

Übungsbegleiter

Inhaltsverzeichnis

Vorwort	3
1 Wahrscheinlichkeitsmaße und Mengensysteme	4
2 Bedingte Wahrscheinlichkeit und der Satz von Bayes	7
3 Zufallsvariablen	10
4 Verteilungen	15
4.1 Diskrete Zufallsvariablen	15
4.2 Absolutstetige Zufallsvariablen	19
5 Zufallsvektoren	23
Bibliografie	26

Vorwort

Dieser Aufschrieb soll als Begleitinformationen zu den Übungsblättern verstanden werden. Ziel ist es, Inhalte aus den Übungen tiefer zu beleuchten als es das Vorrechnen im regulären Übungsbetrieb erlaubt. Insbesondere versuche ich neben dem rigorosen mathematischen Verständnis aus den Übungsaufgaben ein “intuitives” und weniger formellastiges Verständnis zu fördern.

Gemäß der Natur der Sache müssen einige Inhalte aus dem Skript wiederholt werden. Dennoch kann und soll dieser Aufschrieb das Vorlesungsskript (Spodarev 2020) bzw. die ältere, aber nicht ganz deckungsgleiche Version (Spodarev 2018) nicht ersetzen. Ebenso ist dies nicht als eine Art Zusammenfassung der “wichtigen Prüfungsinhalte” zu verstehen. Es ist lediglich mein Anliegen, einzelne grundlegende Aspekte auf eine weniger formale Weise zu beleuchten, da diese Dinge - meiner Erfahrung nach - notgedrungen durch Zeitmangel und Stoffdichte oftmals im Vorlesungs- und Übungsbetrieb zu kurz kommen.

In den vergangenen Semester habe ich versucht, diese Informationen über zusätzliche Folien in den Übungen darzustellen. Damit diese sich allerdings auch ohne meinen zugehörigen Vortrag als Lernunterlage eignen, musste ich einige Inhalte in Textform auf den Folien beschreiben. Ich habe zwar versucht, Texte in den Folien auf ein Minimum zu beschränken, dennoch musste ich - zu meinem eigenen Unwohl - feststellen, dass ich für meinen Geschmack immer noch zu viel Text auf den Folien einbringen musste.

Ich bin einer der größten Kritiker an Vorträgen mit vollgeschriebenen Folien. Leider sind solche Vorträge meiner Erfahrung nach leider eher die Norm als die Ausnahme. Daher bin ich erfreut darüber, diesen Übungsbegleiter als Lernunterlage aushändigen und meine Übungsfolien dadurch schlanker gestalten zu können.

Ob der Übungsbegleiter letztendlich eine sinnvolle Ergänzung zum Übungsbetrieb darstellt, dürft ihr als Übungsteilnehmer selbst entscheiden. Ich bin für Feedback immer offen und nehme auch Kritik (insbesondere bzgl. zu starken Ungenauigkeiten durch die umgangssprachliche Beschreibung mathematischer Inhalte) gerne an. Ihr dürft euch per Mail oder Übungsfeedback im Moodle immer mit euren Anregungen melden.

Abschließend möchte ich mich an dieser Stelle herzlichst bei [Jun.-Prof. Dr. Marco Oesting](#) für die aufmerksame Korrektur dieses Textes bedanken.

1 Wahrscheinlichkeitsmaße und Mengensysteme

Das stochastische Grundgerüst in einem Grundraum E basiert auf Mengen bzw. Ereignissen in einem Wahrscheinlichkeitsraum, der als ein Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ definiert ist, das die Menge von Elementarereignissen bzw. die Grundmenge $\Omega \subset E$ mit einer σ -Algebra \mathcal{F} und einem Wahrscheinlichkeitsmaß \mathbb{P} kombiniert. Dabei sind die zwei hier noch unbekannten Begriffe definiert durch

Definition 1.1 (σ -Algebra). Ein Mengensystem $\mathcal{F} \subset \mathcal{P}(\Omega)$ heißt σ -Algebra, wenn alle drei der folgenden Bedingungen erfüllt sind:

1. Es gilt $\emptyset \in \mathcal{F}$.
2. Falls $A \in \mathcal{F}$, so gilt auch $A^c \in \mathcal{F}$.
3. Falls $A_1, A_2, \dots \in \mathcal{F}$ so gilt auch $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

Definition 1.2 (Wahrscheinlichkeitsmaß). Sei \mathcal{F} eine σ -Algebra auf Ω . Eine Mengenfunktion $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}$ heißt Wahrscheinlichkeitsmaß auf (Ω, \mathcal{F}) , wenn

1. $\mathbb{P}(A) \in [0, 1]$ für alle $A \in \mathcal{F}$,
2. $\mathbb{P}(\Omega) = 1$ und
3. Falls $A_1, A_2, \dots \in \mathcal{F}$ paarweise disjunkt sind, so gilt

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Die σ -Algebra dient dabei als Sammelsurium von Mengen, denen wir eine Wahrscheinlichkeit zuordnen können und die wir daher als messbar bezeichnen. Der zentrale Punkt, warum wir das Konstrukt der σ -Algebra brauchen und uns bei dem zugrundeliegenden Mengensystem nicht mit der Potenzmenge $\mathcal{P}(\Omega)$ begnügen können, ist der Satz von Vitali. Grob formuliert besagt dieser, dass es Mengen gibt, die nicht messbar sind, d.h. diesen Mengen kann man kein sinnvolles Maß/Volumen zuordnen.

Stattdessen beschränkt man sich auf einige wenige Axiome, die messbare Mengen erfüllen sollten und landet damit bei der Definition der σ -Algebra, deren Bedingungen sich aus stochastischer Perspektive übersetzen lassen als

1. Einem unmöglichen Ereignis kann man eine Wahrscheinlichkeit zuordnen.
2. Falls man einem Ereignis eine Wahrscheinlichkeit zuordnen kann, so kann man auch dem Gegenereignis eine Wahrscheinlichkeit zuordnen.
3. Lassen sich einer abzählbaren Anzahl von Ereignissen jeweils eine Wahrscheinlichkeit zuordnen, so lässt sich auch dem Ereignis, dass irgendeines dieser Ereignisse eintritt, eine Wahrscheinlichkeit zuordnen.

Bei den bisherigen Überlegungen haben wir lediglich die Tatsache untersucht, ob man eine Wahrscheinlichkeit zuordnen kann und haben auf diese Weise die σ -Algebra beleuchtet. Jedoch kann man mit der σ -Algebra alleine keine sinnvolle Aussage über die Höhe der Wahrscheinlichkeit eines Ereignisses treffen.

Dafür benötigen wir eine Funktion, die genau diese Aufgabe erledigt und auch hier gehen wir wieder axiomatisch vor, damit die entsprechende Funktion möglichst intuitive Eigenschaften besitzt. Gerade diese Eigenschaften finden sich in der Definition des Wahrscheinlichkeitsmaßes und lassen sich wieder aus stochastischer Perspektive übersetzen als

1. Jede Wahrscheinlichkeit liegt zwischen 0 und 1.
2. Die Wahrscheinlichkeit, dass irgendetwas passiert ist 1.
3. Falls man beliebig viele Ereignisse betrachtet, die nicht gleichzeitig eintreten können, so ergibt sich die Wahrscheinlichkeit für das Ereignis, dass eins dieser Ereignisse eintritt, als die Summe der Wahrscheinlichkeiten der einzelnen Ereignisse.

Betrachten wir nun zwei elementare Wahrscheinlichkeitsräume.

Beispiel 1.1. Sei $\Omega \subset \mathbb{R}$ und A eine nichtleere Teilmenge von Ω . Dann ist $(\Omega, \{\emptyset, \Omega, A, A^c\}, \mathbb{P})$ mit

$$\mathbb{P}(B) = \begin{cases} 0 & , B = \emptyset \\ 1 & , B = \Omega \\ p & , B = A \\ 1 - p & , B = A^c \end{cases}$$

ein Wahrscheinlichkeitsraum, falls $p \in [0, 1]$.

