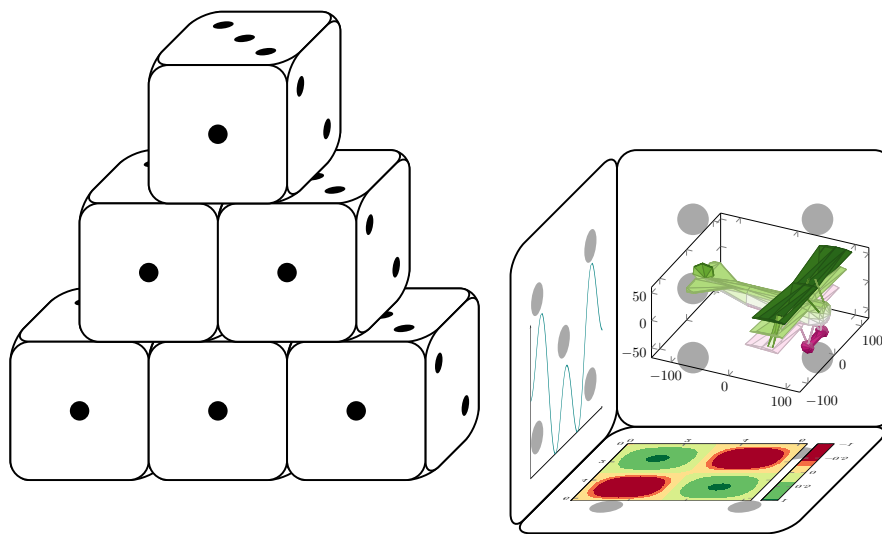


Angewandte Stochastik 1

Dr. Larisa Yaroslavtseva



MITSCHRIEB VON

FLORIAN SIHLER

florian.sihler@uni-ulm.de

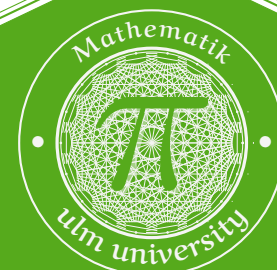
Version vom:

5. August 2020

Zitternd greift er die Feder, mühsam das letzte Blatt.
Legt es auf den Einband aus Leder, mit dem alles begonnen hat.

Er beginnt sein Werk zu vollenden, an diesem schönen Sommertag.
Widmet die Worte den Enten, die er eben noch gesehen hat.

Florian Sihler, 04.08.2016



Ver. 2.2.0



Inhaltsverzeichnis

1	Eine Einleitung	1
2	Wahrscheinlichkeiten	3
2.1	Wahrscheinlichkeitsräume	3
2.2	Ereignisse	3
2.3	Wahrscheinlichkeiten	6
2.4	Endliche Wahrscheinlichkeitsräume	10
2.4.1	Laplace und die Gleichverteilung	10
2.4.2	Abstecher: Urnenmodell	13
2.4.3	Hypergeometrische Verteilung	16
2.4.4	Die Bernoulliverteilung und die Binomialverteilung	17
2.5	Diskrete Wahrscheinlichkeitsräume	18
2.5.1	Die Poisson-Verteilung	20
2.5.2	Die geometrische Verteilung	21
2.6	Stetige Wahrscheinlichkeitsräume	22
2.6.1	Die Borel- σ -Algebra	23
2.6.2	Stetige Gleichverteilung	25
2.6.3	Exponentialverteilung	26
2.6.4	Die Normalverteilung	27
2.7	Bedingte Wahrscheinlichkeiten	28
2.8	Unabhängigkeit	31
2.8.1	Die Unabhängigkeit von zwei Ereignissen	31
2.8.2	Die Unabhängigkeit von mehr als zwei Ereignissen	33
3	Zufallsvariablen	35
3.1	Zufallsvariablen definiert	35
3.2	Die Verteilungsfunktion	39
3.3	Zufallsvektoren	44
3.3.1	Der diskrete Fall	45
3.3.2	Der absolutstetige Fall	47
3.4	Unabhängigkeit von Zufallsvariablen	49
3.5	Kenngrößen von Zufallsvariablen	52
3.5.1	Erwartungswert	52
3.5.2	Varianz	58
3.5.3	Kovarianz und Korrelationskoeffizient	62
3.6	Grenzwertsätze für Folgen und Zufallsvariablen	65
3.6.1	Gesetze der großen Zahlen	65

3.7 Der zentrale Grenzwertsatz	67
Definitionen	69
Sätze	71



EINE EINLEITUNG

WICHTIGE BEGRIFFLICHKEITEN UND GESCHICHTE!

Das Wort *Stochastik* stammt aus dem altgriechischen und bedeutet: Ratekunst, Kunst des Vermutens. Im Modernen ist es ein Oberbegriff für die Wahrscheinlichkeitsrechnung und die Statistik.

Das zentrale Objekt der Stochastik ist das *Zufallsexperiment*, wobei dies für uns ein Experiment ist, bei dem mehrere Ergebnisse eintreten können, aber es nicht vorhersagbar ist, welches Ergebnis eintritt.

Definition 1.0.1 – Zufallsexperiment

Ein Experiment, bei dem mehrere Ergebnisse möglich sind, wobei nicht vorhersagbar ist, welches eintritt.

Betrachten wir hierzu ein paar Beispiele...

Beispiel 1.1 – Münzwurf

Das klassische Beispiel ist seit jeher der Münzwurf, bei dem eine Münze geworfen und zwei Ergebnisse möglich sind: Die Münze zeigt Kopf oder Zahl. Es ist nicht möglich vorherzusagen welche der beiden Seiten eintrifft.

Natürlich noch ein weiterer Klassiker:

Beispiel 1.2 – Würfeln

Beim Würfeln sind die möglichen Ergebnisse die Zahlen 1 – 6 (also 1, 2, 3, 4, 5 und 6). Dieser Würfelwurf hat *keine deterministische Regelmäßigkeit*: Wiederholungen des Zufallsexperiments können also verschiedene Ergebnisse haben. Weiter hat er eine *statistische Regelmäßigkeit*: die relativen Möglichkeiten der einzelnen Ergebnisse stabilisieren sich nach vielen Wiederholungen.

Deswegen sagt man auch: ein *Zufallsexperiment* hat *keine* deterministische, aber dafür eine statistische Regelmäßigkeit. Betrachten wir zum Beispiel 1000 Münzwürfe, so ist anzunehmen, dass die relative Häufigkeit für „Kopf“ ungefähr $\frac{1}{2}$ ist. Es gilt also:

$$\frac{\text{Anzahl „Kopf“}}{1000} \approx \frac{500}{1000} = \frac{1}{2}.$$

Analog ergibt sich die relative Häufigkeit von Zahl zu $\frac{1}{2}$.

Diese statistische Regelmäßigkeit ermöglicht es und Vorhersagen bestimmter Art zu machen. Kehren wir hierzu zu unserem Münzwurf-Beispiel zurück:

Beispiel 1.3 – Vorhersagen dank der statistischen Regelmäßigkeit

So können wir bei 1000 Münzwürfen annehmen, dass „Kopf“ *mindestens* 450 eintritt – zumindest mit einer Wahrscheinlichkeit von 99.8%. Selbstredend können wir nicht sagen, dass die Wahrscheinlichkeit 100% ist. So ist es auch möglich, dass 1000mal „Zahl“ eintritt – auch wenn die Wahrscheinlichkeit dafür lächerlich gering ist; sie existiert! Wir sprechen von einer „sehr hohen Wahrscheinlichkeit“.

Bemerkung 1.1 – Themenaufteilung in Angewandte Stochastik I und II

In *Angewandte Stochastik 1* behandeln wir die Wahrscheinlichkeitsrechnung. Dies beinhaltet die mathematische Modellierung und die Analyse von Zufallsexperimenten. Im anschließenden Semester kümmert sich dann der zweite Teil der Vorlesung um die Statistik. In dieser hat man Daten (Temperatur, Anzahl von Bestellungen, ...) und versucht existierende mathematische Modelle aufgrund dieser empirischen Daten anzupassen.

2

WAHRSCHEINLICHKEITEN

Das Ziel dieses Kapitels ist die mathematische Modellierung von Zufallsexperimenten.

2.1 Wahrscheinlichkeitsräume

Für die mathematische Modellierung gibt es diverse Komponenten die von Relevanz sind, wir beginnen mit der einfachsten, dem Ergebnisraum...

Definition 2.1.1 – Ergebnisraum Ω

Der Ergebnisraum bezeichnet die Menge aller möglichen Ergebnisse ω eines Zufallsexperimentes.

Die Modellierung des Ergebnisraums ist für gewöhnlich nicht schwer, wir festigen das Wissen anhand ein paar Beispielen...

Beispiel 2.1 – Den Ergebnisraum angeben

1. Beim Münzwurf ist der Ergebnisraum das Ereignis „Kopf“ und das Ereignis „Zahl“. Bei der Modellierung können wir hierfür aber verschiedene Bezeichner wählen, so ist $\Omega = \{K, Z\}$ mit $K \hat{=} \text{Kopf}$ und $Z \hat{=} \text{Zahl}$ ebenso valide, wie: $\Omega = \{0, 1\}$, wobei 0 hier den Kopf und 1 die Zahl beschreibe.
2. Bei einem Würfel ist der Ergebnisraum wieder die Menge aller möglicher Ergebnisse und damit aller Zahlen: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
3. Bei der Anzahl der täglichen Bestellungen eines Artikels bei der Firma „Super-Products“, ist die Anzahl der möglichen Ergebnisse unbeschränkt, damit ergibt sich: $\Omega = \{0, 1, 2, \dots\} = \mathbb{N}_0$.
4. Möchten wir die Temperatur bei der Ulmer Adenauerbrücke an einem bestimmten Tag zu einer bestimmten Uhrzeit aufzeichnen so ergibt sich als Ergebnisraum beispielhaft $\Omega = [-50, 50]$ (wenn wir mal für die Menschheit so generell problematische Ausgänge ignorieren), oder eben, präziser als: $\Omega = \mathbb{R}$ (kommt ja auch auf die Skala an).

Bei der Wahl des Ergebnisraums gibt es (naiv-gesprochen) kein „richtig“ oder „falsch“, das Ziel ist es einen einfachen, aber adäquaten Raum zu wählen. Der Ergebnisraum kann hierbei endlich (Punkt 1, Punkt 2), unendlich (Punkt 3), abzählbar (Punkt 3) oder überabzählbar (Punkt 4) sein.

2.2 Ereignisse

Oft ist man nicht an dem tatsächlichen Ergebnis des Zufallsexperiments interessiert, sondern nur daran, ob das Ergebnis zu einer vorgegebenen Menge von Ergebnissen A gehört oder nicht („Ich gewinne!“ vs. „Ich gewinne nicht!“). Machen wir auch hierzu ein paar Beispiele...

Beispiel 2.2 – Ereignisse ausgeführt

1. Beim Würfeln können wir uns zum Beispiel nur für gerade Augenzahlen interessieren, dann ist $A = \{2, 4, 6\}$.
2. Bei der Anzahl täglicher Bestellungen, können wir uns dafür interessieren, dass der Vorrat an 100 nicht überschritten wird, dann ist $A = \{0, 1, 2, \dots, 100\}$.
3. Bei der Temperatur können wir uns dafür interessieren, dass die Temperatur über 25°C . In diesem Fall haben wir mit $\Omega = [-50, 50]$ das Intervall $A = (25, 50]$, oder wenn $\Omega = \mathbb{R}$: $A = (25, \infty)$.

Definition 2.2.1 – Ereignis A

Ein Ereignis ist eine Teilmenge A aus dem **Ergebnisraum** Ω , dem Wahrscheinlichkeiten zugeordnet werden sollen.

Man sagt „das **Ereignis** A tritt ein“, falls das Ergebnis ω des **Zufallsexperiments** in A liegt (also: $\omega \in A$). Mittels Mengenoperationen können die **Ereignisse** zu neuen Ereignissen verknüpft werden. Auch hierzu machen wir wieder ein paar Beispiele...

Beispiel 2.3 – Ereignisse vereinen

Wir betrachten hier die zwei **Ereignisse** A und B :

1. $A \cup B$: das Ereignis A *oder* das Ereignis B tritt ein.
2. $A \cap B$: das Ereignis A *und* das Ereignis B tritt ein.
3. $A \setminus B$: Das Ereignis A , nicht aber das Ereignis B tritt ein.
4. $A^c := \omega \setminus A$: Das Ereignis A tritt nicht ein.
5. $\bigcup_{i=1}^{\infty} A_i$: Mindestens eines der Ereignisse A_1, A_2, \dots tritt ein.
6. $\bigcap_{i=1}^{\infty} A_i$: Alle Ereignisse A_1, A_2, \dots treten ein.

Betrachten wir dazu noch ein weiteres, konkretes Beispiel...

Beispiel 2.4 – Man kann eine Münze zweimal werfen!

Eine Münze wird zweimal geworfen. Der **Ergebnisraum** ist $\Omega = \{KK, KZ, ZK, ZZ\}$.

Sei A_1 nun: „Im ersten Wurf fällt Kopf“ und A_2 : „Im zweiten Wurf fällt Kopf“. Damit gilt: $A_1 = \{KK, KZ\}$, $A_2 = \{KK, ZK\}$. Die Menge $A_1 \cup A_2 = \{KK, KZ, ZK\}$ ist nun das Ereignis: „Es fällt mindestens einmal Kopf.“ Die Menge $A_1 \cap A_2 = \{KK\}$ ist das Ereignis: „Es fällt zweimal Kopf.“

Doch welche Teilmengen A von Ω sollen die Wahrscheinlichkeiten $P(A)$ zugeordnet werden? Ideal wäre eine Definition $P(A)$ für alle Teilmengen $A \subset \Omega$ also für alle Mengen $A \in \mathcal{P}(\Omega)$. Die Potenzmenge, die hier aufgefasst wird mit $\mathcal{P}(\Omega)$ wird also die Menge aller Teilmengen definiert: $\mathcal{P}(\Omega) = \{B : B \subset \Omega\}$. Dies geht, sofern Ω endlich oder abzählbar ist. Im Allgemeinen ist es aber nicht möglich, wenn Ω zum Beispiel überabzählbar ist (also wenn eventuell $\Omega = \mathbb{R}$). In diesen Fällen beschränkt man sich deswegen auf ein *Teilsystem* $\Sigma \subseteq \mathcal{P}(\Omega)$ also eine Teilmenge aus der Potenzmenge der **Ereignissen**. Das Teilsystem muss nun aber der Forderung genügen, dass es gegenüber Mengenoperationen mit Ereignissen abgeschlossen ist. Also liefert eine beliebige Mengenoperation (wie wir sie **definiert** haben) zwischen zwei Ereignissen aus Σ wieder ein Ereignis aus Σ ist. Im Hintergrund dieser Forderung definieren wir uns den Begriff der σ -Algebra:

Definition 2.2.2 – Die σ -Algebra



Sei der **Ergebnisraum** $\Omega \neq \emptyset$ eine beliebige Menge. Eine Menge $\Sigma \subseteq \mathcal{P}(\Omega)$ heißt eine σ -Algebra, falls die folgenden drei Eigenschaften erfüllt sind:

1. $\Omega \in \Sigma$
2. Wenn $A \in \Sigma$ gilt: $A^c \in \Sigma$
3. Wenn $A_1, A_2, \dots \in \Sigma$, dann gilt: $\bigcup_{i=1}^{\infty} A_i \in \Sigma$.

Die Elemente $A \in \Sigma$ der σ -Algebra heißen Ereignisse.

Es gibt aber ein paar Dinge zu beachten:

Bemerkung 2.1

1. Aus der **Definition der σ -Algebra** folgt, dass auch gelten muss, dass $\emptyset \in \Sigma$, da $\emptyset = (\Sigma)^c \in \Sigma$. Diese leere Menge \emptyset trägt den Namen *unmögliches Ereignis*. Ein solches Ereignis tritt nie ein.
2. Aus der dritten Bedingungen folgt auch, dass endliche Vereinigungen wieder in Σ enthaltene Ereignisse sind. Dies erzeugen wir durch $A_i = \emptyset$ für alle $i > n$ mit $\forall A_1, A_2, \dots, A_n \in \Sigma$ erhalten wird dann: $\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n \cup \emptyset \cup \dots \cup \emptyset \in \Sigma$.

Betrachten wir nun erstmal anhand von einigen Beispielen, wie die Komponenten einer Sigma-Algebra aussehen. Wir beginnen wieder mit dem Klassiker, dem Münzwurf. Hier gilt $\Omega = \{K, Z\}$, womit die Mengen: $\Sigma_1 = \{\Omega, \emptyset\}$ sowie $\Sigma_2 = \{\Omega, \emptyset, \{K\}, \{Z\}\}$ beide die drei Eigenschaften der σ -Algebra erfüllen.

Die σ -Algebra Σ_1 trägt auch den Namen „Triviale σ -Algebra“, da es sich immer um eine σ -Algebra über Ω handelt. Es ist damit auch die einfachste σ -Algebra. Auch die drei Eigenschaften lassen sich hier direkt Zeigen. So ist: $\Omega \in \Sigma$ per Definition, $\Omega^c = \emptyset$ sind auch beide in der Menge enthalten und die unendliche Vereinigung ist mit Ω auch in Σ enthalten. Für Σ_2 sind die ersten Beiden Punkte ebenso (so gilt zusätzlich $K^c = Z$). Für den formalen Beweis des dritten Punkte müsste man alle Kombinationen der Elemente durchprobieren, mit $\Sigma_2 = \mathcal{P}(\Omega)$ können wir uns das aber auch direkt sparen ☺. Wichtig ist, dass die Ergebnisse keine Ereignisse sind, da $K \notin \Sigma_2$ sondern nur $\{K\} \in \Sigma_2$.

