ aufaddiert.

$$P(A) := \sum_{n=1}^{\infty} p_n \mathbb{1}_A(\omega_n). \tag{2.2.3}$$

*Man nennt dann $P : 2^\Omega \to [0,1]$ ein diskretes W-Maß auf Ω, daß auf die abzählbare Menge $\{\omega_n \mid n \in \mathbb{N}\}$ **konzentriert** ist.*

Anmerkung. (Diskrete W-Räume) 1. Eine konvergente Reihe mit positiven Summanden kann man beliebig umsortieren. Es kommt also auf die Reihenfolge der Punkte ω_n nicht an. Man schreibt daher auch

$$\sum_{n\in\mathbb{N}} p_n := \sum_{n=1}^{\infty} p_n. \tag{2.2.4}$$

Da es auf die Reihenfolge nicht ankommt, kann man auch andere abzählbare Indexmengen $I = \{i_n \mid n \in \mathbb{N}\}$ zulassen. Für die Summe schreibt man dann $\sum_{i\in I} p_i := \sum_{n=1}^{\infty} p_{i_n}$.

2. In den theoretischen Aussagen werden wir die abzählbar vielen ω_n immer mit $n = 1, 2, \dots$ durchnummerieren. Es können aber auch andere abzählbare Indexmengen auftreten, wie

- die ganzen Zahlen Z. Diese kann man folgendermaßen abzählen: $\mathbb{Z} = 0, 1, -1, 2, -2, 3, -3, \dots$. Man schreibt in diesem Fall für die Summe der Wahrscheinlichkeiten $\sum_{n=-\infty}^{\infty} p_n$.
- doppelt indizierte Punkte $\omega_{m,n}$ mit $m, n \in \mathbb{N}$. Mit dem Cantorschen Diagonalverfahren kann man die Punkte folgendermaßen zeilenweise durchzählen:

$$\begin{array}{lllll} \omega_{11} \\ \omega_{12} & \omega_{21} \\ \omega_{13} & \omega_{22} & \omega_{31} \\ \omega_{14} & \omega_{23} & \omega_{32} & \omega_{41} \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

- Allgemeiner gilt: Wenn I, J abzählbar sind, so ist auch $I \times J$ abzählbar.

3. Häufig wird $\Omega = \{\omega_n \mid n \in \mathbb{N}\}$ sein. Es wird sich als praktisch erweisen, auch solche Fälle zuzulassen, wie $\Omega = \mathbb{R}$ und $\{\omega_1, \omega_2, \dots\} = \mathbb{Z}$.

4. Eigentlich gehören die endlichen W-Räume ebenfalls zu den diskreten W-Räumen. Man sollte die Definition 2.2.1 eigentlich so formulieren:

2.2.2 Def. (alternative Definition: diskrete W-Räume)
Es sei Ω eine beliebige Menge, I eine höchstens abzählbar unendliche Indexmenge und $\{\omega_i \mid n \in I\}$ eine Teilmenge von Ω. Jedem ω_i sei eine Wahrscheinlichkeit p_i zugeordnet, derart daß

$$0 \leq p_i \leq 1 \quad \text{für } i \in I, \tag{2.2.5}$$

$$\sum_{i \in I} p_i = 1. \tag{2.2.6}$$

Man definiert dann die Wahrscheinlichkeit einer Teilmenge $A \in 2^\Omega$, indem man die Wahrscheinlichkeiten der $\omega_i \in A$ aufaddiert.

$$P(A) := \sum_{i \in I \cap A} p_i \mathbb{1}_A(\omega_i). \tag{2.2.7}$$

Man nennt dann $P : 2^\Omega \to [0,1]$ ein diskretes W-Maß auf Ω.

Konvention: Wir werden i. a. die einfachere Formulierung der Definition 2.2.1 benutzen und stillschweigens bemerken, daß alle Folgerungen natürlich auch für endliche W-Räume gelten.

Anmerkung. (diskretes W-Maß ist σ-additiv) Ein diskreter W-Raum erfüllt die beiden Axiome (i) und (ii) aus der Definition 1.1.5 eines endlichen W-Raumes. In einem unendlichen diskreten W-Raum (Ω, P) kann man abzählbare paarweise disjunkte Teilmengen $A_n \in 2^\Omega$, $(n \in \mathbb{N})$, betrachten und die Aussage in Feststellung 1.1.6 Gleichung (1.1.1) über die Additivität von P verschärfen.

2.2.3 Satz (diskretes W-Maß ist σ-additiv)

Es sei (Ω, P) ein diskreter W-Raum. Dann gilt für jede Familie $(A_n)_{n \in \mathbb{N}}$ paarweise disjunkter Mengen $A_n \in 2^\Omega$

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n). \tag{2.2.8}$$

Diese Eigenschaft heißt σ-Additivität.

Anmerkung. Mit dem Zusatz σ Bezeichnet man kurz Regeln über abzählbare Vereinigungen oder Summen. Ebenso ist das Kürzel δ für abzählbare Durchschnitte oder Prodkte im Gebrauch.

Beweis. P sei konzentriert auf die Menge $\{\omega_i \mid i \in \mathbb{N}\}$.

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{i=1} p_i \mathbb{1}_{\bigcup_{n=1}^{\infty} A_n}(\omega_i)$$

Da die A_n paarweise disjunkt sind, gilt

$$= \sum_{i=1} p_i \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega_i)$$

wobei in der Summe $\sum\limits_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega_i)$ alle Summanden bis auf höchstens einen gleich 0 sind. Da alle Terme nichtnegativ sind, darf man die Summationsreihenfolge vertauschen:

$$\begin{aligned} P\Big(\bigcup_{n=1}^{\infty} A_n\Big) &= \sum_{n=1}^{\infty} \sum_{i=1} p_i \mathbb{1}_{A_n}(\omega_i) \\ &= \sum_{n=1}^{\infty} P(A_n). \end{aligned}$$

2.2.4 Festst. (Resultate aus Kap. 1.)

1. *Für einen abzählbar unendlichen diskreten W-Raum (Ω, P), wobei P auf $\{\omega_n \mid n \in \mathbb{N}$ konzentriert ist, gelten alle Definitionen und Ergebnisse, die wir in Abschnitt 1 getroffen bzw. hergeleitet haben sinngemäß weiter, da für ihre Herleitung nur endliche Operationen – Vereinigung, Durchschnitt, Komplement von Mengen, Summe, Produkt, Quotient von Zahlen – verwendet werden; diese gelten mit der Definition 2.2.1, Gleichung (2.2.3) weiter.*

2. *Bei der Bildung des Erwartungswertes einer reellen Zufallsvariablen $X : (\Omega, P) \to \mathbb{R}$ muß man immer die Zusatzvoraussetzung machen, daß die Reihe*

$$E(|X|) := \sum_{n=1}^{\infty} |X(\omega_n)| P\{\omega_n\} \tag{2.2.9}$$

absolut konvergent *ist. Wir schreiben hierfür kurz $E(|X|) < \infty$.*

Dann konvergiert auch die Reihe

$$E(X) := \sum_{n=1}^{\infty} X(\omega_n) P\{\omega_n\} \tag{2.2.10}$$

und die Summanden dürfen ***beliebig umsortiert*** *werden. D.h., man kann mit absolutkonvergenten Reihen* bedenkenlos *rechnen.*

3. *Wenn die Varianz auftritt, muß vorausgesetzt werden, daß $E(|X|^2) < \infty$ ist. Aus der Schwarzschen Ungleichung folgt*

$$E(|X|) = E(|1 \cdot X|) \leq \sqrt{E(1)}\sqrt{E(|X|^2)} < \infty.$$

Dann ist auch

$$\mathrm{Var}(X) = E(X^2) - \big(E(X)\big)^2 < \infty.$$

Die Kovarianz bildet man nur für reelle Zufallsvariablen X, Y, für die

$$E(|X|^2) < \infty \quad \text{und} \quad E|(|Y^2) < \infty$$

ist.

2.2.5 Festst. (Erwartungswert der geomtr. V)

Für eine geometrisch verteilte Zufallsvariable gilt:

(i) *Der Erwartungswert ist*

$$E(X) = \frac{1-p}{p} = \frac{1}{p} - 1.$$

(ii) *Die Varianz ist*

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

Anmerkung. Beim wiederholten Würfeln mit einem fairen Würfel hat man also *im Mittel*

$$E(X) = \Big(\frac{1}{6}\Big)^{-1} - 1 = 5$$

Fehlversuche, bis die erste 6 fällt. Die Varianz von X ist

$$\mathrm{Var}(X) = \frac{5}{6}\Big(\frac{1}{6}\Big)^{-2} = 30.$$

Die Standardabweichung ist $\sigma(X) = \sqrt{30}$.

Beweis. (i) Da man Potenzreihen innerhalb Ihres Konvergenzbereiches gliedweise differenzieren kann, gilt

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} kp(1-p)^k \\ &= -p(1-p)\frac{d}{dp}\sum_{k=0}^{\infty}(1-p)^k \\ &= -p(1-p)\frac{d}{dp}\frac{1}{1-(1-p)} = p(1-p)\frac{1}{p^2} \\ &= \frac{1}{p} - 1. \end{aligned}$$

(ii) Aus Feststellung 1.18.3 (ii) folgt

$$\mathrm{Var}(X) = E\big(X(X-1)\big) + E(X) - \big(E(X)\big)^2.$$

$$\begin{aligned} E\big(X(X-1)\big) &= \sum_{k=0}^{\infty} k(k-1)p(1-p)^k \\ &= p(1-p)^2 \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} \\ &= p(1-p)^2 \frac{d^2}{dp^2} \sum_{k=0}^{\infty} (1-p)^k \\ &= p(1-p)^2 \frac{d^2}{dp^2} \frac{1}{1-(1-p)} \\ &= p(1-p)^2 \frac{2}{p^3} \\ &= \frac{2(1-p)^2}{p^2} \end{aligned}$$

Somit ist

$$\mathrm{Var}(X) = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \frac{(1-p)^2}{p^2} = \frac{1-p}{p^2}.$$

Bsp. (St. Peterburger Paradoxon) Das Rechnen und argumentieren mit Zufallsvariablen X, für die $E(|X|) = \infty$ ist, führt leicht zu paradoxen Ergebnissen. Hier ist ein Beispiel, das auf einen Artikel zurückgeht, den J. Bernoulli 1738 in der Zeitschrift der *St. Petersburger Akademie* publizierte.

In einem Glückspiel wird eine faire Münze geworfen. Fäll Kopf, so erhält der Spieler den doppelten Einsatz, anderenfalls ist der Einsatz verloren. Ein Spieler mit unbegrenztem Kapital (und Zeit) entscheidet sich für die folgende Strategie:

Er setzt beim ersten Spiel 1 Euro; gewinnt er, so hört er auf zu spielen. Anderenfalls verdoppelt er seinen Einsatz. Dies macht er so lange, bis er das erste mal gewinnt.

Gewinnt er beim n-ten Spiel, so hat er bis dahin

$$\sum_{\nu=1}^{n} 2^{\nu} = 2^n - 1$$

Euro gesetzt und 2^n Euro gewonnen. Der Reingewinn ist 1 Euro. die Wahrscheinlichkeit, daß er erstmals im n-ten Spiel gewinnt ist

$$\mathcal{G}_{1/2}\{n-1\} = \frac{1}{2}\Big(\frac{1}{2}\Big)^{n-1} = 2^{-n}.$$

Die Wahrscheinlichkeit, daß er nie gewinnt ist also kleiner als 2^{-n} für alle $(n = 1, 2, \dots)$, d.h. gleich 0.

Obwohl das Spiel fair ist, liefert die Methode mit Sicherheit den Reingewinn 1 Euro!

Das Problem liegt darin begründet, daß bei beliebig langer Spielzeit der Gewinn – ohne Berücksichtigung des Einsatzes – unendlichen Erwartungswert besitzt.

Der Grundraum $\mathbb{N}$ mit der geometrischen Wahrscheinlichkeit $\mathcal{G}_{1/2}$ beschreibt die Situation, daß der Gewinn im n-ten Spiel eintritt. Die Zufallsvariable

$$X : \mathbb{N} \to \mathbb{N} \quad \text{mit } X(n) = 2^n$$

beschreibt den Gewinn. X hat aber unendlichen Erwartungswert:

$$E(X) = \sum_{n=1}^{\infty} X(n)\mathcal{G}_{1/2}\{n\} = \sum_{n=1}^{\infty} 2^n 2^{-n} = \infty.$$

Ebenso hat das Eingesetzte Kapital $Y = X - 1$ unendlichen Erwartungswert.

In einem Spiel ist der Erwartungswert des Gewinnes, sofern er endlich ist, der faire Preis, den man zu Beginn des Spieles setzen kann. Ist der Erwartungswert unendlich, so gibt es keinen sinnvollen fairen Preis. In dem obigen Spiel müßte der Spieler vorab unendlich viel Kapital als Einsatz zahlen!.