Beispiel 1.2 (Borelsche σ -Algebra). Es sei $\mathcal{B}([0, 1])$ die Borel- σ -Algebra auf $[0, 1]$. Grob (und sehr vereinfacht) gesagt beinhaltet die Borel- σ -Algebra fast alle Mengen, die irgendwie von Interesse sein könnten. In diesem Fall beschränken wir uns dabei auf Teilmengen aus dem Intervall $[0, 1]$. Zeitgleich sind diese Mengen so “gutartig”, dass man ihnen einen Volumen bzgl. des Lebesgue-Maßes λ zuordnen kann.

Das Lebesgue-Maß ist hierbei ein elementares Maß, das einem Intervall dessen Länge als Maß zuordnet, d.h. $\lambda([a, b]) = b - a$ für alle $a < b$. Es gilt nun, dass $([0, 1], \mathcal{B}([0, 1]), \lambda)$ ein Wahrscheinlichkeitsraum ist.

2 Bedingte Wahrscheinlichkeit und der Satz von Bayes

Idealerweise ist ein medizinischer Test immer zu 100% zuverlässig. Allerdings wissen wir aus der Praxis, dass es immer wieder zu falsch-positiven oder falsch-negativen Testergebnissen in der Medizin kommt. Das beste worauf wir also hoffen können, ist das ein Test mit sehr großer Wahrscheinlichkeit richtig liegt. Somit sollte ein Test so gestaltet sein, dass die Wahrscheinlichkeit, dass ein Patient tatsächlich krank ist, wenn der Test dies anzeigt, sehr groß ist.

Intuitiv ist das alles sehr einleuchtend. Allerdings sollten wir das auch noch mathematisch formal präzisieren. Offensichtlich können wir die gesuchte Wahrscheinlichkeit nicht einfach mit $\mathbb{P}(K)$ bezeichnen, wobei K für das Ereignis steht, dass ein zufällig ausgewählter Mensch krank ist. Schließlich haben wir zusätzliche Information. Unser Test war ja positiv. Dementsprechend würden wir erwarten, dass es nun wahrscheinlicher ist, dass der Patient krank.

Ein naiver Ansatz wäre nun also zu vermuten, dass die gesuchte Wahrscheinlichkeit $\mathbb{P}(K \cap T)$ entspricht, wobei T dem Ereignis “Test positiv” und der Schnitt der Mengen dem Ereignis “Beide Ereignisse eingetreten” entspricht. Offensichtlich stellt man schnell fest, dass $\mathbb{P}(K \cap T) \leq \mathbb{P}(K)$, was wiederum bedeuten würde, dass wir es für weniger wahrscheinlich halten, krank zu sein, nachdem der Test positiv ausgefallen ist. Das kann nicht unserer Intuition entsprechen.

Allerdings können wir uns überlegen, dass wir zusätzlich zu unseren vorherigen Überlegung den Grundraum auf die Fälle einschränken können, in denen T wahr ist. In diesem Fall muss ich die Wahrscheinlichkeit $\mathbb{P}(K \cap T)$ nur noch entsprechend der Wahrscheinlichkeit des neuen Grundraumes normieren, was uns die gesuchte Wahrscheinlichkeit $\mathbb{P}(K \cap T)/\mathbb{P}(T)$ liefert und zur allgemeinen Definition der bedingten Wahrscheinlichkeit führt.

Definition 2.1 (Bedingte Wahrscheinlichkeit). Es seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$ seien zwei Ereignisse, wobei $\mathbb{P}(B) > 0$. Dann ist die bedingte Wahrscheinlichkeit von A unter der Bedingung B definiert als

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

In diesem Zusammenhang bietet es sich auch an, dass Konzept von Unabhängigkeit zu erwähnen. Intuitiv ergibt es Sinn zwei Ereignisse als (stochastisch) unabhängig voneinander zu betrachten,

wenn die Wahrscheinlichkeit, dass beide Ergebnisse eintreten, sich als Produkt der beiden Einzelwahrscheinlichkeiten ergibt. Formal definieren wir

Definition 2.2 (Unabhängigkeit). Es seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$ seien zwei Ereignisse. Die Ereignisse A, B sind genau dann unabhängig, falls $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Verknüpft man dies mit Definition 2.1 so stellt man fest, dass für unabhängige Mengen A, B gilt, dass $\mathbb{P}(A|B) = \mathbb{P}(A)$ und $\mathbb{P}(B) = \mathbb{P}(B|A)$. In der Praxis stellt man oft fest, dass wir zwar an einer bedingten Wahrscheinlichkeit $\mathbb{P}(A|B)$ interessiert sind, aber nur die Wahrscheinlichkeit $\mathbb{P}(B|A)$ messen können.

Beispiel 2.1. Aus medizinischen Studien wissen wir, dass Rauchen das Risiko erhöht, an Lungenkrebs zu erkranken. Die untersuchte Größe ist hier also $\mathbb{P}(K|R)$, wobei K für das Ereignis “Patient hat Krebs” und R für das Ereignis “Patient ist Raucher” steht. Jedoch können wir in Studien nur Krebspatienten als Raucher bzw. Nichtraucher identifizieren, d.h. durch unsere Beobachtungen erhalten wir nur Aufschluss über die Wahrscheinlichkeit $\mathbb{P}(R|K)$.

Um dennoch die gesuchte Größe zu berechnen, hilft uns der Satz von Bayes.

Theorem 2.1 (Bayes). Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$ zwei Ereignisse mit $\mathbb{P}(A), \mathbb{P}(B) > 0$. Dann gilt

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

Somit stellt der Satz von Bayes, benannt nach dem Pfarrer Thomas Bayes, fest, dass die bedingten Wahrscheinlichkeiten proportional zueinander sind, wenn man Bedingung und zu untersuchendes Ereignis miteinander vertauscht. Zugleich liefert der Satz von Bayes den passenden Proportionalitätsfaktor dazu.

Ein letztes nützliches Werkzeug in diesem Zusammenhang ist der Satz über die totale Wahrscheinlichkeit. Dieser ermöglicht es, die Wahrscheinlichkeit eines Ereignisses durch die Zerlegung des Grundraumes als gewichtetes Mittel der zugehörigen bedingten Wahrscheinlichkeiten zu berechnen.

Theorem 2.2 (Totale Wahrscheinlichkeit). *Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und A, B_1, B_2, \dots seien Ereignisse mit $\mathbb{P}(B_i) > 0$ für alle $i \in \mathbb{N}$. Außerdem sind die B_i paarweise disjunkt und $\cup_{i \in \mathbb{N}} B_i = \Omega$. Dann gilt*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Nach dem Satz von Bayes gilt $\mathbb{P}(A|B) \approx \mathbb{P}(B|A)$ genau dann, wenn $\mathbb{P}(A) \approx \mathbb{P}(B)$. Das heißt im Allgemeinen ist es nicht möglich, die bedingte Wahrscheinlichkeit zu invertieren. Dennoch passiert diese Fehleinschätzung im Alltag so häufig, dass dieser logische Fehler auch als **“Verwechslung der Umgekehrung”**, **“Umkehrungsfehlschluss”** oder **“Trugschluss des Anklägers”** bekannt ist. Eins der bekanntesten Beispiele dafür beruht dabei auf einer Studie mit 100 Ärzten in der folgendes Fallbeispiel durchgespielt wurde:

Beispiel 2.2 ((Eddy 1982)). Eine Patientin hat einen Knoten in der Brust woraufhin eine Mammographie durchgeführt wurde deren Befund zurückliefert, dass die Patienten einen bösartigen Tumor hat. Weiterhin sei bekannt, dass

- 1% aller Brusttumore bösartig sind.
- Mammographien einen bösartigen Tumor in 80 % der Fälle als solchen erkennen und in den sonstigen Fällen als gutartig einstufen.
- Mammographien einen gutartigen Tumor in 90 % der Fälle als solchen erkennen und in den übrigen Fällen als bösartig einstufen.