Betrachten wir dies aber noch einmal an einem anderen Beispiel – hier nun einmal beim ziehen einer von drei nummerierten Kugeln. Damit gilt $\Omega = \{1, 2, 3\}$ und es existieren die folgenden σ -Algebren:

- ◇ $\Sigma_1 = \{\Omega, \emptyset\}$
- ◇ $\Sigma_2 = \{\Omega, \emptyset, \{1\}, \{2, 3\}\}$
- ◇ $\Sigma_3 = \{\Omega, \emptyset, \{2\}, \{1, 3\}\}$
- ◇ $\Sigma_4 = \{\Omega, \emptyset, \{3\}, \{2, 1\}\}$
- ◇ $\Sigma_5 = \{\Omega, \emptyset, \{1\}, \{2\}, \{2, 3\}, \{1, 3\}, \{1, 2\}, \{3\}\} = \mathcal{P}(\Omega)$

Wir treffen nun die Behauptung, dass es sich bei allen diesen Σ_i -Mengen um σ -Algebren handelt – dies lässt sich aber mit etwas Aufwand einfach zeigen.

Wir haben ja zuerst gefordert, dass Σ abgeschlossen ist, die **Definition** fordert aber nur drei Bedingungen. Wir erfassen nun, dass diese Bedingungen ausreichen, um die Abgeschlossenheit zu zeigen!

Satz 2.2.1 – Die σ -Algebren sind abgeschlossen

Für jede σ -Algebra Σ gilt:

1. $\emptyset \in \Sigma$
2. Wenn $A, B \in \Sigma$, dann gilt $A \cup B, A \cap B, A \setminus B \in \Sigma$
3. Wenn $A_1, A_2, \dots, A_n \in \Sigma$, dann gilt $\bigcup_{i=1}^{\infty} A_i \in \Sigma$.

Beweis 2.1 – Zum Satz von eben

Wir zeigen die Punkte einzeln:

1. Die leere Menge ist leicht zu zeigen, dies haben wir bereits erfasst mit den **obigen Bemerkungen** gezeigt.
2. Dies gilt auch aufgrund **obigen Bemerkungen** sowie dem Zusatz, dass per Definition $A^c, B^c \in \Sigma$. Damit ergibt sich nach DeMorgan: $A \cap B = (A^c \cup B^c)^c$ mit $A, B \in \Sigma$ – damit ist der Schnitt gezeigt. Auch die Differenz zeigen wir durch das Umschreiben: $A \setminus B = A \cap B^c$.
3. Seien $A_1, A_2, \dots \in \Sigma$. Dann gilt wieder mit DeMorgan:

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} \underbrace{A_i^c}_{\in \Sigma} \right)^c \in \Sigma$$

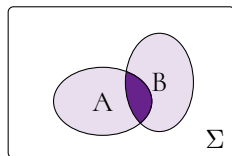
2.3 Wahrscheinlichkeiten

Wir definieren uns nun die Wahrscheinlichkeit $P(A)$ für $A \in \Sigma$.

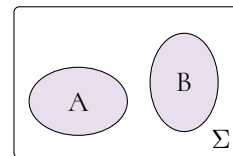
Definition 2.3.1 – Disjunkte Mengen

Wir nennen zwei Mengen A, B *disjunkt*, falls $A \cap B = \emptyset$ gilt, also wenn A und B sich keine Elemente teilen (siehe: **Abbildung 2.1**).

Weiter nennen wir die Mengen A_1, A_2, \dots, A_n *paarweise disjunkt* (p.d.) wenn für alle $i \neq j$ gilt, dass $A_i \cap A_j = \emptyset$. Oder anders: keine zwei Mengen teilen sich Elemente.



(a) Zwei nicht-disjunkte Mengen A und B .



(b) Zwei disjunkte Mengen A und B .

Abb. 2.1: Vergleich von disjunkten und nicht disjunkten Mengen.

Auf dieser Basis möchten wir nun eine Idee konzeptionieren. Diese verlangt, dass P die folgenden drei Eigenschaften erfüllt:

1. Für alle $A \in \Sigma$ gilt $0 \leq P(A) \leq 1$
2. Weiter soll $P(\Omega) = 1$ und $P(\emptyset) = 0$ gelten.
3. Der Zusammenhang $P(A^c) = 1 - P(A)$ soll für alle $A \in \Sigma$ gelten.
4. Wenn $A, B \in \Sigma$ **disjunkt** sind, soll gelten: $P(A \cup B) = P(A) + P(B)$.
5. Wenn $A_1, A_2, \dots, A_i \in \Sigma$ paarweise disjunkt sind gilt: $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Hierrüber motiviert sich die folgende Definition:

Definition 2.3.2 – Wahrscheinlichkeitsmaß

Sei $(\Omega \neq \emptyset)$ und Σ eine σ -Algebra auf Ω . Eine Abbildung $P : \Sigma \rightarrow [0, 1]$ heißt ein *Wahrscheinlichkeitsmaß* (oder kurz: W-Maß) auf Σ , falls:

1. $P(\Omega) = 1$
2. Sind $A_1, A_2, \dots, A_n \in \Sigma$ paarweise disjunkt so gilt: $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ (diese Eigenschaft bezeichnen wir als σ -Additiv).

Mit den nun bekannten Bezeichnern Ω, Σ und P definieren wir:

Definition 2.3.3 – Wahrscheinlichkeitsraum

Das Tripel (Ω, Σ, P) heißt *Wahrscheinlichkeitsraum* (W-Raum).

Bemerkung 2.2 – Direkt folgende Beschränkungen

Wir zeigen nun wieder, dass die getroffenen Einschränkungen für P wieder reichen um die anderen daraus zu folgern:

1. $P(\emptyset) = 0$. Denn, wenn $A_1 = A_2 = \dots = \emptyset$, dann sind $A_1, A_2, \dots \in \Sigma$ alle paarweise disjunkt, damit folgt:

$$P(\emptyset) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \underbrace{P(A_i)}_{=P(\emptyset)} = \sum_{i=1}^{\infty} P(\emptyset)$$

Sei $c := P(\emptyset)$. Dann gilt $c \in [0, 1]$ und wir vermerken:

- a) $c = 0: 0 = \sum_{i=1}^{\infty} 0 = 0 \in [0, 1]$ ✓
- b) $c > 0: c = \sum_{i=1}^{\infty} c = \infty \notin [0, 1]$ ✗

Damit gilt: $c = 0$.

2. Seien $A_1, A_2, \dots, A_n \in \Sigma$ p.d.. Setze nun $A_{n+1} = A_{n+2} = \dots = \emptyset$. Dann ist $A_1, A_2, \dots \in \Sigma$ wieder p.d. und es folgt:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P(A_1 \cup \dots \cup A_n \cup \emptyset \cup \emptyset \cup \dots) \\ &= P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^n P(A_i) \end{aligned}$$

Also gilt: $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ (nach der Additivität). Insbesondere gilt damit übrigens auch: $P(A \cup B) = P(A) + P(B)$, falls $A, B \in \Sigma$ disjunkt sind. Die Wahl $B = A^c$ liefert damit: $A, A^c \in \Sigma$, welche beide disjunkt sind und damit:

$$1 = P(\underbrace{A \cup A^c}_{=\Omega}) = P(A) + P(A^c)$$

Damit folgt aber auch $P(A^c) = 1 - P(A)$.

Beispiel 2.5 – Konstruktion von Wahrscheinlichkeitsräumen

1. Versuchen wir einen Münzwurf vollständig mathematisch zu modellieren. Hierzu konstruieren wir den **W-Raum** (Ω, Σ, P) als:
 Ω ist, da nur zwei Ergebnisse: $\{K, Z\}$
 Σ ist, da endlich, die Potenzmenge: $\Sigma = \mathcal{P}(\Omega) = \{\emptyset, \Omega, \{K\}, \{Z\}\}$
 P ist $\Sigma \rightarrow [0, 1]$ mit $P(\Omega) = 1, P(\emptyset) = 0, P(\{K\}) = P(\{Z\}) = 1/2$.
2. Modellieren wir uns nun einmal einen Würfelwurf wieder mit dem **W-Raum** (Ω, Σ, P) :
 Ω ist, da sechs Seiten: $\{1, 2, 3, 4, 5, 6\}$
 Σ ist, da endlich, die Potenzmenge: $\Sigma = \mathcal{P}(\Omega) = \{\emptyset, \Omega, \{1\}, \{2\}, \dots\}$ (da diese entsprechend Wortreich zu schreiben wäre, belassen wir es bei $\mathcal{P}(\Omega)$)
 P ist $\Sigma \rightarrow [0, 1]$ mit $P(\Omega) = 1, P(\emptyset) = 0, P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = 1/6$ (bei einem fairen Würfel). Wir müssen aber nun noch die weiteren Teilmengen von Σ betrachten. Also: $P(\{1, 2\}) = 2/6 = 1/3$ (analog für die anderen zwei-Elementigen), $P(\{1, 2, 3\}) = 3/6 = 1/2$, analog für alle drei Elementigen und so weiter. Das nun für alle aufzuschreiben wäre recht aufwändig, weswegen wir eine Kurzform verwenden: $P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6}$ für alle $A \in \Sigma$, wobei $|A|$ die Anzahl der Elemente in A (also die *Kardinalität von A* bezeichnet).
3. Modellieren wir uns das (biologische) Geschlecht des nächsten Neugeborenen in einem Land mit dem **W-Raum** (Ω, Σ, P) mit:
 Ω ist, da nur zwei Ergebnisse: $\{W, M\}$ (weiblich und männlich)
 Σ ist, da endlich, die Potenzmenge: $\Sigma = \mathcal{P}(\Omega) = \{\emptyset, \Omega, \{M\}, \{W\}\}$
 P ist $\Sigma \rightarrow [0, 1]$ mit $P(\Omega) = 1, P(\emptyset) = 0, P(\{W\}) = p, P(\{M\}) = 1 - p$ wobei $p \in [0, 1]$ auf Basis relativer Häufigkeiten (mit einer beliebigen Restriktion) erfolgen kann. In Deutschland ist $p \simeq 0.4863$ (im Zeitraum 1970 bis 1999)

Wir möchten im Folgenden nun noch weiter die Eigenschaften von **Wahrscheinlichkeitsmaßen** betrachten...

Satz 2.3.1 – Eigenschaften von Wahrscheinlichkeitsmaßen

Sei (Ω, Σ, P) ein **W-Raum** und seien $A, B, A_1, A_2, \dots \in \Sigma$:

1. Ist $A \subseteq B$, so gilt $P(B) = P(A) + P(B \setminus A)$ sowie $P(A) \leq P(B)$ (Monotonie).
2. Für $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (dies ist eine Verallgemeinerung für die Regel bei **disjunkten** Mengen, da hier $A \cap B = \emptyset$ und damit $P(A \cap B) = P(\emptyset) = 0$ nach der Definition).
3. Zudem gelte die σ -subadditivität: $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$

Beweisen wir nun diese Eigenschaften!

Beweis 2.2 – Zum Satz von oben

Betrachten wir die Komponenten wieder einzeln:

1. Im Fall $A \subseteq B$ gilt: $B = A \cup (B \setminus A)$ mit den **disjunkten Mengen** A und $B \setminus A$. Wir verwenden nun die **Bemerkung von oben**, dass $P(A \cup B) = P(A) + P(B)$ wenn A und B disjunkt,

um zu zeigen, dass:

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A).$$

Da $P(B \setminus A) \geq 0$ folgt $P(B) \geq P(A)$.

2. Es gilt $A \cup B = A \cup (B \setminus A)$ dies lässt sich auch grafisch gut veranschaulichen. In diesem Beispiel zeigt [Abbildung 2.2a](#) die Darstellung von $B \setminus A$ (in Lila markiert). Wir können dies aber noch weiter aufplustern, so gilt $A \cup B = A \cup (B \setminus (A \cap B))$ wobei $A \cap B$ bereits in [Abbildung 2.1a](#) dargestellt wird. Es gilt also: A und $B \setminus (A \cap B)$ sind disjunkt. Damit gilt:

$$\begin{aligned} P(A \cup B) &= P(A \cup (B \setminus (A \cap B))) \\ &= P(A) + P(\underbrace{B \setminus (A \cap B)}_{\subseteq B}) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

3. Seien hierfür $A_1, A_2, \dots \in \Sigma$. Setzen wir nun $A_0 = \emptyset$ und $B_i = A_i \setminus (\bigcup_{j=0}^{i-1} A_j)$ mit $i \in \mathbb{N}$. Es gilt damit auch $B_i \in \Sigma$ (siehe [Abbildung 2.2b](#)). Für alle $i \in \mathbb{N}$ gilt B_1, B_2, \dots sind **paarweise disjunkt** mit $B_j \subseteq A_j$. Uns interessiert hierbei aber, dass $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$ – wobei wir den Beweis über die gegenseitige Teilmengenbeziehung hier einfach als gegeben ansehen. Damit folgt:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} \underbrace{B_i}_{\substack{\text{p.d.} \\ \subseteq A_i}}\right) = \sum_{i=1}^{\infty} P(\underbrace{B_i}_{\subseteq A_i}) \leq \sum_{i=1}^{\infty} P(A_i)$$

Bemerkung 2.3 – Dies gilt auch für endliche Mengen

Die Wahl $A_i = \emptyset$ für alle $i \geq n+1$ liefert, dass für alle $A_1, \dots, A_n \in \Sigma$:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P(A_1 \cup \dots \cup A_n \cup \underbrace{\emptyset}_{=A_{n+1}} \cup \underbrace{\emptyset}_{=A_{n+2}} \cup \emptyset \dots) \\ &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^n P(A_i) \end{aligned}$$

Damit gilt:

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

Die **Subadditivität** ist damit ebenso gezeigt.

Wir betrachten nun verschiedene Wahrscheinlichkeitsräume genauer...

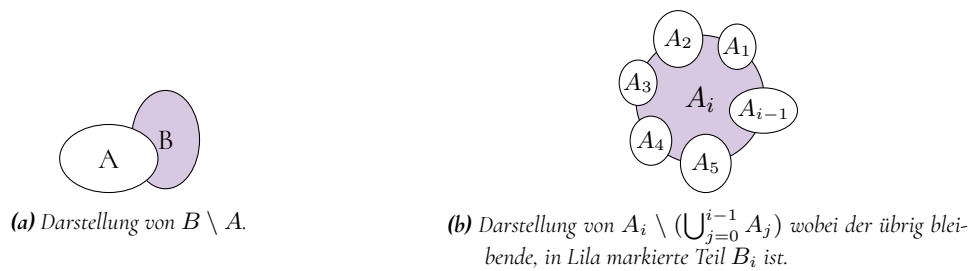


Abb. 2.2: Verschiedene Mengenoperationen skizziert

2.4 Endliche Wahrscheinlichkeitsräume

2.4.1 Laplace und die Gleichverteilung

In diesem Abschnitt betrachten wir ein endliches Ω , das heißt $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ für verschiedene ω_i ist endlich, wir können also $\Sigma = \mathcal{P}(\Omega)$ wählen. Darüber hinaus ist jedes **W-Maß** P bereits durch die Werte $p_i = P(\{\omega_i\})$ mit $i = 1, \dots, n$ von Elementarereignissen $\{\omega_i\}$ *eindeutig bestimmt*. Da es für jedes Ereignis $A \in \Sigma$: $P(A) = P(\bigcup_{\omega_i \in A} \{\omega_i\})$ wobei alle ω_i **p.d.** sind (da sie oben bereits als verschieden definiert wurden), können wir schreiben $P(A) = \sum_{\omega_i \in A} P(\{\omega_i\}) = \sum_{\omega_i \in A} p_i$. Das ist zum Beispiel dann nicht mehr der Fall, wenn $\Omega = \mathbb{R}$ (aber solche Ω betrachten wir in diesem Abschnitt ja nicht).

Beispiel 2.6 – Würfeln

Hier ist $\Omega = \{1, 2, \dots, 6\}$ und $\Sigma = \mathcal{P}(\Omega)$ reicht es zu definieren: $p_i = P(\{\omega_i\}) = 1/6$ für alle $i = 1, \dots, 6$ und $P(A) = \sum_{\omega_i \in A} p_i = \sum_{\omega_i \in A} 1/6 = \frac{|A|}{6}$.

Definieren wir den Begriff der Endlichkeit noch einmal gesondert.

Definition 2.4.1 – Endlicher Wahrscheinlichkeitsraum

Ein **W-Raum** (Ω, Σ, P) heißt *endlicher W-Raum*, falls Ω endlich ist und $\Sigma = \mathcal{P}(\Omega)$ gilt.

Bemerkung 2.4 – Generalisierbarkeit

Sei (Ω, Σ, P) ein **endlicher W-Raum**. Dann ist es im allgemeinen *nicht* der Fall, dass alle Elementarereignisse gleich wahrscheinlich sind. ein solches Beispiel hatten wir bereits bei der Wahl des Geschlechts: Beispiel zu Kindern **Punkt 3**. Es handelt sich hierbei um einen Spezialfall (der aber dennoch ganz nett zu betrachten ist).

Wir betrachten nun diese **endliche W-Räume** bei welchen alle Elementarereignisse die gleiche Wahrscheinlichkeit besitzen genauer.

Definition 2.4.2 – Laplace-Wahrscheinlichkeitsraum

Ein **endlicher Wahrscheinlichkeitsraum** (Ω, Σ, P) mit $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ und $p_i = P(\{\omega_i\}) = \frac{1}{|\Omega|} = 1/n$ für alle $i = 1, \dots, n$ heißt *Laplace-W-Raum* (Beispiel: Münzwurf). Das **Wahrscheinlichkeitsmaß** P trägt dann den Bezeichner *diskrete Gleichverteilung*. Wir schreiben dann vom englischen *Uniform* auch: $P = U(\Omega)$.

Definition 2.4.3 – Laplace-Experiment

Ein **Zufallsexperiment**, welches durch einen **Laplace-Raum** beschrieben ist, nennt man *Laplace-Experiment*.

Sei nun (Ω, Σ, P) ein **Laplace W-Raum**. Dann folgt aus

$$P(A) = \sum_{\omega_i \in A} P(\{\omega_i\}) = \sum_{\omega_i \in A} P_i$$

die Gleichung:

$$\begin{aligned} P(A) &= \sum_{\omega_i \in A} P(\{\omega_i\}) \sum_{\omega_i \in A} \frac{1}{|\Omega|} \\ &= \frac{1}{|\Omega|} \sum_{\omega_i \in A} 1 = \frac{|A|}{|\Omega|} \end{aligned}$$

Es gilt also: $P(A) = \frac{|A|}{|\Omega|}$.

