Theoretische Ergänzung zur Physik II

Priv.-Doz. Dr. Rochus Klesse

Sommersemester 2004

Inhaltsverzeichnis

0.1 Einführung

Wir wollen in den ersten drei Wochen der Theoretischen Ergänzungsvorlesung zur Physik II einige elementare Begriffe und Konzepte der Wahrscheinlichkeitstheorie erarbeiten, denen sie immer wieder im Laufe Ihres Physikstudiums begegnen werden.

Inhalt der ersten drei Vorlesungen:

- Grundbegriffe der Wahrscheinlichkeitstheorie und Statistik
- Zufallsgrößen und Wahscheinlichkeitsverteilungen
- Zentraler Grenzwertsatz

Literaturhinweis

- H. Fischer, H. Kaul, Mathematik für Physiker, Kap. 1.4, Mengen und Wahrscheinlichkeit, Teubner Stuttgart, 1990.

Kapitel 1

Teil 1: Elementare Wahrscheinlichkeitstheorie

1.1 Grundbegriffe der Wahrscheinlichkeitstheorie

Grob gesagt ist eine Wahrscheinlichkeitaussage eine *quantifizierte* Erwartung über den Ausgang eines Versuchs mit ungewissem Ausgang, eines sog. *Zufallsexperiments*. Eine mathematisch exakte aber dennoch praktikable Definition ist schwierig. Wir begnügen uns deshalb mit einer einfachen

Intuitiven Definition der Wahrscheinlichkeit:
Wir sagen ein Versuch (Vorgang, Zufallsexperiment) resultiert mit Wahrscheinlichkeit $p(E)$ in ein Ereignis E genau dann, wenn wir *erwarten*, daß bei N-maliger Wiederholung im Limes $N \to \infty$ das Ereignis E $N(E) = p(E) \cdot N$ mal eintreffen wird.

Damit sehr eng verwandt ist der Begriff der

Relativen Häufigkeit
Tritt bei N-maliger Wiederholung eines Versuchs ein Ereignis E $N(E)$-mal auf, so beträgt seine *relative Häufigkeit*

$$h_N(E) = \frac{N(E)}{N} \,.$$

Daß eine Unterscheidung zwischen Wahrscheinlichkeit und relativer Häufigkeit angebracht sein kann, zeigen folgende zwei Beispiele:

- "Diese Glühbirne brennt mit Wahrscheinlichkeit $0,99$ länger als $50h$."
  ist eine sinnvolle Wahrscheinlichkeitsaussage, auch für eine individuelle Glühbirne. Da die Birne aber nur einmal durchbrennen kann, sind wir nicht in der Lage die Aussage durch eine Messung der relativen Häufigkeit zu bestätigen.

- "Die Zahlen $1, 2, 3, 4, 5$, und 6 werden nächsten Samstag mit Wahrscheinlichkeit $p \approx {}^1\!/_{14\cdot 10^6}$ im Lotto gezogen."

Diese Aussage beruht keinesfalls auf eine Messung der entsprechenden relativen Häufigkeit; dazu ist Zahl der bis heute überhaupt vorgenommenen Lottoziehungen viel zu klein.

Im Gegensatz zur empirisch definierten und bestimmbaren relativen Häufigkeit ist die Wahrscheinlichkeit eher eine abstrakte, idealisierte Größe, die insbesondere von den gemachten Annahmen, der zugrunde gelegten Theorie oder einfach von unserem Kenntnisstand abhängig ist.

In all den Fällen, in denen ein Versuch unter vergleichbaren Bedingungen beliebig oft wiederholbar ist, gilt aber natürlich nach Definition

$$\lim_{N\to\infty} h_N(E) = p(E)\,.$$

Wahrscheinlichkeitsaussagen sind oft nur möglich – aber meistens dann auch von besonderem praktischen Nutzen – weil es eine Theorie für *Beziehungen* zwischen Wahrscheinlichkeiten von Ereignissen gibt: die *Wahrscheinlichkeitstheorie.*

Wir verzichten auf eine axiomatische Formulierung der Wahrscheinlichkeitstheorie [1] zugunsten einer deduktiven Entwicklung aus der oben gegebenen intuitiven Definition der Wahrscheinlichkeit. Dazu sind noch ein paar Begriffe nötig:

Unter einem **atomaren Ereignis** oder einem **Ergebnis** eines Versuchs verstehen wir ein Ereignis, daß sich *nicht* weiter in Unterereignisse aufgliedern läßt.

Beim Würfeln mit einem Spielwürfel ist etwa das Ereignis "Augenzahl 5" ein atomares Ereignis, "ungerade Augenzahl" hingegen nicht, da sich letzteres in die atomaren Ereignisse "Augenzahl 1", "Augenzahl 3" und "Augenzahl 5" zerlegen läßt.

Wir halten noch fest, daß jeder Versuch immer nur in *ein* atomares Ereignis ausgehen kann. Atomare Ereignisse schließen sich einander aus.

Die Menge $\mathcal{S}$ aller atomaren Ereignisse oder Ergebnisse bilden den **Merkmalraum** oder auch **Stichprobenraum** des Versuchs. Seine Elemente, die atomaren Ereignisse, werden wir im folgenden oft mit a_1, a_2, ... bezeichnen.

In unserem Würfelbeispiel ist der Merkmalraum

$$\begin{aligned} \mathcal{S} &= \{\text{“Augenzahl 1”}, \text{“Augenzahl 2”}, \ldots, \text{“Augenzahl 6”}\} \\ &\equiv \{1,\, 2,\, \ldots,\, 6\}\,. \end{aligned}$$

Ist ein Ereignis E *nicht* atomar, muss es sich nach Def. letztendlich in mehrere atomare Ereignisse a_{l_1}, a_{l_2}, ..., a_{l_m} des Merkmalraums $\mathcal{S}$ aufgliedern lassen. Wenn wir dann noch annehmen, daß das Ereignis E durch seine atomaren Bestandteile vollständig beschrieben ist, können wir allgemein definieren:

[1] die im ersten Drittel des letzten Jahrhunderts entwickelt wurde, Kolmogoroff, Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin (Springer) 1933.

Ein **Ereignis** E eines Versuchs mit Merkmalraum $\mathcal{S}$ ist eine *Teilmenge* von S,

$$E = \{a_{l_1},\ a_{l_2},\ \ldots,\ a_{l_m}\} \subset \mathcal{S}\ .$$

Das Ereignis E tritt bei einem Versuch genau dann ein, wenn der Versuch mit einem Ergebnis $a \in E$ ausgeht.

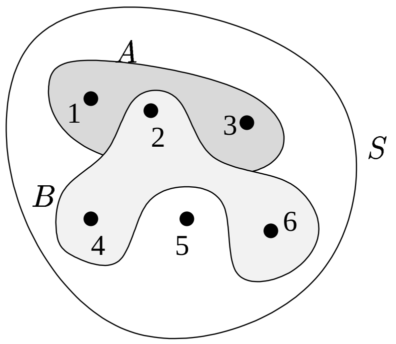
Damit auch die atomaren Ereignisse (Ergebnisse) nach dieser Definition Ereignisse sind, identifizieren wir ggf. das atomare Ereignis a_i mit der einelementigen Teilmenge $\{a_i\}$.

Zur Veranschaulichung betrachten wir zwei nicht-atomare Würfelereignisse A und B als Teilmengen von $\mathcal{S} = \{1, 2, \ldots, 6\}$:

$A = \{1, 2, 3\}$: "Augenzahl ≤ 3",

$B = \{2, 4, 6\}$: "gerade Augenzahl",

Graphisch:



Dank der mengentheoretischen Darstellung können wir jetzt Mengenoperationen der Teilmengen von $\mathcal{S}$ in natürlicher Weise als Ereigniskombinationen auffassen. Etwa für zwei Ereignisse $A, B \subset S$:

- $A \cup B$: "Ereignis A **oder** Ereignis B"
- $A \cap B$: "Ereignis A **und** Ereignis B"
- $A \setminus (B \cap A)$: "Ereignis A **und nicht** Ereignis B"

usw.

$\mathcal{S}$ ist als Teilmenge von $\mathcal{S}$ ebenso ein Ereignis. Es beinhaltet alle überhaupt möglichen Ergebnisse und tritt daher mit Sicherheit ein. S nennen wir daher das **sichere Ereignis**. Die Wahrscheinlichkeit seines Eintretens ist damit eins,

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

Unter dem **Gegenereignis** $\bar{E}$ eines Ereignisses E verstehen wir das Ereignis

$$\bar{E} = S \setminus E\,.$$

$\bar{E}$ tritt offenbar immer genau dann ein, wenn E nicht eintritt.

Unsere intuitive Wahrscheinlichkeitsdefinition zusammen mit der Identifizierung von Ereignissen mit Teilmengen von S erlaubt es nun, die Wahrscheinlichkeit eines *beliebigen* Ereignisses anhand der atomaren Wahrscheinlichkeiten zu berechnen:

Sei $S = \{a_i\}_{i=1,\dots,n}$ der Merkmalraum eines Versuchs mit als bekannt vorausgesetzten Wahrscheinlichkeiten $P(a_j)$ der atomaren Ereignisse $a_j \in S$. $E = \{a_{l_1}, \dots, a_{l_m}\} \subset S$ sei ein beliebiges Ereignis, dessen Wahrscheinlichkeitwir nun aus den $P(a_j)$ bestimmen wollen.