Gesucht ist nun die bedingte Wahrscheinlichkeit $\mathbb{P}(B|P)$, dass $B = \text{“Tumor ist bösartig”}$ gegeben $P = \text{“Test auf bösartigen Tumor ist positiv”}$. In der Studie haben fast alle der Ärzte diese Wahrscheinlichkeit auf ungefähr 75 % geschätzt und liegen damit sehr nahe an $\mathbb{P}(P|B) = 80\%$. Die echte Wahrscheinlichkeit liefert der Satz von Bayes zusammen mit dem Satz der totalen Wahrscheinlichkeit als

$$\begin{aligned} \mathbb{P}(B|P) &= \frac{\mathbb{P}(B)\mathbb{P}(P|B)}{\mathbb{P}(P)} \\ &= \frac{\mathbb{P}(B)\mathbb{P}(P|B)}{\mathbb{P}(P|B)\mathbb{P}(B) + \mathbb{P}(P|B^c)\mathbb{P}(B^c)} \\ &= \frac{0.01 \cdot 0.80}{0.80 \cdot 0.01 + 0.1 \cdot 0.99} = 0.0748. \end{aligned}$$

Anmerkung. Das vorangegangene Beispiel wurde als Verdeutlichung für logische Trugschlüsse aufgeführt. Aus aktuellen Anlässen sei an dieser Stelle erwähnt, dass aus diesem fiktiven Beispiel keine Rückschlüsse auf aktuelle medizinische Praxis in Bezug auf Mammographien oder sonstige Untersuchungen möglich ist.

3 Zufallsvariablen

Im Jahr 1901 hat der italienische Mathematiker Mario Lazzarini das Buffonsche Nadelexperiment durchgeführt, um π zu approximieren. Angeblich baute er dafür eine Maschine, um eine Nadel 3408 mal zu werfen und zu zählen, wie oft die Nadel so fällt, dass sie ein Gitter aus parallelen Linien kreuzt.

Die theoretischen Grundlagen des Experiments sind unumstritten, jedoch werden Lazzarinis Ergebnisse teilweise bezweifelt¹. Hauptkritik besteht in der Frage, ob die Anzahl der Versuche gerade so gewählt wurde, um die damalige Approximation $\pi \approx \frac{355}{113}$ möglichst gut zu treffen. Außerdem wird daran gezweifelt, ob die Nadelwurfmaschine tatsächlich existierte und die Nadelwürfe überhaupt stattgefunden haben.

Letztendlich ist es für uns egal, ob Lazzarini die Nadel (selber) geworfen hat oder nicht. Interessant ist an dieser Stelle dennoch, dass der komplette Versuchsaufbau darauf ausgelegt ist, lediglich eine Zahl X zu ermitteln. Diese Zahl soll uns angeben, wie oft die Nadel das Liniengitter gekreuzt hat. Auf gewisse Weise fasst diese Zahl X den relevanten Teil des Experimentes zusammen und lässt alle irrelevanten Informationen wie z.B. die genaue Nadelposition fallen². Damit diese Zusammenfassung eines Zufallsexperiments durch eine Variable auf einer fundierten mathematischen Basis steht, müssen wir dies sauber definieren.

Definition 3.1 (Zufallsvariable). Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Dann heißt die Abbildung $X : \Omega \rightarrow \mathbb{R}$ Zufallsvariable, falls sie messbar bzgl. der Borel- σ -Algebra $\mathcal{B}(\mathbb{R})$ ist, d.h.

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{F}.$$

für alle $B \in \mathcal{B}(\mathbb{R})$.

Wie eben erwähnt ist eine Zufallsvariable also einfach eine Funktion, die ein Zufallsexperiment / Ereignis aus dem Grundraum nimmt und diesem Versuchsausgang eine reelle Zahl zuweist. Die einzige Bedingung, die wir zusätzlich stellen, ist die Existenz der Wahrscheinlichkeit $\mathbb{P}(X \in B)$ für alle $B \in \mathcal{B}(\mathbb{R})$, d.h. egal welchen Wert die Zufallsvariable annimmt, die Wahrscheinlichkeit, dass dies passiert, ist wohldefiniert. Diese Wohldefiniertheit wird dadurch gesichert, dass die zugehörige Menge in der σ -Algebra \mathcal{F} enthalten ist.

¹siehe [Wikipedia](#).

²no pun intended

Anmerkung. Falls uns mehr als ein einzelner Wert interessiert, so können wir die Werte in einem Vektor zusammenfassen. $X : \Omega \rightarrow \mathbb{R}$ ist dann eine Abbildung in den \mathbb{R}^n und statt Mengen $B \in \mathcal{B}(\mathbb{R})$ zu betrachten, müssen wir nun Mengen $B' \in \mathcal{B}(\mathbb{R}^n)$ untersuchen.

Beispiel 3.1. Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, wobei $\Omega = \{1, 2, 3, 4, 5\}$ und

$$\mathcal{F} = \{\emptyset, \Omega, \{1, 2, 3, 4\}, \{3, 4, 5\}, \{5\}, \{1, 2\}, \{1, 2, 5\}, \{3, 4\}\}.$$

a. Definiere $X : \Omega \rightarrow \mathbb{R}$, $X(\omega) = \omega$. Dann gilt $X^{-1}(\{3\}) = \{3\} \notin \mathcal{F}$ und somit ist X keine Zufallsvariable.

b. Definiere $Y : \Omega \rightarrow \mathbb{R}$, $Y(\omega) = \mathbb{1}_{\{3,4,5\}}(\omega)$ und bestimme die Urbilder

$$Y^{-1}(B) = \begin{cases} \emptyset & , 0 \notin B, 1 \notin B \\ \{1, 2\} & , 0 \in B, 1 \notin B \\ \{3, 4, 5\} & , 0 \notin B, 1 \in B \\ \Omega & , 0 \in B, 1 \in B \end{cases}$$

für $B \in \mathcal{B}(\mathbb{R})$. Also ist Y eine Zufallsvariable.

In diesen Fällen war es einfach, alle relevanten Urbilder händisch zu bestimmen. Jedoch kann dies für komplizierte Zufallsvariablen anders sein. Praktischerweise gibt es ein Theorem, dass uns direkt die Messbarkeit einer Abbildung X gibt, indem nur eine bestimmte Klasse von Urbildern betrachtet wird.

Theorem 3.1. *Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Abbildung auf dem Grundraum in die reelle Zahlen. X ist eine Zufallsvariable genau dann, wenn $\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$ für alle $x \in \mathbb{R}$.*

Üblicherweise interessiert es uns in der Stochastik gar nicht so sehr, wie die Abbildung X in Abhängigkeit von $\omega \in \Omega$ definiert ist. Tatsächlich untersuchen wir hauptsächlich die Frage, welche Werte die Zufallsvariable annimmt und wie wahrscheinlich dies für konkrete Werte oder ganze Bereiche ist. Konkret begnügen wir uns damit “nur” die so genannte Verteilung von X zu kennen.

Definition 3.2 (Verteilung). Es sei X eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Die Funktion $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ mit

$$\mathbb{P}_X(B) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\})$$

heißt Verteilung von X .

Anmerkung. Zwar kommt die Größe $\omega \in \Omega$ in Definition 3.2 noch formal im Rahmen des Urbildes vor, allerdings werden wir schon bald sehen, dass für viele Klassen von Verteilungen, konkrete Wahrscheinlichkeiten in den Vordergrund und ω in den Hintergrund rücken.

Interessanterweise sind die in Theorem 3.1 betrachteten Mengen vollkommen ausreichend, um die Verteilung einer Zufallsvariable komplett zu charakterisieren. Dazu definieren wir uns die Hilfsfunktion

Definition 3.3 (Verteilungsfunktion). Es sei X eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Die Verteilungsfunktion $F_X : \mathbb{R} \rightarrow [0, 1]$ ist definiert als

$$F_X(x) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x\}) = \mathbb{P}(X \leq x)$$

für alle $x \in \mathbb{R}$.

und nun gilt

Theorem 3.2. Die Verteilung \mathbb{P}_X einer Zufallsvariablen X wird eindeutig durch deren Verteilungsfunktion F_X charakterisiert.