Beispiel 2.7 – Die klassischen Laplace-Experimente

1. Ein Münzwurf und ein Würfel werden gewöhnlich als **Laplace-Experimente** angesehen (dies liegt in der Symmetrie der geworfenen Objekte begründet). Es gibt *keinen* Grund warum (im optimalen Fall) eine Seite bevorzugt werden sollte.
2. Betrachten wir nun ein zweimaliges Würfeln. Dieses Experiment kann durch einen **Laplace-W-Raum** modelliert werden, wobei $\Omega = \{1, \dots, 6\}^2 := \{(i, j) \mid i, j \in \{1, \dots, 6\}\}$. Diese Angabe genügt bei einem Laplace-Raum bereits für die Modellierung, da dann automatisch $\Sigma = \mathcal{P}(\Omega)$ und $P = U(\Omega)$. Sei nun $A =$ „Die Augensumme ist ≥ 10 “. Es gilt: $P(A) = \frac{|A|}{|\Omega|}$. Durch die Modellierung ergibt sich $|\Omega| = 6 \cdot 6 = 36$. Für die Kardinalität von A schreiben wir die Menge zuerst einmal auf: $A = \{(4, 6), (5, 6), (6, 6), (6, 5), (6, 4), (5, 5)\}$. Damit gilt $|A| = 6$ und somit $P(A) = \frac{6}{36} = 1/6$.

Bemerkung 2.5 – Nicht alles kann Laplace sein

Experimente mit endliche vielen Ergebnissen versucht man „immer“ (automatisch) durch **Laplace-Experimente** zu beschreiben. Dies kann aber gefährlich sein, da man zuerst prüfen muss, dass auch wirklich alle $\omega \in \Omega$ gleich wahrscheinlich sind.

Diese Bemerkung ist wichtig! Wir werden uns nun ein Beispiel ansehen, bei welchem das gefährlich sein kann:

Beispiel 2.8 – Das Problem mit Laplace

Wir nehmen zwei nicht unterscheidbare Münzen, welche gleichzeitig geworfen werden und suchen nun ein Modell für dieses **Zufallsexperiment**. Wir setzen mit einem Versuche an:

Wir haben drei mögliche Ergebnisse: $\Omega = \{KK, KZ, ZZ\}$ die Reihenfolge der Ergebnisse ist ja beim gleichzeitigen Werfen irrelevant. Modellieren wir dies also als **Laplace-Raum** mit dem obigen Ω . Dann würde gelten: $P(\{KK\}) = P(\{KZ\}) = P(\{ZZ\}) = 1/3$. Das Modell ist so aber *nicht richtig*! Das Ergebnis KZ hat in der Realität die doppelte Wahrscheinlichkeit im Vergleich zu KK und ZZ .

Wir können genauer argumentieren, wenn wir die Münzen zur Identifikation nummerieren. Dann können wir das Zufallsexperiment durch den **Laplace-Raum** $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{P})$ modellieren, wobei $\tilde{\Omega} = \{kk, kz, zk, zz\}$. Also gilt es, dass $\tilde{P}(\{kk\}) = \tilde{P}(\{kz\}) = \tilde{P}(\{zk\}) = \tilde{P}(\{zz\}) = 1/4$. Im ursprünglichen Experiment sind die Elementarereignisse $\{kz\}$ und $\{zk\}$ ^(a) verschmolzen. Wir müssen also $P(\{KZ\}) = 1/2$ sowie $P(\{KK\}) = P(\{ZZ\}) = 1/4$ definieren – womit wir die *Laplace-Eigenschaft* verlieren.

^(a) wir schreiben sie hier nur zur Trennung klein, es handelt sich ja ohnehin nur um abstrakte Symbole.

Unter Umständen kann es nun aber äußerst schwierig sein, die Kardinalitäten von A und Ω zu bestimmen. Wir werden nun also mehr Möglichkeiten und Tricks lernen, für bestimmte Arten von Mengen die Kardinalität zu berechnen.

Satz 2.4.1 – Die Kardinalität des Kreuzproduktes

Seien A_1, A_2, \dots, A_n endliche Mengen, sowie $A = \{(a_1, a_2, \dots, a_n) \mid a_1 \in A_1, a_2 \in A_2, \dots, a_n \in A_n\}$. Sei also A die Menge der Vektoren über $A = A_1 \times A_2 \times \dots \times A_n$. Dann gilt:

$$|A| = |A_1| \cdot |A_2| \cdot \dots \cdot |A_n|$$

Der Beweis für diesen Satz kann der Kombinatorik entnommen werden ☺. Wir werden hier aus Zeitgründen nur Beispiele betrachten.

Beispiel 2.9 – Thorsten, Sie brauchen den Job!

Thorsten hat ein Bewerbungsgespräch und stellt dazu ein passendes Outfit zusammen. Zur Verfügung stehen ihm hierbei 3 Hemden (A_1), 4 Krawatten (A_2) und 2 Anzüge (A_3). Wir fragen uns nun, wie viele verschiedene Kleidungskombinationen er zur Verfügung hat.

Nach dem **Satz von eben** berechnet sich dies zu: $3 \cdot 4 \cdot 2 = 24$ (unter der Annahme, dass er nur ein Hemd, eine Krawatte und einen Anzug gleichzeitig tragen kann).

Wir betrachten neben diesem recht einfachen Beispiel noch ein komplizierteres:

Beispiel 2.10 – Das Geburtstagsproblem

Seien $n \leq 365$ StudentInnen in dieser Vorlesung. Wie groß ist nun die Wahrscheinlichkeit, dass mindestens zwei dieser n StudentInnen am gleichen Tag Geburtstag haben? Intuitiv ist klar, dass die Wahrscheinlichkeit mit dem n skaliert. Aber wie skaliert die Wahrscheinlichkeit? Bevor wir dies formal lösen, stellen wir zwei Annahmen um das Problem transparent zu formulieren:

1. Das Jahr hat $M = 365$ Tage (wir ignorieren also Schalt- und Säkularjahre).
2. Jeder Geburts-Tag ist gleich wahrscheinlich, also statistisch ist jeder Tag als Geburtstag gleich wahrscheinlich.

Warum haben wir die Annahme mit $n \leq 365$ getroffen? Nun mit $n \geq 366$ ist die Wahrscheinlichkeit aufgrund zwangsläufig auftretenden Kollision (im Extremfall sind 365 Körbe mit verschiedenen Kugeln belegt und eine ist „noch übrig“) immer 1.

Modellieren wir uns nun also das **Zufallsexperiment** durch $\Omega = \{(\omega_1, \dots, \omega_n) : \omega_1, \dots, \omega_n \in \{1, \dots, M\}\}$ wobei ω_i den Geburtstag einer StudentIn bezeichne und wir mit $1, \dots, M$ alle Tage von 1 (1. Januar) bis 365 (31. Dezember) durchnummerieren. Der Vektor $(\omega_1, \dots, \omega_n)$ bezeichne hierbei alle Geburtstage. Wir können dies damit auch als Kreuzprodukt schreiben:

$$\Omega = \{1, \dots, M\} \times \{1, \dots, M\} \times \dots \times \{1, \dots, M\} = \{1, \dots, M\}^n.$$

Da alle Elemente des Vektors sind Gleichwahrscheinlich und unabhängig voneinander (dies haben wir ja der Einfachheit halber so angenommen). Damit können wir das Geburtstagsproblem als einen **Laplace-Raum** mit $\Omega = \{1, \dots, M\}^n$ modellieren. Nach dem **Satz von eben** ergibt sich damit:

$$\begin{aligned} |\Omega| &= |\{1, \dots, M\}| \cdot |\{1, \dots, M\}| \cdot \dots \cdot |\{1, \dots, M\}| \\ &= M \cdot M \cdot \dots \cdot M = M^n \end{aligned}$$

Sei das **Ereignis** nun: $A =$ „mindestens zwei der n Studenten haben am gleichen Tag Geburtstag“. Doch was ist nun $P(A)$? Hierfür vermerken wir noch einmal, dass gilt $P(A) = 1 - P(A^c)$. Dies ist

in manchen Fällen (und so auch in diesem) viel einfacher zu berechnen. Hier ist $A^c = \text{„alle Studenten haben an verschiedenen Tagen Geburtstag.“}$ Es gilt also: für den Vektor: $\{(\omega_1, \omega_2, \dots, \omega_n) \in \Omega : \omega_i \neq \omega_j \text{ für } i \neq j\}$.

Damit hat die erste StudentIn n Möglichkeiten für einen Geburtstag, die Zweite hat aber nur noch $n - 1$, da ja bereits einer mit einem Geburtstag belegt, die Dritte nur $n - 2$...

Damit gilt für die Wahrscheinlichkeit:

$$|A^c| = M \cdot (M - 1) \cdot (M - 2) \cdot \dots \cdot (M - n + 1).$$

Also gilt:

$$P(A^c) = \frac{|A^c|}{|\Omega|} = \frac{M \cdot (M - 1) \cdot (M - 2) \cdot \dots \cdot (M - n + 1)}{M^n}$$

und damit gilt für die Wahrscheinlichkeit:

$$P(A) = 1 - \frac{M \cdot (M - 1) \cdot (M - 2) \cdot \dots \cdot (M - n + 1)}{M^n}$$

Wir haben hier überall M anstelle von 365 geschrieben, um die Formel unabhängig von der Anzahl der Tage zu machen (man könnte das Problem ja auch auf Monate beschränken). Für so berechnete Zahlen, siehe [Tabelle 2.1](#).

n	4	16	22	23	50	64
$P(A)$	0.016	0.28	0.48	0.51	0.97	0.99

Tbl. 2.1: Wahrscheinlichkeiten für verschiedene n beim Geburtstagsproblem.

Es gilt zu beachten, dass die Anzahl der Elemente von Mengen Ω und A^c aus dem Geburtstags-Beispiel von oben, lässt sich mit sogenannten „Urnenmodellen“ bestimmen.

2.4.2 Abstecher: Urnenmodell

Definition 2.4.4 – Urnenmodell

Ein Urnenmodell präsentiert eine Urne mit n von 1 bis n nummerierten Kugeln. Aus dieser Urne werden zufällig (nacheinander) k Kugeln gezogen. Das Ergebnis $(\omega_1, \dots, \omega_k)$ bezeichnet die Nummern der gezogenen Kugeln ω_i .

Es gibt verschiedene Möglichkeiten um aus einem **Urnenmodell** zu ziehen:

1. Mit Zurücklegen. Hier werden die Kugeln nach dem Ziehen und Vermerken wieder in die Urne gelegt. Das Ergebnis $(1, 1, 2, 3)$ ist bei vier Ziehungen also möglich.
2. Ohne Zurücklegen. Hier werden die Kugeln mit der Ziehung aus dem Pool der Urne entfernt. Damit ist das Ergebnis $(1, 1, 2, 3)$ bei vier Ziehungen nicht mehr möglich, da die 1-er Kugel nicht mehr in der Urne ist.
3. Mit Beachtung der Reihenfolge: Hier ist die Ziehungsreihenfolge der Kugeln von Relevanz. Wir unterscheiden also $(1, 2, 3, 4)$ und $(2, 3, 4, 1)$.

4. Ohne Beachtung der Reihenfolge: Hier ist die Ziehungsreihenfolge der Kugeln *nicht* von Relevanz. Die Ergebnisse $(1, 2, 3, 4)$ und $(2, 3, 4, 1)$ sind also identisch.

Natürlich können die beiden Ausrichtungen kombiniert werden – so kann man mit Zurücklegen ohne Beachtung der Reihenfolge, mit Zurücklegen und mit Beachtung der Reihenfolge, ... Wir werden nun alle vier so möglichen Fälle betrachten.

Bemerkung 2.6 – Fall 1: Ziehen mit Zurücklegen und Beachtung der Reihenfolge

Hier sei die Menge Ω_1 aller möglichen Ergebnisse durch

$$\Omega_1 = \{(\omega_1, \dots, \omega_k) : \omega_1, \dots, \omega_k \in \{1, \dots, n\}\} = \{1, \dots, n\}^k$$

gegeben. Die Kardinalität dieser Menge haben wir ebenfalls bereits berechnet, so gilt

$$|\Omega_1| = n^k$$

(nach dem **Satz von oben**). Eine solche Menge erhalten wir zum Beispiel beim **Geburtstagsproblem** oder beim zehn-Maligen Würfeln mit: $\Omega_1 = \{1, \dots, 6\}^{10}$.

Definition 2.4.5 – Fakultät

Sei $n \in \mathbb{N}$. Die Zahl $n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1$ heißt die *Fakultät* von n . Wir definieren den Randfall $0! = 1$.

Bemerkung 2.7 – Fall 2: Ziehen ohne Zurücklegen aber mit Beachtung der Reihenfolge

Erstmal muss hier die offensichtliche Beschränkung gelten, dass $k \leq n$. Wir können nicht mehr Kugeln ziehen, als in der Urne vorhanden sind. Nun sei die Menge Ω_2 aller möglichen Ergebnisse durch

$$\Omega_2 = \{(\omega_1, \dots, \omega_k) : \omega_1, \dots, \omega_k \in \{1, \dots, n\} \text{ und } \omega_i \neq \omega_j \text{ für } i \neq j\}$$

gegeben. Wir berechnen nun die Kardinalität: $|\Omega_2| = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$, da wir bei jeder Ziehung ja eine Möglichkeit weniger haben. Für diese Angabe verwenden wir die Fakultät:

$$|\Omega_2| = \frac{n(n-1) \cdot \dots \cdot (n-k+1)(n-k) \cdot (n-k-1) \cdot \dots \cdot 1}{(n-k)(n-k-1) \cdot \dots \cdot 1} = \frac{n!}{(n-k)!}$$

Insbesondere für $n = k$ gilt:

$$|\Omega_2| = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n!.$$

Es handelt sich also um die Anzahl aller Permutationen von $1, \dots, n$. Dieser Fall war mit $A^{\mathbb{C}}$ auch schon beim **Geburtstagsproblem** vertreten. Auch findet man diese Art des **Urnenmodells** bei der Wahl der $k = 5$ Elfmeterschützen am Ende eines Fußballspiels aus $n = 11$ Spielern. Hier gelte zum Beispiel: $\frac{11!}{(11-5)!} = \frac{11!}{6!} = 55\,440$ Möglichkeiten!

Definition 2.4.6 – Binomialkoeffizient

Seien $n \in \mathbb{N}$ und $k \in \{0, \dots, n\}$. Die Zahl

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$$

heißt *Binomialkoeffizient* „ n über k “.

Bemerkung 2.8 – Fall 3: Ziehen ohne Zurücklegen und ohne Beachtung der Reihenfolge

Auch hier muss die offensichtliche Beschränkung gelten, dass $k \leq n$. Wir können nicht mehr Kugeln ziehen, als in der Urne vorhanden sind. Nun sei die Menge Ω_3 aller möglichen Ergebnisse wie folgt notiert. Zu beachten ist hierbei, dass wir zur Einfachheit keine Mengenschreibweise verwenden (obwohl sie auch möglich wäre). Wir werden einfach normieren, dass die Vektoren aufsteigend sortiert sein müssen – damit spielt die Ziehungsreihenfolge keine Rolle:

$$\Omega_3 = \{(\omega_1, \dots, \omega_k) : 1 \leq \omega_1 < \omega_2 < \omega_3 < \dots < \omega_n \leq n\}$$

gegeben. Wir berechnen nun die Kardinalität: $|\Omega_3| \stackrel{\Omega_2}{=} \frac{n!}{k!}$. Dies lässt sich einfach erklären – Ω_2 liefert uns alle möglichen Ergebnisse und beachtet die Reihenfolge. Durch die Division mit $k!$ „fassen“ wir die Anzahl der möglichen Permutationen pro Vektor zusammen. Kurz:

$$|\Omega_3| = \binom{n}{k}.$$

Dieser Fall findet sich zum Beispiel im Lotto, bei dem aus 49 Zahlen $1, \dots, 49$ insgesamt 6 Zahlen gewählt werden. Die Wahrscheinlichkeit beträgt also: $\binom{49}{6} = 13\,983\,816$ verschiedene Möglichkeiten.

Bemerkung 2.9 – Fall 4: Ziehen mit Zurücklegen aber ohne Beachtung der Reihenfolge

Hier gilt, dass die Ereignisse $(1, 2, 2)$, $(2, 1, 2)$ und $(2, 2, 1)$ für uns möglich und gleich sind. In diesem Beispiel sei die Menge Ω_4 aller möglichen Ergebnisse wie folgt notiert. Allerdings konstruieren wir den Vektor in sortierter Reihenfolge. Die Ereignisse $1, 2, \dots$ seien hierin aufsteigend sortiert. Wir schreiben also für

$$\Omega_4 = \{(\omega_1, \dots, \omega_k) : 1 \leq \omega_1 \leq \omega_2 \leq \dots \leq \omega_k \leq n\}.$$

Hier besteht der Unterschied zum Ω_3 im \leq anstelle des $<$, da wir ja mit Zurücklegen arbeiten. Ein Beispiel wäre hier das gleichzeitige Werfen zweier Würfel. Hier ist $n = 6$ und $k = 2$, womit wir insgesamt auf (siehe **nachfolgender Satz**):

$$\binom{6+2-1}{2} = \binom{7}{2} = \frac{7!}{2!} = \frac{7!}{5! \cdot 2!} = \frac{7 \cdot 6}{2} = 21$$

Damit sind 21 verschiedene Ergebnisse möglich.

Satz 2.4.2 – Anzahl der Elemente beim Zurücklegen ohne Beachtung der Reihenfolge

Wir sagen hier:

$$|\Omega_4| = \binom{n+k-1}{k}.$$

Da sich der Beweis hierzu als schwierig gestaltet, werden wir das einfach so „akzeptieren“.

Bemerkung 2.10 – Urnenmodelle und Erinnerung

Die Urnenmodelle helfen uns dabei die Anzahl an möglichen Ereignissen in einem Zufallsexperiment zu berechnen.