Bei N Versuchen erwarten wir nach Def. der Wahrscheinlichkeit

$$\begin{array}{rrcl} & N(a_{l_1}) & = & P(a_{l_1}) \cdot N \text{ mal Ergebnis } a_{l_1}\,, \\ & N(a_{l_2}) & = & P(al_2) \cdot N \text{ mal Ergebnis } a_{l_2}\,, \\ & \vdots & & \vdots \\ \text{und} & N(a_{l_m}) & = & P(al_m) \cdot N \text{ mal Ergebnis } a_{l_m}\,. \end{array}$$

Da mit dem jedem Eintreten eines atomaren Ereignisses $a_{l_1}, a_{l_2}, \dots$ oder a_{l_m} definitionsgemäß auch immer das Ereignis E eintritt, erwarten wir

$$N(E) = \sum_{j=1}^{m} N(a_{l_j})$$

mal das Eintreten von E. Da andererseits $N(E)$ durch $P(E) \cdot N$ gegeben sein soll, erhalten wir

$$P(E) = \frac{\sum_{j=1}^{m} N(a_{l_j})}{N} = \sum_{j=1}^{m} P(a_{l_j})\,.$$

Damit haben wir die grundlegende Erkenntnis gewonnen, daß die Wahrscheinlichkeit$P(E)$ eines Ereignisses E durch die Summe seiner atomaren Wahrscheinlichkeiten $P(a)$, $a \in E$, gegeben ist,

$$P(E) = \sum_{a \in E} P(a)\,. \tag{1.1}$$

Würfelbeispiel: Die atomare Wahrscheinlichkeiten für $a \in S = \{1, 2, \dots, 6\}$ seien je $P(a) = 1/6$. Dann erhalten wir etwa

- $P(\text{“Augenzahl ungerade”}) = P(\{1, 3, 5\}) = P(1) + P(3) + P(5) = \frac{1}{2}$
- $P(\text{“Augenzahl} \geq 5\text{”}) = P(\{5, 6\}) = P(5) + P(6) = \frac{1}{3}$

Normierung der Wahrscheinlichkeiten
Aus $P(S) = 1$ einerseits und $P(S) = \sum_{a \in S} P(a)$ andererseits folgt die **Normierungsbedingung** atomarer Wahrscheinlichkeiten:

$$1 = \sum_{a \in S} P(a)\,.$$

Anhand fundamentalen Gleichung (1.1) können wir nun Beziehungen zwischen den Wahrscheinlichkeiten von kombinierten Ereignissen herstellen:

$$\begin{aligned} P(A \cup B) &= \sum_{a \in A \cup B} P(a) \\ &= \sum_{a \in A} P(a) + \sum_{a \in B} P(a) - \sum_{a \in A \cap B} P(a) \end{aligned}$$

Wenn wir nun die Summen der rechten Seite wieder nach Gl. (1.1) in Wahrscheinlichkeit übersetzen, erhalten wir die Beziehung

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,.$$

Zwei Ereignisse A und B schließen einander aus, wenn sie nicht gemeinsam eintreten können, d.h. genau dann wenn $P(A \cap B) = \emptyset$. In diesem Fall vereinfacht sich obige Gleichung zu $P(A \cup B) = P(A) + P(B)$. Hieraus folgt durch vollständige Induktion die sehr wichtige

Additionsregel
Sind $A_1, \dots, A_n$ sich wechselseitig ausschließende Ereignisse (d.h. $P(A_i \cap A_j) = 0$ für $i \neq j$), dann gilt

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)\,.$$

Nach dem gleichen wie oben bei $A \cup B$ angewendeten Schema erhält man leicht weitere Beziehungen. Beispielsweise

$$P(\underbrace{A \setminus (B \cap A)}_{\text{“}A \text{ und nicht } B\text{”}}) = P(A) - P(B \cap A)\,. \tag{1.2}$$

Insbesondere

$$P(\bar{A}) = P(S \setminus (\underbrace{S \cap A}_{=A})) = P(S) - P(A)$$

aufgrund $P(S) = 1$ also

$$P(\bar{A}) = 1 - P(A) \tag{1.3}$$

Gleichungen wie (1.2), (1.2) oder (1.3) erlauben es, Beziehungen zwischen Wahrscheinlichkeiten aufzustellen, ohne daß man dazu etwas über die zugrundeliegenden atomaren Wahrscheinlichkeiten wissen oder voraussetzen muss. Sie sind von allgemeiner Gültigkeit und deshalb sehr wichtig.

Möchte man darüber hinaus gehen und konkret Wahrscheinlichkeiten von Ereignissen berechnen, muss man auf die atomaren Wahrscheinlichkeiten zurückgreifen. In der Regel läßt sich über diese allerdings nicht allzuviel sagen, außer daß es keinen Grund gibt warum ein bestimmtes atomares Ereignisse wesentlich wahrscheinlicher sein sollte als ein beliebig anderes. Dieser Grundsatz der Gleichberechtigung auf atomarer Ebene führt dann zur Annahme von gleichwahrscheinlichen oder *gleichverteilten* atomaren Ereignissen. Diese Annahme kann manchmal durch Symmetrieargumente gestützt werden (etwa beim Würfel). In vielen Fällen ist sie aber einfach nur die einzig praktikable Lösung, die dennoch oft zu befriedigenden Resultaten führt. Wir kommen daher nun zu dem sehr wichtigen Fall von

Gleichverteilten atomaren Wahrscheinlichkeiten
Alle atomaren Ereignisse eines Versuchs mit Merkmalraum $S = \{a_1, a_2, \ldots, a_n\}$ seien gleichwahrscheinlich. Aufgrund der Normierungsbedingung ist damit für alle $a \in S$

$$P(a) \equiv p = \frac{1}{|\mathcal{S}|} \,,$$

wobei $|\mathcal{S}|$ für die Anzahl der Elemente in S steht. Aus Beziehung (1.1) folgt sofort für ein Ereignis $E \subset S$ die Wahrscheinlichkeit

$$P(E) = \sum_{a \in E} p = |E| \cdot p = \frac{|E|}{|S|} \,.$$

(Wobei $|E|$ die Anzahl der Elemente in E bezeichnet.) Mit anderen Worten:

$$\boxed{P(E) = \frac{|E|}{|S|} = \frac{\text{Anzahl günstiger Ereignisse}}{\text{Anzahl möglicher Ereignisse}}}$$

wobei ein Ereignis a "günstig" ist, wenn wenn es aus E ist. Diese auf Laplace zurückgehende Regel ist in sehr vielen Fällen anwendbar. Insbesondere stützt sich die gesamte statische Physik auf diese einfache Regel.

Zwei Beispiele

Würfeln: aus Symmetriegründen gleichverteilte Wahrscheinlichkeiten:

$$P(\text{“ungerade Augenzahl”}) = P(\{1,3,5\}) = \frac{|\{1,3,5\}|}{|S|} = \frac{3}{6} = \frac{1}{2}$$

Lotto: Ziehung von 6 Kugeln aus 49 ohne Zurücklegen, d.h. $S = \{(n_1, n_2, \dots, n_6) | n_i \neq n_j, n_i \in \{1, \dots, 49\}\}$.
Wie groß ist Wahrscheinlichkeit der Ziehung der Zahlen $1, 2, 3, \dots, 6$ (ohne Beachtung der Reihenfolge) ?
Das Ereignis dessen Wahrscheinlichkeit wir ermitteln wollen ist

$$\begin{aligned} E &= \{(1,2,3,\dots,6), (2,1,3,4,\dots,6), \dots\} \\ &= \text{alle Permutationen von } (1,2,3,\dots,6) \end{aligned}$$

damit

$$P = \frac{|E|}{|S|} = \frac{6!43!}{49!} = \frac{1}{\binom{49}{6}} = \frac{1}{13.983.816} \approx 7 \cdot 10^{-8},$$

denn

$$\begin{aligned} |E| &= 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 6! \\ |S| &= 49 \cdot 48 \cdot \dots \cdot 44 = \frac{49!}{43!}. \end{aligned}$$

[Alternative Lösung:

$$\tilde{S} = \text{ alle } 6 - \text{elementigen Teilmengen von } \{1, \dots, 49\}$$
$$\tilde{E} = \{\{1, 2, \dots, 6\}\}$$

$$\rightarrow P = \frac{|\tilde{E}|}{|\tilde{S}|} = \frac{1}{\binom{49}{6}}.$$

]

Bedingte Wahrscheinlichkeit
Für zwei Ereignisse A und B bezeichnet die **bedingte Wahrscheinlichkeit** $P(B|A)$ die Wahrscheinlichkeit des Ereignisses B unter der Voraussetzung des Eintreffens von A.

D.h. man erwartet, daß bei N Versuchen in $N(A) = P(A) \cdot N$ Fällen A eintritt, unter denen wiederum $N_A(B) = P(B|A) \cdot N(A)$ mal Ereignis B eintritt.

Es gilt:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{1.4}$$

Um dies zu einzusehen, vergegenwärtigen wir uns zuerst, daß nach Def. offenbar

$$N(B \cap A) = P(B|A)N(A) .$$

Mit $N(A) = P(A)N$ und $N(B \cap A) = P(B \cap A)N$ folgt hieraus sofort die Behauptung.

Ein Würfelbeispiel:

$$\begin{aligned} P(\text{“Augenzahl 2”}|\text{“Augenzahl gerade”}) &= \frac{P(\text{“Augenzahl =2 und gerade”})}{P(\text{“Augenzahl gerade”})} \\ &= \frac{1/6}{1/2} = \frac{1}{3} . \end{aligned}$$

Der Begriff der bedingten Wahrscheinlichkeit erlaubt uns nun die Definition von Statistischer Unabhängigkeit:

Zwei Ereignisse A und B heißen **statistisch unabhängig**, wenn das Eintreten von A die Wahrscheinlichkeit des Eintreffens von B *nicht* beeinflusst; oder umgekehrt. In Formeln bedeutet dies:

$$P(B|A) \stackrel{!}{=} P(B) \tag{1.5}$$

$$\text{oder} \quad P(A|B) \stackrel{!}{=} P(A) \tag{1.6}$$

Aufgrund der Beziehung $P(B|A) = \frac{P(B \cap A)}{P(A)}$ ist (1.5) offenbar gleichbedeutend zu

$$P(B \cap A) = P(A) \cdot P(B) \tag{1.7}$$

und dies ist wiederum äquivalent zu (1.6), was das “oder” in der Definition erklärt.

Häufig wird Beziehung (1.7) zur Definition der statistischen Unabhängigkeit benutzt. Diese alternative Definition ist insofern etwas unbefriedigend, als dass in der Praxis gerade diese Beziehung benutzt wird, um die Wahrscheinlichkeit $P(A \cap B)$ des gleichzeitigen Eintreffens zweier *statistisch unabhängiger* Ereignisse A und B zu berechnen.

Demgegenüber erlaubt unsere Definition es diesen Zirkelschluss zu vermeiden: Ist beispielsweise aus physikalischen Gründen sichergestellt, dass Ereignis A nicht das Auftreten von B beeinflusst, so ist die Bedingung (1.5) und damit auch die statistische Unabhängigkeit sichergestellt. Folglich kann Formel (1.7) ohne logischen Widerspruch benutzt werden, um die Wahrscheinlichkeit $P(A \cap B)$ zu berechnen.

Die statistische Unabhängigkeit **mehrerer** Ereignisse $A_1, \ldots, A_n$ könnte nun sehr einfach durch eine zu (1.7) analogen Bedingung

$$P(A_1 \cap A_2 \cap ... \cap A_n) = \prod_{i=1}^{n} P(A_i)$$

definiert werden. Aus den gerade diskutierten Gründen streben wir aber eine Definition an, die *nicht* die Wahrscheinlichkeit $P(A_1 \cap A_2 \cap ... \cap A_n)$ verwendet.