Beispiel 3.2. Nehmen wir an, dass wir einen blauen und einen grünen Würfel haben, die sich abgesehen von der Farbe um nichts unterscheiden. Nun sei G die gewürfelte Augenzahl des grünen Würfels und B die gewürfelte Augenzahl des blauen Würfels. Offensichtlich sind die Verteilungsfunktionen von B und G gleich und wir sprechen davon, dass B und G identisch verteilt sind.

Intuitiv bedeutet das allerdings nicht, dass B und G immer gleich sind. Schließlich würden wir ja nicht erwarten, dass der blaue und grüne Würfel immer die gleiche Zahl zeigen. Bildlich gesprochen bedeutet die identische Verteilung nur, dass die Histogramme der beiden Würfel sich annähern, je öfter gewürfelt wurde (siehe Abbildung 3.1).

Weiterhin überlegt man sich schnell, dass $\mathbb{P}(G \leq x) = 0$ für alle $x < 1$, $\mathbb{P}(G \leq x) = 1$ für alle $x \geq 6$ und $\mathbb{P}(G \leq x) = \mathbb{P}(G \leq \lfloor x \rfloor)$ für alle $x \in [1, 6)$, da G sinnvollerweise nur die Werte $x = 1, \dots, 6$ annehmen kann.

In dieser Aussage erkennen wir wieder, dass uns die $\omega \in \Omega$ wenig interessieren. Wir brauchen den Ausdruck $G(\omega)$ nicht konkret definieren, um zu wissen, dass nach Möglichkeit $G(\omega) \in \{1, \dots, 6\}$ für alle $\omega \in \Omega$ sein sollte, wenn G tatsächlich die Augenzahlen des grünen Würfels beschreiben soll.



Abbildung 3.1: Die Histogramme des grünen und roten Würfels nähern sich an, je öfter man würfelt. Deren Verteilung sind identisch.

An dieser Stelle stoßen wir allerdings auf eine mathematische Spitzfindigkeit. Auch wenn uns die Rolle des ω nicht wirklich interessiert, so wollen wir die Funktion G möglichst wenig einschränken. In diesem Fall können wir die Anforderungen an G noch ein wenig aufweichen und stattdessen fordern, dass $G(\omega)$ andere Werte annehmen darf, solange die Wahrscheinlichkeit, dass $G(\omega) \in \{1, \dots, 6\}$ gleich 1 ist, also in Symbolen ausgedrückt $\mathbb{P}(G \in \{1, \dots, 6\}) = 1$. Wir sprechen hier davon, dass G *fast sicher* in $\{1, \dots, 6\}$ liegt.

Dies bestärkt nochmal die Tatsache, dass uns die ω gar nicht so sehr interessieren, da die Funktion R auf einer Menge A machen kann was sie will, solange $\mathbb{P}(A) = 0$. Dies ist am Anfang zunächst etwas gewöhnungsbedürftig, bietet aber im späteren Verlauf einige Vorteile.

Abschließend berechnen wir noch anhand des Beispiels $\mathbb{P}(G \leq 2.4) = \mathbb{P}(G = 1) + \mathbb{P}(G = 2)$ und aus der Tatsache, dass $\mathbb{P}(G = k) = \frac{1}{6}$ für alle $k \in \{1, \dots, 6\}$ und $\mathbb{P}(G = k) = 0$ sonst, dass $\mathbb{P}(G \leq x) = \frac{\lfloor x \rfloor}{6}$ für alle $x \in [1, 6)$. Zusammenfassend haben wir also

$$F_G(x) = \mathbb{P}(G \leq x) = \begin{cases} 0 & , x < 1 \\ \frac{\lfloor x \rfloor}{6} & , 1 \leq x < 6 \\ 1 & , x \geq 6 \end{cases}$$

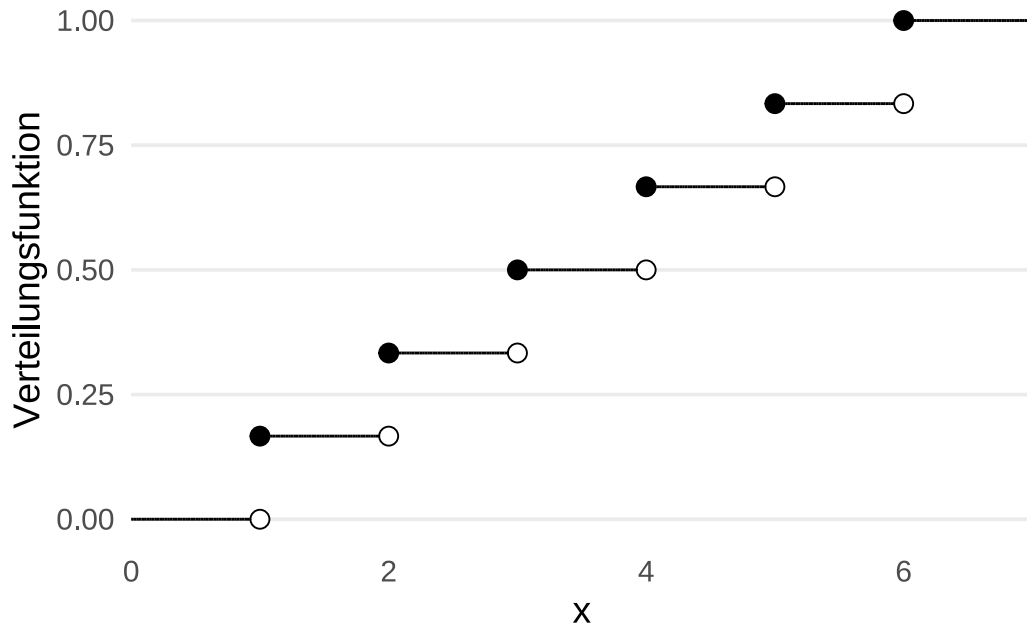


Abbildung 3.2: Verteilungsfunktion eines Würfels

Der Vollständigkeit halber noch ein Theorem bzgl. den Eigenschaften einer Verteilungsfunktion und der Existenz einer entsprechenden Zufallsvariable.

Theorem 3.3. *Es gelten die folgenden zwei Aussagen.*

a. *Es sei X eine beliebige Zufallsvariable. Dann besitzt deren Verteilungsfunktion $F_X : \mathbb{R} \rightarrow [0, 1]$ die Eigenschaften*

1. Asymptotik: $\lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow \infty} F_X(x) = 1.$
2. Monotonie: $F_X(x) \leq F_X(x + h)$ für alle $x \in \mathbb{R}$ und $h \geq 0.$
3. Rechtsseitige Stetigkeit: $\lim_{x \rightarrow x_0 + 0} F_X(x) = F_X(x_0)$ für alle $x \in \mathbb{R}.$

b. *Erfüllt eine Funktion $F : \mathbb{R} \rightarrow [0, 1]$ die Eigenschaften 1 bis 3, so existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Zufallsvariable X auf diesem Wahrscheinlichkeitsraum dessen Verteilungsfunktion F ist.*

4 Verteilungen

Im letzten Kapitel sind wir darauf eingegangen, dass man Zufallsexperimente mit einer Zufallsvariable und dessen Verteilung zusammenfassen kann. Natürlich eignet sich nicht jede Verteilung, um jedes Zufallsexperiment zu beschreiben. In Beispiel 3.2 wäre es zum Beispiel wenig situationsbezogen gewesen, die Augenzahl eines klassischen Würfels über eine Zufallsvariable Y mit $\mathbb{P}(Y = 7) > 0$ oder $\mathbb{P}(Y = 3.5) > 0$ zu beschreiben.

Dementsprechend müssen wir darauf achten, dass der kleinste Wertebereich $C \subset \mathbb{R}$ einer Zufallsvariable X , sodass $\mathbb{P}(X \in C) = 1$, zum entsprechenden Zufallsexperiment passt. Insbesondere gibt es gute Gründe, zwischen einem abzählbaren und überabzählbaren Wertebereich C zu unterscheiden. Ist der Bereich C abzählbar, so sprechen wir von einer *diskreten Zufallsvariable*¹.