Sei (Ω, Σ, P) ein **endlicher W-Raum**. P heißt diskrete Gleichverteilung wenn gilt, dass $P(\omega_i) = \frac{1}{|\Omega|}$ für alle $i = 1, \dots, n$ wobei $\Omega = \{\omega_1, \dots, \omega_n\}$.

Wir werden nun weitere Verteilungen kennenlernen...

2.4.3 Hypergeometrische Verteilung

Wir nehmen uns hierzu wieder das **Urnenmodell** zur Hilfe. Befinden sich in einer Urne also n Kugeln, mit zwei verschiedenen Farben. Seien die Farben hier einmal *blau* und *weiß* und weiter B der Kugeln blau (somit sind $n - B$ der Kugeln weiß), wobei $B \in \{0, \dots, n\}$. Im Experiment ziehen wir nun $k \in \{1, \dots, n\}$ aus dieser Urne (ohne Zurücklegen und ohne Beachtung der Reihenfolge, also: **Fall 3**). Nun interessiert uns die Frage „Bestimme die Wahrscheinlichkeit des Ereignisses A_b für $b \in \{0, \dots, k\}$.“ Das Ereignis A_b sei hierbei definiert als $A_b :=$ „ b der k Kugeln sind blau“.

Wir beginnen damit unser Experiment zu modellieren und versuchen es mit einem **Laplace-Experiment**. Hierzu beginnen wir die Kugeln zu nummerieren (mit $1, 2, \dots, n$) und modellieren dieses Zufallsexperiment durch den **Laplace-W-Raum** über $\Omega (= \Omega_3) = \{(\omega_1, \dots, \omega_k) : 1 \leq \omega_1 < \omega_2 < \dots < \omega_k \leq n\}$. Da diese Modellierung gelingt, können wir nun bereits über Wahrscheinlichkeiten reden und erhalten nach Laplace: $P(A_b) = \frac{|A_b|}{|\Omega|}$. Nach dem bereits angesprochenen **Fall 3** gilt für die Anzahl der Elemente in Ω : $|\Omega| = \binom{n}{k}$. Zudem gilt für $b > B$ (es sollen mehr blaue Kugeln gezogen werden als vorhanden) oder $k - b > n - B$ (es sollen mehr weiße Kugeln gezogen werden, als vorhanden), dass $|A_b| = 0$. In allen anderen Fällen berechnen wir:

$$|A_b| = \binom{B}{b} \cdot \binom{n-B}{k-b}.$$

Doch wie kommt diese Wahrscheinlichkeit zustande? $\binom{B}{b}$ bezeichnet die Wahrscheinlichkeit b der B blauen Kugeln zu ziehen. Analog bezeichnet $\binom{n-B}{k-b}$ die Wahrscheinlichkeit $k - b$ weiße Kugeln aus den $n - B$ weißen Kugeln zu ziehen.

Damit folgt für $P(A_b)$ mit dieser Formel:

$$P(A_b) = \begin{cases} 0, & \text{falls } b > B \text{ oder } b < k - (n - B) \\ \frac{\binom{B}{b} \cdot \binom{n-B}{k-b}}{\binom{n}{k}}, & \text{sonst.} \end{cases}$$

Wir können das Zufallsexperiment aber auch durch einen anderen Wahrscheinlichkeitsraum modellieren. Sei hierzu $\Omega = \{0, \dots, k\}$ und $\Sigma = \mathcal{P}(\Omega)$. Nun ist

$$P(\{b\}) = \begin{cases} 0, & \text{falls } b > B \text{ oder } b < k - (n - B) \\ \frac{\binom{B}{b} \cdot \binom{n-B}{k-b}}{\binom{n}{k}}, & \text{sonst.} \end{cases}$$

Damit ist $P(A) = \sum_{b \in A} P(\{b\})$ für $A \in \Sigma$. Dieses Wahrscheinlichkeitsmaß nennen wir *hypergeometrische Verteilung* mit den Parametern n, B, k (mit n für die Gesamtanzahl an Kugeln, B für die Anzahl an blauen Kugeln und k für die Anzahl der zu ziehenden Kugeln). Wir notieren dies als: $P = H(n, B, k)$.

Beispiel 2.11 – Lotto mit „6 aus 49“

Wie wahrscheinlich ist es, mindestens 5 Richtige beim Lotto „6 aus 49“ zu wählen?

Mit $n = 49$, $B = 6$ und $k = 6$ ist die Wahrscheinlichkeit für $P(\text{„mindestens 5 richtige“})$ beschrieben als: $P(\text{„mindestens 5 richtige“}) = P(\text{„genau 5 richtige“}) + P(\text{„genau 6 richtige“})$. Mit der *hypergeometrischen Verteilung* modellieren wir:

$$P(\text{„mindestens 5 richtige“}) = H(49, 6, 6)(\{5\}) + H(49, 6, 6)(\{6\})$$

$$= \frac{\binom{6}{5} \cdot \binom{49-6}{6-5}}{\binom{49}{6}} + \frac{\binom{6}{6} \cdot \binom{49-6}{6-6}}{\binom{49}{6}}$$

$$\approx 0.000017.$$

2.4.4 Die Bernoulliverteilung und die Binomialverteilung

Eine Bernoulliverteilung ist eine einfache Verteilung, welche nur zwei Ergebnisse hat: $\Omega = \{0, 1\}$ wobei wir hier einmal 0 mit „Misserfolg“ und 1 mit „Erfolg“ verbinden. Da es auch hier eine endliche Menge ist gilt $\Sigma = \mathcal{P}(\Omega)$. Sei nun $p \in [0, 1]$ und sei P gegeben durch $P(\{1\}) = p$ und $P(\{0\}) = 1 - p$ wobei $P(\emptyset) = 0, P(\Omega) = 1$. Das **Wahrscheinlichkeitsmaß** P bezeichnen wir als *Bernoulliverteilung* mit Parameter p und schreiben $P = B(1, p)$.

Definition 2.4.7 – Bernoulliverteilung

Die Bernoulliverteilung ist eine Verteilung mit zwei Ergebnissen $\Omega = \{0, 1\}$ für Erfolg und Misserfolg. Mit $p \in [0, 1]$ für die Erfolgswahrscheinlichkeit schreiben wir für $P = B(1, p)$ für die Wahrscheinlichkeit.

Bei der Binomialverteilung wiederholen wir ein **Zufallsexperiment** mit zwei möglichen Ergebnissen (1 für Erfolg, 0 für Misserfolg) n -Mal. Bezeichne p die Wahrscheinlichkeit für einen Erfolg und $1 - p$ die Wahrscheinlichkeit für einen Misserfolg (wir haben also eine Bernoulliverteilung). Wir modellieren dies durch den **Wahrscheinlichkeitsraum** (Ω, Σ, P) mit

$$\Omega = \{(\omega_1, \dots, \omega_n) : \omega_1, \dots, \omega_n \in \{0, 1\}\} = \omega \{0, 1\}^n$$

und $\Sigma = \mathcal{P}(\Omega)$ (da endlich). Nun sei für $\omega = (\omega_1, \dots, \omega_n)$ und die Anzahl der Erfolge $k = \sum_{i=1}^n \omega_i$ die folgende Wahrscheinlichkeit beschrieben

$$P(\{\omega\}) = p^k \cdot (1 - p)^{n-k}.$$

Für ein Ereignis gilt damit $P(A) = \sum_{\omega \in A} P(\{\omega\})$ und $A \in \sigma$.

Die Aufgabe ist es nun die Wahrscheinlichkeit für genau k Erfolge zu bestimmen. Es gilt:

$$\begin{aligned} P(A_k) &= \sum_{\omega \in A_k} P(\{\omega\}) = \sum_{\omega \in A_k} p^k (1 - p)^{n-k} \\ &= p^k (1 - p)^{n-k} \sum_{\omega \in A_k} 1 = p^k (1 - p)^{n-k} |A - k| \\ &= p^k (1 - p)^{n-k} \binom{n}{k} \end{aligned}$$

wobei $A_k =$ „genau k Erfolge“, da wir mit dieser Verteilung wieder bei **Fall 3** gelandet sind. Man kann dieses Zufallsexperiment auch durch den folgenden **Wahrscheinlichkeitsraum** modellieren: $\Omega = \{0, 1, \dots, n\}$ sowie $\Sigma = \mathcal{P}(\Omega)$. Damit gilt $P(\{k\}) = p^k (1 - p)^{n-k} \binom{n}{k}$ und $P(A) = \sum_{k \in A} P(\{k\})$, $A \in \Sigma$. Dieses **Wahrscheinlichkeitsmaß** nennen wir die *Binomialverteilung* mit den Parametern n und p . Man schreibt $P = B(n, p)$.

Definition 2.4.8 – Binomialverteilung

Die Binomialverteilung bezeichnet das n -fache Durchführen eines **Bernoulli-Experiments**. Wir schreiben $P = B(n, p)$.

Beispiel 2.12 – Zur Binomialverteilung

- Wir versuchen es mit 9 maligem Würfeln und suchen die Wahrscheinlichkeit für *genau* eine 6. Wir modellieren also durch (Ω, Σ, P) mit $\Omega = \{0, 1, \dots, 9\}$, $\Sigma = \mathcal{P}(\Omega)$ sowie $P = B(9, 1/6)$ (da die Wahrscheinlichkeit für einen Erfolg, also eine 6 in einem Wurf $1/6$ betrifft). Wir interessieren uns in der **Binomialverteilung** für das Ereignis „genau eine 6“. Damit gilt:

$$P(\text{„genau eine Sechs“}) = \underbrace{B(9, 1/6)}_{=P}(\{1\}) = \left(\frac{1}{6}\right)^1 \cdot \left(\frac{5}{6}\right)^8 \cdot \binom{9}{1} \approx 0.35$$

- Nehmen wir Überraschungseier, wobei in jedem 7. Ei eine Sammelfigur steckt. Wie hoch ist nun die Wahrscheinlichkeit dafür, genau zwei Sammelfiguren zu erhalten beziehungsweise mindestens zwei Sammelfiguren zu erhalten, wenn wir uns 10 Eier kaufen?

Hier Modellieren wir das Konstrukt durch (Ω, Σ, P) mit $\Omega = \{0, 1, \dots, 10\}$ für die Anzahl der gefundenen Sammelfiguren sowie $\Sigma = \mathcal{P}(\Omega)$. Für die Wahrscheinlichkeit ergibt sich mit der **Binomialverteilung** wieder: $P = B(10, 1/7)$. Dann ist die Wahrscheinlichkeit für $P(\text{„genau 2 Sammelfiguren“}) = B(10, 1/7)(\{2\}) = \left(\frac{1}{7}\right)^2 \left(\frac{6}{7}\right)^8 \binom{10}{2} \approx 0.27$.

Für $P(\text{„mindestens 2 Sammelfiguren“})$ gilt:

$$\begin{aligned} P(\text{„mindestens 2 Sammelfiguren“}) &= B(10, \frac{1}{7})(\{2, 3, \dots, 10\}) \\ &= \sum_{k=2}^{10} \left(\frac{1}{7}\right)^k \cdot \left(\frac{6}{7}\right)^{10-k} \cdot \binom{10}{k} \end{aligned}$$

Alternativ können wir auch über das Komplement arbeiten. Sei $A = \text{„mindestens 2 Sammelfiguren“}$. Dann gilt:

$$\begin{aligned} P(A) &= 1 - P(A^c) \\ &= 1 - (P(\text{„keine Sammelfigur“}) + P(\text{„genau eine Sammelfigur“})) \\ &= 1 - B(10, 1/7)(\{0\}) - B(10, 1/7)(\{1\}) \\ &= 1 - \left(\frac{1}{7}\right)^0 \left(\frac{6}{7}\right)^{10} \binom{10}{0} - \left(\frac{1}{7}\right)^1 \left(\frac{6}{7}\right)^9 \binom{10}{1} \\ &\approx 0.43 \end{aligned}$$

2.5 Diskrete Wahrscheinlichkeitsräume

Bisher haben wir ein endliches Ω betrachtet. Wir konnten Ω also darstellen als $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ für ein $n \in \mathbb{N}$. Nun betrachten wir Ω allerdings als *abzählbare, unendliche* Menge, wir können also Ω immernoch als Menge schreiben: $\Omega = \{\omega_1, \omega_2, \dots\}$ und die Elemente auch noch durchnummerieren, nun fehlt uns aber das „maximale“ Element. Betrachten wir ein Beispiel:

Beispiel 2.13 – Anzahl der täglichen Bestellungen eines Artikels

In diesem Fall ist $\Omega = \{0, 1, 2, 3, \dots\}$ also $\Omega = \mathbb{N}_0$. Dieses Beispiel haben wir auch **schon betrachtet**, es dient hier nur noch einmal dazu das System ins Gedächtnis zu rufen.

In diesem Fall können wir $\mathcal{P}(\Omega)$ wieder als Sigma wählen, also $\Sigma = \mathcal{P}(\Omega)$. Weiter ist jedes **Wahrscheinlichkeitsmaß** P auf Σ ebenso wie im Fall eines endlichen Ω durch die Werte $p_i = P(\{\omega_i\})$ mit $\omega_i \in \Omega$ eindeutig bestimmt, da für jedes **Ereignis** $A \in \Sigma$ gilt, dass

$$P(A) = P\left(\bigcup_{\omega_i \in A} \{\omega_i\}\right) = \sum_{\omega_i \in A} P(\{\omega_i\}) = \sum_{\omega_i \in A} p_i$$

da die Mengen $\{\omega_i\}$ jeweils **paarweise disjunkt** sind. Um nun aber die unendlichen Summen zu berechnen müssen wir uns an die *Geometrische Reihe* aus den anderen Mathematik-Veranstaltungen erinnern. Hierzu betrachten wir die geometrische Reihe...

Definition 2.5.1 – Geometrische Reihe

Für alle $x \in (0, 1)$ gilt:

$$\sum_{k=0}^n x^k = \frac{1 - x^{n+1}}{1 - x}$$

für ein $n \in \mathbb{N}$. Dies ist bereits als die geometrische Summe bekannt. Lässt man n nun gegen Unendlich laufen so ergibt sich:

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x}.$$

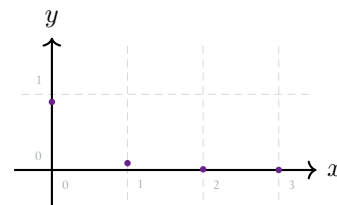
Betrachten wir hierzu ein Beispiel!

Beispiel 2.14 – Berechnen der Wahrscheinlichkeit bei Bestellungen

Sei die Wahrscheinlichkeit, dass wir pro Tag genau $k \in \mathbb{N}_0$ -Bestellungen für das Super-duper-Produkt erhalten, gegebene durch $p(k) = 9 \cdot \left(\frac{1}{10}\right)^{k+1}$ für alle $k \in \mathbb{N} \setminus \{0\}$. Wie groß ist dann die Wahrscheinlichkeit, dass pro Tag mehr als 10 Super-duper-Produkte bestellt werden?

Wir modellieren uns die Menge wie folgt $\Omega = \{0, 1, 2, 3, \dots\} = \mathbb{N}_0$. Wie gegeben gilt damit auch $P(\{k\}) = 9 \cdot \left(\frac{1}{10}\right)^{k+1}$ für alle $k \in \Omega$. Für ein Ereignis gilt weiter $P(A) = \sum_{k \in A} P(\{k\})$ für $A \in \Sigma$. Abgesehen von dem „unendlich“ hat sich also nichts geändert. Veranschaulichen wir uns mal die Wahrscheinlichkeiten für verschiedene k :

- ◇ $k = 0 : P(\{0\}) = 9 \cdot \frac{1}{10} = 0.9$
- ◇ $k = 1 : P(\{1\}) = 9 \cdot \frac{1}{100} = 0.09$
- ◇ $k = 2 : P(\{2\}) = 9 \cdot \frac{1}{1000} = 0.009$
- ◇ ...



Nun ist $A = \text{„Pro Tag werden mehr als 10 Produkte bestellt“}$ und wir suchen $P(A) = P(\{11, 12, 13, \dots\})$. Hier ist es einfacher über das Komplement von A zu argumentieren: $A^c = \{0, 1, 2, \dots, 10\}$. Damit gilt:

$$\begin{aligned} P(A^c) &= \sum_{k \in A^c} P(\{k\}) = \sum_{k=0}^{10} 9 \cdot \left(\frac{1}{10}\right)^{k+1} \\ &= 9 \cdot \sum_{k=0}^{10} \left(\frac{1}{10}\right)^{k+1} = 9 \cdot \frac{1}{10} \cdot \sum_{k=0}^{10} \left(\frac{1}{10}\right)^k \end{aligned}$$

$$= \frac{9}{10} \cdot \frac{1 - \left(\frac{1}{10}\right)^{11}}{1 - \frac{1}{10}} = \frac{9}{10} \cdot \frac{1 - \frac{1}{10^{11}}}{\frac{9}{10}} = 1 - \frac{1}{10^{11}}$$

Damit gilt $P(A) = 1 - P(A^c) = 1 - (1 - \frac{1}{10^{11}}) = \frac{1}{10^{11}}$.

Wir werden uns nun zwei relevante Wahrscheinlichkeitsmaße genauer ansehen!

2.5.1 Die Poisson-Verteilung

Wir definieren zuerst:

Definition 2.5.2 – Poisson-Verteilung

Sei $\Omega = \mathbb{N}_0$, $\Sigma = \mathcal{P}(\Omega)$ sowie ein Parameter $\lambda > 0$ gegeben. So bezeichnet

$$P(\{k\}) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad (k \in \mathbb{N}_0)$$

mit:

$$P(A) = \sum_{k \in A} P(\{k\}), \quad A \in \Sigma$$

die Poisson-Verteilung. Wir schreiben $P = P(\lambda)$.

Wir wollen nun einmal zeigen, dass es sich auch wirklich um ein gültiges Wahrscheinlichkeitsmaß handelt! Es ist klar, dass $P(A) \geq 0$ gilt. Weiter gilt:

$$P(\Omega) = \sum_{k=0}^{\infty} P(\{k\}) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{=e^{\lambda}} = 1$$

Die letzte Vereinfachung erfolgt darüber, dass es sich bei $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$ um die Taylor-Darstellung von e^{λ} handelt. Damit gilt $P(\lambda) = 1$ und $P(A) \leq 1$ für $A \in \Sigma$. Weiter lässt sich übrigens noch zeigen, dass auch die **σ -Additivität** gilt, das wollen wir hier aber nicht tun. Damit gilt, dass es sich bei P um ein **Wahrscheinlichkeitsmaß** handelt!