Dazu führen wir die *statistische Unabhängigkeit* **mehrerer** Ereignisse $A_1, \ldots, A_n$ *rekursiv* auf die statistische Unabhängigkeit **zweier** Ereignisse zurück und definieren:[2]

n Ereignisse $A_1, \ldots, A_n$ sind *statistisch unabhängig* genau dann, wenn entweder

$$n = 2 \text{ und } A_1 \text{ und } A_2 \text{ statistisch unabhängig sind,}$$

oder für alle $i \in \{1, \ldots, n\}$

$$\begin{array}{rl} & \text{1) die } (n-1) \text{ Ereignisse } A_1, \ldots, A_{i-1}, A_{i+1}, \ldots A_n \textit{ stat. unabh.}, \\ \text{und} & \text{2) } A_i \text{ und } A_1 \cap \cdots \cap A_{i-1} \cap A_{i+1} \cap \cdots \cap A_n \text{ stat. unabh. sind} \end{array}$$

Durch einfache vollständige Induktion beweist man nun leicht die zu (1.7) analoge und sehr oft anwendbare

Produktregel:
Sind die Ereignisse $A_1, ..., A_n$ statistisch unabhängig, so gilt

$$P(\underbrace{A_1 \cap A_2 \cap ... \cap A_n}_{\text{“}A_1 \text{ und } A_2 \text{ und ... und } A_n\text{”}}) = \prod_{i=1}^{n} P(A_i)\,.$$

Beispiele:

[2] Im Folgenden schreiben wir zur besseren Unterscheidung die *statistische Unabhängigkeit* **mehrerer** Ereignisse kursiv, die wie oben def. statistische Unabhängigkeit **zweier** Ereignisse dagegen normal.

a) Wahrscheinlichkeit, dass beim Würfeln mit 10 Würfeln 10 mal eine 1 erscheint?
– Die Würfelereignisse sind *physikalisch* unabhängig
$\rightarrow$ Die Würfelereignisse sind statistisch unabhängig
–Wahrscheinlichkeit, daß der i-te Würfel eine 1 zeigt, ist $p_i = \frac{1}{6}$
$\rightarrow p = \prod_{i=1}^{10} p_i = (\frac{1}{6})^{10} \approx 1,6 \cdot 10^{-8} \approx \frac{1}{60 \cdot 10^6}$

b) Glühbirnen: Wahrscheinlichkeit, dass i-te Glühbirne innerhalb 50h durchbrennt:

$$p_i = 1 - 0,99 = \frac{1}{100}$$

– Durchbrennen *physikalisch* unabhängig (Annahme)
$\rightarrow$ Durchbrennen statistisch unabhängig
$\rightarrow$ Wahrscheinlichkeit, dass alle drei innerhalb von 50h durchbrennen:

$$p = p_1 \cdot p_2 \cdot p_3 = \left(\frac{1}{100}\right)^3 = \frac{1}{10^6}$$

1.2 Zufallsgrößen und Wahrscheinlichkeitsverteilungen

Zur Motivation von Zufallsgrößen und deren Wahrscheinlichkeitsverteilungen betrachten wir zwei Beispiele:

Glücksspiel: Eine Münze wird geworfen. Für “Kopf” erhalten wir einen Euro, für “Zahl” müssen wir einen Euro zahlen. Wie sieht die Gewinn/Verlustverteilung nach N Würfen aus?

Random Walk: ein Teilchen bekommt pro Zeiteinheit einen zufälligen Stoß nach links oder rechts und hüpft daher zufällig um eine Längeneinheit nach links oder rechts (Brownsche Bewegung, Diffusion). Wie sieht die Ortsverteilung nach N Zeitschritten aus?

Zur mathematischen Beschreibung derartiger Probleme benötigen wir den Begriff der **Zufallsgröße**. Sie ordnet jedem Ergebnis eines Versuchs eine reelle Zahl zu,

$$\begin{aligned} X : S &\rightarrow \mathbf{R} \\ a &\rightarrow X(a) \end{aligned}$$

mit Wertemenge $W = X(S) = \{x_1, x_2, ..., x_n\}$

In unseren zwei Beispielen ist jedesmal $S = \{(a_1, ..., a_n) \,|\, a_i \in \{+1, -1\}\}$ und $X(a_1, ..., a_n) = \sum_{i=1}^{n} a_i$.

Uns interessiert die Wahrscheinlichkeit $f(x)$, mit der die Zufallsgröße X einen Wert

$x \in W$ annimmt. Die Funktion $f(x)$ heißt die **Wahrscheinlichkeitsverteilung** von X,

$$\begin{aligned} f : W &\rightarrow \mathbf{R}_+ \cup \{0\} \\ x &\rightarrow f(x) \end{aligned}$$

f kann anhand der Wahrscheinlichkeiten für Ereignisse $E \subset S$ bestimmt werden: Offenbar nimmt die Zufallsgröße X den Wert x immer genau dann an, wenn der Versuch mit einem Ergebnis aus dem Urbild $E_x \equiv X^{-1}(\{x\})$ von x unter X endet. E_x als Teilmenge von S definiert demnach ein Ereignis, dessen Wahrscheinlichkeit genau $f(x)$ ist. In Formeln:

$$f(x) = P(E_x) = P\left(X^{-1}(\{x\})\right) \; . \tag{1.8}$$

Zur Illustration dieses Zusammenhangs betrachten wir obiges Beispiel für $N = 2$: In diesem Fall ist $S = \{--, -+, +-, ++\}$, wobei die Ergebnisse gleichwahrscheinlich seien. Die Zufallsgröße X (Gewinn/Verlust bzw. Teilchenort) ist gegeben durch

$$\begin{array}{ll} X(--) = -2 & \equiv \; x_1 \\ \left.\begin{array}{l} X(-+) = 0 \\ X(+-) = 0 \end{array}\right\} & \equiv \; x_2 \\ X(++) = +2 & \equiv \; x_3 \end{array}$$

Hieraus ergibt sich nach Gl. (1.8) die Wahrscheinlichkeitsverteilung

$$\begin{aligned} f(x_1) &= P(\{--\}) = \frac{1}{4} \\ f(x_2) &= P(\{-+, +-\}) = \frac{1}{2} \\ f(x_3) &= P(\{++\}) = \frac{1}{4} \; . \end{aligned}$$

Wir stellen fest, das $f(x)$ im Unterschied zu den atomaren Wahrscheinlichkeiten nicht gleichverteilt ist. Dies ist der Regelfall.

Aus der Normierung der atomaren Wahrscheinlichkeiten $P(a)$, $a \in S$, folgt die **Normierung** der Wahrscheinlichkeitsverteilung:

$$\sum_{x \in W} f(x) = \sum_{x \in W} P\left(X^{-1}\left(\{x\}\right)\right) = P\left(\overset{\circ}{\bigcup_{x \in W}} X^{-1}(\{x\})\right) = P(S) = 1$$

wobei wir beim zweiten Schritt die Additionsregel sowie die Tatsache ausgenutzt haben, dass die Urbilder aller $x \in W$ disjunkt sind und ganz S überdecken.

Wir fassen zusammen:
Eine diskrete Zufallsgröße X mit Werten $W \subset \mathbf{R}$ ($|W| < \infty$) wird durch eine normierte ($\sum_x f(x) = 1$), nicht-negative Wahrscheinlichkeitsverteilung $f : W \rightarrow \mathbf{R}$ beschrieben. $f(x)$ ist die Wahrscheinlichkeit, dass X bei einem Versuch den Wert x

annimmt.

Zufallsgrößen sind reelwertig und erlauben daher die Definition von Mittelwert oder Erwartungswert $\langle X \rangle$ als arithmetisches Mittel über die in N Versuchen von X angenommenen Werte $\xi_1, \ldots, \xi_N \in W$, im Grenzfall $N \to \infty$. D.h.

$$\langle X \rangle \equiv \mu_X = \lim_{N\to\infty} \frac{1}{N} \sum_{l=1}^{N} \xi_l = \lim_{N\to\infty} \frac{1}{N} \sum_{x \in X} N_x\, x = \sum_{x \in X} \underbrace{\left(\lim_{N\to\infty} \frac{N_x}{N} \right)}_{\equiv f(x)} x \,,$$

wobei N_x die Anzahl der Versuche mit $X = x$ bezeichnet.

Wir definieren daher den **Mittelwert** oder **Erwartungswert** einer Zufallsgröße durch

$$\langle X \rangle \equiv \mu_X \equiv \sum_{x \in W} x\, f(x) \,.$$

Ebenso können wir den Erwartungswert einer beliebigen Funktion $A : \mathbf{R} \to \mathbf{R}$ bezüglich X definieren:

$$\langle A(X) \rangle \equiv \langle A \rangle_X \equiv \sum_{x \in W} A(x) \cdot f(x)$$

Die Bildung des Erwartungswerts ist **linear**, d.h. für $A_1, A_2 : \mathbf{R} \to \mathbf{R}$, $\lambda \in \mathbf{R}$ gilt

$$\begin{aligned} \langle A_1 + A_2 \rangle &= \langle A_1 \rangle + \langle A_2 \rangle \\ \langle \lambda A_1 \rangle &= \lambda \langle A_1 \rangle \,, \end{aligned}$$

wie man leicht nachrechnet.

Für $A = x^n$ erhalten wir das **n-te Moment** der Zufallsgröße X:

$$\langle X^n \rangle \equiv M_n(X) = \sum_{x \in W} f(x) \cdot f(x)$$

Wir wissen:

$$\begin{aligned} M_0(X) &= \left\langle X^0 \right\rangle = \langle 1 \rangle = \sum_x f(x) = 1 \text{ (Normierung)} \\ M_1(X) &= \langle X \rangle = \mu(X) \text{ (Erwartungswert)} \end{aligned}$$

Die **Varianz** einer Zufallsgröße ist definiert als

$$\sigma^2 \equiv \left\langle (X - \mu_X)^2 \right\rangle = \left\langle X^2 - 2\mu X + \mu^2 \right\rangle = \left\langle X^2 \right\rangle - 2\mu \underbrace{\langle X \rangle}_{\mu} + \mu^2 = \left\langle X^2 \right\rangle - \mu^2$$

d.h.

$$\sigma^2 \equiv \left\langle (X - \langle X \rangle)^2 \right\rangle = \left\langle X^2 \right\rangle - \langle X \rangle^2$$

Die Varianz σ^2 ist ein Maß für die **Breite** oder **Streuung** der Verteilung einer Zufallsgröße. Nach Tschebyschew lässt sich etwa die Wahrscheinlichkeit dafür, dass X einen Wert x annimmt, der um mehr als $\varepsilon > 0$ vom Mittelwert μ abweicht, durch die *Tschebyschewsche Ungleichung*

$$P(|X-\mu| > \varepsilon) < \frac{\sigma^2}{\varepsilon^2}$$

abschätzen (vgl. Übung).