Anmerkung. Das Produkt endlich vieler diskreter W-Räume ist wieder diskret:

2.2.6 Bez. (endl. Produkte diskreter WR.)

(i) Es seien (Ω_i, P_i), $(i = 1, \dots, k)$ endlich viele diskrete W-Räume. Die W-Maße seien jeweils auf die Mengen

$$M_i = \{\omega_n^{(i)} \mid n \in \mathbb{N}\}$$

konzentriert.

$$\Omega := \Omega_1 \times \cdots \times \Omega_k.$$

Man erklärt das Produktmaß $P := P_1 \otimes \cdots \otimes P_k$ auf

$$\Omega := \Omega_1 \times \cdots \times \Omega_k.$$

durch

$$P(A) = \sum_{(n_1,\dots,n_k)\in\mathbb{N}^k} p_{n_1} \cdots p_{n_k} \mathbb{1}_A(\omega_{n_1}^{(1)}, \dots, \omega_{n_k}^{(k)}) \tag{2.2.11}$$

$$= \sum_{n_1=1}^{\infty} \cdots \sum_{n_k=1}^{\infty} p_{n_1} \cdots p_{n_k} \mathbb{1}_A(\omega_{n_1}^{(1)}, \dots, \omega_{n_k}^{(k)}) \tag{2.2.12}$$

für $A \in 2^{\Omega}$.

(ii) Aus Gleichung (2.2.12) folgt die Produkteigenschaft:

$$P(A_1 \times \dots A_k) = P_1(A_1) \cdots P_k(A_k)$$

für $A_i \in 2^{\Omega_i}$, $(i = 1, \dots, k)$.

(iii) $M := M_1 \times \cdots \times M_k$ eine abzählbare Teilmenge von $\Omega_1 \times \cdots \times \Omega_n$ und das Produktmaß ist auf M konzentriert.

(iv) (iv) Mit dieser Definition des Produktmaßes gelten alle Resultate aus Kapitel 1 unveränder fort.

2.2.7 Bez. (Bildverteilung einer ZV)

Es sei (Ω, P) ein diskreter W-Raum und P sei konzentriert auf die abzählbare Menge $\{\omega_n \mid n \in \mathbb{N}\}$. Für eine Zufallsvariable $X : (\Omega, P) \to M$ ist dann das Bildmaß

$$P_X(B) := P\{X \in B\} \quad \text{für } B \in 2^M$$

auf die **höchsten abzählbare** Menge $\{\mathcal{X} := X(\omega_n) \mid n \in \mathbb{N}\}$ konzentriert; diese kann auch endlich sein. Also ist (M, P_X) wieder ein diskreter W-Raum.

2.2.8 Folg. (Bildverteilungen)

Mit dieser Definition des Bildmaßes (→ Bezeichnung 2.2.7) gelten alle Resultate aus Kapitel 1 weiter. Insbesondere gilt:

(i) *Endlich viele Zufallsvariable $X = (X_1, \dots, X_n)$ sind unabhängig, wenn ihre gemeinsame Verteilung P_X gleich dem Produkt ihrer Randverteilungen ist:*

$$P_X = P_{X_1} \otimes \cdots \otimes P_{X_n}.$$

(ii) *Für den Erwartungswert einer reellen Zufallsvariablen $(\Omega, P) \to \mathbb{R}$ mit $E(|X|) < \infty$ gilt*

$$E(X) = \sum_{x\in\mathcal{X}} x P_X\{x\}$$

Anmerkung. Daß das Bild eines diskreten W-Raumes wieder ein diskreter W-Raum ist, sieht sehr ansprechend aus, schränkt aber die Anwendbarkeit der diskreten W-Räume zur Modellbildung stark ein:

Es gibt auf einem diskreten W-Raum keine abzählbare Folge $(X_n)_n$ identisch verteilter unabhängiger Zufallsvariabler.

Wir wollen das Begründung kurz skizzieren: Annahme, $(X_n)_n$ sei eine abzählbare Folge identisch verteilter unabhängiger Zufallsvariabler. Man wähle $B_n \in X_n(\Omega)$ so, daß $P\{X_n\}(B_n) = p \in (0,1)$ für $n \in \mathbb{N}$. Die Zufallsvariablen $\mathbb{1}_{X_n^{-1}(B_n)}$ bilden eine abzählbare Folge von unabhängigen Bernoulli-Variablen mit Parameter $p \in (0,1)$.

Ähnlich wie im Beispiel 2.1.2 des unendlichen Münzwurfs zeigt man, daß der projektive Limes dieses System das Intervall $[0,1)$ mit der uniformen Verteilung $\mathfrak{U}$ ist. Der Raum $([0,1), \mathfrak{U})$ ist aber kein diskreter W-Raum.

2.3 Konvergenz gegen geomtr. Vertlng.

Anmerkung. Wir interessieren uns für konvergente Folgen diskreter W-Maße, die alle auf dieselbe abzählbare menge konzentriert sind. Dann ist der Grenzwert wieder ein diskretes W-Maß, In diesem Abschnitt betrachten wir ein relativ einfaches und anschauliches Beispiel, in dem der Grenzwert die geometrische Verteilung ist. In den folgenden Abschnitten untersuchen wir dann die Konvergenz der Binomialverteilung bei wachsendem n und zeigen:

Bei geeigneter Parameterwahl konvergiert die Binomialverteilung $\mathcal{B}_{n,p}$ einerseits gegen die

→ *Poisson-Verteilung auf $\mathbb{N}$, also wieder eine diskrete Verteilung.*

und bei anderer Wahl der Parameter gegen die

→ *Normalverteilung auf $\mathbb{R}$, also eine Verteilung mit stetiger Dichte.*

2.3.1 Bsp. (Konvergenz gegen die geom. V.)

Wir betrachten zunächst Zufallsexperimente mit einem endlichen Ergebnisraum. Aus einer Urne mit K schwarzen Kugeln und $N-K$ roten Kugeln wird eine geordnete Stichprobe $\omega := (\omega_1, \dots, \omega_n)$ vom Umfang N gezogen. Es werden also alle Kugeln gezogen. Der Raum Ω_{ord} der geordneten Stichproben sei mit Laplace-Wahrscheinlichkeit versehen. Es gibt $\binom{N}{K}$ mögliche Plätze für die schwarzen Kugeln, die roten kommen auf die restlichen Plätze. Also ist:

$$|\Omega_{\text{ord}}| = \binom{N}{K}.$$

Die Zufallsvariable $X : \Omega_{\text{ord}} \to \mathbb{N}$ gibt an, wieviele rote Kugeln gezogen werden, bevor die erste schwarze kommt:

$$X : \omega \mapsto \max\{i \mid 0 \le i \le N-K,\ \omega_1, \dots, \omega_i \text{ rot}\}$$

Wenn die erste schwarze Kugel im $k+1$-ten Zug erschien, sind bereits k rote Kugeln und eine schwarze gezogen worden. Es verbleiben also noch $\binom{N-k-1}{K-1}$ mögliche Plätze für die verbleibenden $K-1$ schwarzen Kugeln. Also ist

$$P_{N,K}\{X = k\} = \frac{\binom{N-k-1}{K-1}}{\binom{N}{K}} \quad \text{für } k = 0, \dots N-K.$$

Wir betrachten den Raum $\mathbb{N}_0$ mit dem diskreten W-Maß $P_{N,K}$. Es gilt also

$$\begin{aligned} &P_{N,K}\{X = k\} \\ &\quad = \frac{(N-k-1)!}{(K-1)!(N-K-k)!} \frac{K!(N-K)!}{N!} \\ &\quad = \frac{K}{N-k} \prod_{i=0}^{k-1} \frac{N-K-i}{N-i} \end{aligned} \tag{2.3.1}$$

für $k = 0, \dots N-K$ und 0 sonst.

Man kann die Gleichung (2.3.1) auch direkt interpretieren, als die Wahrscheinlichkeit, zunächst k rote Kugeln zu ziehen und dann eine schwarze.

Grenzübergang: Wie verhält sich die Verteilung $P_{N,K}$ wenn man immer mehr Kugeln in der Urne hat?

Man führe einen Grenzübergang $N, K \to \infty$ derart aus, daß der Anteil der schwarzen Kugeln $\frac{K}{N} \to p \in (0,1)$ konvergiert. D.h. man wähle eine Folge $(K_N)_N$ so daß

$$\lim_{N\to\infty} \frac{K_N}{N} = p \quad \text{mit } 0 < p < 1.$$

Aus der Gleichung (2.3.1) erhält man in diesem Fall

$$\lim_{N\to\infty} P_{N,K_N}\{k\} = p(1-p)^k.$$

Der Grenzwert ist die geometrische Wahrscheinlichkeit zum Parameter p:

$$\lim_{N\to\infty} P_{N.K} = \mathcal{G}_p \tag{2.3.2}$$

Das Ergebnis ist anschauulich zu verstehen. Bei festem k wähle man N und $K_N \approx pN$ sehr groß. Dann ändert sich beim Ziehen von k Kugeln aus der Urne der Restbestand praktisch nicht und folglich sind die Wahrschenlichkeiten für die Farbe der nächsten Kugeln annähernd konstant p bzw. $1-p$. Die Wahrscheinlichkeit, erst k rote und dann eine schwarze Kugel zu ziehen ist also $\approx (1-p)^k p$.

Anmerkung. (Bose-Einstein-Statistik) Man kann das W-Maß in dem obigen Beispiel 2.3.1 auch als die Verteilung der Anzahl der Objekte in einer Zelle bei einer Belegung mit nichtunterscheidbaen Objekten interpretieren (→ Bezeichnung 1.2.5).

R rote Kugeln werden durch einen Zufallsmechanismus auf $K+1$ Zellen mit den Nummern $0, 1, \dots, K$ verteilt. Als Zufallsmechanismus verwende man ein Urne mit K schwarzen Kugeln und $R = N-K$ roten Kugeln, aus der man alle Kugeln der Reihe nach entnimmt. Man hat also eine geordnete Stichprobe von roten und schwarzen Kugeln. Die roten Kugeln werden dann nach folgender Regel auf die $K+1$ Zellen verteilt:

Begint die Stichprobe mit einer roten Kugel, so kommt diese und die folgenden roten Kugeln solange in die Zelle mit der Nummer 0, bis eine schwarze Kugel gezogen wird, andernfalls bleibt die Zelle mit der Nummer 0 leer. Die schwarzen Kugeln bilden nun die Trennwände zwischen den folgenden Zellen mit den Nummern $1, \dots, K-1$. Die roten Kugeln zwischen der i-ten und der $i+1$-ten schwarzen Kugel kommen in die Zelle mit der Nummer i. Die roten Kugeln, die auf die K-te schwarze Kugel folgen, kommen in die letzte Zelle mit der Nummer K.

Ein Beispiel sagt mehr als tausend Worte: Es sei $K = 5$, $R = 6$. Die Stichprobe

$$r, r, s, s, r, r, r, s, s, r, s$$

ergibt die Belegungszahlen $2, 0, 3, 0, 1, 0$.

Da alle $\binom{R+K}{R}$ Anordnungen von R roten und K schwarzen Kugeln gleichwahrscheinlich sind, bewirkt dieser Zufallsmechanismus, daß alle so erzeugten Belegungen der Zellen gleichwahrscheinlich sind. Die Zufallsvariable X mit der durch die Gleichung (2.3.1) gegebene Verteilung gibt die Anzahl der roten Kugeln in der Zelle mit der Nummer 0 an.

$$P_{R,K}\{X = k\} := \frac{K}{R+K-k} \prod_{i=0}^{k-1} \frac{R-i}{R+K-i} \quad \text{für } k = 0, \dots, R.$$

Für $R, K \to \infty$ mit $\frac{R}{K} \to \lambda$ ist

$$\lim_{R,K\to\infty} P_{R,K}\{k\} = \frac{1}{\lambda+1} \Big(\frac{\lambda}{1+\lambda}\Big)^k = \mathcal{G}_p\{k\}, \tag{2.3.3}$$

wobei $p = \frac{\lambda}{1+\lambda}$ und λ der Erwartungswert von $\mathcal{G}_p$ ist.

Aus Symmetriegründen hat die Anzahl der Kugeln in jeder anderen Zelle die gleiche Verteilung wie X und ist also aprroximativ geometrisch verteilt.

$\binom{R+K}{R}$ ist die Anzahl Belegungen von $K+1$ Zellen mit R nicht unterscheidbaren Objekten (→ Folgerung 1.2.6). In der Physik heißt diese Laplace-Wahrscheinlichkeit auf den Belegungen die **Bose-Einstein Statistik**. Man findet sie in der Quantenmechanik bei der Beschreibung der sogenannten Bose-Teilchen. Elementarteilchen von diesem Typ sind **ununterscheidbar**, d.h. der Austausch zweier Teilchen ändert nichts am Zustand des Systems.

Es können mehrere Bose-Teilchen in der selben Zelle sein oder physikalisch ausgedrückt, denselben Zustand haben. Ein anderer Typ sind die Fermi-Dirac-Teilchen, die ebenfalls ununterscheidbar sind, von denen aber immer höchsten eines in einer Zelle sein kann. Die Gleichverteilung auf den möglichen Belegungen impliziert also, das die die Zahl der Teilchen in der gleichen Zelle annähernd geometrisch verteilt ist.

Anmerkung. Im allgemeinen hat man bei Belegungenmit mit unterscheidbaren Objekten nicht die Gleichverteilung (Bose-Einstein-Statistik), sondern die Multinomialverteilung. Die Anzahl der Objekte in einer Zelle ist dann die Randverteilung, also binomialverteilt (→ Beispiel 1.8.5).

Wir werden analog zum Vorgehen im Beispiel 2.3.1 für Zufallsvariable X_n, die $\mathcal{B}_{n,p}$ verteilt sind, den Grenzwert der Verteilung für $n \to \infty$ unter der Nebenbedingung $E(X_n) \to \lambda$ untersuchen. Dabei stoßen wir die auf die → Poissonverteilung auf $\mathbb{N}$.