Nach dieser Bezeichnung ist die Zufallsvariable R aus Beispiel 3.2 eine diskrete Zufallsvariable. Tatsächlich ist R sogar eine der einfachsten Zufallsvariablen, da sie jedem möglichen Wert $1, \dots, 6$ die gleiche Wahrscheinlichkeit zuordnet.

Innerhalb der Klassen der diskreten Zufallsvariablen gibt es allerdings noch eine Vielzahl an weiteren Verteilungen, die unterschiedlich Wahrscheinlichkeiten zulassen. In diesem Kapitel werden wir einige davon kennenlernen. Außerdem werden wir auch die Klasse der *absolutstetigen Zufallsvariablen* erkunden. Diese führen uns in den Bereich eines überabzählbaren Wertebereichs C .

4.1 Diskrete Zufallsvariablen

Um die unterschiedlichen diskreten Verteilungen kennenzulernen, führen wir den Begriff der Zähldichte ein. Da die Zähldichte auf sehr simple Weise mit der Verteilungsfunktion F_X einer diskreten Zufallsvariable X zusammenhängt, charakterisiert auch die Zähldichte die Verteilung von X eindeutig.

Definition 4.1 (Zähldichte). Es sei X eine diskrete Zufallsvariable, d.h. es gibt eine abzählbare Menge $C = \{x_1, x_2, \dots\}$, sodass $\mathbb{P}(X \in C) = 1$ und $\mathbb{P}(X = x_k) > 0$ für alle $k \in \mathbb{N}$ bzw.

¹Streng genommen ist die Verteilung der jeweiligen Zufallsvariable diskret. Jedoch werden die Zufallsvariablen häufig auch so genannt.

$k = 1, \dots, N$, falls C endlich mit $|C| = N$. Dann heißt die Funktion $p(k) = \mathbb{P}(X = x_k)$, $k \in \mathbb{N}$ bzw. $k = 1, \dots, N$, Zähldichte von X .

Beispiel 4.1 (Diskrete Gleichverteilung). Die Zähldichte der diskreten Gleichverteilung auf einer Menge $C = \{x_1, \dots, x_N\}$ ist gegeben als $p(k) = \frac{1}{N}$ für alle $k = 1, \dots, N$. In Zeichen schreiben wir $X \sim U\{x_1, \dots, x_N\}$. Für $C = \{1, \dots, 6\}$ entspricht diese Verteilung der Verteilung in Beispiel 3.2.

Beispiel 4.2 (Geometrische Verteilung). Die geometrische Verteilung besitzt die Zähldichte $p(k) = p(1-p)^{k-1}$ für alle $k \in \mathbb{N}$, wobei $p \in [0, 1]$. Diese Verteilung beschreibt die Anzahl der unabhängigen Versuche eines Zufallsexperiments bis zum ersten ‘Erfolg’. Wichtig dabei ist, dass jeder Versuch des Experiments unter gleichbleibenden Bedingungen stattfindet.

Nehmen wir an, dass ein Paar versucht, ein Mädchen zu bekommen. Außerdem ist es für dieses Paar scheinbar logistisch² möglich so lange Kinder zu kriegen, bis es ein Mädchen bekommt. Wir interessieren uns nun für die Fragestellung, wie wahrscheinlich es ist, dass das Paar genau 3 Jungs bekommt, bevor das erste Mädchen geboren wurde und modellieren dies mit einer geometrischen Verteilung.

Es sei also X die Anzahl der Kinder des Paares nachdem es aufhört, Kinder zu bekommen. Dann gilt $X \sim \text{Geo}(p)$ und wir berechnen die gesuchte Wahrscheinlichkeit als $\mathbb{P}(X = 4) = p(1-p)^3$. Nach Angaben der WHO³ liegt die natürliche ‘sex ration at birth’ bei 105 Jungen zu 100 Mädchen. Somit können wir annehmen, dass $p \approx 0.4878$ und wir berechnen $\mathbb{P}(X = 4) = 0.0655$.

Beispiel 4.3 (Binomialverteilung). Die Binomialverteilung besitzt die Zähldichte $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$ für alle $k \in \{0, \dots, n\}$, wobei $n \in \mathbb{N}$ und $p \in [0, 1]$. Diese Verteilung beschreibt die Anzahl der Erfolge bei n unabhängigen Versuchen eines unter gleichbleibenden Bedingungen stattfindenden Zufallsexperiments. Für $n = 1$ nennen wir die Binomialverteilung auch Bernoulli-Verteilung und schreiben $X \sim \text{Ber}(p)$. Im Allgemeinen schreiben wir $X \sim \text{Bin}(n, p)$.

Betrachten wir wieder das Paar aus Beispiel 4.2 und nehmen an, dass das Paar unabhängig vom Geschlecht nun 4 Kinder bekommen hat. Nun wollen wir untersuchen, wie wahrscheinlich

²Anlässlich der Anmerkung meiner Freundin sei an dieser Stelle erwähnt, dass mir bewusst ist, dass zum Kinderkriegen mehr als nur ‘logistische’ Überlegungen dazugehören. Das ist hier natürlich alles nur grob vereinfacht dargestellt.

³[Link](#)

es ist, dass das Paar mindestens 2 Mädchen bekommen hat. Dazu definieren wir X als die Anzahl der Mädchen und nehmen an, dass $X \sim \text{Bin}(n, p)$ mit $n = 4$ und $p = 0.4878$. Dann gilt

$$\begin{aligned}\mathbb{P}(X \geq 2) &= \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) \\ &= \binom{4}{2} p^2 (1-p)^2 + \binom{4}{3} p^3 (1-p)^1 + \binom{4}{4} p^4 (1-p)^0 = 0.8388.\end{aligned}$$

Beispiel 4.4 (Poisson Verteilung). Die Poissonverteilung besitzt die Zähldichte $p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, \dots$, wobei $\lambda > 0$. Die Poissonverteilung beschreibt die Anzahl der Ereignisse, die mit gleichbleibender Rate unabhängig voneinander in einem festen Zeitfenster eintreten. Wir schreiben $X \sim \text{Poi}(\lambda)$. In dem populärwissenschaftlichen und meiner Meinung nach empfehlenswerten Buch (Spiegelhalter 2019) demonstriert der Statistiker David Spiegelhalter, wie verblüffend gut⁴ eine Poissonverteilung die Anzahl der Morde in England und Wales zwischen 2014 und 2016 modellieren kann.

Insgesamt gab es 1.545 Vorfälle über einen Zeitraum von 1.095 Tagen. Dies führt zu einer durchschnittlichen Rate von $\lambda = 1.41$ Vorfällen pro Tag führt. Dazu nimmt Spiegelhalter an, dass die Anzahl der Vorfälle X an einem beliebigen Tag $\text{Poi}(\lambda)$ -verteilt ist. Multipliziert man die Wahrscheinlichkeit $\mathbb{P}(X = k)$ für $k = 0, \dots, 5$ und $\mathbb{P}(X > 5)$ mit der Anzahl der beobachteten Tage, so erhält man die erwartete Anzahl von Tagen mit $k = 0, \dots, 5$ bzw. mehr als 5 Vorfällen und kann dies mit den beobachteten Daten vergleichen (siehe Tabelle 4.1 und Abbildung 4.1).

Tabelle 4.1: Vergleich der Poissonverteilung mit Parameter 1.41 und der beobachteten Anzahl von Morden in England und Wales zwischen April 2014 und März 2016, siehe Table 10.3 in (Spiegelhalter 2019)

Morde an einem Tag	Beobachtete Anzahl der Tage	Erwartete Anzahl der Tage
0	259	267.10
1	387	376.80
2	261	265.90
3	131	125.00
4	40	44.10
5	13	12.40
Mehr als 5	3	3.60

⁴Für mich persönlich ist hier verblüffend, wie gut man Beobachtungen durch geeignete mathematische Modelle beschreiben kann. An sich ist es allerdings auch gut, dass die Poissonverteilung so gut passt, da dies als Indiz dafür dienen kann, dass die Mordrate halbwegs konstant ist und kein Serienmörder o.ä. unterwegs ist.