Doch warum betrachten wir die Verteilung überhaupt?

Satz 2.5.1 – Die Poisson-Approximation

Sei p_1, p_2, \dots eine Folge mit $p_j \in (0, 1)$ für $j \in \mathbb{N}$, wobei gilt:

$$\lim_{n \rightarrow \infty} n \cdot p_n = \lambda > 0.$$

So konvergiert die Folge $p_n = \frac{\lambda}{n}$ für $\lambda = 2$ gegen 2. Dann gilt (mit der **Binomialverteilung**)

$$B(n, p_n)(\{k\}) \rightarrow P(\lambda)(\{k\}) \quad \text{für alle } k \in \mathbb{N}_0.$$

Das heißt die Binomialverteilung $\binom{n}{k} p_n^k (1 - p_n)^{n-k}$ konvergiert gegen $\frac{\lambda^k}{k!} e^{-\lambda}$ für $n \rightarrow \infty$ für alle $k \in \mathbb{N}_0$.

Wie interpretiert sich dies? Die **Possion-Verteilung** taucht dann auf, wenn *viele* Versuche mit *kleiner* Erfolgswahrscheinlichkeit durchgeführt werden, das heißt, falls n „groß“ und p „klein“ ist. Für diese Approximation machen wir ein Beispiel!

Beispiel 2.15 – Isch bin a wunderschöner Schmedderling

Ein einzelner Schmetterling einer bestimmten Spezies zeugt 2000 Nachkommen. Diese Nachkommen werden unabhängig voneinander mit einer Wahrscheinlichkeit von 0.001 das geschlechtsreife Alter erreichen. Wie groß ist nun die Wahrscheinlichkeit, dass mindestens 2 der Nachkommen das geschlechtsreife Alter erreichen?

Wir modellieren hierzu wieder ein Zufallsexperiment mit $n = 2000$ und damit $\Omega = \{0, 1, \dots, 2000\}$ als **Ergebnisraum**. Damit gilt auch $\Sigma = \mathcal{P}(\Omega)$. Doch wie modellieren wir nun P ? Dies lässt sich mit der **Binomialverteilung** ganz einfach als $P = B(n; p)$ präsentieren. Unser A ist nun „mindestens 2 Nachkommen erreichen das geschlechtsreife Alter“. Mathematisch: $A = \{2, 3, \dots, n\}$. Wir arbeiten hier wieder mit der Gegenwahrscheinlichkeit ($A^c = \{0, 1\}$) und damit $P(A) = 1 - P(A^c)$. Damit gilt:

$$\begin{aligned} P(A) &= 1 - P(\{0\}) - P(\{1\}) \\ &= 1 - \left(\binom{2000}{0} p^0 (1-p)^{2000} \right) - \left(\binom{2000}{1} p^1 (1-p)^{1999} \right) \end{aligned}$$

Dies lässt sich sehr schwer berechnen („0.999¹⁹⁹⁹ =?“). Alternativ lässt sich die Wahrscheinlichkeit durch die **Poisson-Verteilung** approximieren! Hier gilt $n \cdot p = 2000 \cdot 0.001 = 2$. Damit gilt $P(\{0\}) = B(n, p)(\{0\}) \approx \frac{2^0}{0!} \cdot e^{-2} = e^{-2}$ sowie $P(\{1\}) = B(n, p)(\{1\}) \approx \frac{2^1}{1!} \cdot e^{-2} = 2 \cdot e^{-2}$. Damit erhalten wir:

$$P(a) \approx 1 - e^{-2} - 2e^{-2} \approx 0.593.$$

2.5.2 Die geometrische Verteilung

Wir betrachten zunächst das folgende Zufallsexperiment, welches wieder mit „Erfolg“ und „Misserfolg“ arbeitet. Die Erfolgswahrscheinlichkeit sei hier $p \in (0, 1]$ und das Experiment soll solange durchgeführt werden, bis der erste Erfolg eintritt. Die Frage ist hier an welcher Stelle dieser Erfolg eintritt!

Doch wie modellieren wir dies?

hier ist $\Omega = \mathbb{N}$, da der Erfolg ja an jeder Stelle auftreten kann. Weiter ist $\Sigma = \mathcal{P}(\Omega)$ und damit $P(\{k\}) = (1-p)^{k-1} \cdot p$ für $k \in \mathbb{N}$. Für einen Erfolg an der Stelle k haben wir ja $k-1$ Misserfolge und dann einen Erfolg! Dass es sich bei P um ein **Wahrscheinlichkeitsmaß** handelt, können wir prüfen! So gilt $P(A) \geq 0$ für $A \in \sigma$ und

$$\begin{aligned} P(\Omega) &= \sum_{k=1}^{\infty} P(\{k\}) = \sum_{k=1}^{\infty} (1-p)^{k-1} \cdot p \\ &= p \cdot \sum_{k=1}^{\infty} (1-p)^{k-1} \stackrel{\ell=k-1}{=} p \sum_{\ell=0}^{\infty} (1-p)^{\ell} \\ &= p \cdot \frac{1}{1-(1-p)} = 1. \end{aligned}$$

Zudem gilt für $\forall A \in \Sigma : P(A) \leq P(\Omega) = 1$. Auch die σ -Additivität lässt sich zeigen, damit handelt es sich bei P auch wirklich um ein **Wahrscheinlichkeitsmaß**. Dieses Maß besitzt auch einen besonderen Bezeichner!

Definition 2.5.3 – Geometrische Verteilung

Die geometrische Verteilung mit Parameter $p \in (0, 1]$ (geschrieben $P = G(p)$) definiert sich als:

$$P(\{k\}) = (1 - p)^{k-1} \cdot p, \quad (\text{für } k \in \mathbb{N})$$

Und bezeichnet die Wahrscheinlichkeit für $k - 1$ Misserfolge und einen Erfolg.

Wir betrachten ein Beispiel...

Beispiel 2.16 – Roll until six

Wir Würfeln für dieses Beispiel mit einem fairen Würfel solange, bis die erste 6 eintrifft. Wir suchen nun die Wahrscheinlichkeit dafür, dass die erste 6 nach einer geraden Anzahl an Würfeln kommt (also eine nicht-Sechs und eine Sechs oder drei nicht-Sechsen und eine Sechs).

Wir modellieren hierzu wieder $\Omega = \mathbb{N}$, da die erste 6 an jeder Stelle kommen kann. Weiter gilt damit $\Sigma = \mathcal{P}(\Omega)$. Für die Wahrscheinlichkeit ergibt sich, da wir es ja als Erfolg- ($p_6 = 1/6$) und Misserfolg-Experiment ($p_{\bar{6}} = 5/6$) betrachten können: $P = G(\frac{1}{6})$. Damit gilt:

$$P(\{2, 4, 6, \dots\}) = \sum_{k=1}^{\infty} P(\{2 \cdot k\}) = \sum_{k=1}^{\infty} (1 - p)^{2 \cdot k - 1} \cdot p.$$

Bei dieser Aufgabe vereinfacht uns nun auch das Kompletz den Rechenaufwand nicht weiter. Also rechnen wir weiter:

$$\begin{aligned} P(\{2, 4, 6, \dots\}) &= p \sum_{k=1}^{\infty} (1 - p)^{2k-1} = \frac{1}{6} \cdot \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{2(k-1)} \cdot \frac{5}{6} \\ &= \frac{5}{36} \sum_{k=1}^{\infty} \left(\left(\frac{5}{6}\right)^2\right)^{k-1} = \frac{5}{36} \sum_{k=1}^{\infty} \left(\frac{25}{36}\right)^{k-1} \\ &\stackrel{\ell=k-1}{=} \frac{5}{36} \sum_{\ell=0}^{\infty} \underbrace{\left(\frac{25}{36}\right)^{\ell}}_{= \frac{1}{1 - \frac{25}{36}}} = \frac{5}{36} \cdot \frac{1}{\frac{11}{36}} \\ &= \frac{5}{11} \end{aligned}$$

Wichtig ist zu erkennen, dass diese Wahrscheinlichkeit kleiner ist als die Hälfte! Anders herum gilt für die Wahrscheinlichkeit einer 6 an einer ungeraden Stelle: $P(A^c) = 1 - P(A) = 1 - \frac{5}{11} = \frac{6}{11}$.

Wir definieren nun noch:

Definition 2.5.4 – Diskreter Wahrscheinlichkeitsraum

Ein **Wahrscheinlichkeitsraum** (Ω, Σ, P) heißt *diskret*, falls Ω endlich oder abzählbar endlich ist (höchstens abzählbar) und $\Sigma = \mathcal{P}(\Omega)$ gilt.

2.6 Stetige Wahrscheinlichkeitsräume

In diesem Fall werden wir $\Omega = \mathbb{R}$ betrachten, also einen überabzählbaren **Ergebnisraum**. Diesen haben wir schon mit der **Temperatur** kennengelernt, wobei wir hier den Raum auf ein Intervall eingeschränkt

haben. Hier können wir aber mit \mathbb{R} rechnen, da alle diese Intervalle ja Teilmengen von \mathbb{R} definieren und wir sie durch die Ausweitung auf \mathbb{R} nur vergrößern (was bei Ergebnisräumen ja erlaubt ist).

2.6.1 Die Borel- σ -Algebra

Wie sehen hier nun Σ und P aus? Beginnen wir mit der Auswahl von Σ . Hier haben wir mit $\Sigma = \mathcal{P}(\Omega) = \mathcal{P}(\mathbb{R})$ ein Problem, da diese **Sigma-Algebra** zu groß ist und viele pathologische Ereignisse enthält. Stattdessen verwenden wir eine kleinere σ -Algebra auf \mathbb{R} . Die kleinste Sigma-Algebra...

Definition 2.6.1 – Borel- σ -Algebra

Die kleinste σ -Algebra auf \mathbb{R} , die alle Intervalle $[a, b]$ mit $a \leq b$ enthält, heißt die *Borel- σ -Algebra* und wird mit $\mathcal{B}(\mathbb{R})$ bezeichnet.

Die kleinste Sigma-Algebra bedeutet hier folgendes. Für alle σ -Algebren Σ auf \mathbb{R} muss gelten, dass sie, sofern sie alle $[a, b]$ mit $a \leq b$ enthalten, eine Obermenge von der **Borel- σ -Algebra** sein müssen, also das $\mathcal{B}(\mathbb{R}) \subseteq \Sigma$ gilt. Eine solche Algebra ist gar nicht so einfach zu kreieren, die folgende Bemerkung kann uns aber hierbei helfen...

Bemerkung 2.11 – Noch mehr Borel

1. $\mathcal{B}(\mathbb{R})$ enthält *alle* „relevanten“ Teilmengen von \mathbb{R} und weiter die Mengen $[a, b]$, $[a, b)$, $(a, b]$, (a, b) , (a, ∞) , $-\infty, a$, $[a, \infty)$, $(-\infty, a]$ sowie $\{a\}$ für $a \leq b$.
2. Wichtig ist aber, dass $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$ die Algebra enthält also nicht *alle* Teilmengen (sonst wäre ihre Formulierung ja auch witzlos.)

Nun möchten wir aber auch noch das **Wahrscheinlichkeitsmaß** sinnvoll definieren. Die Idee ist hier, dass wir aus der Folge p_1, p_2, \dots eine Funktion $p : \mathbb{R} \rightarrow \mathbb{R}$ machen und aus der Summe ein Integral (©)! Wir definieren...

Definition 2.6.2 – Wahrscheinlichkeitsdichte

Eine Funktion $p : \mathbb{R} \rightarrow \mathbb{R}$ heißt *Wahrscheinlichkeitsdichte* (Dichte), falls gilt:

1. $p(x) \geq 0$ für alle $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} p(x) dx = 1$.

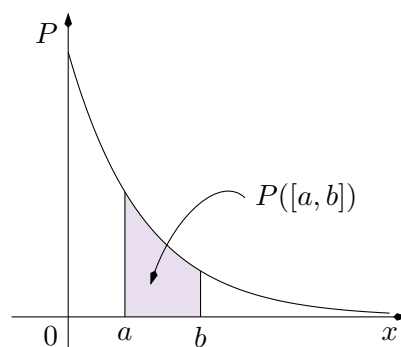


Abb. 2.3: Die Veranschaulichung eines Integrals über ein Intervall.

Beispiel 2.17 – Eine beispiellose Dichte!

Wir möchten nun zeigen, dass die Funktion

$$p(x) = \begin{cases} 0, & \text{falls } x < 0 \\ e^{-x}, & \text{falls } x \geq 0 \end{cases}$$

eine **Dichte** ist. Wir zeigen dies durch den Nachweis der beiden Eigenschaften:

1. $p(x) \geq 0$, für alle $x \in \mathbb{R}$.
2. Nun behandeln wir das Integral (hierbei ist der Exponent $-\infty$ nur in Anführungszeichen gesetzt, da wir dies hier als Vereinfachung für den Limes notieren):

$$\begin{aligned} \int_{-\infty}^{\infty} p(x) \, dx &= \int_0^{\infty} e^{-x} \, dx = -e^{-x} \Big|_0^{\infty} \\ &= \underbrace{-(e^{-\infty})}_0 - \underbrace{(-e^{-0})}_{=1} = 1 \end{aligned}$$

Die Idee ist es nun, P so zu definieren, dass für alle $[a, b]$ mit $a \leq b$ so zu definieren, dass gilt

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

dies ist auch in **Abbildung 2.3** veranschaulicht.

Sei P also zu berechnen für $a = 0$ und $b = 1$. Dann gilt

$$\begin{aligned} P([0, 1]) &= \int_0^1 p(x) \, dx = \int_0^1 e^{-x} \, dx \\ &= -e^{-x} \Big|_0^1 = -e^{-1} - (-e^{-0}) = 1 - \frac{1}{e}. \end{aligned}$$

Doch können wir dies überhaupt machen? Die **Borel- σ -Algebra** enthält ja vielmehr Mengen als nur Intervalle. Hier hilft uns nun das folgende Theorem...

Satz 2.6.1 – Die Dichte ist eindeutig für das Wahrscheinlichkeitsmaß

Sei $p : \mathbb{R} \rightarrow \mathbb{R}$ eine **Dichte**. Es existiert genau ein (!) **Wahrscheinlichkeitsmaß** P auf $\mathcal{B}(\mathbb{R})$ mit

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

für alle $a \leq b$.

Dies wollen wir nun mit dem Begriff *absolut stetig* fundieren...

Definition 2.6.3 – Absolut stetig

Sei $p : \mathbb{R} \rightarrow \mathbb{R}$ eine **Dichte** und sei P ein **Wahrscheinlichkeitsmaß** auf $\mathcal{B}(\mathbb{R})$ mit der Eigenschaft von oben: $P([a, b]) = \int_a^b p(x) \, dx$ für alle $a \leq b$. Dann heißt das Wahrscheinlichkeitsmaß P **absolutstetig** mit **Dichte** p , man sagt auch „ P besitzt Dichte p “.

Ist $\Omega = \mathbb{R}$ wählen wir also $\Sigma = \mathcal{B}(\mathbb{R})$ und p als **Dichte**, die **eindeutig** das Wahrscheinlichkeitsmaß definiert. Hierzu haben wir noch ein Theorem...

Satz 2.6.2 – Das absolutstetige Wahrscheinlichkeitsmaß

Sei P absolutstetig mit der Dichte p . Dann gilt:

1. $P(\{a\}) = 0$ für alle $a \in \mathbb{R}$!
2. $P([a, b]) = P((a, b)) = P([a, b)) = P((a, b]) = \int_a^b p(x) dx$, da $P(\{a\})$, also von einem einzelnen Element Null ist.

Dieses Theorem liefert auch, dass P nicht mehr eindeutig durch die Werte der einzelnen Elementarereignisse $P(\{a\})$ mit $a \in \mathbb{R}$ definiert, da diese 0 sind! Wie angemerkt, wird P durch die Dichte p eindeutig beschrieben!

2.6.2 Stetige Gleichverteilung

Für ein endliches Ω wissen wir, dass $P = U(\Omega)$ gilt und $P(\{\omega\}) = \frac{1}{|\Omega|}$ für $\omega \in \Omega$. Dies können wir für $[0, 1]$ nun aber nicht mehr definieren, da die Kardinalität der Menge unendlich ist und damit jede Wahrscheinlichkeit 0. Stattdessen betrachten wir ein Intervall und sagen, dass zwei Intervalle mit „gleicher Länge“ die „gleiche Wahrscheinlichkeit“ besitzen.

Wir definieren uns nun mit $a < b$ die folgende Dichte:

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{sonst.} \end{cases}$$

Diese ist auch in Abbildung 2.4 veranschaulicht! Wir prüfen, ob es sich auch wirklich um eine Dichte handelt:

1. $p(x) \geq 0$ für alle $x \in \mathbb{R}$ (nach Definition).
2. $\int_{-\infty}^{\infty} p(x) dx$ lässt sich auf $[a, b]$ beschränken, da nur hier nicht 0. Damit gilt:

$$\int_{-\infty}^{\infty} p(x) dx = \int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} \cdot x \Big|_a^b = \frac{b-a}{b-a} = 1$$

Wenn P die Dichte p besitzt, nennen wir P eine stetige Gleichverteilung.

Definition 2.6.4 – Stetige Gleichverteilung

Eine stetige Gleichverteilung ist das Wahrscheinlichkeitsmaß P welches über die folgende Dichte p eindeutig auf dem Intervall $[a, b]$ mit $a < b$ definiert wird:

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{sonst.} \end{cases}$$

Man schreibt $P = \mathcal{U}([a, b])$.

Nun haben wir aber einfach behauptet, dass ein Intervall gleicher Länge auch die gleiche Wahrscheinlichkeit besitzt. Dies wollen wir nun auch beweisen!