2.4 Poisson-Verteilung

Anmerkung. Die Exponentialfunktion e^x gilt

$$e^x = \lim_{n\to\infty} \Big(1 + \frac{x}{n}\Big)^n \quad \text{für } x \in \mathbb{R}. \tag{2.4.1}$$

e^x hat die Potenzreihenentwicklung

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad \text{für } x \in \mathbb{R}. \tag{2.4.2}$$

Für $\lambda > 0$ ist $e^{-\lambda}\frac{\lambda^n}{n!} > 0$ und

$$\sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!} = 1. \tag{2.4.3}$$

Man kann also mit diesen Werten ein diskretes W-Maß auf $\mathbb{N}_0$ bilden.

2.4.1 Bez. (Poisson-Verteilung)

1. Für $\lambda > 0$ heißt das Wahrscheinlichkeitsmaß $\mathcal{P}_\lambda$ auf $\mathbb{N}_0$ mit

$$\mathcal{P}_\lambda\{k\} = e^{-\lambda}\frac{\lambda^k}{k}$$

die Poisson-Verteilung zum Parameter λ

2. Eine Zufallsvariable X heißt Poisson-verteilt, wenn sie ihre Werte in $\mathbb{N}_0$ annnimt und

$$P\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k}$$

ist.

2.4.2 Festst. (Eigenschaften: Poisson-Verteilung)

Eine Poisson-verteilte Zufallsvariable X hat den Erwartungswert

$$E(X) = \lambda$$

und die Varianz

$$Var(X) = \lambda.$$

λ ist also zugleich der Erwartungswert und die Varianz ein $\mathcal{P}_\lambda$-verteilten Zufallsvariablen.

Beweis.

$$E(X) = \sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

Aus Feststellung 1.18.3 (ii) folgt

$$\mathrm{Var}(X) = E\big(X(X-1)\big) + E(X) - \big(E(X)\big)^2.$$

$$\begin{aligned} E\big(X(X-1)\big) &= \sum_{k=0}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} \\ &= \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2 \end{aligned}$$

und somit

$$\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

2.4.3 Satz (Konvergenz gegen Poisson-Verteilung)

Für eine Folge von Binomialverteilungen $\mathcal{B}_{n,p_n}$, $(n \in \mathbb{N})$, deren Mittelwerte $\mu_n := np_n$ konvergieren:

$$\lim_{n\to\infty} \mu_n =: \lambda \tag{2.4.4}$$

gilt

$$\lim_{n\to\infty} \mathcal{B}_{n,p_n}\{k\} = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{2.4.5}$$

Unter der Nebenbedingung (2.4.4) konvergiert die Binomialverteilung gegen die Poissonverteilung.

Beweis. Es ist

$$\begin{aligned}\mathcal{B}_{n,p_n}\{k\} &= \binom{n}{k} p_n^k (1-p_n)^{n-k}\\ &= \frac{1}{k!} n(n-1)\cdots(n-k+1) p_n^k (1-p_n)^{n-k}\\ &= \frac{1}{k!}\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}\mu_n^k \big(1-\frac{\mu_n}{n}\big)^{n-k}\\ &= \frac{1}{k!}\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}\mu_n^k \big(1-\frac{\mu_n}{n}\big)^{-k}\big(1-\frac{\mu_n}{n}\big)^{n}\end{aligned}$$

Letzteres konvergiert für $n\to\infty$ gegen

$$\mathcal{B}_{n,p_n}\{k\} \to \frac{1}{k!}e^{-\lambda}.$$

Dabei haben wir das folgende Bemerkung 2.4.4 benutzt:

2.4.4 Bem. Man kann die Gleichung (2.4.1) noch etwas verschärfen: Für $0 \le \mu_n \to \lambda$ gilt

$$\big(1-\frac{\mu_n}{n}\big)^n \to e^{-\lambda}$$

Dies sieht man folgendermaßen ein: Da der natürliche Logarithmus $\log x$ die Ableitung $\frac{d}{dx}\log x = 1/x$ hat, folgt aus dem Mittelwertsatz der Differentialrechnung

$$\log\big(1-\frac{\mu_n}{n}\big) - \log 1 = -\frac{\mu_n}{n}\frac{1}{1-\xi_n}$$

mit $0<\xi_n<\mu_n/n$. Also gilt $\xi_n\to 0$ für $n\to\infty$. Da $\log 1 = 0$ ist folgt hieraus

$$\begin{aligned}\big(1-\frac{\mu_n}{n}\big)^n &= \exp\big(n\log\big(1-\frac{\mu_n}{n}\big)\big)\\ &= \exp\big(n\big(-\frac{\mu_n}{n}\frac{1}{1-\xi_n}\big)\big)\\ &\to \exp(-\lambda).\end{aligned}$$

Anmerkung. Die Approximation der Binomialverteilung $\mathcal{B}_{n,p}$ durch die Poisson-Verteilung $\mathcal{P}_\lambda$ mit $\lambda = np$ findet ihre Anwendung für große Werte von n und sehr kleine Werte von p. Das ist aus folgenden Gründen angebracht:

– Die Binomialkoeffizienten $\binom{n}{k}$ werden schnell sehr groß und die Werte der Binomialverteilung lassen sich für große n und kleine p nicht gut berechnen. Die Poisson-Verteilung kann gut rekursiv berechnen:

$$\mathcal{P}_\lambda(0) = e^{-\lambda}, \mathcal{P}_\lambda(k) = \frac{\lambda}{k}\mathcal{P}_\lambda k-1$$

– Manchmal sind n und p nicht genau bekannt, man weiß nur, daß n groß und p klein ist. Wenn man den Erwartungswert $\lambda = np$ schätzen schätzen, kann man die Verteilung trotzdem gut approximieren.

2.4.5 Bsp. (Verteilung der Schadensmeldungen)

Die Anzahl der Schadensmeldungen in einem Zeitraum $(0,t]$, $t>0$, sei proportional zur Länge des Intervalls $(0,t]$, also gleich αt mit einem empirisch zu ermittelnden Faktor α. Die Schadensmeldungen treffen zufällig ein, z.B. bei einer Versicherung. Wie ist die Verteilung X der Schadensmeldungen in einem festen Zeitraum $(0,t]$ anzusetzen?

Klar ist, X nimmt seine Werte in $\mathbb{N}_0$ an. Wir zerlegen das Intervall $(0,t]$ in n gleichlange Teilintervalle. Wenn n groß ist und somit die Teilintervalle kurz sind, ist anzunehmen, daß in jedem Teilintervall höchstens ein Schadenfall eintritt. Weiterhin nehmen wir an, daß das Auftreten eines Schadens in einem Teilintervall nicht davon abhängt, ob und in welchen anderen Teilintervallen Schäden auftreten. D.h., man kann so tun, als ob die Schadensfälle in den einzelnen gleichlangen Teilintervallen n-fach Bernoulli-verteilt mit dem Parameter $p=\alpha t/n$ sind Die Zufallsvariable X ist die Summe der *„Erfolge“*. und folglich annähernd binomialverteilt mit den Parametern n und $p\approx \alpha t/n$.

Diese Überlegung kann man für alle hinreichen großen $n\in\mathbb{N}$ anstellen. Das liefert den Ansatz für die Verteilung von X:

$$\begin{aligned}P\{X=k\} &= \lim_{n\to\infty}\mathcal{B}_{n,\alpha t/n}\{k\}\\ &= \mathcal{P}_{\alpha t}\{k\} = e^{-\alpha t}\frac{(\alpha t)^k}{k!} \quad \text{für } k\in\mathbb{N}_0.\end{aligned} \tag{2.4.6}$$

für $k\in\mathbb{N}_0$.

Anmerkung. Wir halten fest, die Poisson-Verteilung $\mathcal{P}_\lambda$ ist ein Modell für die Anzahl rein zufällig auftretender Zeitpunkte in einem Zeitintervall. Man denke an die Zeitpunkte, in denen im Beispiel 2.4.5) ein Schaden eintritt. Andere typische Beispiele sind die in einem Callcenter eingehenden Anrufe, die über einen Mail-Server weiterzuleitenden Emails, die Anzahl der Kunden an einem Schalter, die Anzahl der Zeitpunkte, in denen ein Atom einer radioaktiven Substanz zerfällt. Das Intervall kann auch eine andere Dimension haben, z.B. die Anzahl der Fahrzeuge auf einem Straßenabschnitt.

Anmerkung. (zum Beispiel 2.4.5 **Schadensmeldungen)** In der obigen Diskussion haben wir darauf verzichtet, einen formalen W-Raum (Ω,P) für $X:(\Omega,P)\to\mathbb{N}_0$ anzugeben. Wir gehen mal davon aus, daß einen solchen Raum gibt und wollen die Herleitung der Gleichung (2.4.6) etwas präzisieren:

1. Man zerlegt das Intervall $(0,t]$ in n-Teile und bezeichnet mit $X_{n,j}$, $(j=1,\dots,n)$ die Anzahl der Schadensmeldungen im j-ten Teilintervall. Dann ist

$$X = X_{n,1}+\dots X_{n,n}.$$

Wir haben vorausgesetzt:

$$X_{n,1},\dots,X_{n,n} \quad \textbf{unabhängig und identisch verteilt}. \qquad (*)$$

2. Dann haben wir $X_{n,j}$ durch die Indikatorvariable $\mathbb{1}_{\{X_{n,j}\ge 1\}}$ ersetzt und X durch die Summe

$$S_n := \sum_{j=1}^n \mathbb{1}_{\{X_{n,j}\ge 1\}}.$$

angenähert. Da die Ereignisse $\{X_{n,j\}} > 1\}$ unabhängig sind und die gleiche Wahrscheinlichkeit $p_n := P\{X_{n,1}\ge 1\}$ besitzen, ist S_n eine $\mathcal{B}_{n,p_n}$ verteilte Zufallsvariable. Aus

$$\mathbb{1}_{\{X_{n,1}\ge 1\}} \le X_{n,1}$$

folgt

$$p_n = E(\mathbb{1}_{\{X_{n,1}\ge 1\}}) \le E(X_{n,1}) = \frac{\alpha t}{n}.$$

3. Wir haben dann mit *„munteren Worten“* $p_n\approx \alpha t/n$ gesetzt, d.h. wir setzen voraus, daß

$$\lim_{n\to\infty} np_n = \lim_{n\to\infty} E(S_n) \overset{!}{=} E(X) = \alpha t \qquad (**)$$

Hiermit folgt aus Satz 2.4.3

$$\lim_{n\to\infty} P\{S_n = k\} = \lim_{n\to\infty}\mathcal{B}_{n,p_n}\{k\} = \mathcal{P}_{\alpha t}. \tag{2.4.7}$$

Wir wollen nun schließen, daß X ebenfalls Poissson-verteilt mit dem Parameter αt ist.

4. Dazu zeigen wir die folgende Aussage:
Aus $(**)$ folgt

$$\lim_{n\to\infty} P\Big(\bigcup_{j=1}^n\{X_{n,j}\ge 2\}\Big) = 0. \tag{2.4.8}$$

D.h. bei immer feiner werdender Zerlegung wird das Auftreten von mehr als einem Schaden immer unwahrscheinlicher. Dies sieht man folgendermaßen ein:

Wir benutzen in der folgenden Gleichung die σ-Additivität des nicht genauer angegebenen Wahrscheinlichkeitsmaßes P. Im Falle diskreter W-Räume hatten wir die σ-Additivität in Satz 2.2.3 gezeigt.

$$\begin{aligned} E(S_n) &= \sum_{j=1}^{n} P\{\mathbb{1}_{\{X_{n,j}\geq 1\}} = \sum_{j=1}^{n}\sum_{k=1}^{\infty} P\{X_{n,j}=k\} \\ &\leq \sum_{j=1}^{n}\sum_{k=1}^{\infty} kP\{X_{n,j}=k\} \stackrel{!}{=} \sum_{j=1}^{n} E(X_{n.j}) = E(X). \end{aligned}$$

Aus $(**)$ folgt somit

$$\lim_{n\to\infty}\sum_{j=1}^{n}\sum_{k=2}^{\infty}(k-1)P\{X_{n,j}=k\} = 0.$$

Aus der σ-Additivität von P folgt nun

$$\begin{aligned} P\Big(\bigcup_{j=1}^{n}\{X_{n,j}\geq 2\}\Big) &\leq \sum_{j=1}^{n}\sum_{k=2}^{\infty} P\{X_{n,j}=k\} \\ &\leq \sum_{j=1}^{n}\sum_{k=2}^{\infty}(k-1)P\{X_{n,j}=k\} \to 0 \end{aligned}$$

und somit die die Gleichung (2.4.8).

5. Aus den Gleichungen (2.4.7) und (2.4.8) erhalten wir nun die Verteilung von X. Wir zerlegen dazu die Menge $\{X=k\}$ in die folgenden Fälle:

$$\begin{aligned} P\{X=k\} &= P(\{X=k\}\cap\{X=S_n\}) + P(\{X=k\}\cap\{X\neq S_n\}) \\ &= P(\{S_n=k\}\cap\{X=S_n\}) + P(\{X=k\}\cap\{X\neq S_n\}) \\ &= P\{S_n=k\} - P(\{S_n=k\}\cap\{X\neq S_n\}) \\ &\quad + P(\{X=k\}\cap\{X\neq S_n\}) \end{aligned}$$

Da

$$\{X\neq S_n\}\subset\bigcup_{j=1}^{n}\{X_{n,j}\geq 2\}$$

ist, folgt aus (2.4.8)

$$\begin{aligned} \lim_{n\to\infty} P(\{S_n=k\}\cap\{X\neq S_n\}) &= 0, \\ \lim_{n\to\infty} P(\{X=k\}\cap\{X\neq S_n\}) &= 0. \end{aligned}$$

Also ist $P\{X=k\} = \lim_{n\to\infty} P\{S_n=k\} = \mathcal{P}_{\alpha t}\{k\}$.