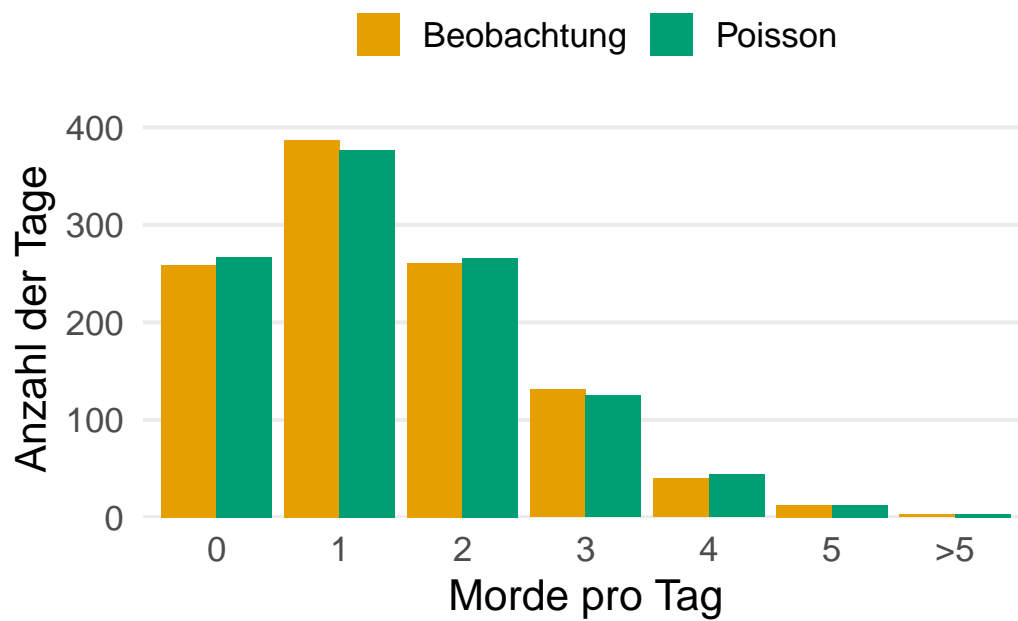


Abbildung 4.1: Vergleich der Poissonverteilung mit Parameter 1.41 und der beobachteten Anzahl von Morden in England und Wales zwischen April 2014 und März 2016, siehe Table 10.3 in (Spiegelhalter 2019)

Annahme 4.1.

- *All diese Zähldichten haben zwei Dinge gemeinsam:*
 1. $p(k) \in [0, 1]$ für alle $k \in \mathbb{N}$, da $p(k)$ immer eine Wahrscheinlichkeit ist.
 2. Die Summe über alle $p(k)$ ist gleich 1, da die zu den Wahrscheinlichkeiten gehörenden Ereignisse, eine disjunkte Vereinigung des Grundraumes darstellt.
- *Nach dem Poissonschen Approximationssatz bzw. dem Gesetz der seltenen Ereignisse lässt sich die Binomialverteilung für große n und kleine p durch eine Poissonverteilung mit Parameter $\lambda = np$ approximieren.*

4.2 Absolutstetige Zufallsvariablen

Die bisherige Klasse von Verteilungen ist leider keine adäquate Wahl für viele Anwendungsfälle. Wollen wir beispielsweise die (zufällige) Ankunftszeit eines Freundes an einem vereinbarten Treffpunkt beschreiben, so stellen wir fest, dass es überabzählbar viele mögliche Ankunftszeiten gibt (sofern wir Zeit als ein Kontinuum auffassen).

Somit brauchen wir eine Zufallsvariable X , die überabzählbar viele Werte annehmen kann. Allerdings stellt uns das vor die Herausforderung, dass wir nicht analog zu Annahme 4.1 fordern können, dass $\mathbb{P}(X = k) > 0$ für überabzählbar viele Werte k . Dies wäre ein Widerspruch, da da sich sonst alle Wahrscheinlichkeiten der Form $\mathbb{P}(X = k)$ nicht zu 1 “summieren” können⁵.

Deswegen müssen wir uns mit einer Hilfsfunktion begnügen, die ein Analogon zur Zähldichte $\mathbb{P}(X = k)$ für überabzählbare Werte darstellt. Außerdem sollte sich das Integral dieser Funktion über alle möglichen Werte 1 ergeben (was wir als Analogon zur überabzählbaren Summation betrachten). Auf diese Weise können wir absolutstetige Zufallsvariablen sinnvoll definieren.

Definition 4.2 (Absolutstetigkeit, Dichte). Die Verteilung einer Zufallsvariable X heißt *absolutstetig*, falls die Verteilungsfunktion F_X von X folgender Darstellung besitzt:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad x \in \mathbb{R},$$

wobei $f_X : \mathbb{R} \rightarrow [0, \infty)$ eine (Lebesgue)integrierbare Funktion auf \mathbb{R} ist. Die Funktion f wird dabei als *Dichte von X* bezeichnet.

⁵Außerdem hätte die Verteilungsfunktion solch einer Zufallsvariable überabzählbar viele Sprungstellen. Dies ist aber unmöglich, da eine monotone Funktion höchstens abzählbar viele Sprungstellen haben kann. Danke an Marco Oesting für diesen Hinweis zur technischen Untermalung der obigen Begründung.

Genau wie es im diskreten Fall mit der Zähldichte funktioniert hat, können wir im absolutstetigen Fall eine Verteilung eindeutig über die Dichtefunktion charakterisieren. Zudem können wir beliebige Wahrscheinlichkeiten mittels

$$\mathbb{P}(X \in B) = \int_B f_X(y) \, dy = \int_{\mathbb{R}} \mathbb{1}_B(y) f_X(y) \, dy$$

für alle $B \in \mathcal{B}(\mathbb{R})$ berechnen.

Annahme 4.2.

- *Absolutstetige Zufallsvariablen sind nicht die einzigen Zufallsvariablen mit überabzählbarem Wertebereich C . Schließlich kann man sich auch Zufallsvariablen als Mischung von diskreten und absolutstetigen Zufallsvariablen konstruieren. Diese haben dann einen überabzählbaren Wertebereich, sind aber nicht absolutstetig.*
- *Analog zur Annahme 4.1 muss die Zähldichte die folgenden Eigenschaften erfüllen:*

1. $f_X(x) \geq 0$ für alle $x \in \mathbb{R}$
2. $\int_{\mathbb{R}} f_X(x) \, dx = 1$

Beispiel 4.5 (Normalverteilung). Die Normalverteilung zählt zu den wichtigsten Verteilungen in der Stochastik. Ihre Dichte ist gegeben durch

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}, \quad x \in \mathbb{R},$$

wobei $\mu \in \mathbb{R}$ und $\sigma > 0$ Parameter der Verteilung sind (siehe Abbildung 4.2).

Betrachtet man die Dichten der Normalverteilung in Abbildung 4.2, so lässt sich erkennen, dass ein Großteil der Fläche zwischen Funktion und x -Achse im Intervall $[\mu - a, \mu + a]$ liegt, wobei $a \in \mathbb{R}$ eine von σ abhängige Zahl ist. Konkreter gilt für eine normalverteilte Zufallsvariable X die sogenannte k - σ -Regel:

$$\mathbb{P}(\mu - k\sigma \leq X \leq \mu + k\sigma) \approx \begin{cases} 68\% & , \, k = 1 \\ 95\% & , \, k = 2 \\ 99.7\% & , \, k = 3 \end{cases} \quad (4.1)$$

Es lässt sich erahnen, dass dies eine gute Eigenschaft ist, die es ermöglicht viele natürliche Phänomene wie Körpergröße, Blutdruck, Messfehler in der Physik und noch viel mehr zu modellieren.