Unter der **Standardabweichung** oder **Schwankung** einer Zufallsgröße versteht man die Wurzel der Varianz,

$$\sigma = \sqrt{\langle X^2\rangle - \langle X\rangle^2}$$

Für unser Beispiel des zweimaligen Münzwurfes erhalten wir

$$\begin{aligned} \langle X\rangle &= \mu = \sum_{x\in W} x f(x) = -2\cdot\frac{1}{4} + 0\cdot\frac{1}{2} + 2\cdot\frac{1}{4} = 0 \\ \left\langle X^2\right\rangle &= \sum_{x\in W} x^2 f(x) = 4\cdot\frac{1}{4} + 0\cdot\frac{1}{2} + 4\cdot\frac{1}{4} = 2 \\ \text{also}\quad \sigma^2 &= \left\langle X^2\right\rangle - \langle X\rangle^2 = 2\,. \end{aligned}$$

1.2.1 Binomialverteilung

Eine wichtige Wahrscheinlichkeitsverteilung ist die **Binominalverteilung**. Sie basiert auf einen binären Zufallsversuch mit Merkmalraum $S = \{0,1\}$. Die Eintrittswahrscheinlichkeiten seien $P(1) \equiv p$ und $P(0) = 1-p \equiv q$. Nach N-maliger Ausführung dieses Versuchs erhalten wir als Ergebnis eine Sequenz aus $S_N = \{0,1\}^N = \{(a_1,...,a_N)\,|\,a_i \in \{0,1\}\}$. Für diese Ergebnisse definieren wir die Zufallsgröße:

$$X_N(a_1,\ldots,a_N) := \sum_{i=1}^{N} a_i \in W = \{0,1,...,N\}$$

Gesucht ist nun die Wahrscheinlichkeitsverteilung $f_{N,p}$ für X_N:

X_N nimmt den Wert x genau dann an, wenn in der Sequenz $(a_1,\ldots,a_N)$ genau x-mal eine 1 erscheint. Es gibt offenbar $\binom{N}{x}$ solcher Sequenzen in S_N, die jeweils mit Wahrscheinlichkeit $p^x q^{N-x}$ auftreten. Nach dem Additionssatz für die Wahrscheinlichkeit sich ausschließender Ereignisse müssen wir diese Wahrscheinlichkeiten aufsummieren und erhalten so

$$f_{N,p}(x) = \binom{N}{x} p^x (1-p)^{N-x}\,. \tag{1.9}$$

Dies ist die Binominalverteilung. Ihre Normierung folgt aus dem Binomialsatz:

$$\sum_{x=0}^{N} f_{N,p}(x) = \sum_{x=0}^{N}\binom{N}{x} p^x (1-p)^{N-x} = (p+(1-p))^N = 1$$

Es lässt sich zeigen (vgl. Übungen), dass

$$\langle X \rangle = p \cdot N$$

$$\left\langle X^2 \right\rangle = pN + p^2 N(N-1)$$

$$\sigma = (\left\langle Y^2 \right\rangle - \langle Y \rangle^2)^{\frac{1}{2}} = \sqrt{p(1-p) \cdot N} = \sqrt{pq \cdot N}$$

Zusammenfassend:
Die Zufallsgröße $X_N := \sum_{i=1}^{N} a_i$ mit $P(\text{“}a_i = 1\text{”}) = p$ und $P(\text{“}a_i = 0\text{”}) = 1 - p$ ist binomialverteilt,

$$f_{N,p}(x) = \binom{N}{x} p^x (1-p)^{N-x} \ .r$$

Ihr Mittelwert ist $p \cdot N$ und ihre Standardabweichung ist $\sqrt{p(1-p)}\sqrt{N}$.

Anwendung: „$\sqrt{n}$-Regel “ Wir zählen Regentropfen, die während einer festgelegten Zeitdauer auf eine Fläche A fallen. Ihre Anzahl sei n. Niemand erwartet, dass bei Wiederholung des Versuchs *exakt* die gleiche Anzahl Tropfen ermittelt wird. Vielmehr erwarten wir, dass die Tropfenzahl n um einen Mittelwert $\langle n \rangle$ schwankt. Wie groß ist die Schwankung, d.h. die Standardabweichung der Zufallsgröße n?

Wir betrachten ein sehr großes Ensemble von N Regentropfen mit $N \gg \langle n \rangle$. Tropfen dieses Ensembles fallen demnach mit sehr geringer Wahrscheinlichkeit $p \ll 1$ auf die Fläche A, und mit Wahrscheinlichkeit $1 - p \approx 1$ daneben. Die Tropfenzahl n ist dann die binomialverteilte Zufallsgröße

$$n = \sum_{i=1}^{N} a_i \ ,$$

wobei $a_i = 1$ mit Wahrscheinlichkeit p und $a_i = 0$ mit Wahrscheinlichkeit $1 - p$. Ihr Mittelwert ist $\langle n \rangle = pN$ ihre Standardabweichung ergibt sich wegen $p \ll 1$ in guter Näherung zu

$$\sigma = \sqrt{p(1-p)N} \approx \sqrt{pN} = \sqrt{\langle n \rangle} \ .$$

Unter den gemachten Annahmen ist die Schwankung durch die Wurzel des Mittelwerts gegeben, – unabhängig von Ensemblegröße N und Eintrittswahrscheinlichkeit p! Das ist der Inhalt der „$\sqrt{n}$-Regel “, nach der die Anzahl n eingetroffener unabhängiger Ereignisse mit $\sigma = \sqrt{\langle n \rangle}$ schwankt.

1.2.2 Kontinuierliche Zufallsgrößen

Bisher haben wir nur sog. **diskrete** Zufallsgrößen betrachtet, bei denen die Wertemenge X aus endlich vielen, diskreten Werten x_i besteht. Oft ist es dagegen angebracht, von einen Kontinuum möglicher Werte einer Zufallsgröße auszugehen, d.h. der Wertebereich ist die Menge der reellen Zahlen, $W = \mathbf{R}$. In diesem Fall sprechen wir von einer **kontinuierlichen** Zufallsgröße .

Wir können uns diese als Grenzfall einer diskreten Zufallsgröße vorstellen, in dem der Abstand δx zwischen den möglichen diskreten Werten gegen 0 geschickt wird.

Wir definieren daher analog zur obigen Feststellung:
Eine **kontinuierliche** Zufallsgröße X ist durch eine nicht-negative, normierte Wahrscheinlichkeitverteilung $f : \mathbf{R} \to \mathbf{R}$ gegeben.
$f(x) \cdot \delta x$ ist die Wahrscheinlichkeit, dass bei einem Versuch die Zufallsgröße X einen Wert ξ in $[x, x + \delta x]$ annimmt.

Es folgt, dass die Wahrscheinlichkeit für das Vorliegen eines Wertes x aus einem Intervall $[a, b]$ durch

$$P(x \in [a, b]) = \int_a^b dx\, f(x)$$

gegeben ist. Damit ergibt sich die Normierungsbedingung:

$$1 \stackrel{!}{=} P(x \in \mathbf{R}) = \int_{-\infty}^{\infty} dx\, f(x)\ .$$

Erwartungswerte, Momente und Varianz sind wie für diskrete Zufallsgrößen definiert, wobei jedesmal die Summation $\sum_{x\in W}$ durch eine Integration $\int_{-\infty}^{\infty} dx$ zu ersetzen ist. Also etwa

$$\begin{aligned} \langle X^n \rangle &= \int_{-\infty}^{\infty} dx\, x^n f(x) \\ \langle A(X) \rangle &= \int_{-\infty}^{\infty} dx\, A(x) f(x) \end{aligned}$$

Von zentraler Bedeutung ist die kontinuierliche **Gaußverteilung**. Sie ist die Wahrscheinlichkeitsverteilung einer kontinuierlichen Zufallsgröße X mit Erwartungswert μ, Varianz σ^2 und ist durch

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \equiv g_{\mu,\sigma}(x)$$

definiert. Der zentrale Grenzwertsatz, den wir später kennenlernen werden, besagt, dass in sehr vielen Fällen die Wahrscheinlichkeitsverteilung einer Zufallsgröße durch eine Gaußverteilung approximiert werden kann.

Mit $\int e^{-t^2} dt = \sqrt{\pi}$ und $\int t^2 e^{-t^2} dt = \frac{\sqrt{\pi}}{2}$ und geeignete Substitutionen verifiziert man leicht:

$$\int g_{\mu,\sigma}(x)\, dx = 1; \qquad \int x g_{\mu,\sigma}(x)\, dx = \mu; \qquad \int (x-\mu)^2 g_{\mu,\sigma}(x)\, dx = \sigma^2$$

Anwendung: Für große N ist der exakte Ausdruck (1.9) für die Binomialverteilung $f_{N,p}$ wegen des Binomialkoeffizienten recht unhandlich. Falls auch $pN \gg 1$ kann sie aber sehr gut durch eine Gaußverteilung mit $\mu = pN$ und $\sigma = \sqrt{p(1-p)N}$ genähert werden:

$$f_{N,p}(x) \approx g_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi p(1-p)N}} e^{-\frac{(y-pN)^2}{2p(1-p)N}} \,. \tag{1.10}$$

Die Gültigkeit dieser Näherung läßt sich mittels der Stirlingsche Näherung

$$n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

zeigen (vgl. Übungen), sie folgt aber auch aus dem zentralen Grenzwertsatz (vgl. Vorlesung).