Anmerkung. Bei der Herleitung der Poisson-Verteilung in der obigen Anmerkung haben wir vorausgesetzt, daß sich X als Summe von n unabhängigen, identisch verteilten Zufallsvariablen schreiben läßt. Wir wollen zeigen, daß dies für eine Poisson-verteilte Zufallsvariable für jedes n einzeln möglich ist, indem wir X durch eine passende Zufallsvariable $\widetilde{X}$ mit der gleichen Verteilung ersetzen. Eigentlich benötigen wir einen W-Raum, auf dem dies für alle n zugleich möglich ist.

In der Übung wird gezeigt, daß für unabhängig Zufallsvariable X und Y, die Poisson-verteilt sind mit den Parametern λ bzw μ, die Summe $X+Y$ wieder Poisson-verteilt ist mit dem Parameter $\lambda+\mu$.

Es sei X Poisson-verteilt mit dem Parameter λ. Man wähle eine Zufallsvariable Y die Poisson-verteilt mit dem Parameter λ/n ist. Z.B. die identische Abbildung von $(\Omega,P) := (\mathbb{N}_0,\mathcal{P}_{\lambda/n})$.

Man bilde nun das n-fache Produkt $(\Omega^n, P^{\otimes n})$ mit den Projektionen $\mathrm{pr}_\nu : \Omega^n\to\Omega$ und setze

$$X_\nu = Y\circ \mathrm{pr}_\nu \quad \text{für } \nu = 1,\ldots,n.$$

Die X_ν sind unabhängig und $\mathcal{P}_{\lambda/n}$ verteilt. Also ist

$$\widetilde{X} := \sum_{\nu=1}^{n} X_\nu$$

$\mathcal{P}_\lambda$ verteilt.

Da X und $\widetilde{X}$ dieselbe Verteilung haben, kann man X durch $\widetilde{X}$ ersetzen. Die letztere Zufallsvariable ist die Summe von n identisch verteilten unabhängigen Zufallsvariablen.

2.5 Allgemeine W-Räume

Anmerkung. Wir geben die formale Definition eines W-Raumes. Die diskreten W-Räume ordnen sich hier als wichtiges Beispiel ein. Wir werden keine tieferliegenden Ergebnisse über allgemeine W-Räume herleiten, sondern benötigen die Definition im wesentlichen, um in den folgenden Beispielen die übliche Terminologie verwenden zu können.

Es erweist sich als zweckmäßig und vielfach aus Gründen der Logik als notwendig, daß man bei überabzählbaren Ereignisräumen, wie z.B. bei den reellen Zahlen $\mathbb{R}$, als Ereignisse nicht mehr alle Teilmengen des Grundraumes zuläßt, sondern sich auf eine kleineres Mengensystem beschränkt, daß alle *„interessierenden"* Ereignisse enthält. Im Fall der reellen Zahlen wird man z. B. die Wahrscheinlichkeit von Intervallen berechnen wollen.

Wie wir bei der Untersuchung diskreter W-Räume bereits gesehen haben, ist das wesentliche theoretische und rechnerische Hilfsmittel die Approximation durch einfachere Modelle. Der Übergang zu Grenzwerten vereinfacht einerseits die Theorie aber auch die praktische Berechnung. Der Grenzwert ist der ideale genaue Wert, mit den man jede Approximation vergleichen kann. Die Kenntnis des Grenzwertes ermöglicht es oft andere, bessere Näherungen zu finden, als die, mit denen man zunächst die Existenz des Grenzwertes und seine Eigenschaften gefunden hat.

Um Grenzwerte bilden zu können, müssen wir für die Ereignisse abzählbare Mengenoperationen (Vereinigung, Durchschnitt, Komplement) zulassen. Im Falle der reellen zahlen kommt man also nicht umhin, auch all die Teilmengen zu betrachten, die sich durch *„beliebig wiederholte"* abzählbare Mengenoperationen aus Intervallen bilden lassen. Man nennt dies das System der *Borel-Mengen* $\mathcal{B}(\mathbb{R})$ auf $\mathbb{R}$. Es wird sehr aufwendig und ist auch nicht nötig, dieses Mengensystem durch Konstruktionsvorschriften zu beschreiben. Man charakterisiert die Borel-Mengen $\mathcal{B}(\mathbb{R})$ als das kleinste Mengensystem, das alle Intervalle enthält und aus dem abzählbaren Mengenoperationen nicht hinausführen.

Um diese Gedanken zu präzisieren gibt man eine axiomatische Charakterisierung und gelangt so zum Begriff der σ-Algebra:

2.5.1 Def. (σ-Algebra)

Es sei Ω eine Menge. eine Teilmenge $\mathcal{A} \subset 2^\Omega$ heißt ein Mengensystem auf Ω.

Ein Mengensystem $\mathcal{A}$ heißt eine σ-Algebra, wenn folgendes gilt:

1. $\Omega \in \mathcal{A}$

2. *Mit $A \in \mathcal{A}$ ist auch das komplementäre Menge $A^c \in \mathcal{A}$.*

3. *Ist $(A)_{n\in\mathbb{N}}$ in $\mathcal{A}$, so ist auch*

$$\bigcup_{n=1}^\infty A_n \in \mathcal{A}.$$

*In der W-Theorie nennt das Paar $(\Omega, \mathcal{A})$ einen **Ereignisraum** und die Elemente von $\mathcal{A}$ Ereignisse. In der Maßtheorie sagt man **Meßraum** oder **meßbarer Raum** dazu.*

Anmerkung.

1. Man beachte, es nicht gefordert wird, daß eine σ-Algebra für *beliebige Familien* $(A_i)_{i\in I}$ die Vereingungen $\bigcup_{i\in I} A_i$ enthält. Diese Eigenschaft ist i.a. auch falsch.

Ein einfaches Beispiel ist die – sonst nicht weiter interessante – σ-Algebra

$$\mathcal{A} := \{A \subset \mathbb{R} \mid A \text{ oder } A^c \text{ höchstens abzählbar}\}.$$

Es sind alle einpunktigen Mengen $\{a\} \in \mathcal{A}$. Jede Menge ist Vereinigung ihrer Punkte. Da $\mathbb{R}$ überabzählbar ist, ist $\mathcal{A} \neq 2^\mathbb{R}$, z.B. ist das Intervall $[0, \infty) \notin \mathcal{A}$.

2. In den Axiomen einer σ-Algebra steht auch nicht, daß die einpunktigen Mengen zu $\mathcal{A}$ gehören. Das wird aber häufig der Fall sein.

2.5.2 Festst. (Rechenregeln: σ-Algebren)

Es sei $\mathcal{A}$ eine σ-Algebra auf der Menge Ω.

(i) *Es ist $\emptyset = \Omega^c \in \mathcal{A}$.*

(ii) *Sind $A_1, \dots, A_n \in \mathcal{A}$, so ergänze man sie durch $\emptyset = A_{n+1} = A_{n+2} = \dots$ zu einer Folge. Somit ist die endliche Vereinigung $A_1 \cup \dots \cup A_n = \bigcup_{n=1}^\infty A_n \in \mathcal{A}$.*

(iii) *Es sei $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$. Nach der de Morganschen Regel gilt:*

$$\bigcap_{n=1}^\infty A_n = \Big(\bigcup_{n=1}^\infty A_n^c\Big)^c \in \mathcal{A}.$$

(iv) *Analog zum Punkt* (ii) *ist mit $A_1, \dots, A_n \in \mathcal{A}$ auch der Durchschnitt $A_1 \cap \dots \cap A_n \in \mathcal{A}$.*

(v) *Ist $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ eine fallende Folge, d.h.*

$$A_1 \supset A_2 \cdots A_n \supset A_{n+1} \cdots,$$

dann kann man A_k als disjunkte Vereinigung schreiben:

$$A_k = \Big(\bigcap_{n=1}^\infty A_n\Big) \cup \bigcup_{n=k}^\infty (A_n \setminus A_{n+1}),$$

Das System $\bigcap_{j=1}^\infty A_j$ und $(A_n \setminus A_{n+1})_n$ besteht aus paarweise disjunkten Mengen. Ist $(A_n)_n$ eine beliebige Folge in $\mathcal{A}$, dann kann man die Vereinigung der A_n als disjunkte Vereinigung schreiben:

$$\bigcup_{n=k}^\infty A_n = \bigcup_{n=1}^\infty \Big(A_n \setminus \bigcup_{i=1}^{n-1} A_i\Big)$$

Also

$$\bigcup_{n=k}^\infty A_n = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2)) \cup \dots$$

2.5.3 Bsp. (σ-Algebren)

1. Für eine beliebige Menge ist die Potenzmenge 2^Ω eine σ-Algebra auf Ω. 2^Ω ist offensichtlich die **größte** σ-Algebra auf Ω. Sie ist i.a. viel zu groß.

Die **kleinste** σ-Algebra auf Ω ist $\{\Omega, \emptyset\}$.

2. Der Durchschnitt einer beliebigen Familie von σ-Algebren auf Ω ist eine σ-Algebra.

3. Wenn $\Omega = \{\omega_n \mid n \in \mathbb{N}\}$ höchstens abzählbar unendlich ist, so ist 2^Ω die kleinste σ-Algebra auf Ω, die alle einelementigen Teilmengen (Punkte) von Ω enthält. Aus diesem Grunde rechnet man bei diskreten W-Räumen immer mit der vollen Potenzmenge.

4. Zu $\mathcal{M} \subset 2^\Omega$ gibt es eine **kleinste $\mathcal{M}$ umfassende** σ-Algebra, die man mit $\sigma(\mathcal{M})$ bezeichnet.

Beweis. Die Potenzmenge 2^Ω ist eine σ-Algebra ist, die $\mathcal{M}$ enthält. Nun bilde man den Durchschnitt aller σ-Algebren, die $\mathcal{M}$ enthalten. Dieser Durchschnitt ist nicht leer, ist eine σ-Algebra, und enthält $\mathcal{M}$.

5. Ist insbesondere $\Omega = \mathbb{R}$, so bezeichnet man mit $\mathcal{B}(\mathbb{R})$ die kleinste σ-Algebra auf $\mathbb{R}$, die alle offenen Mengen enthält. $\mathcal{B}(\mathbb{R})$ heißt die **Borel-Algebra** von $\mathbb{R}$.

Jedes abgeschlossene oder halbabgeschlossene Intervall ist Durchschnitt von abzählbar vielen offenen Intervallen:

$$[a,b] = \bigcap_{n=1}^{\infty} (a - \tfrac{1}{n}, b + \tfrac{1}{n}),$$
$$[a,b) = \bigcap_{n=1}^{\infty} (a - \tfrac{1}{n}, b).$$

Weiterhin ist jede offene Menge in $\mathbb{R}$ die Vereinigung von höchstens abzählbar vielen offenen Intervallen. Daher ist $\mathcal{B}(\mathbb{R})$ auch die kleinste σ-Algebra die alle beschränkten Intervalle enthält.

$\mathcal{B}(\mathbb{R})$ ist auch die kleinste σ-Algebra, die alle Intervalle der Form $(-\infty, a]$ enthält.

Weiterhin ist jede einpunktige Menge $\{a\} = [a,a] \in \mathcal{B}(\mathbb{R})$ und somit auch jede abzählbare Menge in $\mathcal{B}(\mathbb{R})$).

Wenn man von dem Ereignisraum $\mathbb{R}$ spricht, ist immer $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ gemeint.

6. Für ein Intervall $I \subset \mathbb{R}$ sei

$$\mathcal{B}(I) := \{A \mid A \in \mathcal{B}(\mathbb{R}) \text{ und } A \subset I\}.$$

$\mathcal{B}(I)$ ist die kleinste σ-Algebra auf I, die alle Teilintervalle von I enthält.

7. Die Borel-Algebra $\mathcal{B}(\mathbb{R}^d)$ ist die kleinste σ-Algebra, die alle offenen Teilmengen des $\mathbb{R}^d$ enthält. Wie im Falle $n = 1$ zeigt man, daß $B(R^n)$ die kleinste σ-Algebra ist, die alle offenen oder alle halboffenen oder alle abgeschlossenen achsenparallelen Quader enthält.

Weiterhin ist $\mathcal{B}(\mathbb{R}^d)$ die kleinste σ-Algebra, die alle Mengen der Form

$$(-\infty, a_1] \times (-\infty, a_2] \times \cdots \times (-\infty, a_n]$$

enthält.

2.5.4 Bem. (Bild und Urbild einer σ-Algebra)

1. *Es seien $(\Omega, \mathcal{A})$ ein Meßraum, $\widetilde{\Omega}$ eine Menge und $X : \Omega \to \widetilde{\Omega}$ eine Abbildung. Dann ist*

$$\widetilde{\mathcal{A}} := \{B \in 2^{\widetilde{\Omega}} \mid X^{-1}(B) \in \mathcal{A}\}$$

eine σ-Algebra.

Man beachte, das das Mengensystem $\{X(A) \mid A \in \mathcal{A}\}$ i.a. keine σ-Algebra ist, selbst wenn X surjektiv ist, da i.a. $X(A)^c \neq X(A^c)$ ist.

2. *Es seien Ω eine Menge, $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ eine Meßraum und $X : \Omega \to \widetilde{\Omega}$ eine Abbildung. Dann ist*

$$X^{-1}(\mathcal{A}) := \{A \in 2^{\Omega} \mid A = X^{-1}(B) \text{ für ein } B \in \widetilde{\mathcal{A}}\}$$

eine σ-Algebra.