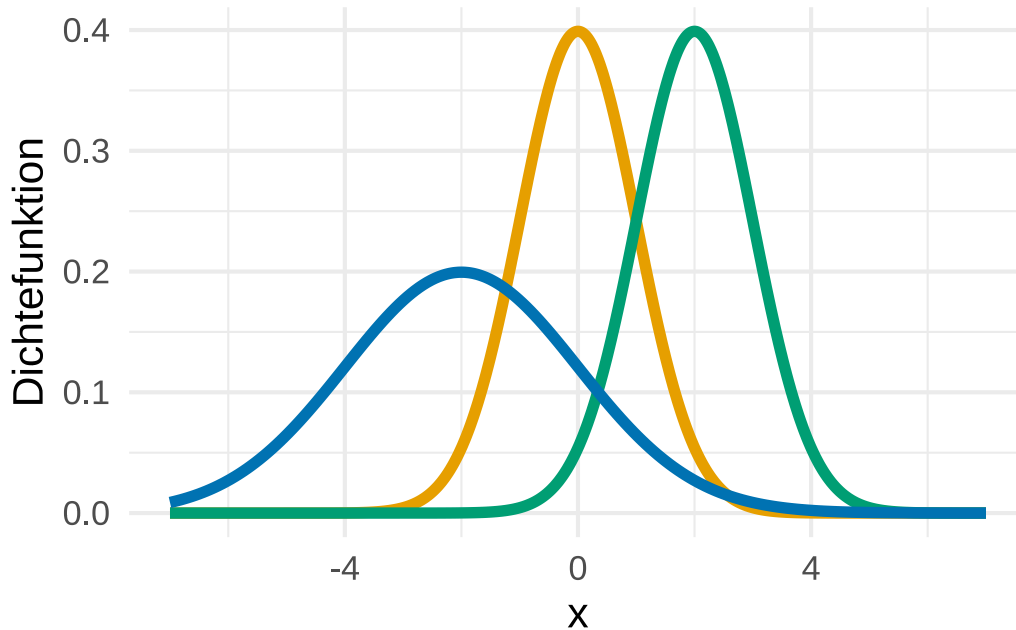


Abbildung 4.2: Dichte der Normalverteilung für Parameter $\mu = 0, \sigma = 1$ (orange), $\mu = 2, \sigma = 1$ (grün), $\mu = -2, \sigma = 2$ (blau)

Ein weiterer Grund für die Wichtigkeit der Normalverteilung ist der zentrale Grenzwertsatz (ZGWS), den ich hier nur anschnitten möchte. Wir werden in einem späteren Kapitel ausführlicher über den ZGWS reden. Grob formuliert besagt der Satz, dass ich unter sehr milden Bedingungen eine Summe von Zufallsvariablen $X_1 + \dots + X_n$ nach einer Transformation $(X_1 + \dots + X_n - a_n)/b_n$, wobei $a_n \in \mathbb{R}$ und $b_n > 0$ geeignete Konstanten (in Abhängigkeit von n) sind, mithilfe einer Standardnormalverteilung ($\mu = 0, \sigma = 1$) approximieren kann.

Die Folge a_n übernimmt dabei die sogenannte Zentrierung und die Folge b_n übernimmt die Normierung/Skalierung. Auch auf die genaue Wahl von a_n und b_n werden wir später noch genauer eingehen.

Nehmen wir nun an, dass wir einen Würfel $n \in \mathbb{N}$ mal werfen und $X_i, i = 1, \dots, n$, die gewürfelte Augenzahl des i -ten Versuches darstellt. Dann ist nach Beispiel 3.2 X_i diskret gleichverteilt. Wählen wir nun $a_n = 3.5n$ und $b_n = \sqrt{35n}/12$, so “näht” sich die Verteilung von $(X_1 + \dots + X_n - a_n)/b_n$ für steigendes n an die Standardnormalverteilung an (siehe Abbildung 4.3).

Beispiel 4.6 (Exponentialverteilung). Eine weitere wichtige absolutstetige Verteilung ist die

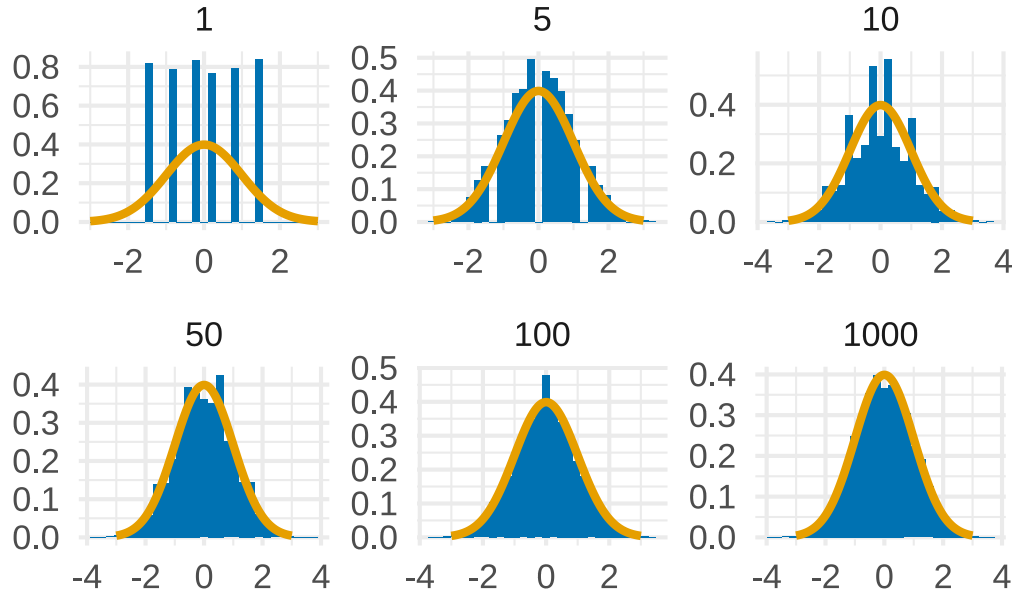


Abbildung 4.3: Histogramm von jeweils $N = 10.000$ Realisierungen von $(X_1 + \dots + X_n - a_n)/b_n$ für $n \in \{1, 5, 10, 50, 100, 1000\}$.

Exponentialverteilung mit Dichtefunktion

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{I}_{\{x > 0\}},$$

wobei $\lambda > 0$ ein Parameter dieser Verteilung ist. Sei T die Restlebensdauer einer Person oder eines Gegenstandes zum Zeitpunkt $t > 0$. Außerdem nehmen wir an, dass T nicht-negativ ist und Verteilungsfunktion F und Dichte f besitzt. Nach Kapitel 3 in (Norberg 2020) lässt sich die sogenannte Sterblichkeitsintensität definieren als

$$\mu(t) := \frac{f(t)}{1 - F(t)}.$$

Mithilfe dieser Größe lässt sich die Wahrscheinlichkeit, einer t -jährigen Person in (einem kurzen) Intervall $(t, t + dt)$ zu sterben bzw. eines t -jährigen Gegenstandes kaputt zu gehen, approximativ durch $\mu(t)dt$ ermitteln. Nimmt man nun an, dass $T \sim \text{Exp}(\lambda)$ mit $\lambda > 0$, so gilt $\mu(t) = \lambda$. Also ist die Sterblichkeitsintensität unabhängig vom Alter t , was zwar nicht unbedingt eine sinnvolle Modellierung der Lebensdauer eines Menschen darstellt, aber für die Lebensdauer von technischen Geräten dennoch geeignet ist.

5 Zufallsvektoren

Nehmen wir einmal an, dass in nächster Zeit eine Wahl anstünde und wir die Wahl mit stochastischen Mitteln prognostizieren wollen. Beispielsweise könnte man die Antwort eines Teilnehmers einer Wahlumfrage mit einer Zufallsvariable X_1 modellieren. Üblicherweise reicht es allerdings nicht aus, nur eine Person zu befragen.

Also brauchen wir für die Antworten weiterer Umfrageteilnehmer ebenfalls Zufallsvariablen X_2, \dots, X_n . Wollen wir diese Zufallsvariablen nicht einzeln betrachten und unsere Stichprobe lieber als Ganzes betrachten, so können wir die Ergebnisse in einem Vektor $X = (X_1, \dots, X_n)$ bündeln. Dies bringt uns auf natürliche Weise zu dem Konzept von Zufallsvektoren.