Analog zu den bisher betrachteten eindimensionalen Zufallsgrößen definieren wir eine n-**dimensionale Zufallsgröße** als Abbildung

$$\begin{aligned} \underline{X} = (X_1, \ldots, X_n) : S &\rightarrow W \subset \mathbf{R}^n \\ a &\mapsto (x_1(a), \ldots, x_n(a)) \,. \end{aligned}$$

Eine n-dimensionale Zufallsgröße kann als ein n-dimensionaler Vektor aufgefasst werden, dessen Komponenten eindimensionale Zufallsgrößen sind. Sie wird durch eine Wahrscheinlichkeitverteilung $f : \mathbf{R}^n \rightarrow \mathbf{R}$, $\underline{x} \rightarrow f(\underline{x})$ beschrieben, wobei

$$f(x_1, \ldots, x_n) \quad = \quad \text{Wahrscheinlichkeit, dass } \underline{X} \text{ den Wert } \underline{x} \text{ annimmt,}$$

falls es sich um eine diskrete Zufallsgröße handelt, und

$$\begin{aligned} f(x_1, \ldots, x_n) \cdot (\delta x)^n \quad = \quad & \text{Wahrscheinlichkeit, dass } \underline{X} \text{ einen Wert} \\ & \text{in } [x_1, x_1 + \delta x] \times \cdots \times [x_n, x_n + \delta x] \text{ annimmt,} \end{aligned}$$

wenn es eine kontinuierliche Zufallsgröße ist.

Auch die n-dim. Wahrscheinlichkeitverteilungen sind nicht-negativ und normiert:

$$\begin{aligned} 1 &= \sum_{\underline{x} \in W} f(\underline{x}) 1 \,, \qquad \text{(diskret)} \\ 1 &= \int dx_1 \ldots \int dx_n f(\underline{x}) \equiv \int d^n \underline{x} f(\underline{x}) \qquad \text{(kontinuierlich)} \end{aligned}$$

Erwartungswerte von Funktionen $A : \mathbf{R}^n \rightarrow \mathbf{R}^m, \mathbf{C}$ sind wie vorher definiert. Für kontinuierliche Zufallsgrößen etwa

$$\langle A(\underline{X}) \rangle = \int d^n \, \underline{x} A(\underline{x}) \, f(\underline{x}) \,.$$

n Zufallsgrößen $X_1, \ldots, X_n$ sind **statistisch unabhängig**, wenn für alle $\underline{x} = (x_1, \ldots, x_n) \in W_1 \times \cdots \times W_n$ die Urbilder $E_{x_i} = X_i^{-1}(\{x_i\})$, $i = 1, \ldots, n$ statistisch unabhängige Ereignisse sind. Im Falle der statistischen Unabhängigkeit gilt für die Wahrscheinlichkeitsverteilung der kombinierten n-dim. Zufallsgröße $\underline{X} := (X_1, \ldots, X_n)$ die Relation

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) \,.$$

Dies folgt direkt aus der Produktregel $P(A_1 \cap \ldots A_n) = \prod_i P(A_i)$ stat. unabhängiger Ereignisse A_i.

Zufallsgrößen, die auf physikalisch unabhängige Versuche basieren, sind in der Regel statistisch unabhängig.

Als Beispiel einer zweidimensionalen Zufallsgröße betrachten wird die Brownsche Bewegung (Diffusion, random walk) in der Ebene: Ein Teilchen (*random walker*) befindet sich zur Zeit $t = 0$ am Ursprung und führt nach jedem diskreten Zeitschritt $t \to t+1$ je einen zufälligen Hüpfer um ± 1 in x_1-Richtung und in x_2-Richtung mit gleichgroßer Wahrscheinlichkeit 1/2 aus. Der Ort $\underline{\mathrm{X}}_t = (X_1, X_2)_t$ zur dikretisierten Zeit $t \in \mathbf{N}$ ist die uns interessierende Zufallsgröße. Wir nehmen an, dass die Hüpfer in x_1- und x_2-Richtung physikalisch unabhängig, und somit die Komponenten X_1 und X_1 von $\underline{\mathrm{X}}$ statistisch unabhängige Zufallsgrößen sind.

Die Zufallsgröße X_1 ist gegeben durch:

$$X_1 = \sum_{i=1}^{t} a_i \quad \text{mit} \quad a_i = \begin{cases} +1 : \text{Wahrscheinlichkeit } 1/2 \\ -1 : \text{Wahrscheinlichkeit } 1/2 \end{cases}$$

(Den Index t schreiben wir nicht immer explizit hin.) Die zweite Komponente X_2 ist analog definiert. Statt X_1 (oder X_2) direkt zu betrachten, ist es zweckmässiger, sie gemäß

$$Y_1 := \frac{1}{2}(X_1 + t) = \sum_{i=1}^{t} \underbrace{\frac{1}{2}(a_i + 1)}_{\in\{1,0\}} = \sum_{i=1}^{t} b_i$$

auf eine offenbar binomialverteilte Zufallsgröße Y_1 zu transformieren. Damit ergibt sich die Wahrscheinlichkeitsverteilung für X_1 aus der Binomialverteilung $f^{\text{bin}}_{N,p}$ mit Parametern $N = t$, $p = 1/2$ (siehe Gl. (1.9)). Für große Zeiten $t \gg 1$ können wir die gaußsche Näherung (1.10) verwenden, wobei $\mu = t/2$ und $\sigma = \sqrt{t}/2$. Wir erhalten somit

$$f_t(x_1) = f^{\text{bin}}_{t,1/2}\left(\frac{1}{2}(x_1 + t)\right) \simeq g_{t/2,\sqrt{t}/2}\left(\frac{1}{2}(x_1 + t)\right) = \sqrt{\frac{2}{\pi t}} \exp\left(\frac{-x_1^2}{2t}\right) .$$

Für große Zeiten $t \gg 1$ sind damit die identischen Wahrscheinlichkeitsverteilungen $f^{(1)}$ und $f^{(2)}$ für die Komponenten X_1 und X_2 gegeben durch

$$f^{(1)}_t(x_1) = \sqrt{\frac{2}{\pi t}} \exp\left(\frac{-x_1^2}{2t}\right) \qquad \text{und} \qquad f^{(2)}_t(x_2) = \sqrt{\frac{2}{\pi t}} \exp\left(\frac{-x_2^2}{2t}\right) .$$

Die Erwartungswerte verschwinden, $\langle X_1 \rangle = \langle X_2 \rangle = 0$, die Varianzen σ_i^2, $i = 1, 2$, betragen für beide Komponenten (für große t)

$$\begin{aligned} \sigma_i^2 &= \left\langle X_i^2 \right\rangle = \sum_{x,\,\delta x = 2} x^2 f_t(x) = \sum_{x,\,\delta x = 2} x^2 \sqrt{\frac{2}{\pi t}} e^{\frac{-x^2}{2t}} \\ &= \sum_{x,\,\delta x = 1} x^2 \sqrt{\frac{1}{2\pi t}} e^{\frac{-x^2}{2t}} = \sqrt{\frac{1}{2\pi t}} \int dx x^2 e^{\frac{-x^2}{2t}} = t . \end{aligned}$$

(Man beachte, dass zu einer (un)geraden Zeit t die Zufallsgrößen X_i nur (un)gerade Werte annehmen können. Daher $\delta x = 2$ in den ersten zwei Summen. In der dritten summieren wir über alle x und teilen dafür durch zwei, damit wir die Summe durch das folgende Integral approximieren können.)

Die Wahrscheinlichkeitsverteilung der zweidimensionalen Zufallsgröße ist nun das Produkt

$$f_t(\underline{x}) = f_t(x_1) f_t(x_2) = \frac{2}{\pi t} \exp\left(\frac{-1}{2t} \underbrace{(x_1^2 + x_2^2)}_{=r^2} \right) .$$

Wir halten fest, dass

1) die Aufenthaltswahrscheinlichkeit $f_t(\underline{x})$ am Ort $\underline{x}$ zur Zeit t für Abstände $|\underline{x}| \equiv r \gg \sqrt{t}$ exponentiell klein ist,

2) die Wahrscheinlichkeit, dass sich das Teilchen zur Zeit t wieder am Ursprung befindet (*Rückehrwahrscheinlichkeit zur Zeit* t), durch $f_t(\underline{0})$ gegeben ist und sich für große Zeiten wie

$$f(\underline{0}) = \frac{2}{\pi t} \propto \frac{1}{t}$$

verhält.

Das mittlere Entfernungsquadrat $\langle r^2 \rangle_t$ zur Zeit t ist durch

$$\langle r^2 \rangle_t = \langle X_1^2 + X_2^2 \rangle_{\underline{X}_t}$$

definiert und wird in den Übungen berechnet.

1.3 Zentrale Grenzwertsatz

Zentrale Grenzwertsatz
Sei $X_1, X_2, X_3, \ldots$ eine Sequenz statistisch unabhängiger Zufallsgrößen mit $\langle |X_i|^3 \rangle < c$ für ein $c \in \mathbf{R}$ für alle i. Dann ist die Zufallsgröße

$$X := \frac{1}{N} \sum_{i=1}^{N} X_i$$

im Limes $N \to \infty$ **gaußverteilt** mit Mittelwert

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \mu_i$$

und Varianz

$$\sigma^2 = \frac{1}{N} \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right) .$$

Anmerkungen:

In Worten gefasst besagt der zentrale Grenzwertsatz, dass eine sich additiv aus vielen zufälligen Beiträgen zusammensetzende Zufallsgröße in guter Näherung gaußverteilt ist. Wann immer eine Zufallsgröße diese Bedingung erfüllt, kann nach dem zentralen Grenzertsatz mit einer Gaußverteilung gerechnet werden. Beispielsweise der Mittelwert einer Zufallsgröße aus $N \gg 1$ Messungen, die Position eines Random Walkers, oder binomialverteilte Größen im Allgemeinen.

Mit der Phrase "additiv aus vielen Beiträgen" ist hier gemeint, dass der tatsächliche Wert der Zufallsgröße *nicht* von nur einigen wenigen Beiträgen stammt. Im zentralen Grenzwertsatz wird diesem Umstand durch die Beschränkung $\langle |X_i|^3 \rangle < c$ der dritten Momente und dem Limes $N \to \infty$ Rechung getragen: Mit dem dritten Moment von $|X_i|$ ist auch das erste Moment beschränkt. Im Limes $N \to \infty$ wird daher der einzelne Beitrag X_i/N einer herausgegriffenen $i$ten Zufallsgröße sehr klein ausfallen.

Um die Konvergenz gegen eine Gaußverteilung im Limes $N \to \infty$ sicher zu stellen fordern wir hier die Beschränktheit der dritten Momente von $|X_i|$. Diese Bedingung ist etwas stärker als notwendig, vereinfacht aber den Beweis.

Aufgrund der Linearität des Erwartungswert ist leicht einzusehen, dass der Mittelwert von X gegen das Mittel $\frac{1}{N}\sum_i \langle X_i \rangle$ der Mittelwerte $\langle X_i \rangle$ konvergiert.