Anmerkung. Wir definieren nun den allgemeinste Version eines W-Raumes.

2.5.5 Def. (W-Maß auf einer σ-Algebra)

Es sei $(\Omega, \mathcal{A})$ ein Meßraum. Eine Abbildung $P : \mathcal{A} \to [0,1]$ heißt ein W-Maß, wenn die Axiome von Kolmogorov gelten:

1. $P(\Omega) = 1$.

2. *Für eine Folge $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkter Ereignisse in $\mathcal{A}$ gilt*

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n) \quad (\sigma\textbf{-additiv})$$

Das Tripel $(\Omega, \mathcal{A}, P)$ heißt ein ***W-Raum****.*

2.5.6 Satz (Rechenregeln: W-Maße)

Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum ($\to$ Definition 2.5.5). Dann gilt

(i) *Für zwei disjunkte Mengen $A, B \in \mathcal{A}$ ist*

$$P(A \cup B) = P(A) + P(B) \quad (\textbf{additiv})$$

Es gelten also alle Rechenregeln aus Feststellung 1.1.6 über endliche Mengenoperationen.

(ii) *Für eine beliebige Folge $(A_n)_n$ in $\mathcal{A}$ gilt*

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} P(A_n) \quad (\sigma\textbf{-subadditiv})$$

(iii) *Für eine monoton fallende Folge $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ gilt*

$$P\Big(\bigcap_{n=1}^{\infty}\Big) = \lim_{n\to\infty} P(A_n). \quad (\textbf{von oben stetig})$$

(iv) *Für eine monoton wachsende Folge $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ gilt*

$$P\Big(\bigcup_{n=1}^{\infty}\Big) = \lim_{n\to\infty} P(A_n). \quad (\textbf{von unten stetig})$$

2.5.7 Bsp. (Uniforme Verteilung)

1. Man zeigt in allen gängigen Lehrbüchern zur Analysis III, daß es genau eine W-Maß $\mathcal{U}$ auf den Borel-Mengen $\mathcal{B}([0,1])$ gibt, so das

$$\mathcal{U}[a,b] := b - a \quad \text{für } 0 \leq a \leq b < 1.$$

gilt. Aus Stetigkeit von oben bzw. unten folgt dann

$$\mathcal{U}[a,b) = \mathcal{U}(a,b] = \mathcal{U}(a,b) = \mathcal{U}[a,b]$$

Man nennt diese W-Maß die uniforme Verteilung auf $[0,1]$. In den Lehrbüchern zur Analysis heißt diese W-Maß das **Lebesgue-Maß** und wird mit λ bezeichnet.

2. Das Lebesgue-Maß λ existiert auch auf jedem anderen echten, beschränkten Intervall $I \subset \mathbb{R}$. Um die uniforme Verteilung auf I zu erhalten, muß man λ normieren. Man setzt

$$\mathcal{U}_I(A) := \frac{\lambda(A)}{\lambda(I)} \quad \text{für } A \in \mathcal{B}(I).$$

Ist $I = [a,b]$, so gilt also

$$\mathcal{U}_I([c,d]) = \frac{d-c}{b-a} \quad \text{für } a \leq c < d \leq b.$$

2.6 Zufallsvariable und ihre Verteilung

2.6.1 Def. (meßbare Abb, Zufallsvariable)

1. *Es seien $(\Omega, \mathcal{A})$ und $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ Meßräume. Eine Abbildung $X : \Omega \to \widetilde{\Omega}$ heißt* ***meßbar****, wenn*

$$X^{-1}(B) \in \mathcal{A} \quad \text{für } B \in \widetilde{\mathcal{A}}.$$

2. *Ist $(\Omega, \mathcal{A}, P)$ ein W-Raum und $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ ein Meßraum, dann heißt eine meßbare Abbildung $X : \Omega \to \widetilde{\Omega}$ eine Zufallsvariable.*

Wir schreiben hierfür kurz $X : (\Omega, \mathcal{A}, P) \to (\widetilde{\Omega}, \widetilde{\mathcal{A}})$.

Anmerkung. Für jede Zufallsvariable $X : (\Omega, \mathcal{A}, P) \to (\widetilde{\Omega}, \widetilde{\mathcal{A}})$ definiert jedes $B \in \widetilde{\mathcal{A}}$ ein Ereignis $\{X \in B\} = X^{-1}(B) \in \mathcal{A}$, dessen Wahrscheinlichkeit $P\{X \in B\}$ man bilden kann. Wie im endlichen und diskreten Fall ($\rightarrow$ Satz 1.6.3) heißt

$$P_X : B \mapsto P\{X \in B\} \quad \text{für } B \in \widetilde{\mathcal{A}}$$

die Bildverteilung von X.

2.6.2 Festst. (Regeln: Zufallsvariable)

(i) *Es seien $(\Omega, \mathcal{A}, P)$ ein W-Raum, $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ und $(\widehat{\Omega}, \widehat{\mathcal{A}})$ Meßräume. Ist $X : \Omega \to \widetilde{\Omega}$ eine Zufallsvariable und $\varphi : \widetilde{\Omega} \to \widehat{\Omega}$ meßbar, dann ist $\varphi \circ X$ wieder Meßbar, also eine Zufallsvariable.*

(ii) *Es seien $(\Omega, \mathcal{A}, P)$ ein W-Raum, $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ ein Meßraum und Ist $\mathcal{D}$ ein Erzeugendensystem von $\widetilde{\mathcal{A}}$, d.h., $\sigma(\mathcal{D}) = \mathcal{A}$.*

Eine Abbildung $X : \Omega \to \widetilde{\Omega}$ ist bereits eine Zufallsvariable, wenn

$$X^{-1}(D) \in \mathcal{A} \quad \text{für alle } D \in \mathcal{D} \text{ gilt.}$$

Beweis. (i) Nach Vorausetzung ist für alle $C \in \widehat{\mathcal{A}}$

$$(\varphi \circ X)^{-1}(C) = X^{-1}(\varphi^{-1}(C)) \in \mathcal{A}.$$

(ii) Das Mengensystem

$$\widehat{\mathcal{A}} := \{A \in 2^{\widetilde{\Omega}} \mid X^{-1}(A) \in \mathcal{A}\}$$

ist eine σ-Algebra, die $\mathcal{D}$ umfaßt. Also ist $\widetilde{\mathcal{A}} = \sigma(\mathcal{D}) \subset \widehat{\mathcal{A}}$.

2.6.3 Satz (Verteilung einer Zufallsvariablen)

Es seien $(\Omega, \mathcal{A}, P)$ ein W-Raum, $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ ein Meßraum und $X : \Omega \to \widetilde{\Omega}$ eine Zufallsvariable. Dann ist die Abbildung

$$P_X : \widetilde{\mathcal{A}}) \to [0, 1],$$
$$P_X : B \mapsto P(X^{-1}(B)) \quad \text{für } B \in \widetilde{\mathcal{A}}$$

ein W-Maß auf $\widetilde{\mathcal{A}}$.

P_X heißt die Verteilung von X oder das Bildmaß von P unter X. $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, P_X)$ heißt der von X induzierte W-Raum.

Beweis. Wir weisen die Eigenschaften aus Definition 2.5.5 nach. P_X ist eine Abbildung von $\widetilde{\mathcal{A}}$ in $[0, 1]$ mit $P_X(\widetilde{\Omega}) = 1$. Wir zeigen die σ-Additivität: Für eine Folge $(B_n)_n$ paarweise disjunkter Mengen in $\widetilde{\mathcal{A}}$ gilt

$$\begin{aligned} P_X\Big(\bigcup_{n=1}^{\infty} B_n\Big) &= P\Big(X^{-1}\big(\bigcup_{n=1}^{\infty} B_n\big)\Big) \\ &= P\Big(\bigcup_{n=1}^{\infty} X^{-1}(B_n)\Big) \\ &= \sum_{n=1}^{\infty} P(X^{-1}(B_n)) = \sum_{n=1}^{\infty} P_X(B_n). \end{aligned}$$

Man beachte, wenn B_1, B_2 disjunkt sind, so sind auch die Urbilder $X^{-1}(B_1)$ und $X^{-1}(B_2)$ disjunkt.

Anmerkung. Die Feststellung 1.6.4 gilt sinngemäß weiter:

2.6.4 Bem. (Funktionen von Zufallsvariablen)

Gegeben seien ein W-Raum $(\Omega, \mathcal{A}, P)$, Meßräume $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ $(\widehat{\Omega}, \widehat{\mathcal{A}})$ und Zufallsvariable

$$X : \Omega \to \widehat{\Omega}, \quad Y : \Omega \to \widetilde{\Omega}$$

und eine meßbare Abbildung $Z : \widetilde{\Omega} \to \widehat{\Omega}$ so, daß

$$X = Z \circ Y$$

ist. Man bilde den W-Raum $(\widetilde{\Omega}, P_Y)$. Dann haben die Zufallsvariablen

$$X : (\Omega, P) \to \widehat{\Omega} \quad \text{und} \quad Z : (\widetilde{\Omega}, P_Y) \to \widehat{\Omega}$$

die gleiche Verteilung: $(\widehat{\Omega}, P_X) = (\widehat{\Omega}; P_Z)$ oder kurz

$$(P_Y)_Z = P_{Z \circ Y}. \tag{2.6.1}$$

2.7 Eindeutigkeit eines W-Maßes

Anmerkung. (Problem: W-Funktion zu $\mathcal{U}$ identisch 0)

1. Für die Uniforme Verteilung auf $[0,1]$ hat jede einpunktige Menge das Maß Null:

$$\mathcal{U}\{a\} = \lim_{n\to\infty} \mathcal{U}([a-\frac{1}{n}, a+\frac{1}{n}]) = 0.$$

Bei endlichen und diskreten W-Räumen (Ω, P) ist die W-Funktion $p : \omega \mapsto P\{\omega\}$ ein nützliches Hilfsmittel ($\to$ Satz 1.1.7). Im Falle der uniformen Verteilung ist die W-Funktion identisch 0 und nützt nichts bei der Berechnung von $\mathcal{U}$.

2. Da es i.a. nicht einfach ist, die Werte $P(A)$ eines W-Maßes P für alle $A \in \mathcal{A}$ zu berechnen, sucht man nach einem möglichst einfachen Mengensystem $\mathcal{D} \subset \mathcal{A}$, derart daß P durch seine Werte auf $\mathcal{D}$ eindeutig festgelegt ist. Das obige Beispiel zeigt, daß die einpunktigen Mengen dazu i.a. nicht geeignet sind.

Der folgende Satz aus der Maßtheorie, den wir ohne Beweis angeben, sagt, wie ein solches Mengensystem $\mathcal{D}$ beschaffen sein muß. Für einen Beweis siehe z.B. [2, Anhang M.4].

3. In den Anwendungen rechnet man die meiste Zeit mit der Wahrscheinlichkeit von Ereignissen aus $\mathcal{D}$. Aber ab und zu ist es doch wichtig, daß man nicht auf diese kleine Mengensystem beschränkt ist, sondern sich gedanklich in der von $\mathcal{D}$ erzeugten σ-Algebra frei bewegen kann.

2.7.1 Satz (Dynkin: Eindeutigkeit v. W-Maßen)

Es seien $(\Omega, \mathcal{A})$ ein Meßraum und P, Q zwei W-Maße auf $\mathcal{A}$. Ist $\mathcal{D} \subset \mathcal{A}$ ein $\bigcap$-stabiles Mengensystem und stimmen P und Q auf $\mathcal{D}$ überein, so stimmen sie auch auf der erzeugten σ-Algebra $\sigma(\mathcal{D})$ überein. Wenn also $\sigma(\mathcal{D}) = \mathcal{A}$ ist, so folgt $P = Q$.

Dabei bedeutet, $\mathcal{D}$ ist $\bigcap$-stabil:

$$D \cap E \in \mathcal{D} \quad \text{für alle } D, E \in \mathcal{D}.$$

2.7.2 Bsp. (zum Satz von Dynkin)

Das Mengensystem

$$\mathcal{D} := \{(-\infty, x_1] \times \cdots \times (-\infty, x_n] \mid x \in \mathbb{R}^d\}$$

ist $\bigcap$-stabil und erzeugt die Borel-Algebra $\mathcal{B}(\mathbb{R}^d)$. Satz 2.7.1 führt zu der Definition 2.8.1 der Verteilungsfunktion $F(x) := P((-\infty])$ eines W-Maßes auf $\mathbb{R}$.

Bsp. Sei $\Omega = \{1,2,3,4\}$ und $\mathcal{D} = \{\{1,2\},\{2,3\}\}$. Da

$$\begin{aligned}
\{1\} &= \{1,2\} \setminus \{2,3\},\\
\{2\} &= \{1,2\} \cap \{2,3\},\\
\{3\} &= \{2,3\} \setminus \{1,2\},\\
\{4\} &= \Omega \setminus (\{1,2\} \cup \{2,3\})
\end{aligned}$$

gilt, ist die von $\mathcal{D}$ erzeugte σ-Algebra gleich der Potenzmenge 2^Ω.

$\mathcal{D}$ ist nicht $\bigcap$-stabil. Hier sind zwei verschieden W-Maße auf Ω, die auf $\mathcal{D}$ übereinstimmen:

$$\begin{aligned}
&P \text{ Laplace-W. auf } \Omega,\\
&Q := P(\cdot|\{2,4\}).
\end{aligned}$$

Dieses Beispiel zeigt, daß im Satz 2.7.1 die Voraussetzung, das erzeugende System $\mathcal{D}$ ist $\bigcap$-stabil, wirklich benötigt wird.