Definition 5.1 (Zufallsvektor). Es seien X_1, \dots, X_n Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ gegeben. Dann nennen wir $X = (X_1, \dots, X_n)$ Zufallsvektor mit Komponenten X_1, \dots, X_n . Wir nennen X diskret bzw. absolutstetig, falls alle Komponenten diskret bzw. absolutstetig sind.

Glücklicherweise übertragen sich viele Konzepte aus dem Ein- ins Mehrdimensionale. Im diskreten Fall sprechen wir nun statt einer Zähldichte $\mathbb{P}(X_1 = k_1)$ von einer gemeinsamen Zähldichte $\mathbb{P}(X_1 = k_1, \dots, X_n = k_n)$. Analog dazu gibt es im absolutstetigen Fall auch eine gemeinsame Dichtefunktion $f_X(x_1, \dots, x_n)$. Ebenso gibt es auch wieder die Verteilungsfunktion

$$\begin{aligned} F_X(x_1, \dots, x_n) &= \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= \begin{cases} \sum_{k_1 \leq x_1} \cdots \sum_{k_n \leq x_n} \mathbb{P}(X_1 = k_1, \dots, X_n = k_n) \\ \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_X(t_1, \dots, t_n) dt_1 \cdots dt_n \end{cases} \end{aligned}$$

Für unabhängige Zufallsvariablen lässt sich dies noch vereinfachen.

Theorem 5.1. *Es sei X ein Zufallsvektor mit Komponenten X_1, \dots, X_n . Nun sind X_1, \dots, X_n genau dann unabhängig, wenn*

$$F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

für alle $x_1, \dots, x_n \in \mathbb{R}$.

Annahme 5.1. *Im diskreten Fall bedeutet dies, dass die gemeinsame Zählzähldichte von X in das Produkt der (Rand-)Zähldichten von X_1, \dots, X_n zerfällt. Im absolutstetigen Fall passiert das gleiche bzgl. der gemeinsamen Dichtefunktion und den Randdichtefunktionen von X_1, \dots, X_n . Üblicherweise arbeitet man im Mehrdimensionalen überwiegend mit der Zähldichte und Dichtefunktion, da diese einfacher zu handhaben sind und die Verteilungsfunktion eindeutig charakterisieren.*

Falls die Zufallsvariablen abhängig sind, so lässt sich auf den ersten Blick keine sinnvolle Aussage bzgl. dem Zusammenhang zwischen gemeinsamer Verteilungsfunktion und den Randverteilungsfunktionen treffen. Insbesondere lässt sich kein einfacher Schluss von den Randverteilungen auf die gemeinsame Verteilung ziehen. Dazu fehlt uns nämlich die Abhängigkeitsstruktur zwischen den einzelnen Zufallsvariablen. Andersherum lässt sich nach Kenntnis der gemeinsamen Verteilung glücklicherweise viel über die Randverteilungen sagen.

Theorem 5.2. *Es sei X ein Zufallsvektor mit Komponenten X_1, \dots, X_n und gemeinsamer Zähldichte $p(k_1, \dots, k_n)$ bzw. gemeinsamer Dichtefunktion $f_X(x_1, \dots, x_n)$. Dann gilt für alle $i = 1, \dots, n$*

$$\mathbb{P}(X_i = x) = \sum_{k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_n} p(k_1, \dots, k_{i-1}, x, k_{i+1}, \dots, k_n) \quad (1)$$

bzw.

$$f_{X_i}(x) = \int_{\mathbb{R}^{n-1}} f_X(t_1, \dots, t_{i-1}, x, t_{i+1}, \dots, t_n) dt_1, \dots, dt_{i-1} dt_{i+1} \dots, dt_n. \quad (2)$$

Letztendlich bedeutet Gleichung (1), dass wenn man sich für die Randwahrscheinlichkeit $\mathbb{P}(X_i = x)$ interessiert, lediglich die Wahrscheinlichkeiten aller Ereignisse summieren muss, die mit dem Ereignis $\{X_i = x\}$ konform sind. Dementsprechend lässt sich Gleichung (2) als infinitesimale Version von Gleichung (1) verstehen. Die beiden Gleichungen kann man auch so interpretieren, dass man die überflüssigen Variablen $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n$ “rausintegriert” bzw. im diskreten Fall $k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_n$ “raussummiert”.

Die Idee, die Wahrscheinlichkeit eines beliebigen Ereignisses A über die Summation bzw. Integration von mit A konformen Mengen und der gemeinsamen (Zähl-)dichte zu berechnen, findet man oft in der Stochastik. Beispielsweise wird diese Technik auch für die Bestimmung der (Zähl-)Dichte von Funktionen von Zufallsvariablen genutzt.

Beispiel 5.1. Es seien X_1 und X_2 zwei beliebige $\text{Poi}(\lambda)$ -verteilte Zufallsvariablen, wobei $\lambda > 0$. Da $X_1, X_2 \in \{0, 1, \dots\}$ f.s., ist ebenfalls $Z := X_1 + X_2 \in \{0, 1, \dots\}$ f.s..

Wollen wir nun die Dichte der Zufallsvariable Z , d.h. die Wahrscheinlichkeit $\mathbb{P}(Z = k) = \mathbb{P}(X_1 + X_2 = k)$ für alle $k \in \mathbb{N}_0$, berechnen, so müssen wir lediglich die Wahrscheinlichkeiten aller möglichen Kombinationen $\{X_1 = i\} \cap \{X_2 = j\}$ summieren, in denen $i + j = k$. Somit berechnen wir

$$\mathbb{P}(X_1 + X_2 = k) = \sum_{i=0}^{\infty} \mathbb{P}(X_1 = i, X_2 = k - i). \quad (3)$$

Da wir an dieser Stelle die Abhängigkeitsstruktur von X_1 und X_2 nicht kennen (da wir nur die Randverteilungen angegeben haben), können wir leider nicht weiterrechnen. Nehmen wir zusätzlich an, X_1 und X_2 wären unabhängig, so zerfällt die gemeinsame Zähldichte in das Produkt der Randdichten und wir könnten den Ausdruck in Gleichung (3) weiter vereinfachen.

Die in Beispiel 5.1 illustrierte Vorgehensweise ist natürlich nicht nur für Poissonverteilte Zufallsvariablen gültig. Im Allgemeinen können wir das folgende Theorem formulieren:

Theorem 5.3. *Seien X_1 und X_2 zwei absolutstetige Zufallsvariablen mit Dichte f_{X_1} und f_{X_2} und gemeinsamer Dichte f . Dann gilt:*

$$\begin{aligned} f_{X_1+X_2}(z) &= \int_{-\infty}^{\infty} f(z-t, t) \, dt \\ f_{X_1-X_2}(z) &= \int_{-\infty}^{\infty} f(z+t, t) \, dt \\ f_{X_1 \cdot X_2}(z) &= \int_{-\infty}^{\infty} \frac{1}{|t|} f\left(\frac{z}{t}, t\right) \, dt \\ f_{X_1/X_2}(z) &= \int_{-\infty}^{\infty} |t| f(z t, t) \, dt \end{aligned}$$

Bibliografie

- Eddy, David M. 1982. „Probabilistic reasoning in clinical medicine: Problems and opportunities“. In *Judgment under Uncertainty: Heuristics and Biases*, herausgegeben von Daniel Kahneman, Paul Slovic, und Amos Tversky, 249–67. Cambridge University Press. <https://doi.org/10.1017/CBO9780511809477.019>.
- Norberg, Ragnar. 2020. *Basic Life Insurance Mathematics*. <http://web.math.ku.dk/~mogens/liefebook.pdf>.
- Spiegelhalter, David. 2019. *The Art of Statistics: Learning from Data*. Pelican.
- Spodarev, Evgeny. 2018. *Vorlesungsskript zur Elementaren Wahrscheinlichkeitsrechnung und Statistik*.
- . 2020. *Vorlesungsskript zur Elementaren Wahrscheinlichkeitsrechnung und Statistik*.