Demgegenüber konvergiert die Varianz σ^2 von X nicht einfach gegen das Mittel der Varianzen σ_i^2 der Zufallsgrößen X_i, sondern gegen das Mittel $\frac{1}{N}\sum_i \sigma_i^2$ *dividiert durch* N. Im Falle von gleichverteilten X_i bedeuted dies, dass die Standardabweichung von X gegenüber der Standardabweichung σ_0 der X_i um einen Faktor $\sqrt{N}$ reduziert ist; eine Variante der $\sqrt{N}$-Regel. Diese Tatsache wird bei der Mittelwertbildung von schwankenden Größen benutzt (vgl. Übungen).

Binomialverteilung $\to$ Gaußverteilung
Wir wollen nun noch mit Hilfe des zentralen Grenzwertsatzes die gaußsche Näherung der Binomialverteilung begründen (vgl. Gl. (1.10)). Dazu betrachten wir statt der binomialverteilten Zufallsgröße $X = \sum_{i=1}^N a_i$, mit a_i wie in Abschnitt 1.2.1, die Größe

$$\tilde{X} := \frac{1}{N}X = \frac{1}{N}\sum_{i=1}^N a_i \,.$$

Die Zufallsgrößen a_i sind gleichverteilt mit Mittelwert $\langle a_i \rangle = p$, Varianz $\sigma_0^2 = \langle a_i^2 \rangle - \langle a_i \rangle^2 = p(1-p)$. Das dritte Moment $\langle |a_i|^3 \rangle = p$ ist offenbar beschränkt. Nach dem Grenzwertsatz ist im Limes $N \to \infty$ die Zufallsgröße $\tilde{X}$ gaußverteilt mit Mittelwert $\tilde{\mu} = p$ und Varianz $\tilde{\sigma}^2 = p(1-p)/N$. Wir transformieren nun die Zufallsgröße $\tilde{X}$ durch Multiplikation mit N wieder auf X zurück. Durch diese einfache Umskalierung ändert sich offenbar nicht die funktionale Form der Wahrscheinlichkeitsverteilung, d.h. $X = N\tilde{X}$ ist ebenfalls gaußverteilt. Es bleibt noch Mittelwert und Varianz von X zu bestimmen: Der Mittelwert ergibt sich zu

$$\mu = \langle X \rangle = \left\langle N\tilde{X} \right\rangle = N\left\langle \tilde{X} \right\rangle = Np \,,$$

die Varianz zu

$$\sigma^2 = \left\langle X^2 \right\rangle - \langle X \rangle^2 = N^2(\left\langle \tilde{X}^2 \right\rangle - \left\langle \tilde{X} \right\rangle^2) = N^2\tilde{\sigma}^2 = p(1-p)N \,.$$

(Hier haben wir wieder die Linearität des Erwartungswertes ausgenutzt.) Dies sind genau die Parameter die wir in Abschnitt 1.2.1 angegeben haben.

Beweis
Zum Beweis des zentralen Grenzwertsatzes benotigen wir ein außerordentlich wichtiges mathematisches Werkzeug der Physik: die **Fouriertransformation**. Aufgrund ihrer Bedeutung erlauben wir uns jetzt einen Exkurs, in dem wir einige Grundlagen der Fouriertranformation erarbeiten. Danach können wir ohne Schwierigkeiten den Grenzwertsatz beweisen und sind zudem gut für die Elektrodynamik gerüstet, in der die Fouriertransformation sehr effektiv eingesetzt werden kann.

1.4 Fouriertransformation

Die Fouriertransformation einer Funktion $f(x)$ liefert eine Zerlegung in die harmonischen Funktionen $\sin(kx)$ und $\cos(kx)$:

$$f(x) \stackrel{!}{=} \sum_k a_k \sin(kx) + b_k \cos(kx) \ , \tag{1.11}$$

wobei geeignete Koeffizienten a_k, b_k zu geeignenten Wellenzahlen k zu wählen sind. Abb. 1.1 zeigt ein Beispiel. Bequemer ist eine komplexe Darstellung, die durch Substitutionen $\sin(kx) = \frac{1}{2i}(e^{ikx} - e^{-ikx})$, $\cos(kx) = \frac{1}{2}(e^{ikx} + e^{-ikx})$ und Umsortieren aus (1.11) hervorgeht:

$$f(x) = \sum_k \hat{f}_k e^{ikx} \tag{1.12}$$

Die Koeffizienten $\hat{f}_k$ sind i.A. komplex.

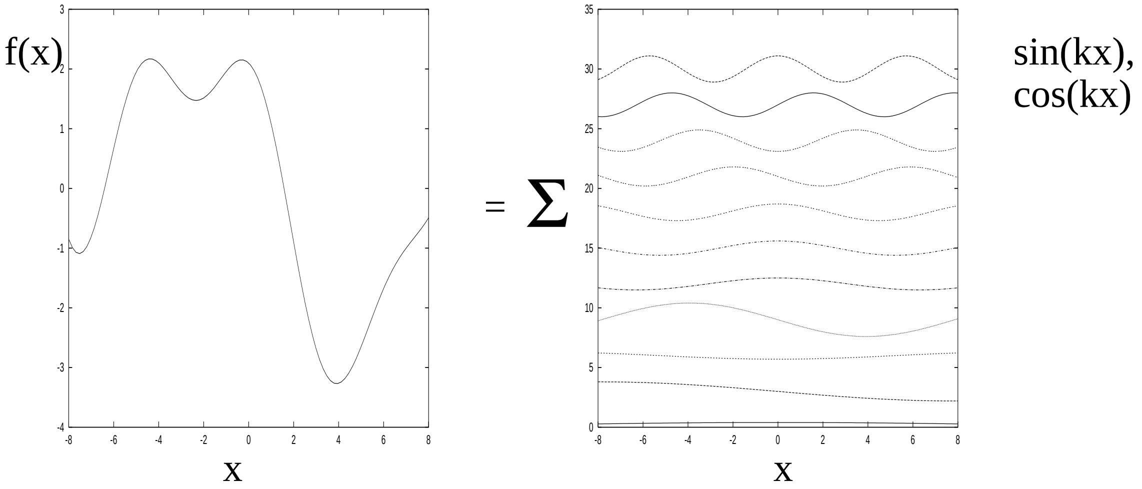
Eine Fourierzerlegung ist immer dann günstig, wenn die harmonischen Funktionen in irgendeinerweise ausgezeichnet sind und eine besonders einfache mathematische Behandlung zulassen. Dies ist in sehr vielen physikalischen Theorien der Fall. In der Physik I Vorlesung haben Sie das schon bei der Untersuchung des getriebenen harmonischen Oszillators ausgenutzt, indem Sie eine harmonisch oszillierende Kraft $f(t) = f_0 \cos(\omega t)$ betrachtet haben.

Wir wollen zeigen, dass eine Fourierzerlegung für alle Funktionen möglich sind, die im Unendlichen verschwinden, $\lim_{x\to\pm\infty} f(x) = 0$. Dabei werden wir eine Formel für die Koeffizienten herleiten.

Wir beginnen damit, dass wir die eigentlich kontinuierliche Funktion $f(x)$ diskretisieren. Dazu geben wir uns ein großes $N \in \mathbf{N}$ vor und definieren eine Menge $\mathbf{X}$ von diskreten Stützstellen im Abstand $a = L/2N$ auf dem Intervall [-L/2, L/2],

$$\mathbf{X} = \{-L/2, -L/2 + a, \ldots, L/2 - 2a, L/2 - a\} \ .$$

Am Ende unserer Diskussion führen wir den Limes $L \to \infty$ und zugleich $a \to 0$ aus und gelangen so wieder zu kontinuierlichen Funktionen auf ganz $\mathbf{R}$. Desweiteren verlangen wir, dass $f(-L/2) = f(L/2)$. Diese periodische Randbedingung erleichtert



AbbildungŸ1.1: Die Funktion $f(x)$ links ist die Überlagerung der rechts abgebildetetn harmonischen Funktionen. (Zur besseren Sichtbarkeit sind zu den harmonischen Funktion verschiedene Konstanten addiert.)

die mathematische Argumentation, stellt im Limes $L \to \infty$ neben der Bedingung $\lim_{x\to\pm\infty} f(x) = 0$ aber keine weitere Einschränkung dar.

Analog zur Menge $\mathbf{K}$ von $2N$ Stützstellen definieren wir eine Menge von $2N$ diskreten Wellenzahlen,

$$\mathbf{K} = \{-K, -K + \delta k, -K + 2\delta k, \ldots, K - 2\delta k, K - \delta k\} ,$$

wobei die maximale Wellenzahl K und der Wellenzahlabstand δk durch

$$K = \frac{\pi}{a} = \frac{2\pi N}{L} \quad \text{und} \quad \delta k = \frac{2\pi}{L}$$

gegeben sind. Die Wahl von δk resultiert aus der periodischen Randbedingung, $f(-L/2) = f(L/2)$, die auf der rechten Seite von (1.12) erfüllt ist, wenn

$$e^{-ikL/2} = e^{ikL/2} \quad \Leftrightarrow \quad 1 = e^{ikL} \quad \Leftrightarrow \quad kL = 2\pi n, \quad \text{mit ganzzahligem } n \in \mathbf{Z} .$$

Hieraus folgt, dass nur Wellenzahlen der Form $k = \frac{2\pi}{L} n = \delta k n$ erlaubt sind. Die maximale Wellenzahl K ergibt sich aus der Einsicht, dass bei vorgegebender räumlicher Diskretisierungslänge $a = L/2N$ eine Oszillation mit Wellenlänge $2\pi/k$ kleiner $2a$ nicht mehr aufgelöst werden kann. Daher $|k| \leq \pi/a \equiv K$.

Mit den genauen Definitionen von Stützstellen $\mathbf{X}$ und Wellenzahlen $\mathbf{K}$ ist die meiste Arbeit getan. Wir behaupten und werden gleich beweisen, dass mit Fourierkoeffizienten

$$\hat{f}(k) := a \sum_{x \in \mathbf{X}} f(x) e^{-ikx}, \quad k \in \mathbf{K}, \tag{1.13}$$

für alle $x \in X$ die Gleichung

$$f(x) = \frac{\delta k}{2\pi} \sum_{k \in \mathbf{K}} \hat{f}(k) e^{ikx} \tag{1.14}$$

erfüllt ist. Dies ist die gesuchte Fourierzerlegung von $f(x)$. Gleichung (1.13) beschreibt die **diskrete Fouriertransformation** von $f(x)$ nach $\hat{f}(k)$. Entsprechend beschreibt Gl. (1.14) die inverse diskrete Fouriertransformation von $\hat{f}(k)$ zurück auf die Funktion $f(x)$. Die diskrete Fouriertransformion wird in numerischen Anwendungen oft benutzt.