2.8 Verteilungsfunktion

Anmerkung. Das Beispiel 2.7.2 führt zu der folgenden Definition:

2.8.1 Def. (Verteilungsfunktion)

Ein W-Maß P auf $\mathcal{B}(\mathbb{R}^d)$ ist eindeutig durch seine Verteilungsfunktion

$$\begin{aligned}
F &: \mathbb{R}^d \to [0,1],\\
F &: x \;\mapsto P\big((-\infty, x_1] \times \cdots \times (-\infty, x_n]\big)
\end{aligned}$$

festgelegt.

Manchmal nennt man F auch kurz die *Verteilung* von P. Die Bezeichnung *„Verteilung“* wird aber ebenfalls für das Bildmaß einer Zufallsvariablen benutzt. Bei einer Zufallsvariablen $X : (\Omega, P) \to \mathbb{R}^d$ muß man sorgfältig zwischen dem Bildmaß P_X und der Verteilungs*funktion* F_X von P_X unterscheiden! Zur Unterscheidung wird für F_X auch die Bezeichnung *kumulative Verteilung* verwendet.

2.8.2 Festst. (Verteilungsfkt. rechtsseitig stetig)

Es sei P ein W-Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfunktion F. Dann gilt

(i) *Für jede monoton fallende Folge $(x_n)_n$, mit $\lim_{n\to\infty} x_n = b$ gilt*

$$\lim_{n\to\infty} F(x_n) = P\Big(\bigcap_{n=1}^{\infty} (-\infty, x_n]\Big) = P\big((-\infty, b]\big) = F(b).$$

F ist also rechtsseitig stetig.

(ii) *Für jede monoton fallende Folge $(x_n)_n$, mit $\lim_{n\to\infty} x_n = -\infty$ gilt*

$$\lim_{n\to\infty} F(x_n) = P\Big(\bigcap_{n=1}^{\infty} (-\infty, x_n]\Big) = P(\emptyset) = 0.$$

Es ist also $\lim_{x\to-\infty} F(x) = 0$.

(iii) *Für jede monoton wachsende Folge $(x_n)_n$, mit $\lim_{n\to\infty} x_n = \infty$ gilt*

$$\lim_{n\to\infty} F(x_n) = P\Big(\bigcup_{n=1}^{\infty} (-\infty, x_n]\Big) = P\big((-\infty, \infty\big) = 1.$$

Es ist also $\lim_{x\to\infty} F(x) = 1$.

Anmerkung. In der Maßtheorie zeigt man die Umkehrung zu Feststellung 2.8.2.

2.8.3 Satz (Maß zu einer Verteilungsfkt.)

Sei $F : \mathbb{R} \to [0,1]$ eine monoton wachsende, rechtsseitig stetige Funktion mit

$$\lim_{x\to-\infty} F(x) = 0 \quad \textit{und} \quad \lim_{x\to\infty} F(x) = 1.$$

Dann gibt es genau eine W-Maß P auf der Borel-Algebra $\mathcal{B}(\mathbb{R})$, so daß F die Verteilungsfunktion von P ist:

$$P\big((-\infty, x]\big) = F(x) \quad \textit{für } x \in \mathbb{R}.$$

Anmerkung. (Zum Bew. von Satz 2.8.3**)** 1. Wir konstruieren das gesuchte Maß P als die Verteilung einer Zufallsvariablen

$$X : (0,1) \to \mathbb{R},$$

die man die **Quantilfunktion** von F nennt.

Dabei trägt $(0,1)$ die uniforme Verteilung $\mathcal{U}$ (→ Beispiel 2.5.7). und P ist das Bildmaß $\mathcal{U}_X$. Die Abbildung X ist eine Art *„Umkehrfunktion"* zu F. Die genaue Konstruktion, die noch weitere Anwendungen findet, formulieren wir in Feststellung 2.8.4.

2. Wenn $F : \mathbb{R} \to (0,1)$ stetig und streng monoton wachsend ist, und

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{und} \quad \lim_{x\to\infty} F(x) = 1.$$

gilt, so folgt aus dem Zwischenwertsatz, daß das Bild $F(\mathbb{R}) = (0,1)$ ist. F ist also bijektiv und hat eine stetige Umkehrfunktion $X : (0,1) \to \mathbb{R}$, die die gewünschte Gleichung (2.8.2) erfüllt.

3. Wenn F nur monoton wachsend ist oder Sprungstellen hat, definiert man eine verallgemeinerte Umkehrfunktion, die sogenannte Quantilfunktion:

2.8.4 Festst. (Quantilfunktion)

Sei $F : \mathbb{R} \to [0,1]$ eine monoton wachsende, rechtsseitig stetige Funktion mit

$$\lim_{x\to-\infty} F(x) = 0 \quad \textit{und} \quad \lim_{x\to\infty} F(x) = 1.$$

(i) *Die Quantilfunktion (Quantil-Transformation) X zu F wird definiert durch*

$$X(t) := \min\{x \in \mathbb{R} \mid F(x) \geq t\} \quad \textit{für } t \in (0,1). \tag{2.8.1}$$

Man beachte, da F rechtsseitig stetig ist, wird das Minimum in Gleichung (2.8.1) wirklich angenommen. Man zeichne ein Bild, das zeigt

– was in einer Sprungstelle von F passiert,

– was passiert, wenn F auf einem Teilintervall konstant ist

und wie X jeweils aussieht.

(ii) *Nach Konstruktion der Quantilfunktion X gilt für $t \in (0,1)$, $x \in \mathbb{R}$*

$$X(t) \leq x \quad \Leftrightarrow \quad t \leq F(x)$$
$$X^{-1}((-\infty, x]) = (0, F(x)] \in \mathcal{B}((0,1), \tag{2.8.2}$$

Aus Gleichung (2.8.2) folgt, daß $X : \big((0,1), \mathcal{B}((0,1)\big) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ meßbar ist.

Anmerkung. Aus Gleichung (2.8.2) folgt für das Bildmaß der Quantilfunktion X:

2.8.5 Folg. (Verteilung der Quantilfunktion)

Die Quantilfunktion X ist eine reelle Zufallsvariable auf dem W-Raum $(0,1)$ mit der uniformen Verteilung $\mathcal{U}$ und es gilt

$$\mathcal{U}\big(X^{-1}(-\infty, x]\big) = \mathcal{U}\big((0, F(x)]\big) = F(x).$$

D.h. F ist die Verteilungsfunktion ihrer Quantilfunktion X.

Anmerkung. Da jedes W-Maß P auf $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ durch seine Verteilungsfunktion F eindeutig bestimmt ist (→ Definition 2.8.1), kann man die Folgerung 2.8.5 auch so aussprechen:

2.8.6 Folg.

Jedes W-Maß P auf $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ ist Bildmaß einer reellen Zufallsvariablen auf dem Wahrscheinlichkeitsraum $\big((0,1), \mathcal{B}(0,1), \mathcal{U}_{(0,1)}\big)$.

Anmerkung. Da viele Programmiersprachen eine Funktion (random) bieten, die eine uniform auf $(0,1)$ verteilte Zufallsvariable nachbilden, ermöglicht die Konstruktion in Folgerung 2.8.6 die Simulation beliebiger reellwertiger Zufallsprozesse.

Anmerkung. (Quartile und Median) Mit der Quantilfunktion bestimmt man die sogenannten Quantile eines W-Maßes P, bzw einer Verteilungsfunktion F, bzw einer Zufallsvariablen X. Eine Zahl $q \in \mathbb{R}$ heißt ein α-Quantil, $0 < \alpha < 1$, von P, wenn

$$P\big((-\infty, q]\big) \geq \alpha \quad \text{und} \quad P\big([q, \infty)\big) \geq 1 - \alpha.$$

Ein $\frac{1}{2}$-Quantil $q_{1/2}$ heißt ein **Median**. Ein Median ist ein Punkt in R, an dem die Verteilungsfunktion das Niveau 1/2 gerade übersteigt oder überspringt. Im Fall einer Sprungstelle zählt dieser Punkt sowohl bei der linken wie bei der rechten Hälfte mit. Wenn die Verteilungsfunktion in diesem Punkt nicht strikt monoton wächst, ist der Median nicht eindeutig bestimmt.

Ist P die Verteilung einer reellen Zufallsvariablen X, so besagt *„ q ist ein α-Quantil von X"*, daß die Wahrscheinlichkeit für Beobachtungen von X,

die $\leq q$ sind, mindestens α ist,

die $\geq q$ sind, mindestens $(1-\alpha)$ ist.

Für eine reelle Zufallsvariable X auf einem Laplaceraum teilt das 1/4-Quantil $q_{1/4}$ grob gesprochen die Anzahl der Beobachtungen von X in einer langen Beobachtungsreihe im Verhältnis 1/4 zu 3/4 auf.

2.9 Dichtefunktion, stetige Verteilungen

Anmerkung. (Lebesgue-Maß und Lebesgue-Integral)

1. Auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ existiert eindeutig ein Maß λ so, daß das Maß $\lambda(I)$ eines jeden beschränkten Intervalls I gleich seiner Länge ist. Dieses Maß λ heißt das **Lebesgue-Maß** auf den Borel-Mengen $\mathcal{B}(\mathbb{R})$.

Definiert wird das Lebesgue-Maß auf $\mathcal{B}(\mathbb{R})$ wie folgt: Ist $G \subset \mathbb{R}$ offen, so ist G Vereinigung von abzählbar vielen offenen Intervallen $G = \bigcup_{n=1}^{\infty} (a_n, b_n)$, wobei $-\infty \le a_n < b_n \le \infty$ gilt. Man setzt dann

$$\lambda(G) = \begin{cases} \sum_{n=1}^{\infty} (b_n - a_n) & \text{wenn alle } a_n, b_n \in \mathbb{R}, \\ \infty & \text{sonst.} \end{cases}$$

Für eine beliebige Menge $A \in \mathcal{B}(\mathbb{R})$ betrachtet man alle offenen Mengen G, die A umfassen und setzt dann

$$\lambda(A) = \inf_{A \subset G} \lambda(G).$$

Nun kann man zeigen, daß so ein eindeutiges, *wohldefiniertes* Maß λ auf $\mathcal{B}(\mathbb{R})$ erklärt wird, und, daß für Intervalle I das Lebesgue-Maß $\lambda(I)$ die Länge des Intervalls ist.

2. Analog gibt es auf dem $(\mathbb{R}^d, \mathcal{B}(R^n))$ eindeutig das **d-dimensionale Lebesgue-Maß** λ^d, welches für Quader und andere elementare Körper den elementaren Inhalt angibt.

3. Da $\lambda(\mathbb{R}) = \infty$ ist, kann man durch Normierung aus $\mathbb{R}$ keinen W-Raum machen. Um weitere Beispiel zu gewinnen, gehen wir folgendermaßen vor: Wenn $f : \mathbb{R} \to [0, \infty)$ eine *intergrierbare*, nichtnegative Funktion mit

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad (*)$$

ist, so setze man für Intervalle $(a, b]$, $-\infty \le a < b < \infty$,

$$P_f((a, b]) = \int_a^b f(x)\,dx. \qquad (**)$$

Dies erzeugt dann ein eindeutig bestimmtes W-Maß P_f auf den Borel-Mengen $\mathcal{B}(\mathbb{R})$. Man nennt f die **Dichtefunktion** dieses Maßes P_f.

4. Wir müssen noch ein paar Worte darüber verlieren, was *integrierbare* Funktion bedeutet und was das das Integral in der Gleichung $(*)$ bzw. $(**)$ eigentlich ist. Für eine tragkräftige Theorie braucht man hier das **Lebesgue-Integral**, Dieses ist eine Erweiterung des Riemann-Integrals mit vielen schönen und praktischen Eigenschaften, die man bei dem enger gefaßten Riemann-Integral noch nicht findet.

In vielen praktischen Fällen ist die Funktion f stetig oder stückweise stetig, so daß man die Integrale $(*)$ und $(**)$ mit Hilfe einer Stammfunktion F bilden kann:

$$P_f((a, b]) = \int_a^b f(x)\,dx \quad = F(b) - F(a) \qquad (**')$$

$$P_f(\mathbb{R}) = \int_{-\infty}^{\infty} f(x)\,dx = \lim_{x\to\infty} F(x) - \lim_{y\to-\infty} F(y) \qquad (*')$$

Aber auch in diesem einfachen Fall ist das Lebesgue-Integral, das natürlich den selben Wert hat, ein nützlicher Helfer.

5. Wir geben die allgemeine Definition einer Dichtefunktion, beschränken uns in den Beispielen dann aber auf den stückweise stetigen Fall.

2.9.1 Def. (Dichtefunktion, stetiges W-Maß)

Eine integrierbare, nicht-negative Funktion $f : \mathbb{R} \to [0, \infty)$ heißt Dichtefunktion eines Wahrscheinlichkeitsmaßes P auf $\mathcal{B}(\mathbb{R})$, wenn für alle a, b mit $-\infty \le a < b < \infty$ gilt

$$P((a, b]) = \int_a^b f(x)\,dx. \qquad (2.9.1)$$

Man nennt f auch Wahrscheinlichkeitsdichte oder kurz Dichte von P.

*Ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R})$ mit Dichtefunktion heißt ein **stetiges W-Maß**.*

Anmerkung.

1. Wir benötigen im folgenden nur den Fall, daß f stetig oder stückweise ist.

2. Wenn ein W-Maß P auf $\mathcal{B}(\mathbb{R})$ eine Dichtefunktion f hat, so ist diese nicht ganz eindeutig bestimmt. So kann man f in endlich vielen Punkten abändern, ohne den Wert des Integrals (2.9.1) zu verändern.