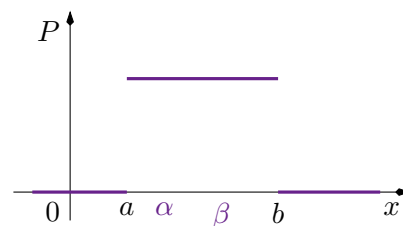


Abb. 2.4: Die stetige Gleichverteilung.

Bemerkung 2.12 – Gleich lang gleich wahrscheinlich ☺

Sei $[\alpha, \beta] \subseteq [a, b]$ (siehe **Abbildung 2.4**). Dann gilt

$$\begin{aligned} P([\alpha, \beta]) &= \int_{\alpha}^{\beta} p(x) \, dx = \int_{\alpha}^{\beta} \frac{1}{b-a} \, dx \\ &= \frac{1}{b-a} \cdot x \Big|_{\alpha}^{\beta} = \frac{\beta - \alpha}{b-a} \\ &= \frac{\text{Länge von } [\alpha, \beta]}{\text{Länge von } [a, b]} \end{aligned}$$

Damit gilt: wenn das Intervall $[\alpha, \beta]$ gleich lang ist wie das Intervall $[a, b]$, so sind auch ihre Wahrscheinlichkeiten identisch (zumindest aus einer intuitiven Betrachtung).

Beispiel 2.18 – Ein Stock, ein langer Stock. Der gebrochen wird

Ein Stock der Länge $L > 0$ wird zufällig in zwei Teile gebrochen. Wie groß ist die Wahrscheinlichkeit, dass eines der beiden Teile mindestens doppelt so lang ist wie das andere?

Zur Lösung modellieren wir die Position der Bruchteile durch den **Wahrscheinlichkeitsraum** $\Omega = \mathbb{R}$, $\Sigma = \mathcal{B}(\mathbb{R})$, $P = \mathcal{U}([0, L])$ wobei L wie angemerkt der Länge entspricht.

Sei A nun definiert als „Einer der beiden Teile ist mindestens doppelt so lange wie der andere.“ Was ist nun $P(A)$?

Hierzu ist es notwendig, A mathematisch modellieren: $A = [0, L/3] \cup [2L/3, L]$. Doch warum? Ein Beispiel, in dem der eine Stock doppelt so lang ist wie der andere ist $L_1 = 1/3$ und $L_2 = 2/3$ (wobei L_1, L_2 die Längen der beiden Teile sind). Damit ist die Bedingung erfüllt, wenn die Länge des Stocks entweder $\leq L/3$ lang ist oder $\geq 2L/3$, wobei er natürlich nicht weniger lang als 0 und länger als L sein kann. Wir können hier die Randfälle $= 0$ und $= L$ hinzufügen, da diese Intervalle nach **der Bemerkung zu Borel** gleich wahrscheinlich sind, wir müssen hier also gar nicht verhandeln, ob das jetzt Sinn macht oder nicht ☺. Mit der Definition von A ist die Berechnung, da die Intervalle weiter **disjunkt** sind, sehr leicht:

$$\begin{aligned} P(A) &= P([0, L/3] \cup [2L/3, L]) \\ &= P([0, L/3]) + P([2L/3, L]) \\ &= \frac{L/3 - 0}{L - 0} + \frac{L - 2L/3}{L - 0} \\ &= \frac{L/3}{L} + \frac{L/3}{L} \\ &= \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \end{aligned}$$

2.6.3 Exponentialverteilung

Diese Verteilung lässt sich nur schwer motivieren. Wir akzeptieren vorerst, dass sie in der Realität häufig Anwendung findet.

Sei $\lambda > 0$ und

$$p(x) = \begin{cases} 0, & \text{falls } x < 0 \\ \lambda e^{-\lambda x}, & \text{falls } x \geq 0 \end{cases}$$

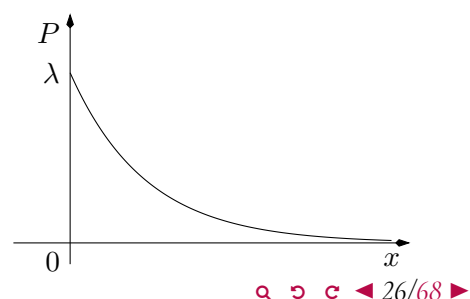


Abb. 2.5: Die Exponentialverteilung.

Dann ist p eine **Dichte**, da wir die Eigenschaften zeigen können:

1. $p(x) \geq 0$ für alle $x \in \mathbb{R}$ (nach der Definition und den Eigenschaften der Exponentialverteilung).
2. Für das Integral gilt:

$$\begin{aligned} \int_{-\infty}^{\infty} p(x) \, dx &= \int_0^{\infty} \lambda e^{-\lambda x} \, dx \\ &= -e^{-\lambda x} \Big|_0^{\infty} \\ &= \underbrace{-e^{-\lambda \cdot \infty}}_{=0} - \underbrace{(-e^{-\lambda \cdot 0})}_{=1} = 1 \end{aligned}$$

Besitzt P die Dichte p , so nennen wir P eine *Exponentialverteilung* mit Parameter $\lambda > 0$ wir schreiben auch $P = \text{Exp}(\lambda)$.

Definition 2.6.5 – Exponentialverteilung

Die *Exponentialverteilung* ist das **Wahrscheinlichkeitsmaß** P welches über die folgende **Dichte** p mit dem Parameter $\lambda > 0$ **eindeutig** definiert wird:

$$p(x) = \begin{cases} 0, & \text{falls } x < 0 \\ \lambda e^{-\lambda x}, & \text{falls } x \geq 0 \end{cases}$$

Man schreibt $P = \text{Exp}(\lambda)$.

Die **Dichte** des Beispiels **zu Beginn des Abschnittes** ist die Dichte von $\text{Exp}(1)$.

Wir werden später eine wichtige Eigenschaften der **Exponentialverteilung** kennenlernen! Die sogenannte „Gedächtnislosigkeit“. Wir benutzen die Exponentialverteilung oft zur Modellierung von Lebensdauern und Wartezeiten, da sie hier ein gutes Modell abgibt.

2.6.4 Die Normalverteilung

Die Normalverteilung besitzt zwei Parameter $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Setze nun

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

für alle $x \in \mathbb{R}$.

Die Normalverteilung ist in **Abbildung 2.6** dargestellt und auch hier gilt $p(x) \geq 0$ und (wir werden dies nur ohne Beweis betrachten):

$$\int_{-\infty}^{\infty} p(x) \, dx = 1$$

Damit erfüllt p die Eigenschaften einer **Dichte**. Besitzt P die Dichte p , so nennt man P die *Normalverteilung* mit Parametern μ und σ^2 . Anschaulich betrachtet ist μ die Verschiebung auf der x -Achse und μ bezeichnet die „Konzentration“ der Funktion um dieses μ .

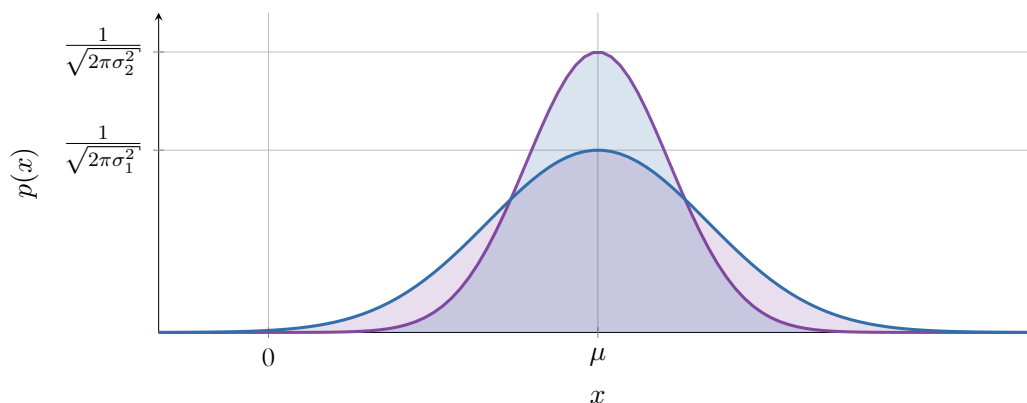


Abb. 2.6: Normalverteilung mit Mittelwert μ und Standardabweichung σ , wobei hier $\sigma_2^2 < \sigma_1^2$.

Definition 2.6.6 – Normalverteilung

Die Normalverteilung ist das **Wahrscheinlichkeitsmaß** P welches über die folgende **Dichte** p mit den Parametern μ und σ^2 **eindeutig** definiert wird:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Man schreibt $P = N(\mu, \sigma^2)$.

Bei der **Normalverteilung** gilt insbesondere für $\mu = 0$ und $\sigma^2 = 1$:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ für alle } x \in \mathbb{R}.$$

Man nennt $P = N(0, 1)$ die **Standardnormalverteilung**.

Bemerkung 2.13 – Dinge, die es bei der Normalverteilung zu beachten gilt.

1. Die Standardnormalverteilung $N(0, 1)$ spielt eine zentrale Rolle in der Stochastik (wir betrachten hierzu später die zentrale Grenzwertsätze).
2. Man verwendet die **Normalverteilung** oft zur Modellierung von Messfehlern, Blattlängen, Blutdruckwerten, Temperaturen, ...

2.7 Bedingte Wahrscheinlichkeiten

Wir haben nun die Modellierung von Wahrscheinlichkeitsräumen abgeschlossen und befassen uns nun noch mit einigen Begrifflichkeiten...

Zur Motivation betrachten wir einen zweimaligen Würfelwurf. Diesen Modellieren wir über einen **Laplace-W-Raum** mit $\Omega = \{1, \dots, 6\}^2$. Betrachten wollen wir das Ergebnis

$$A = \text{„Augensumme ist 8.“}$$

Mathematisch ausgeschrieben gilt hier:

$$A = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}.$$

Damit können wir bereits die Wahrscheinlichkeit berechnen:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{5}{6^2} = \frac{5}{36}.$$

So weit, so bekannt. Doch nun nehmen wir einmal an, wir wissen, dass beim ersten Würfeln eine 6 geworfen wird. Doch wie groß ist nun die Wahrscheinlichkeit von A ?

Mit $\Omega \rightarrow B = \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$ und $A \rightarrow A \cap B = \{(6, 2)\}$ erhalten wir die Wahrscheinlichkeit von A unter B gegeben:

$$\frac{|A \cap B|}{|B|} = \frac{1}{6} \neq \frac{5}{36}.$$

Hier gilt zu beachten, dass:

$$\frac{|A \cap B|}{|B|} = \frac{|A \cap B| / |\Omega|}{|B| / |\Omega|} = \frac{P(A \cap B)}{P(B)}$$

Diese Formel ergibt auch für allgemeine **Wahrscheinlichkeitsräume** Sinn, wie **Abbildung 2.7** veranschaulicht.

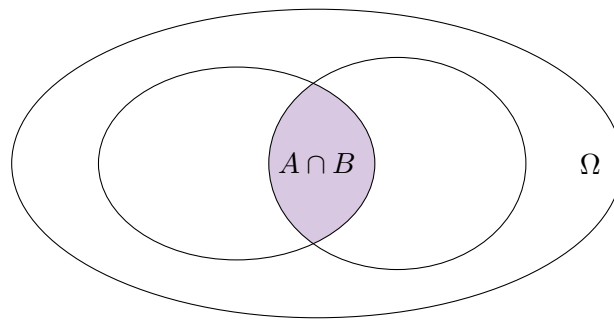


Abb. 2.7: Die Veranschaulichung der Reduktion bei mehreren Mengen.

Mit diesem Wissen definieren wir uns die bedingte Wahrscheinlichkeit...

Definition 2.7.1 – Bedingte Wahrscheinlichkeit

Sei (Ω, Σ, P) ein **W-Raum** und sei $B \in \Sigma$ mit $P(B) \geq 0$. Dann definiert sich die *bedingte Wahrscheinlichkeit von A gegeben B* als:

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

Hierzu erfassen wir uns ein weiteres Theorem

Satz 2.7.1 – Zwei Sätze bei paarweise disjunkten Mengen

Sei (Ω, Σ, P) ein **W-Raum** und seien weiter $B_1, \dots, B_n \in \Sigma$ **paarweise disjunkt** mit $P(B_i) > 0$ für $i = 1, \dots, n$ sowie $B_1 \cup \dots \cup B_n = \Omega$ (betrachte **Abbildung 2.8**), dann ergeben sich die folgenden beiden Sätze:

Satz von der totalen Wahrscheinlichkeit: Für alle $A \in \Sigma$ gilt:

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

Satz von Bayes: für alle $A \in \Sigma$ mit $P(A) > 0$ und alle $i = 1, \dots, n$ gilt:

$$P(B_i | A) = \frac{P(A | B_i) \cdot P(B_i)}{P(A)} = \frac{P(A | B_i) \cdot P(B_i)}{\sum_{j=1}^n P(A | B_j) \cdot P(B_j)}.$$

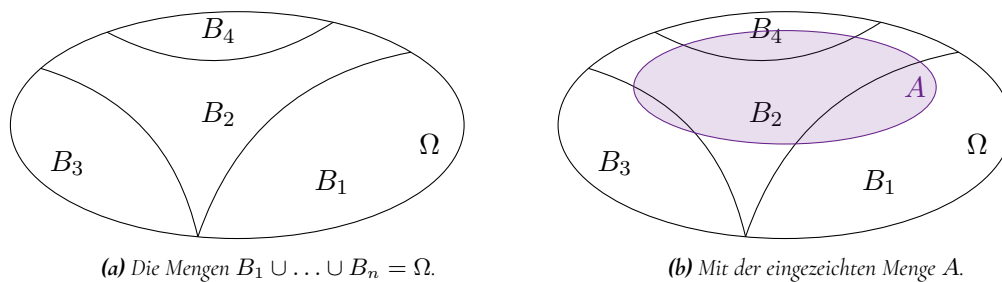


Abb. 2.8: Darstellung der Wahrscheinlichkeitsräume.

Machen wir hierzu mal wieder ein Beispiel, ...

Beispiel 2.19 – Beispiel zur bedingten Wahrscheinlichkeit

Die Maschinen M_1 , M_2 und M_3 produzieren ein Bauteil. Dabei produziert:

M_1 : produziert 500 pro Tag mit einer Ausschussquote von 5%.

M_2 : produziert 200 pro Tag mit einer Ausschussquote von 10%.

M_3 : produziert 100 pro Tag mit einer Ausschussquote von 2%.

Wir kaufen nun ein solches Bauteil (ohne zu wissen, welcher Maschine es entspringt). Hier stellen sich nun die Fragen:

1. Wie groß ist die Wahrscheinlichkeit, dass das Bauteil defekt ist?
2. Angenommen, dass das Bauteil defekt ist, wie groß ist die Wahrscheinlichkeit, dass es aus M_1 kommt?

Diesen Wahrscheinlichkeitsraum könnten wir nun modellieren. Mit den Mitteln der bedingten Wahrscheinlichkeit ist dies aber nicht nötig, da sie immer gelten, egal wie das Experiment aussieht. Wir betrachten hierzu nun also die Ereignisse, die die Menge aller Teile konstruieren:

B_1 = „Bauteil kommt aus M_1 “,

B_2 = „Bauteil kommt aus M_2 “,

B_3 = „Bauteil kommt aus M_3 “,

A = „Bauteil ist defekt.“

Mit dieser Betrachtung gilt: B_1, B_2, B_3 sind disjunkt und $B_1 \cup B_2 \cup B_3 = \Omega$. Ferner gilt (für die Wahrscheinlichkeitsraum, dass ein Teil überhaupt aus einer der gegebenen Maschinen kommt):

$$P(B_1) = \frac{500}{800} = \frac{5}{8}, \quad P(B_2) = \frac{2}{8}, \quad P(B_3) = \frac{1}{8},$$

sowie (dies sind nun die Wahrscheinlichkeiten, dass eine gewisse Maschine einen Ausschuss produziert): $P(A | B_1) = 0.05, P(A | B_2) = 0.1, P(A | B_3) = 0.02$. Nun können wir die Fragen bearbeiten:

1. Hier ist $P(A)$ gesucht. Mit dem **Satz von der totalen Wahrscheinlichkeit** gilt

$$\begin{aligned} P(A) &= \sum_{i=1}^3 P(A | B_i) \cdot P(B_i) \\ &= 0.05 \cdot \frac{5}{8} + 0.1 \cdot \frac{2}{8} + 0.02 \cdot \frac{1}{8} \\ &= \frac{47}{800} = 0.05875 (\approx 6\%) \end{aligned}$$

2. Hier suchen wir $P(B_1 | A)$. Mit dem **Satz von Bayes** gilt:

$$P(B_1 | A) = \frac{P(A | B_1) \cdot P(B_1)}{P(A)} = \frac{0.05 \cdot \frac{5}{8}}{47/800} = \frac{25}{47}$$

Es ist zu beachten, dass für $A, B \in \Sigma$ mit $P(B) > 0$ gilt, dass

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B) - P(A^c \cap B)}{P(B)} = 1 - P(A^c | B)$$

da $B = (A \cap B) \cup (A^c \cap B)$ da A und A^c **disjunkt** sind.

2.8 Unabhängigkeit

2.8.1 Die Unabhängigkeit von zwei Ereignissen

Sei (Ω, Σ, P) ein **W-Raum** und $A, B \in \Sigma$ sowie $P(A), P(B) > 0$. Um den Begriff von Unabhängigkeit zu definieren (also dass zwei Ereignisse unabhängig sind), fassen wir die folgende Idee: Die Unabhängigkeit von A und B soll bedeuten, dass Eintreten von A (beziehungsweise von B) keinen Einfluss auf die Wahrscheinlichkeit des Eintretens von B (beziehungsweise A) hat. Das heißt, dass gilt:

$$\begin{aligned} P(A | B) &\stackrel{\text{Def}}{=} \frac{P(A \cap B)}{P(B)} \stackrel{!}{=} P(A) \text{ und} \\ P(B | A) &\stackrel{\text{Def}}{=} \frac{P(A \cap B)}{P(A)} \stackrel{!}{=} P(B) \end{aligned}$$

Es muss also $P(A \cap B) = P(A) \cdot P(B)$ gelten!

Auf dieser Basis fassen wir die folgende Definition...