Als mathematisches Werkzeug ist die **kontinuierliche Fouriertransformation** wesentlich besser geeignet. Wir erhalten sie aus der diskreten Transformation im Grenzfall $L \to \infty$ und zugleich $a \to 0$. Da a der Abstand der Stützstellen in $\mathbf{X}$ ist und δk der Abstand der Wellenzahlen in $\mathbf{K}$, können im betrachten Grenzfall die Summationen durch folgende Integrale ersetzt werden:

$$a \sum_{x \in \mathbf{X}} \longrightarrow \int dx\,, \qquad \frac{\delta k}{2\pi} \sum_{k \in \mathbf{K}} \longrightarrow \int \frac{dk}{2\pi}\,.$$

Damit erhalten wir aus Gl. (1.13) die Fouriertransfomierte $\hat{f}(k)$ zu

$$\hat{f}(k) = \int dx f(x) e^{-ikx}, \quad k \in \mathbf{R},$$

und die inverse Transformation aus Gl. (1.14) zu

$$f(x) = \int \frac{dk}{2\pi} \hat{f}(k) e^{ikx}\,, \quad x \in \mathbf{R}\,.$$

Diese Gleichungen definieren die kontinuierliche Fouriertransformation einer auf ganz $\mathbf{R}$ definierten Funktion $f(x)$ in eine ebenfalls auf ganz $\mathbf{R}$ definierten Funktion $\hat{f}(k)$ [3].

Fouriertransformation

$$f(x) \quad \xrightarrow{F.T.} \quad \hat{f}(k) = \int \mathrm{d}x f(x) e^{-ikx} \quad \xrightarrow{(F.T.)^{-1}} \quad f(x) = \int \frac{\mathrm{d}k}{2\pi} \hat{f}(k) e^{ikx}$$

Nun zum Beweis der durch die Gleichungen (1.13) und (1.14) definierten diskreten Fouriertransformation: Für ein beliebiges $x_0 \in X$ haben wir offenbar zu zeigen, dass

$$f(x_0) \stackrel{!}{=} \frac{1}{L} \sum_{k \in \mathbf{K}} \left(a \sum_{x \in \mathbf{X}} f(x) e^{-ikx} \right) e^{ikx_0}\,.$$

Nach Vertauschung von x- und k-Summation wird die rechte Seite zu

$$\sum_{x \in \mathbf{X}} f(x) \cdot \underbrace{\frac{a}{L} \sum_{k \in \mathbf{K}} e^{ik(x_0 - x)}}_{\stackrel{!}{=} \delta_{x_0, x}}$$

[3] In einer anderen gängigen Definition der Fouriertransformation wird der Faktor $\frac{1}{2\pi}$ gleichmäßig auf beide Integrale aufgeteilt: statt dx und $\frac{dk}{2\pi}$ dann $\frac{dx}{\sqrt{2\pi}}$ und $\frac{dk}{\sqrt{2\pi}}$ in den jeweiligen Integralen. Beide Konventionen haben ihre Vor- und Nachteile.

Es bleibt zu zeigen, dass die k-Summe mit dem Vorfaktor a/L genau eins ergibt falls $x = x_0$ und andernfalls verschwindet.

Zuerst betrachten wir den Fall $x_0 = x$:

$$\frac{a}{L}\sum_{k\in\mathbf{K}} e^{ik\cdot 0} = \frac{a}{L}\sum_{k\in\mathbf{K}} 1 = \frac{a}{L}2N = 1\,,$$

da $\mathbf{K}$ genau $2N$ Wellenzahlen enthält und $a = L/2N$.

Für $x \neq x_0$ ist die Differenz ein ganzahliges Vielfachtes von a, also $x_0 - x = an$ mit $n \in \mathbf{Z} - \{0\}$. Damit

$$\begin{aligned}\sum_{k\in\mathbf{K}} e^{ik(x_0-x)} = \sum_{k\in\mathbf{K}} e^{ikan} &= \sum_{l=0}^{2N-1} e^{i(-K+l\cdot\delta k)an} \\ &= e^{-iKan}\sum_{l=0}^{2N-1}(\underbrace{e^{i\delta k\cdot an}}_{q})^l \\ &= e^{-iKan}\frac{1-q^{2N}}{1-q} \stackrel{!}{=} 0\,.\end{aligned}$$

Im letzten Schritt zeigen wir, dass $q^{2N} = 1$ und daher obige Summe verschwindet:

$$q^{2N} = e^{i\delta k\, an\, 2N} = e^{i\frac{2\pi}{L}\frac{L}{2N}n\,2N} = e^{i2\pi n} = \cos(2\pi n) + i\sin(2\pi n) = 1\,.$$

Damit ist die Behauptung gezeigt.

Beispiele:

(1) Fouriertransformierte der Delta-Distribution $f(x) = \delta(x)$:
Zur Erinnerung: Die Delta-Distribution $\delta(x)$ hat die definierende Eigenschaft $\int \delta(x-a)h(x)dx = h(a)$ für beliebige (stetige) Funktionen $h(x)$ (vgl. Skript zur Physik I 2003/4, Def. 9.4.). Wir erhalten die Fouriertransformierte

$$\hat{f}(k) = \int \mathrm{d}x\delta(x)\exp(-ikx) = e^{-ik\cdot 0} = 1\,.$$

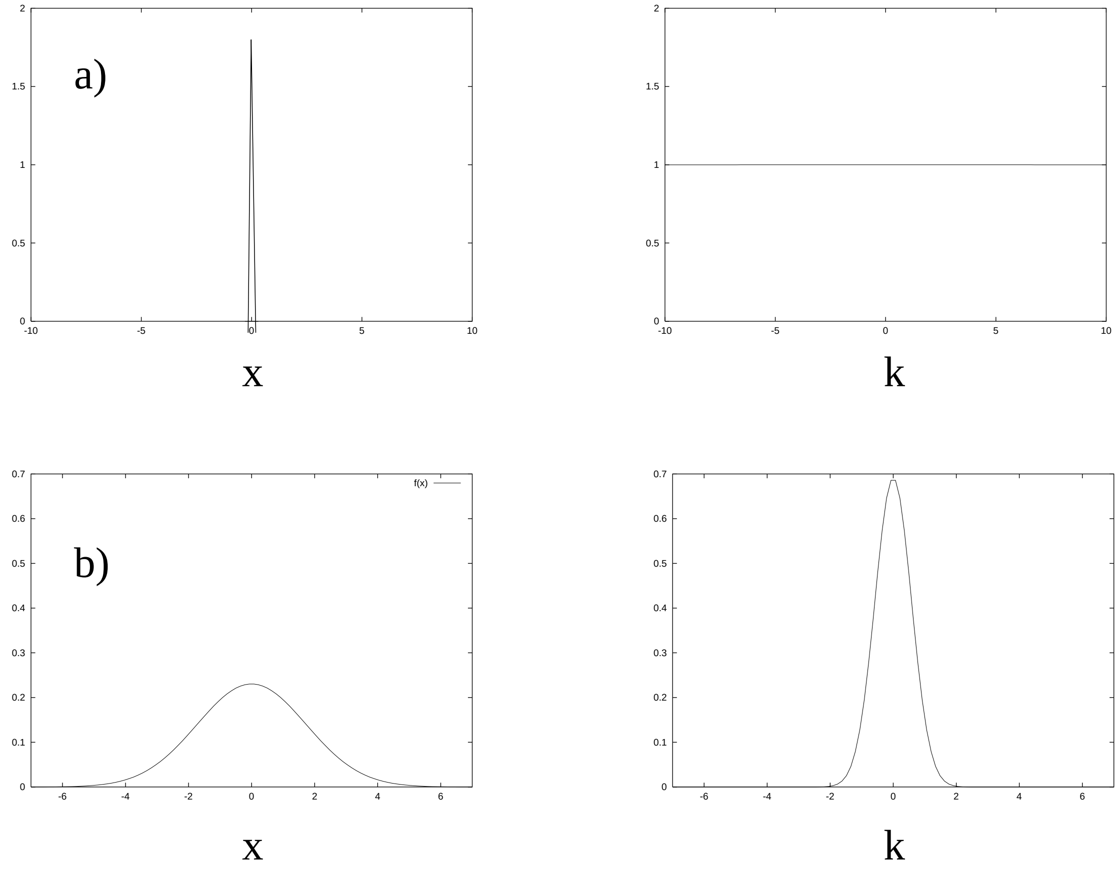
Die Fouriertansformation bildet also den unendlich schmalen δ-Peak bei $x = 0$ auf die konstante Funktion 1 ab (siehe Abb. ??).

(2) Verschobene Delta-Distribution $f_a(x) = \delta(x-a)$:

$$\hat{f}_a(k) = \int \mathrm{d}x\delta(x-a)\exp(-ikx) = e^{-ika}\,.$$

Die Fouriertransformierte $\hat{f}_a$ der um a in x-Richtung verschobenen Funktion f_a unterscheidet sich von der Fouriertransformierten $\hat{f}$ der ursprünglichen Funktion f um einen Faktor e^{-ika}. Diese hier am Spezialfall der Delta-Distribution gefundene Regel hat allgemeine Gültigkeit. Denn:

$$\begin{aligned}\hat{f}_a(k) &= \int dx f_a(x)e^{-ikx} = \int dx f(x-a)e^{-ikx} = \int dx f(x)e^{-ik(x+a)} \\ &= e^{-ika}\int dx f(x)e^{-ikx} = e^{-ika}\hat{f}(k)\,.\end{aligned}$$



AbbildungŸ1.2: a) Die Delta-Distribution $\delta(x)$ (links) wird unter Fouriertransformation auf die konstante Funktion $\hat{\delta}(k) = 1$ (rechts) abgebildet. b) Eine Gaußfunktion mit Varianz $\sigma = 3$ (links) wird unter Fouriertransformation auf eine Gaußfunktion mit Varianz $\hat{\sigma}^2 = 1/3$ abgebildet (rechts).