Man kann zeigen, sind f_1 und f_2 Dichtefunktionen von P, so ist die Menge $\{f_1 \ne f_2\}$ eine Lebesgue-Nullmenge:

$$\lambda\{f_1 \ne f_2\} = 0.$$

Dann ist auch $P\{f_1 \ne f_2\} = 0$.

2.9.2 Festst. (Verteilungs- und Dichtefunktion)

Es sei P ein W-Maß auf $\mathbb{R}$ mit Verteilungsfunktion F.

1. *f ist genau dann die Dichtfunktion von P, wenn*

$$F(x) = \int_{-\infty}^{x} f(\xi)\,d\xi \quad \text{für } x \in \mathbb{R} \qquad (2.9.2)$$

gilt.

2. Man kann den Begriff der Ableitung einer Funktion dahingehend verallgemeinern, daß man die Gleichung (2.9.2) als $F' = f$ schreiben kann. Wir wollen diese Verallgemeinerung nicht weiter verfolgen. Aus dem Hauptsatz der Differential und Integralrechnung folgt:

Ist die Verteilungsfunktion F bis auf endlich viele Ausnahmepunkte stetig differenzierbar und gilt $F' = f$, so ist f die Dichtefunktion von P.

2.9.3 Bsp. (Dichtfnkt. der uniform. Verteilng.)

1. Ein W-Maße P auf einem Intervall I kann man als W-Maße auf ganz $\mathbb{R}$ auffassen. Man setze

$$P(A) := P(A \cap I) \quad \text{für } A \in \mathcal{B}(\mathbb{R}).$$

Das so fortgesetzte W-Maß P ist auf I konzentriert und die bedingte Wahrscheinlichkeit $P(.|I) = P$.

2. Faßt man die uniforme Verteilung $\mathcal{U}_{[0,1]}$ im obigen Sinne als W-Maß auf $\mathbb{R}$ auf, so hat $\mathcal{U}_{[0,1]}$ die Verteilungsfunktion

$$F(x) := \begin{cases} 0 & \text{für } x < 0, \\ x & \text{für } 0 \le x \le 1, \\ 1 & \text{für } x > 1. \end{cases}$$

und die Dichtefunktion

$$f(x) := \begin{cases} 0 & \text{für } x < 0, \\ 1 & \text{für } 0 \le x \le 1, \\ 0 & \text{für } x > 1. \end{cases}$$

Da es auf einzelne Werte der Dichtefunktion nicht ankommt, kann man auch f in den Punkten 0 und 1 auf den Wert 0 setzen. D.h, $f = \mathbb{1}_{[0,1]}$ oder $f = \mathbb{1}_{[0,1)}$ usw.

2.9.4 Bem. (Dichtefunktion auf $\mathbb{R}^d$)

Es sei P ein W-Maß auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ mit Verteilungsfunktion

$$F(x_1, \ldots, x_d) = P\big((-\infty, x_1] \times \cdots \times (-\infty, x_d]\big)$$

für $x = (x_i)_{i=1}^d \in \mathbb{R}^d$. Eine integrierbare, nichtnegative Funktion $f : \mathbb{R}^d \to [0, \infty)$ ist die Dichtefunktion des W-Maßes P, wenn für alle $x = (x_i)_{i=1}^d \in \mathbb{R}^d$ gilt

$$F(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(\xi_1, \dots, \xi_d)\, d\xi_1 \cdots d\xi_d.$$

Dann gilt für alle Quader $A = (a_1, b_1] \times (a_d, b_d]$

$$P(a) = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f(\xi_1, \dots, \xi_d)\, d\xi_1 \cdots d\xi_d.$$

2.10 Exponentialverteilung

2.10.1 Bez. (Exponentialverteilung)

Die Exponentialverteilung ist das kontinuierliche Analogon zur geometrischen Verteilung (→ Beispiel 2.1.3).

(i) Die Exponentialverteilung mit dem Parameter $\lambda > 0$ hat die Dichtefunktion

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0, \\ 0 & \text{für } x \leq 0 \end{cases}$$
$$= \lambda e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x).$$

und die Verteilungsfunktion

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x > 0, \\ 0 & \text{für } x \leq 0 \end{cases}$$
$$= (1 - e^{-\lambda x}) \mathbb{1}_{(0,\infty)}(x).$$

Für die Quantilfunktion erhält man

$$X(t) := -\frac{\log(1-t)}{\lambda} \quad \text{für } t \in (0, 1).$$

(ii) Die Bedeutung des Parameters $\lambda > 0$ erklärt sich aus dem → Erwartungswert und der → Varianz einer exponentialverteilten Zufallsvariablen X:

$$E(X) = \int_0^\infty x \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda},$$
$$\mathrm{Var}(X) = E(X^2) - (EX)^2 = \frac{1}{\lambda^2}.$$

Anmerkung. In den folgenden beiden Bemerkungen diskutieren wir Zufallsexperimente, in denen die Exponentialverteilung typischerweise auftritt.

2.10.2 Bem. (geom. Vert. → exp.Vert.)

Um den Zusammenhang zur geometrischen Verteilung (→ Beispiel 2.1.3) herzustellen, sei $(X_n)_n$ eine Folge geometrisch verteilter Zufallsvariabler mit den Parametern p_n. Wir setzen voraus, daß $\lim_{n\to\infty} np_n = \lambda$ ist. Der Grenzwert der Erwartungswerte der Variablen X_n/n ist dann (→ Feststellung 2.2.5)

$$E\Big(\frac{X_n}{n}\Big) = \frac{1}{np_n} - \frac{1}{n} \to \frac{1}{\lambda}.$$

Für $x > 0$ betrachten wir die Wahrscheinlichkeit der Menge

$$\Big\{\frac{X_n}{n} \leq x\Big\} = \{X_n \leq nx\}$$

Da die geometrische Verteilung auf $\mathbb{N}_0$ konzentriert ist, haben $\{X_n \leq nx\}$ und $\{X_n \leq \lfloor nx \rfloor\}$ die gleiche Wahrscheinlichkeit, wobei $\lfloor nx \rfloor$ die größte ganze Zahl $\leq nx$ ist. Für die Verteilungsfunktion von X_n/n folgt nun

$$\begin{aligned} P_{X_n/n}((-\infty, x]) &= P_{X_n}((-\infty, \lfloor nx \rfloor]) \\ &= \sum_{k=0}^{\lfloor nx \rfloor} \Big(1 - p_n\Big)^k p_n \\ &= 1 - \Big(1 - p_n\Big)^{\lfloor nx \rfloor + 1} \\ &= 1 - \Big(1 - \frac{np_n}{n}\Big)^{\lfloor nx \rfloor + 1}. \end{aligned}$$

Für $n \to \infty$ konvergiert dies gegen $1 - e^{-\lambda x}$ ($\to$ Bemerkung 2.4.4) Somit ist X_n/n für große n annähernd exponentialverteilt mit dem Parameter $\lambda = \lim_{n\to\infty} np_n$.

Anmerkung. (Interpretation der Rechnung in Bem. 2.10.2) Die geometrische Verteilung $\mathcal{G}_p\{k\}$, $k \in \mathbb{N}_0$, gibt die Wahrscheinlichkeit an, daß in einem oft genug wiederholten Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p der erste Erfolg nach k Mißerfolgen eintritt. Man stelle sich vor, daß die einzelnen Bernoulli-Experimente in einem festen Takt mit Taktzeit 1 durchgeführt werden. Verkürzt man die Taktzeit auf $1/n$, $n \in \mathbb{N}$, so sei die Erfolgswahrscheinlichkeit p_n. Die Zufallsvariable X_n gebe an, wieviele Takte der Länge $1/n$ vergangen sind, bis der erste Erfolg eintritt. Die Zufallsvariable X_n/n gibt grob gesagt die Zeit an, die verstrichen ist, bis der erste Erfolg eintritt.

Die getroffene Voraussetzung $np_n \to \lambda$ besagt, daß der Erwartungswert für die Zeit bis zum ersten Erfolg gegen $1/\lambda$ konvergiert. Die Rechnung zeigt, daß dann die Verteilungsfunktion von X_n/n gegen die Verteilungsfunktion der Exponentialverteilung zum Parameter λ konvergiert.

2.10.3 Bem.

Wir erinnern an das Beispiel 2.4.5 der Anzahl der Schadensmeldungen, die bei einer Versicherung nach dem Zeitpunkt 0 bis einschließlich zum Zeitpunkt $t > 0$ eingehen. Wenn $\lambda > 0$ die durchschnittliche Zahl der Schadensmeldungen in einem Zeitraum der Länge 1 ist, so ist die Anzahl der Schadensmeldungen im Zeitraum $(0, t]$ Poisson-verteilt mit dem Parameter λt.

Wir fragen nun nach der Wahrscheinlichkeit $P\big((0, t]\big)$, daß im Zeitraum $(0, t]$, $(t > 0)$, mindestens ein Schaden gemeldet wurde:

$$P\big((0,t]\big) = \sum_{k=1}^{\infty} \mathcal{P}_{\lambda t}\{k\} = 1 - \mathcal{P}_{\lambda t}\{0\} = 1 - e^{-\lambda t}.$$

Da wir erst die Schäden nach dem Zeitpunkt 0 registrieren, ist $P\big((-\infty, 0]\big) = 0$. Die Wahrscheinlichkeit für mindestens einen Schadensfall bis zum Zeitpunkt t ist also exponentialverteilt mit Parameter λ:

$$P\big((-\infty,t]\big) = \begin{cases} 1 - e^{-\lambda t} & (t > 0), \\ 0 & (t \le 0). \end{cases}$$

Anmerkung. Wir halten fest, bei einem Experiment mit rein zufällig auftretenden Zeitpunkten, ist die Anzahl der Zeitpunkte im Zeitintervall $(0, t]$ Poisson-verteilt mit Parameter λt und die Wartezeit auf den ersten Zeitpunkt exponentialverteilt mit Parameter λ. Dabei ist λ die durchschnittliche Anzahl der Zeitpunkte pro Zeiteinheit und $1/\lambda$ die durchschnitliche Wartezeit.

Eine exakte Untersuchung findet man in den Lehrbüchern zur Wahrscheinlichkeitstheorie unter dem Stichwort **Poisson-Prozess**.

2.11 Normalverteilung

Anmerkung. (Formeln zur Normalverteilung) Die Funktion

$$\varphi : x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{für } x \in \mathbb{R} \tag{2.11.1}$$

ist positiv und symmetrisch um den Nullpunkt. Die Funktion φ fällt sehr schnell, so daß für alle Potenzen $k \in \mathbb{N}_0$

$$\lim_{x\to\pm\infty} x^k \varphi(x) = 0 \tag{2.11.2}$$

ist. Man kann zeigen, daß diese Funktion keine Stammfunktion innerhalb der elementaren Funktionen $(x^y, \log x, e^x, \sin x, \cos x, \dots)$ hat. Man bezeichnet die Stammfunktion mit

$$\Phi : x \mapsto \int_{-\infty}^{x} e^{-\xi^2/2}\, d\xi \quad \text{für } x \in \mathbb{R}. \tag{2.11.3}$$

und nennt sie die **Gaußsche Fehlerfunktion** oder wegen der Form ihres Graphens auch die **Gaußsche Glockenfunktion**. Dabei ist die Integrationskonstante so gewählt, daß

$$\lim_{x\to-\infty} \Phi(x) = 0 \tag{2.11.4}$$

ist. Der Faktor $1/\sqrt{2\pi}$ in der Definition von φ ist so gewählt, daß

$$\lim_{x\to\infty} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\xi^2/2}\, d\xi = 1 \tag{2.11.5}$$

wird. Φ ist stetig differenzierbar mit Ableitung $\Phi' = \varphi$. Da $\varphi > 0$ ist, ist

$$\Phi \quad \text{strikt monoton wachsend.} \tag{2.11.6}$$

Die Gleichung (2.11.4),(2.11.4) und (2.11.6) besagen, daß Φ Verteilungsfunktion eines W-Maßes auf $\mathbb{R}$ ist ($\to$ Satz 2.8.3).

Da φ symmetrisch um den Nullpunkt ist, ist der Mittelwert dieses W-Maßes

$$\int_{-\infty}^{\infty} x\varphi(x)\, dx = 0. \tag{2.11.7}$$

Es ist $\frac{d}{dx}\varphi(x) = -x\varphi(x)$. Durch partielle Integration erhält man

$$\int_{-\infty}^{\infty} x^2\varphi(x))\, dx = -\Big\{x \cdot \varphi(x)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \varphi(x)\, dx = 1 \tag{2.11.8}$$

und somit ist die Varianz $= 1$.

Einige Werte von Φ

x	0	0.67	1,00	1,28	1,64	1,96	2,33	3,008
$\Phi(x)$	0,5	0,75	0,84	0,90	0,95	0,975	0,99	0,999

(2.11.9)

Die Werte für $x < 0$ erhält man aus der Symmetrie $\Phi(-x) = 1 - \Phi(x)$.

2.11.1 Bez. (Standard-Normalverteilung)

Das Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit der Dichtefunktion

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{für } x \in \mathbb{R}$$

und der Verteilungsfunktion

$$\Phi : x \mapsto \int_{-\infty}^{x} e^{-\xi^2/2}\, d\xi \quad \text{für } x \in \mathbb{R}.$$

heißt die Standard-Normalverteilung auf $\mathbb{R}$.