Definition 2.8.1 – Unabhängig

Sei (Ω, Σ, P) ein **W-Raum** und $A, B \in \Sigma$. Dann heißen A und B *unabhängig*, falls gilt:

$$P(A \cap B) = P(A) \cdot P(B).$$

Machen wir auch hierzu mal ein Beispiel... Wir betrachten den zwei-maligen Wurf einer fairen Münze und modellieren diesen als **Laplace-W-Raum** mit $\Omega = \{K, Z\}^2$. Nun setzen wir:

$$A_1 = \text{„1. Wurf ist Kopf“}$$

$$A_2 = \text{„2. Wurf ist Kopf“}$$

$$A_3 = \text{„Die Würfe sind verschieden“}$$

Bereits intuitiv können wir sagen, dass das Ergebnis des zweiten Münzwurfes nicht mit dem zweiten Münzwurf zu tun hat. Intuitiv ist aber nicht klar, ob A_1 oder A_2 einen Einfluss auf A_3 haben. Wir werden dies nun aus Formalitätsgründen überprüfen...

Hierzu arbeiten wir wieder über die mathematische Formulierung:

$$A_1 = \{(K, K), (K, Z)\},$$

$$A_2 = \{(K, K), (Z, K)\},$$

$$A_3 = \{(K, Z), (Z, K)\}$$

Somit folgt $P(A_1) = P(A_2) = P(A_3) = \frac{2}{4} = \frac{1}{2}$. Weiter gilt:

$$A_1 \cap A_2 = \{(K, K)\}$$

$$A_1 \cap A_3 = \{(K, Z)\}$$

$$A_2 \cap A_3 = \{(Z, K)\}$$

Somit erhalten wir:

$$P(A_1 \cap A_2) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_2).$$

Damit haben wir schon einmal die Intuition bestätigt, dass der erste Münzwurf nichts mit dem zweiten zu tun hat. Für den Schnitt mit A_3 erhalten wir:

$$P(A_1 \cap A_3) = \frac{1}{4} = P(A_1) \cdot P(A_3),$$

$$P(A_2 \cap A_3) = \frac{1}{4} = P(A_2) \cdot P(A_3),$$

Also sind A_1 und A_3 sowie A_2 und A_3 jeweils unabhängig voneinander. definieren wir uns nun aber noch:

$$A_4 = \text{„Die Würfe sind gleich“}$$

Dann gilt:

$$A_4 = \{(K, K), (K, K)\}$$

und somit: $P(A_4) = \frac{1}{2}$. Nun ist aber $A_3 \cap A_4 = \emptyset$ und somit $P(A_3 \cap A_4) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} = P(A_3) \cdot P(A_4)$. Damit sind A_3 und A_4 nicht unabhängig, was wir uns aber auch (in diesem simplen

Fall) intuitiv erklären können. Wenn wir wissen, dass A_3 eingetreten ist, dann kann A_4 *nicht* mehr eintreten (die Oberseiten der Münze sind ja schon verschieden) und umgekehrt.

Wir betrachten nun die **Unabhängigkeit** von mehr als zwei Ereignissen...

2.8.2 Die Unabhängigkeit von mehr als zwei Ereignissen

Um dies zu zeigen benötigen wir eine weitere Definition, ...

Definition 2.8.2 – Paarweise Unabhängig

Sei (Ω, Σ, P) ein **W-Raum** und $A_1, A_2, \dots, A_n \in \Sigma$ also **Ereignisse**. Dann heißen A_1, A_2, \dots, A_n :

paarweise unabhängig: wenn A_i und A_j für alle $i \neq j$ **unabhängig** sind. Also:

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j).$$

unabhängig: sofern für alle $T \subseteq \{1, \dots, n\}$ mit $|T| \geq 2$ gilt:

$$P\left(\bigcap_{i \in T} A_i\right) = \prod_{i \in T} P(A_i)$$

Unabhängig sind A_1, A_2, \dots, A_n also dann, wenn für die Menge $\{1, 2, 3, 4\}$ gilt, dass zum Beispiel die Teilmenge $T = \{1, 2, 4, n\}$ die folgende Forderung erfüllt (jede Teilmenge muss diese Forderung für sich erfüllen):

$$P(A_1 \cap A_2 \cap A_4 \cap A_n) = P(A_1) \cdot P(A_2) \cdot P(A_4) \cdot P(A_n).$$

Beispiel 2.20 – Zur Unabhängigkeit

Betrachten wir den **Begriff** noch einmal für n_3 . Dann bedeutet *paarweise unabhängig*, dass:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2),$$

$$P(A_1 \cap A_3) = P(A_1) \cdot P(A_3),$$

$$P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)$$

Also die Ereignisse jeweils unabhängig sind. *Unabhängig* sind sie nur dann wenn (für dieses Beispiel) *zusätzlich* noch folgende Gleichheit gelten (für die Teilmengen $|T| = 2$ haben wir schon mit dem paarweise unabhängig geprüft, damit fehlt nur noch $|T| = 3$, da $n = 3$):

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3).$$

Ist das **Münzwurfbeispiel** nun *unabhängig*? Im Beispiel zeigen wir schon, dass die Ereignisse A_1, A_2, A_3 **paarweise unabhängig** sind. Wir prüfen mit $|T| = 3$ nun noch für die Unabhängigkeit:

$$P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_2) \cdot P(A_3).$$

Damit sind A_1, A_2, A_3 *nicht* unabhängig⁽¹⁾!

⁽¹⁾ Auch hier hätten wir uns wieder die Intuition verwenden könne. Wenn die erste Münze und die zweite Münze kopf sind (A_1 und A_2), so können sie *nicht* verschieden sein (A_3).

Bemerkung 2.14 – Es ist keine Äquivalenz!

Die *Unabhängigkeit* impliziert über $|T| = 2$ die *paarweise Unabhängigkeit*, umgekehrt gilt dies, wie soeben gezeigt, nicht!

Beispiel 2.21 – Warum muss die Unabhängigkeit so kompliziert sein?

Betrachten wir einen *Laplace-W-Raum* mit $\Omega = \{1, \dots, 12\}$ und seien $A_1 = \{1, \dots, 9\}$, $A_2 = \{6, \dots, 9\}$ und $A_3 = \{9, \dots, 12\}$ (es könnte ein Urnen-Szenario sein, das spielt für die Betrachtung hier aber keine Rolle es ist einfach ein „beliebiger“ Laplace-W-Raum).

Sind A_1, A_2, A_3 *unabhängig*? Es gilt:

$$P(A_1) = \frac{9}{12} = \frac{3}{4}, \quad P(A_2) = \frac{4}{12} = \frac{1}{3}, \quad P(A_3) = \frac{4}{12} = \frac{1}{3}.$$

Für den Schnitt gilt also:

$$P(A_1 \cap A_2 \cap A_3) = P(\{9\}) = \frac{1}{12}$$

$$P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{4}$$

Allerdings sind A_1, A_2 und A_3 *nicht* unabhängig, da:

$$P(A_1 \cap A_2) = P(A_2) = \frac{1}{3}$$

allerdings gilt:

$$P(A_1) \cdot P(A_2) = \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{4} \neq \frac{1}{3}.$$

Es ist also unbedingt wichtig, den Fall für alle Teilmengen T von $\{1, \dots, n\}$ zu überprüfen.

3

ZUFALLSVARIABLEN

3.1 Zufallsvariablen definiert

Bei einem Zufallsexperiment (Ω, Σ, P) interessiert man sich oft nicht für das tatsächliche Ergebnis $\omega \in \Omega$, sondern für eine Kennzahl $\mathcal{X}(\omega)$, die von ω abhängt. Wir betrachten hier das Beispiel des n -maligen Würfels.

Hier ist der Ergebnisraum

$$\Omega = \{1, \dots, 6\}^n = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in \{1, \dots, 6\} \text{ für alle } i = 1, \dots, n\}$$

Die möglichen Kennzahlen sind hier zum Beispiel $\mathcal{X}(n) = \min_{k=1, \dots, n} \omega_k$, also was ist das Minimum oder $\mathcal{X}(n) = \max_{k=1, \dots, n} \omega_k$ also was ist das Maximum beziehungsweise $\mathcal{X}(n) = \sum_{k=1}^n \omega_k$. Damit haben wir auch schon Zufallsvariablen kennengelernt...

Definition 3.1.1 – Zufallsvariable

Sei (Ω, Σ, P) ein ω -Raum und $\mathcal{X} : \Omega \rightarrow \mathbb{R}$ eine Abbildung. \mathcal{X} heißt eine *Zufallsvariable*, falls für alle $B \in \mathcal{B}(\mathbb{R})$ (also der *Borel-Sigma-Algebra*) gilt:

$$\underbrace{\{\omega \in \Omega : \mathcal{X}(\omega) \in B\}}_{\subset \Omega} \in \Sigma$$

Hierzu gibt es ein paar Bemerkungen...

Bemerkung 3.1 – Was Zufallsvariablen wirklich sind

1. *Zufallsvariablen* sind Abbildungen.
2. Die Aussage $\{\omega \in \Omega : \mathcal{X}(\omega) \in B\} \in \Sigma$ für alle $B \in \mathcal{B}(\mathbb{R})$ muss in dieser Vorlesung *nicht* überprüft werden – auch wenn es natürlich Fälle gibt, in denen die Bedingung nicht erfüllt ist. Wir können hier davon ausgehen, dass alle uns präsentierten Abbildungen, die in der Vorlesung auftauchen, (gültige) Zufallsvariablen sind.
3. Wir schreiben oft auch $\{\mathcal{X} \in B\}$ vereinfachtend/verkürzend für das Ereignis $\{\omega \in \Omega \mid \mathcal{X}(\omega) \in B\}$.

Betrachten wir hierzu auch ein Beispiel!

Beispiel 3.1 – Der bestellt ja schon wieder, ...

Wir betrachten wieder die Anzahl an Bestellungen eines super duper Artikels an 100 Tagen. Wir vernachlässigen für diesen Fall die Wahrscheinlichkeiten und betrachten nur den *Ergebnisraum*:

$$\Omega = \{\omega = (\omega_1, \dots, \omega_{100}) \mid \omega_i \in \mathbb{N}_0 \text{ für alle } i = 1, \dots, 100\} = \mathbb{N}_0^{100}.$$

Wir führen nun ein paar Zufallsvariablen ein, wie zum Beispiel die „Anzahl der Bestellungen an Tag i “:

$$\mathcal{X}_i : \Omega \rightarrow \mathbb{R} \text{ mit } \mathcal{X}_i(\omega) = \omega_i.$$

Die Gesamtanzahl der Bestellungen ergibt sich als:

$$\mathcal{X} : \Omega \rightarrow \mathbb{R} \text{ mit } \mathcal{X}(\omega) = \sum_{i=1}^{100} \omega_i.$$

Natürlich lassen sich noch weitaus mehr Zufallsvariablen konstruieren. Wir interessieren uns in diesem Fall aber nur für das Ereignis: „Die Gesamtanzahl der Bestellungen liegt zwischen 1000 und 2000“:

$$A = \{\omega \in \Omega : 1000 \leq \mathcal{X}(\omega) \leq 2000\} = \{1000 \leq \mathcal{X} \leq 2000\}.$$

Bei letzteren Komponente haben wir nun unsere Kurzschreibweise abgewandt (einfach um uns Schreibarbeit zu sparen).

Die **Zufallsvariablen** zuvor haben wir nun nicht umsonst definiert. Wir sind in der Lage neue Zufallsvariablen durch die Verknüpfung von gegebenen Zufallsvariablen erzeugen. Hier zum Beispiel:

$$\mathcal{X} = \sum_{i=1}^{100} \mathcal{X}_i.$$

Doch mit welchen Wahrscheinlichkeiten haben die Zufallsvariablen nun welchen Wert? Dies werden wir im folgenden erfassen, ...

Definition 3.1.2 – Die Verteilung einer Zufallsvariable

Sei (Ω, Σ, P) ein **Wahrscheinlichkeitsraum**. Sei weiter $\mathcal{X} : \Omega \rightarrow \mathbb{R}$ eine **Zufallsvariable**. Dann heißt das **Wahrscheinlichkeitsmaß**

$$P_{\mathcal{X}} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$$

mit

$$\begin{aligned} P_{\mathcal{X}}(B) &= P(\{\mathcal{X} \in B\}) \\ &= P(\{\omega \in \Omega : \mathcal{X}(\omega) \in B\}), \quad B \in \mathcal{B}(\mathbb{R}) \end{aligned}$$

die Verteilung von \mathcal{X} .

Hierzu mal wieder ein Beispiel!

Beispiel 3.2 – Die Verteilung beim zweimaligen Würfeln

Wir modellieren den **Laplace-Wahrscheinlichkeitsraum** des zweimaligen Würfeln mit $\Omega = \{1, \dots, 6\}^2 = \{\omega = (\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, \dots, 6\}\}$. Sei $\mathcal{X} : \Omega \rightarrow \mathbb{R}$ nun mit $\mathcal{X}(\omega) = \omega_1 + \omega_2$ als **Zufallsvariable** zu betrachten. Dann gilt:

$$\mathcal{X}(\omega) \in \{2, 3, \dots, 12\} \text{ für alle } \omega \in \Omega.$$

Weiter gilt für die Wahrscheinlichkeit folgendes:

$$P_{\Omega}(\{2\}) = P(\{\mathcal{X} \in \{2\}\})$$

$$\begin{aligned}
 &= P(\{\omega \in \Omega \mid \mathcal{X}(\omega) = 2\}) \\
 &= P(\{\omega \in \Omega \mid \omega_1 + \omega_2 = 2\}) \\
 &= P(\{(1, 1)\}) = \frac{|\{(1, 1)\}|}{|\Omega|} = \frac{1}{36}.
 \end{aligned}$$

Analog machen wir dies nun noch für 3, wir merken wie wir unser Wahrscheinlichkeitsmaß auf das des ursprünglichen Wahrscheinlichkeitsraums herunterbrechen können:

$$\begin{aligned}
 P_{\mathcal{X}}(\{3\}) &= P(\{\mathcal{X} \in \{3\}\}) \\
 &= P(\{(1, 2), (2, 1)\}) = \frac{2}{36}.
 \end{aligned}$$

Insgesamt erhalten wir die folgenden Verteilung:

\mathcal{X}	2	3	4	5	6	7	8	9	10	11	12
$P_{\mathcal{X}}(\{x\})$	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

Doch wie berechnen wir dies nun für alle $B \in \mathcal{B}(\mathbb{R})$ also alle B in der **Borel- σ -Algebra**? Wir berechnen dies wie folgt (eine tolle lange Rechnung):

$$\begin{aligned}
 P_{\mathcal{X}}(B) &= P(\{\mathcal{X} \in B\}) = P(\{\mathcal{X} \in B\} \cap \Omega) \\
 &= P(\{\mathcal{X} \in B\} \cap \{\mathcal{X} \in \{2, 3, \dots, 12\}\}) \\
 &= P(\{\omega \in \Omega : \mathcal{X}(\omega) \in B \text{ und } \mathcal{X}(\omega) \in \{2, 3, \dots, 12\}\}) \\
 &= P(\{\omega \in \Omega \mid \mathcal{X}(\omega) \in B \cap \{2, 3, \dots, 12\}\}) \\
 &= P(\{\omega \in B \cap \{2, 3, \dots, 12\}\}) \\
 &= P\left(\bigcup_{x \in \{2, \dots, 12\}} \{\mathcal{X} \in B \cap \{x\}\}\right) \\
 &= \sum_{x \in \{2, \dots, 12\}} P(\{x \in B \cap \{x\}\}) = \sum_{x \in B \cap \{2, \dots, 12\}} P(\{\mathcal{X} = x\}) \\
 &= \sum_{x \in B \cap \{2, \dots, 12\}} P_{\mathcal{X}}(\{x\})
 \end{aligned}$$

Die Umwandlung in die Summe hat übrigens funktioniert, da die Mengen **paarweise disjunkt** sind! So berechnet sich beispielsweise $P_{\mathcal{X}}((3, 6])$ wie folgt:

$$\begin{aligned}
 P_{\mathcal{X}}((3, 6]) &= \sum_{\substack{x \in (3, 6] \cap \{2, \dots, 12\} \\ = \{4, 5, 6\}}} = P_{\mathcal{X}}(\{4\}) + P_{\mathcal{X}}(\{5\}) + P_{\mathcal{X}}(\{6\}) \\
 &= 3/36 + 4/36 + 5/36 = 12/36 = 1/3
 \end{aligned}$$

Mit

$$P_{\mathcal{X}}(B) = \sum_{x \in B \cap \{2, \dots, 12\}} P_{\mathcal{X}}(\{x\})$$

haben wir also unser Wahrscheinlichkeitsmaß definiert ☺.

Es gilt zu beachten, dass wir mit (Ω, Σ, P) angefangen haben und durch die Definition von \mathcal{X} einen neuen **Wahrscheinlichkeitsraum** erzeugt haben: $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{\mathcal{X}})$ wobei $\mathcal{X}: \Omega \rightarrow \mathbb{R}$. Diese Eigenschaft werden wir uns später noch einmal zunutze machen.

Nun, da wir die Grundlagen der Zufallsvariablen erfasst haben, lernen wir sie voneinander abzugrenzen...

Definition 3.1.3 – Diskrete Zufallsvariable und Träger-Menge

Eine **Zufallsvariable** \mathcal{X} nennen wir *diskret* falls es ein $D \subseteq \mathbb{R}$ gibt, welches *abzählbar* ist und die Eigenschaft $P(\{\mathcal{X} \in D\}) = 1$ erfüllt.

Ist \mathcal{X} diskret, so nennen wir die Menge

$$D_{\mathcal{X}} = \{x \in \mathbb{R} : P(\{\mathcal{X} = x\}) > 0\}$$

den *Träger* von \mathcal{X} . Auf Basis der Forderung ergibt sich auch $P(\{\mathcal{X} \in D_{\mathcal{X}}\}) = 1$.

Um uns Schreibarbeit zu sparen, führen wir eine Reihe an Sprech- und Schreibweisen ein. Sei nun \mathcal{X} eine **Zufallsvariable**. Wir bezeichnen \mathcal{X} als...