(3) Gaußfunktion $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2})$**:**

$$\hat{f}(k) = \int dx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} e^{-ikx} = \frac{1}{\sqrt{2\pi\sigma^2}} \int dx e^{-\frac{1}{2\sigma^2}x^2 - ikx} = \exp(-\frac{\sigma^2 k^2}{2}) \,. \quad (1.15)$$

[Hierbei wurde benutzt:

$$\begin{aligned} \int dx e^{-ax^2+bx} &\overset{x=t/\sqrt{a}}{=} \int \frac{dt}{\sqrt{a}} e^{-t^2+\frac{b}{\sqrt{a}}t} = \frac{1}{\sqrt{a}} e^{\frac{b^2}{4a}} \int e^{-(t-\frac{b}{2\sqrt{a}})^2} \\ &\overset{t=s+\frac{b}{2\sqrt{a}}}{=} \frac{1}{\sqrt{a}} e^{\frac{b^2}{4a}} \int ds e^{-s^2} = \sqrt{\frac{\pi}{a}}\, e^{\frac{b^2}{4a}} \,.] \end{aligned}$$

Die Fouriertransfomierte einer Gaußfunktion ist also wieder proportional einer Gaußfunktion. Dabei ist die Varianz $\hat{\sigma}^2 = 1/\sigma^2$ von $\hat{f}(k)$ reziprok zur Varianz σ^2 von $f(x)$.

1.5 Charakteristische Funktion

Zum Beweis des Zentralen Grenzwertsatzes benötigen wir die **charakteristische Funktion** $\phi(k)$ einer Zufallsgröße X: Sie ist für $k \in \mathbf{R}$ definiert durch

$$\phi(k) := \left\langle e^{ikX} \right\rangle .$$

Ist $f(x)$ die Wahrscheinlichkeitsverteilung von X, so ist $\phi(-k)$ gerade die Fouriertransformierte von X:

$$\phi(-k) = \left\langle e^{-ikX} \right\rangle \equiv \int dx\, f(x) e^{-ikX} \equiv \hat{f}(k) .$$

Aufgrund der Eindeutigkeit der Fouriertransformation ist mit der charakteristische Funktion $\phi(k)$ die Wahrscheinlichkeitwverteilung $f(x)$ und somit die Zufallsgröße X eindeutig bestimmt.

Vertauschen wir die Summation der Exponentialreihe $e^x = \sum_n x^n/n!$ und die Erwartungswertbildung, so erhalten wir

$$\phi(k) = \left\langle \sum_{n=0}^{\infty} \frac{(ikX)^n}{n!} \right\rangle = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \mu_n .$$

Die Momente $\mu_n = \langle X^n \rangle$ legen die charakteristische Funktion $\phi(k)$ fest, die, wie gerade festgestellt, wiederum die Zufallsgröße X bestimmt.

Wir halten fest:

Eine Zufallsgröße X ist durch seine Momente $\mu_n = \langle X^n \rangle$ eindeutig bestimmt.

[Die Darstellung von $\phi(k)$ durch die Momente μ_n ist allerdings *nicht* mehr möglich, wenn die μ_n divergieren!]

Wir beweisen nun den Zentralen Grenzwertsatz (vgl. Abschnitt (1.3)) indem wir die charakteristische Funktion Φ_N der Zufallsgröße

$$\tilde{X} = X - \mu = \frac{1}{N} \sum_{i=1}^{N} (\underbrace{X_i - \mu_i}_{\tilde{X}_i})$$

im Limes $N \to \infty$ bestimmen. X, X_i und μ_i sind wie im Abschnitt 1.3 definiert. Die Zufallsgrößen $\tilde{X}_i := X_i - \mu_i$ sind mit den X_i nach Voraussetzung statistisch unabhängig. Da sich $\tilde{X}$ von X nur um eine triviale Verschiebung des Mittelwerts um μ auf $\langle \tilde{X} \rangle = 0$ unterscheidet, können wir später ohne weiteres von der Verteilung von $\tilde{X}$ problemlos auf die Verteilung von X schließen.

Die charakteristische Funktion lautet

$$\begin{aligned}
\Phi_N(k) &= \left\langle e^{ik\tilde{X}} \right\rangle = \left\langle e^{\frac{ik}{N}\sum_j \tilde{x}_j} \right\rangle = \int d^N\underline{x}\tilde{f}(x_1, x_2, \ldots, X_N) e^{\frac{ik}{N}\sum_j X_j} \\
&\overset{s.U.}{=} \int d^N\underline{x}\tilde{f}(x_1)\tilde{f}(x_2)\ldots\tilde{f}(x_N) e^{\frac{ik}{N}x_1} e^{\frac{ik}{N}x_2}\ldots e^{\frac{ik}{N}x_N} \\
&= \prod_{j=1}^{N} \int dx_j \tilde{f}(x_j) e^{i\frac{k}{N}x_j} = \prod_{j=1}^{N} \left\langle e^{i\frac{k}{N}\tilde{X}_j} \right\rangle = \prod_{j=1}^{N} \phi_j(\frac{k}{N})\,. \qquad (1.16)
\end{aligned}$$

Hierbei haben wir die statistische Unabhängigkeit der Zufallsgrößen $\tilde{X}_i$ ausgenutzt. Wir sehen, dass $\Phi_N(k)$ das Produkt der charateristischen Funktionen ϕ_j der Zufallsgrößen $\tilde{X}_j$ mit Argument k/N ist. Im Folgenden beschaffen wir uns mittels einer Taylerentwicklung einen genäherten Ausdruck für die charakteristischen Funktionen $\phi(k/N)$, der für $N \to \infty$ exakt wird.

[**Taylorentwickclung:** Für eine $n+1$ mal differenzierbare Funktion gilt folgende Formel:

$$\begin{aligned}
f(y+h) &= f(y) + \frac{f'(y)}{1!}h + \frac{f''(y)}{2!}h^2 + \cdots + \frac{f^{(n)}(y)}{n!}h^n + \\
&\quad + \frac{1}{(n+1)!} f^{(n+1)}(y+\vartheta h)\, h^{n+1}\,,
\end{aligned}$$

wobei $\vartheta \in]0,1[$. Diese Gleichung ist exakt, enthält dafür aber den unbekannten Parameter ϑ im letzten Term, dem sog. Restglied. Es erlaubt die Fehlerabschätzung, wenn die Entwicklung nach dem Term n-ter Ordnung ohne Restglied abgebrochen wird. Diese Formel wird in der Analysisvorlesung bewiesen.]

In unserem Fall entwicklen wir $e^{i\frac{k}{N}\tilde{X}_j} \equiv e^{i(y+h)}$ um $y = 0$ und $h = \frac{k}{N}\tilde{X}_j$ bis in zweiter Ordnung plus Restglied:

$$e^{i\frac{k}{N}\tilde{X}_j} = 1 + i\frac{k\tilde{X}_j}{N} - \frac{1}{2}\frac{(k\tilde{X}_j)^2}{N^2} + \underbrace{\frac{-ie^{i\frac{k}{N}\tilde{X}_j\vartheta}}{6}\frac{(k\tilde{X}_j)^3}{N^3}}_{=:\Delta}\,. \qquad (1.17)$$

Vernachlässigen wir die Terme von höherer als zweiter Ordnung, handeln wir uns einen Fehler Δ ein, dessen mittlerer Betrag wir unter Zuhilfenahme der Voraussetzung $\left\langle |\tilde{X}_j|^3 \right\rangle < c$ durch

$$\langle|\Delta|\rangle = \left\langle \frac{1}{6}\frac{|k\tilde{X}_j|^3}{N^3} \right\rangle < \frac{1}{6}\frac{|k|^3}{N^3}c$$

abschätzen können. Für hinreichend große N ist dieser Fehler beliebig klein im Vergleich zum gemittelten ersten und dritten Term in (1.17); er kann daher guten Gewissens vernachlässigt werden.

In dieser Näherung erhalten wir folgenden Ausdruck für die charakteristischen Funktionen ϕ_j:

$$\phi_j(\frac{k}{N}) = \left\langle 1 + i\frac{k\tilde{X}_j}{N} - \frac{1}{2}\frac{(k\tilde{X}_j)^2}{N^2} \right\rangle = \langle 1\rangle + i\frac{k}{N}\langle\tilde{X}_j\rangle - \frac{1}{2}\frac{(k)^2}{N^2}\langle\tilde{X}_j^2\rangle\,.$$

Wir nutzen nun aus, dass $\langle \tilde{X}_j \rangle = 0$, $\langle \tilde{X}_j^2 \rangle = \langle (X_j - \mu_j)^2 \rangle = \sigma_j^2$, und erhalten

$$\phi_j(\frac{k}{N}) = 1 - \frac{k^2\sigma_j^2}{2N^2} .$$

Dies setzen wir in die oben gefundene Beziehung (1.16) ein:

$$\Phi_N(k) \quad = \quad \prod_{j=1}^{N} (1 - \frac{k^2\sigma_j^2}{2N^2}) = 1 - \sum_{j=1}^{N} \frac{k^2\sigma_j^2}{2N^2} + \mathcal{O}(1/N^3)$$

Mit der mittleren Varianz $\bar{\sigma}^2 := \frac{1}{N}\sum_{j=1}^{N} \sigma_j^2$ schreibt sich dies als

$$\Phi_N(k) = 1 - \frac{\bar{\sigma}^2 k^2}{2N} + \mathcal{O}(1/N^3) \stackrel{!}{=} \exp\left(-\frac{\bar{\sigma}^2 k^2}{2N}\right) + \mathcal{O}(1/N^3) .$$

Im Limes $N \to \infty$ erhalten wir damit die Fouriertransformierte $\hat{f}_N(k)$ der Wahrscheinlichkeitsverteilung $\tilde{f}_n(x)$ zu

$$\tilde{f}_N(k) \equiv \Phi_N(-k) = \exp\left(-\frac{\bar{\sigma}^2 k^2}{2N}\right) .$$

Vergleich mit der Fouriertransformierten einer Gausßschen Wahrscheinlichkeitsverteilung (1.15) zeigt uns, dass die gesuchte Wahrscheinlichkeitsverteilung $\tilde{f}_N(x)$ von $\tilde{X}$ durch

$$\tilde{f}_N(x) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2/N}} \exp\left(-\frac{x^2}{2\bar{\sigma}^2/N}\right)$$

gegeben ist; eine Gaußsche Verteilung mit $\sigma = \frac{1}{\sqrt{N}}\bar{\sigma}$. Die Wahrscheinlichkeitsverteilung der Zufallsgröße $X = \tilde{X} + \mu$ ist demnach

$$f_N(x) = \tilde{f}_N(x - \mu) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2/N}} \exp\left(-\frac{(x-\mu)^2}{2\bar{\sigma}^2/N}\right)$$

Damit ist der zentrale Grenzwertsatz bewiesen.

Fortsetzung folgt!