Der Erwartungswert der Standard-Normalverteilung ist 0 und die Varianz ist 1. Daher wird die Standard-Normalverteilung mit $\mathcal{N}_{0,1}$ bezeichnet.

Anmerkung. Durch Substitution

$$x = \frac{\xi - \mu}{\sigma} \quad \text{und} \quad dx = \frac{1}{\sigma} d\xi$$

folgt aus den Gleichungen (2.11.5) –(2.11.8)

$$\int_{-\infty}^{\infty} \frac{1}{\sigma}\varphi\Big(\frac{\xi-\mu}{\sigma}\Big)\, d\xi = 1,$$

$$\int_{-\infty}^{\infty} \xi \frac{1}{\sigma}\varphi\Big(\frac{\xi-\mu}{\sigma}\Big)\, d\xi = \mu$$

$$\int_{-\infty}^{\infty} \xi^2 \frac{1}{\sigma}\varphi\Big(\frac{\xi-\mu}{\sigma}\Big)\, d\xi \sigma^2 + \mu^2.$$

Also ist $x\frac{1}{\sigma} \mapsto \varphi\Big(\frac{\xi-\mu}{\sigma}\Big)$ die Dichtefunktion eines W-Maßes mit Mittelwert μ und Varianz σ^2.

Für die zugehörige Verteilungsfunktion erhält man mit der Substitution $\xi = \sigma t + \mu$

$$\int_{-\infty}^{x} \frac{1}{\sigma}\varphi\Big(\frac{\xi-\mu}{\sigma}\Big)\, d\xi = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(t)\, dt = \Phi\Big(\frac{x-\mu}{\sigma}\Big). \tag{2.11.10}$$

Durch Verschiebung und Skalierung erhält man aus der Standard-Normalverteilung eine ganze Familie von W-Maßen.

2.11.2 Bez. (Normalverteilungen $\mathcal{N}_{\mu,\sigma^2}$)

Das W-Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit der Dichtefunktion

$$\frac{1}{\sigma}\varphi\Big(\frac{\xi-\mu}{\sigma}\Big) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)$$

heißt die Normalverteilung mit Mittelwert μ und Varianz σ^2. Es wird mit $\mathcal{N}_{\mu,\sigma^2}$ bezeicnet. Die zugehörige Verteilungsfunktion ist

$$\int_{-\infty}^{x} \frac{1}{\sigma}\varphi\Big(\frac{\xi-\mu}{\sigma}\Big)\, d\xi = \Phi\Big(\frac{x-\mu}{\sigma}\Big),$$

wobei Φ die Verteilungsfunktion von $\mathcal{N}_{0,1}$ ist.

2.11.3 Festst. (affine Transform. normalvertl. ZV)

(i) *Es sei X eine Zufallsvariable mit der Verteilung $\mathcal{N}_{\mu,\sigma^2}$. Für $\alpha, \beta \in \mathbb{R}$, $\alpha \neq 0$ hat dann $Y := \alpha X + \beta$ die Verteilung $\mathcal{N}_{\alpha\mu+\beta,(\alpha\sigma)^2}$.*

(ii) *Es seien $\alpha, \beta \in \mathbb{R}$, $\alpha > 0$. Dann gilt für $x \in \mathbb{R}$*

$$N_{\mu,\sigma^2}\{-\infty, \frac{x-\beta}{\alpha}\} = N_{\alpha\mu+\beta,(\alpha\sigma)^2}\{-\infty, t\}.$$

(iii) *Wenn X normalverteilt mit den Parametern μ und σ^2 ist, dann ist die zentrierte normalisierte Zufallsvariable $X^* = (X-\mu)/\sigma$ standard-normalverteilt ($\to$ Bezeichnung 2.12.1)*

(iv) *Für $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$*

$$N_{0,1}\{-\infty, \frac{x-\mu}{\sigma}\} = N_{\mu,\sigma^2}\{-\infty, x\}.$$

Beweis. (i) Klar ist, daß

$$E(\alpha X + \beta) = \alpha E(X) + \beta,$$
$$\mathrm{Var}(\alpha X + \beta) = \alpha^2 \mathrm{Var}(X)$$

ist. Zu zeigen ist, daß Y normalverteilt ist. Es sei zunächst $\alpha > 0$

Da

$$\{Y \le y\} = \{\alpha X + \beta \le y\} = \{X \le \frac{y-\beta}{\alpha}\}$$

hat die Verteilungsfunktion F_Y die Gestalt

$$F_Y(y) = \int_{-\infty}^{\frac{y-\beta}{\alpha}} \frac{1}{\sigma}\varphi\Big(\frac{\xi-\mu}{\sigma}\Big)\, d\xi$$
$$= \int_{-\infty}^{y} \frac{1}{\alpha\sigma}\varphi\Big(\frac{\eta-(\alpha\mu+\beta)}{\alpha\sigma}\Big)\, d\eta$$

wobei $\xi = \frac{\eta-\beta}{\alpha}$ substituiert wurde. Die Dichtefunktion von F_Y ist also

$$y \mapsto \frac{1}{\alpha\sigma}\varphi\Big(\frac{y-(\alpha\mu+\beta)}{\alpha\sigma}\Big)$$

und folglich ist Y normalverteilt mit Erwartungswert $\alpha\mu + \beta$ und Varianz $(\alpha\sigma)^2$.

Die Aussagen (ii) , (iii) und (iv) folgen nun aus (i)

2.12 Grenzwertsatz von Moivre-Laplace

2.12.1 Bez. (normierte zentrierte ZV)

Um Zufallsvariable bsser vergleichen zu können verschiebt und skaliert man ihre Werte so, daß der Erwartungswert zu 0 und die Varianz zu 1 wird:

zentrieren: Zu einer Zufallsvariablen X mit Erwartungswert μ bilde man $Y = X - \mu$. Dann ist $E(Y) = 0$.

normieren: Zu einer Zufallsvariablen Y mit Varianz σ^2 bilde man $Z = Y/\sigma$. Dann ist $\operatorname{Var}(Z) = 1$.

normieren-zentrieren Zu einer Zufallsvariablen X mit Erwartungswert μ und Varianz σ^2 bilde man

$$X^* := \frac{X-\mu}{\sigma}.$$

X^* ist dann zentriert und normiert:

$$E(X^*) = 0, \quad \operatorname{Var}(X^*) = 1.$$

2.12.2 Satz (Moivre-Laplace)

Es sei $(X_k)_k$ eine Folge binomialverteilter Zufallsvariabler mit dem jeweiligen Parameter $n_k \in \mathbb{N}$, $p_k \in (0,1)$. Wenn

$$\lim_{k\to\infty} \operatorname{Var}(X_k) = \infty$$

ist, dann konvergieren die Verteilungsfunktionen $F_{X_k^*}$ der normierten und zentrierten Zufallsvariablen

$$X_k^* := \frac{X_k - E(X_k)}{\sqrt{\operatorname{Var}(X_k)}} = \frac{X_k - n_k p_k}{\sqrt{n_k p_k (1-p_k)}}$$

gegen die Standard-Normalverteilung. Die Konvergenz ist gleichmäßig auf $\mathbb{R}$:

$$\lim_{k\to\infty} \sup_{x\in\mathbb{R}} \big| F_{X_k^*}(x) - \Phi(x) \big| = 0.$$

2.12.3 Folg. (Moivre-Laplace)

Es seien $0 < p < 1$ und $(S_n)_n$ eine Folge $\mathcal{B}_{n,p}$-verteilter Zufallsvariablen. Dann gilt

$$\lim_{n\to\infty} \sup_{a<b} \Big| P\{a < S_n^* \le b\} - \Phi(b) - \Phi(a) \Big| = 0.$$

Anmerkung. (zum Satz von Moivre-Laplace) 1. Der Satz von Moivre-Laplace ist ein *Spezialfall* des zentralen Grenzwertsatzes.

2. Man benötigt nur die Voraussetzung, daß die X_k binomialverteilt sind und ihre Varianzen

$$\operatorname{Var}(X_k) = n_k p_k (1-p_k) \to \infty \quad \text{für } k \to \infty.$$

Dann konvergieren auch die Erwartungswerte

$$E(X_k) = n_k p_k \ge n_k p_k (1-p_k) \to \infty$$

und die Anzahlen $n_k \to \infty$. Die p_k müssen nur $0 < p_k < 1$ erfüllen, ansonsten ist ihr Verhalten für $k \to \infty$ unwichtig.

3. Man beachte, daß die Verteilungen $F_{X_k^*}$ konvergieren. Es wird nichts über die Konvergenz der Zufallsvariablen X_k^* gesagt. Letzteres macht auch keinen Sinn, denn die X_k haben zwar alle ihre Werte in $\mathbb{N}_0 \subset \mathbb{R}$, sie können aber ganz unterschiedliche Definitionsbereiche haben: $X_k : (\Omega_k, P_k) \to \mathbb{N}_0$.

4. Der Satz sagt etwas über die Konvergenz der Bildmaße $(P_k)_{X_k^*}$ aus. Die Zufallsvariablen sind eigentlich überflüssig, machen aber die Formulierung anschaulicher.

Man kann den Satz von Moivre-Laplace auch als Grenzwertsatz für Folgen $(\mathcal{B}_{n_k,p_k})_k$ von Binomialverteilungen formulieren. Bei dem Bewesis eines vorbereitenden Hilfssatzes, des lokalen Grenzwertsatzes, werden wir diesen Standpunkt auch einnehmen.

2.12.4 Folg. (Approximation der BinomialV.)

Es sei $0 < p < 1$. Dann gilt für $n \to \infty$

$$\sup_{t\in\mathbb{R}} \Big| \mathcal{B}_{n,p}\big((-\infty,t]\big) - \mathcal{N}_{np,np(1-p)}\big((-\infty,t]\big) \Big| \to 0.$$

Also ist für $k, l \in \mathbb{N}_0$

$$\mathcal{B}_{n,p}\big((-\infty,k]\big) \approx \Phi\Big(\frac{k-np}{\sqrt{np(1-p)}}\Big)$$

und

$$B_{n,p}\{k,\dots,l\} = \Phi\Big(\frac{l-np}{\sqrt{np(1-p)}}\Big) - \Phi\Big(\frac{k-np}{\sqrt{np(1-p)}}\Big).$$

Die Approximation wird noch etwas besser, wenn man die sogenannte *Stetigkeitskorrektur* vornimmt:

$$\mathcal{B}_{n,p}\big((-\infty,k]\big) \approx \Phi\Big(\frac{k+\frac12-np}{\sqrt{np(1-p)}}\Big)$$

und

$$B_{n,p}\{k,\dots,l\} = \Phi\Big(\frac{l+\frac12-np}{\sqrt{np(1-p)}}\Big) - \Phi\Big(\frac{k-\frac12-np}{\sqrt{np(1-p)}}\Big).$$

Mit heutigen Computeralgebraprogrammen kann man $\mathcal{B}_{n,p}$ auch für große n recht genau berechnen. Für formelmäßige Rechnungen ist aber die Näherung durch die Gaußsche Fehlerfunktion überlegen.

Anmerkung. Das folgende Bsp. ist dem Lehrbuch [1] von Krengel entnommen:

2.12.5 Bsp. (Bestimmung: Stichprobenumfang)

Wir wollen den Prozentsatz der Wähler einer Partei A schätzen. Werden n Wähler befragt und sind darunter S_n Wähler der Partei A, so sei S_n/n der Schätzer für die Wahrscheinlichkeit p, daß ein zufällig ausgewählter Wähler für die Partei A stimmt.

Wie groß muß n sein, damit die Wahrscheinlichkeit eines Irrtums in der Schätzung von p von mehr als 1% kleiner als 0,05 ist?.

Es soll also gelten

$$P\Big\{ -0{,}01 \le \frac{S_n}{n} - p \le 0{,}01 \Big\} \approx 0{,}95.$$

Wir nehmen an, daß die Befragungen unabhängig sind und somit die S_n $\mathcal{B}_{n,p}$ verteilt sind.

Mit $\mu_n = np$ und $\sigma_n = \sqrt{np(1-p)}$ ergibt sich mit Folgerung 2.12.3

$$\begin{aligned} 0{,}95 &\approx P\Big\{ -\frac{0{,}01n}{\sigma_n} \le S_n^* \le \frac{0{,}01n}{\sigma_n} \Big\} \\ &\approx \Phi\Big(\frac{0{,}01n}{\sigma_n}\Big) - \Phi\Big(-\frac{0{,}01n}{\sigma_n}\Big) \\ &= 2\Phi\Big(\frac{0{,}01n}{\sigma_n}\Big) - 1. \end{aligned}$$

Also $\Phi(0,01n/\sigma_n) \approx 0,975$.

Aus der Tabelle (2.11.9) entnimmt man $\Phi^{-1}(0,975) = 1.96$. Also muß

$$\frac{0,01}{\sqrt{np(1-p)}} \approx 1,96 \quad \Leftrightarrow \quad n \approx p(1-p)10.000 \cdot 1,96^2$$

sein.

Nun ist p leider unbekannt. Da aber $\max\limits_{0<p<1} p(1-p) = 0,25$ ist kommt man in jedem Fall mit

$$n = 0.25 \cdot 10.000 \cdot 1.96^2 = 9600$$

Befragungen aus. Hat man von vorneherein die Information, daß $p \leq 0,1$ ist, so ist $\max\limits_{0<p<0,1} p(1-p) = 0,1 \cdot 0,9 = 0,09$. Man kommt nun mit

$$n = 0,09 \cdot 10.000 \cdot 1,96^2 = 3450$$

Befragungen aus.