Bernoulliverteilt mit dem Parameter $p \in [0, 1]$ falls gilt⁽¹⁾:

$$P(\mathcal{X} = 1) = p, P(\mathcal{X} = 0) = 1 - p$$

In diesem Fall gilt dann auch

$$P(\{\mathcal{X} \in \underbrace{\{0, 1\}}_{=D}\}) = p + 1 - p = 1.$$

Weiter gilt:

$$D_{\mathcal{X}} = \begin{cases} \{0, 1\}, & \text{falls } p \in (0, 1) \\ \{0\}, & \text{falls } p = 0 \\ \{1\}, & \text{falls } p = 1 \end{cases}$$

Wir schreiben hierfür kurz: $\mathcal{X} \sim B(1, p)$.

Binomialverteilt mit Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$ falls⁽²⁾

$$P(\{\mathcal{X} = k\}) = \binom{n}{k} \cdot p^k (1 - p)^{n-k} \text{ für alle } k \in \{0, \dots, n\}.$$

Wir schreiben hier kurz: $\mathcal{X} \sim B(n, p)$.

Gleichverteilt auf G für $G \subset \mathbb{R}$ mit $G \neq \emptyset$ und G ist endlich, falls

$$P(\{\mathcal{X} = x\}) = \frac{1}{|G|} \text{ für alle } x \in G.$$

Hier schreiben wir kurz: $\mathcal{X} \sim U(G)$.

⁽¹⁾ Die Definition ähnelt der der **Bernoulliverteilung**

⁽²⁾ Diese Definition ähnelt der der **Binomialverteilung**

Hypergeometrisch verteilt mit den Parametern $n \in \mathbb{N}$, $B \in \{0, \dots, n\}$, $k \in 1, \dots, n$, falls:

$$P(\{\mathcal{X} = b\}) = \frac{\binom{B}{b} \binom{n-B}{k-b}}{\binom{n}{k}} \text{ für alle } \max(0, k - (n - B)) \leq b \leq \min(B, k).$$

Hier schreiben wir kurz: $X \sim H(n, B, k)$.

Poissonverteilt mit den Parametern $\lambda > 0$, falls

$$P(\{\mathcal{X} = k\}) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ für alle } k \in \mathbb{N}_0.$$

Die Kurzschreibweise lautet hier: $\mathcal{X} \sim P(\lambda)$.

Geometrisch verteilt mit Parameter $p \in (0, 1]$ falls

$$P(\{\mathcal{X} = k\}) = p(1 - p)^{k-1} \text{ für alle } k \in \mathbb{N}$$

die Kurzschreibweise hier ist: $X \sim G(p)$.

Dies sind die wichtigsten Zufallsvariablen, die übrigens auch alle **diskret** sind! Wir definieren nun die **absolutstetigen** Zufallsvariablen...

Definition 3.1.4 – Absolut stetig verteilte Zufallsvariable

Eine **Zufallsvariable** \mathcal{X} heißt **absolutstetig verteilt**, falls $P_{\mathcal{X}}$ **absolutstetig** ist, das heißt, dass $P_{\mathcal{X}}$ eine **Dichte** besitzt.

Auch hierfür führen wir nun wieder eine ganze Reihe an Sprech- und Schreibweisen ein. Sei jeweils \mathcal{X} eine **Zufallsvariable**, dann nennen wir $\mathcal{X} \dots$

Gleichverteilt auf $[a, b]$ mit $a < b$, falls $P_{\mathcal{X}} = U([a, b])$. Wir schreiben auch $\mathcal{X} \sim U([a, b])$.

Exponentialverteilt mit Parameter $\lambda > 0$, falls $P_{\mathcal{X}} = \text{Exp}(\lambda)$. Hier schreiben wir auch: $\mathcal{X} \sim \text{Exp}(\lambda)$.

Normalverteilt mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ falls $P_{\mathcal{X}} = N(\mu, \sigma^2)$ also der **Normalverteilung** (mit den jeweiligen Parametern) entspricht. Wir schreiben hier auch $\mathcal{X} \sim N(\mu, \sigma^2)$.

In dieser Vorlesung betrachten wir übrigens nur **diskrete** sowie **absolut stetige Zufallsvariablen** betrachten, auch wenn es durchaus noch Mischungen gibt!

3.2 Die Verteilungsfunktion

Die Verteilung von \mathcal{X} ist eine Abbildung $P_{\mathcal{X}} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$. Sie ist eindeutig bestimmt durch die Werte:

$$P_{\mathcal{X}}(\underbrace{(-\infty, t]}_{\in \mathcal{B}(\mathbb{R})}) = P(\{\mathcal{X} \leq t\}), \quad t \in \mathbb{R}$$

Damit können wir auch diese Abbildung als eine Verteilung beschreiben! Dies motiviert die folgende Definition...

Definition 3.2.1 – Verteilungsfunktion

Sei \mathcal{X} eine **Zufallsvariable** dann heißt die Funktion $F_{\mathcal{X}} : \mathbb{R} \rightarrow [0, 1]$ mit

$$F_{\mathcal{X}}(t) = P(\{\mathcal{X} \leq t\}), \quad t \in \mathbb{R}$$

die **Verteilungsfunktion** von \mathcal{X} .

Wir konstruieren uns also einfach eine Funktion, die von dem gegebenen t abhängt: $F_{\mathcal{X}}(t)$. Interessant ist nun die Betrachtung, welche Gestalt $F_{\mathcal{X}}$ eigentlich besitzt. Wir betrachten zuerst den Fall, dass \mathcal{X} **diskret** ist. Dann gilt für alle $t \in \mathbb{R}$ folgendes:

$$\begin{aligned} F_{\mathcal{X}}(t) &= P(\{\mathcal{X} \leq t\}) = P(\underbrace{\{x \leq t\} \cap \{x \in D_{\mathcal{X}}\}}_{\text{W-keit:1}}) \\ &= P(\{x \in P_{\mathcal{X}} \cap (-\infty, t]\}) \\ &= \sum_{x \in D_{\mathcal{X}}, x \leq t} P(\{\mathcal{X} = x\}). \end{aligned}$$

Machen wir hierzu erstmal zwei Beispiele, bevor wir uns dem zweiten Fall widmen.

Beispiel 3.3 – Verteilungsfunktion bei einer diskreten Zufallsvariable

1. Sei $X \sim U(\{1, 2, 3\})$. Dann gilt

$$P(\{\mathcal{X} = 1\}) = P(\{\mathcal{X} = 2\}) = P(\{\mathcal{X} = 3\}) = \frac{1}{3}.$$

Also ist der **Träger** auch $D_{\mathcal{X}} = \{1, 2, 3\}$ und

$$F_{\mathcal{X}}(t) = \sum_{x \in D_{\mathcal{X}} : x \leq t} 1/3 = \frac{1}{3} \cdot \sum_{x \in D_{\mathcal{X}} : x \leq t} 1$$

Dies berechnet sich nun aber zu:

$$F_{\mathcal{X}}(t) = \frac{1}{3} \cdot |\{1, 2, 3\} \cap (-\infty, t]| = \begin{cases} 0, & \text{falls } t < 1 \\ 1/3, & \text{falls } t \in [1, 2) \\ 2/3, & \text{falls } t \in [2, 3) \\ 1, & \text{falls } t \geq 3 \end{cases}$$

Der Graph der Funktion befindet sich in **Abbildung 3.1a**. Zu bemerken ist, dass der Abstand hier drei mal $\frac{1}{3}$ ist.

2. Sei nun $\mathcal{X} \sim P(\lambda)$ mit $\lambda > 0$. Dann gilt $P(\{\mathcal{X} = k\}) = \frac{\lambda^k}{k!} e^{-\lambda}$ für alle $k \in \mathbb{N}_0$. Damit ist der **Träger** $D_{\mathcal{X}} = \mathbb{N}_0$. Für die Verteilung gilt

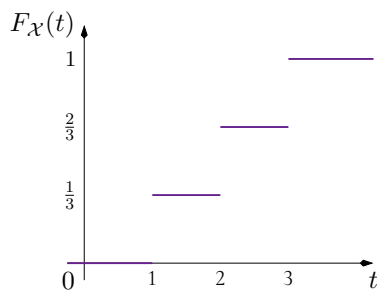
$$F_{\mathcal{X}}(t) = \sum_{\mathcal{X} \in \mathbb{N}_0 : x \leq t} \frac{\lambda^k}{k!} e^{-\lambda}.$$

Der Graph der Verteilung findet sich in **Abbildung 3.1b** – er nähert sich asymptotisch der 1 an der Abstand des Sprungs ergibt sich aus der Verteilung.

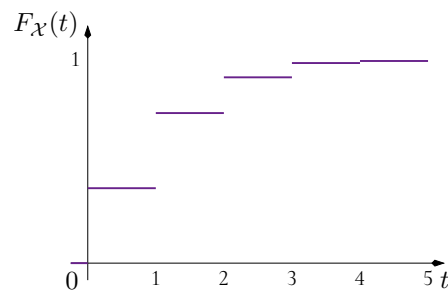
Betrachten wir nun den zweiten Fall. Hier soll \mathcal{X} **absolutstetig** sein. In diesem Fall gilt für alle $t \in \mathbb{R}$, dass

$$F_{\mathcal{X}}(t) = \int_{-\infty}^t p(x) dx,$$

wobei p die **Dichte** von $P_{\mathcal{X}}$ ist. Auch hierzu möchten wir uns ein paar Beispiele genehmigen...



(a) Die Funktion $F_X(t)$ für $X \sim U(\{1, 2, 3\})$.



(b) Die Funktion $F_X(t)$ für $X \sim P(\lambda)$. Die Abstände betragen $e^{-\lambda}, \lambda e^{-\lambda}, \frac{\lambda^2 e^{-\lambda}}{2}$.

Abb. 3.1: Ein paar Verteilungen zur diskreten Zufallsvariable.

Beispiel 3.4 – Die Verteilungen bei absolutstetigen Zufallsvariablen

1. Sei $X \sim U([a, b])$ mit $a < b$. Dann gilt

$$p(x) = \begin{cases} \frac{1}{b-a}, & \text{falls } x \in [a, b] \\ 0, & \text{sonst.} \end{cases}$$

Den Graphen der Dichte kennen wir bereits, betrachte hierzu [Abbildung 2.4](#) im Rahmen der stetigen Gleichverteilung (dort befinden sich noch die hier nicht relevanten Marker α und β). Damit gilt für $F_X(t)$, welches ja (nach der Definition) über das Integral entsteht, folgendes:

$$F_X(t) = \int_{-\infty}^t p(x) dx = \begin{cases} 0, & \text{falls } t < a \\ \int_a^t \frac{1}{b-a} dx = \frac{t-a}{b-a}, & \text{falls } t \in [a, b] \\ \int_a^b \frac{1}{b-a} dx = 1 & \text{falls } t > b \end{cases}$$

Der Graph von $F_X(t)$ findet sich in [Abbildung 3.2a](#).

2. Sei X nun Exponentialverteilt als $X \sim \text{Exp}(\lambda)$. Dann gilt

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

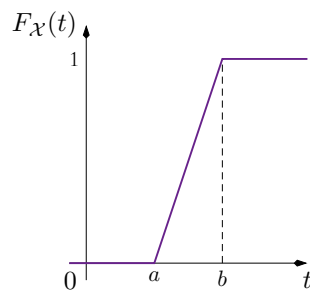
Diese kennen wir bereits von der Exponentialverteilung ([Abbildung 2.5](#)). Für das Integral gilt nun:

$$F_X(t) = \int_{-\infty}^t p(x) dx = \begin{cases} 0, & \text{falls } t < 0 \\ \int_0^t \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_0^t = 1 - e^{-\lambda t}, & t \geq 0 \end{cases}$$

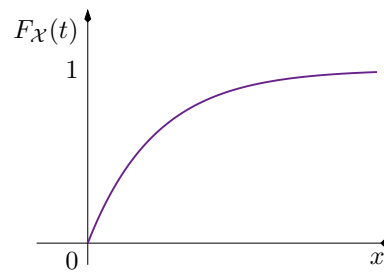
Der Graph findet sich in [Abbildung 3.2b](#), er nähert sich asymptotisch der 1 an. Die Funktion steigt, wie sicherlich auffällt und wie auch all die Anderen, monoton an und konvergiert gegen 1.

Satz 3.2.1 – Die Verteilungsfunktion einer Zufallsvariablen

Die **Verteilungsfunktion** F_X einer (beliebigen) Zufallsvariablen X besitzt die folgenden Eigenschaften:



(a) Die Verteilung für eine stetige Gleichverteilung Zufallsvariable.



(b) Die Verteilung für eine exponentialverteilte Zufallsvariable.

Abb. 3.2: Ein paar Verteilungen zur absolutstetigen Zufallsvariable.

1. F_X ist monoton wachsend, das heißt für alle $t_1 \leq t_2$ gilt: $F_X(t_1) \leq F_X(t_2)$.
2. Für die Grenzwerte gilt: $\lim_{t \rightarrow -\infty} F_X(t) = 0$, $\lim_{t \rightarrow +\infty} F_X(t) = 1$.
3. F_X ist rechtsseitig stetig, das heißt: $\lim_{t_n \downarrow t} F_X(t_n) = F_X(t)$.
4. Für jedes $t \in \mathbb{R}$ existiert der linksseitige Grenzwert $F_X(t^-) := \lim_{t_n \uparrow t} F_X(t_n)$ und es gilt:

$$P(\{X = t\}) = F_X(t) - F_X(t^-).$$

Wir werden ab jetzt t^- im Kontext von Grenzwerten schreiben, wenn es sich um einen Grenzwert „von links“ beziehungsweise „von unten“ handelt.

Wir fassen weitere Eigenschaften in der folgenden Bemerkung...

Bemerkung 3.2 – Wie kann man die Verteilung für Wahrscheinlichkeiten verwenden?

Für alle $a < b$ gilt:

$$\begin{aligned} P(\{X \in (a, b]\}) &= P(\{X \leq b\} \setminus \{X \leq a\}) \\ &= P(\{X \leq b\}) - P(\{X \leq a\}) \\ &= F_X(b) - F_X(a) \end{aligned}$$

sowie

$$\begin{aligned} P(\{X \in [a, b]\}) &= P(\{X = a\}) + P(\{X \in (a, b]\}) \\ &= F_X(a) - F_X(a^-) + F_X(b) - F_X(a) \\ &= F_X(b) - F_X(a^-) \end{aligned}$$

Wir können die Wahrscheinlichkeiten also durch Differenzen der Verteilungsfunktionen ausdrücken. Hier bezeichnet a^- übrigens wieder den „linksseitigen“ Grenzwert.

Mit diesem Wissen können wir uns nun der identischen Verteilung widmen...

Definition 3.2.2 – Identische Verteilung

Seien $(\Omega_1, \Sigma_1, P_1)$ und $(\Omega_2, \Sigma_2, P_2)$ zwei **Wahrscheinlichkeitsräume** und seien $X_1 : \Omega_1 \rightarrow \mathbb{R}$ und $X_2 : \Omega_2 \rightarrow \mathbb{R}$ **Zufallsvariablen**. Dann heißen X_1 und X_2 *identisch verteilt*, falls

$$F_{X_1} = F_{X_2}$$

das heißt: $\forall t \in \mathbb{R} : F_{\mathcal{X}_1}(t) = F_{\mathcal{X}_2}(t)$ für $F_{\mathcal{X}} : \mathbb{R} \rightarrow [0, 1]$. Die Wahrscheinlichkeitsräume können hier komplett verschieden sein!

Machen wir hierzu ein Beispiel, bei dem wir uns zwei solcher Wahrscheinlichkeitsräume modellieren...

Beispiel 3.5 – Identische Verteilung

- Wir betrachten das einmalige Würfeln, modelliert als **Laplace-Wahrscheinlichkeitsraum** mit $\Omega_1 = \{1, \dots, 6\}$. Weiter sei die Zufallsvariable $\mathcal{X}_1 : \Omega_1 \rightarrow \mathbb{R}$ gegeben durch:

$$\mathcal{X}_1(\omega) = 1_{\{2,4,6\}}(\omega) = \begin{cases} 1, & \text{falls } \omega \in \{2, 4, 6\} \\ 0, & \text{sonst.} \end{cases}$$

wobei für alle $A \subset \Omega_1$ die *Indikatorfunktion* von A wie folgt definiert ist:

$$1_A(\omega) = \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{sonst} \end{cases}.$$

- Wir betrachten den zweimaligen Münzwurf wieder als **Laplace-Wahrscheinlichkeitsraum** mit $\Omega_2 = \{0, 1\}^2$. Wir betrachten die Zufallsvariable $\mathcal{X}_2 : \Omega_2 \rightarrow \mathbb{R}$ welche wie folgt gegeben ist:

$$\mathcal{X}_2(\omega) = \mathcal{X}_2((\omega_1, \omega_2)) = |\omega_1 - \omega_2| = \begin{cases} 0, & \text{falls } \omega_1 = \omega_2 \\ 1, & \text{falls } \omega_1 \neq \omega_2 \end{cases}$$

Wir zeigen nun, dass \mathcal{X}_1 und \mathcal{X}_2 *identisch* verteilt sind. Hierzu berechnen wir zuerst für P_1 :

$$\begin{aligned} P_1(\{\mathcal{X}_1 = 1\}) &= P(\{\omega \in \Omega_1 : \mathcal{X}_1(\omega) = 1\}) \\ &= P(\{2, 4, 6\}) = \frac{3}{6} = \frac{1}{2}, \\ P_1(\{\mathcal{X}_1 = 0\}) &= P(\{1, 3, 5\}) = \frac{1}{2}. \end{aligned}$$

Es folgt:

$$P(\{\mathcal{X}_1 \in \{0, 1\}\}) = \frac{1}{2} + \frac{1}{2} = 1$$

da D endlich ist. Also ist \mathcal{X}_1 **diskret** und $P_{\mathcal{X}_1} = \{0, 1\}$. Berechnen wir nun die gleichen Operationen für P_2 :

$$\begin{aligned} P_2(\{\mathcal{X}_2 = 1\}) &= P_2(\{\omega \in \Omega_2 : \mathcal{X}_2(\omega) = 1\}) \\ &= P_2(\{(0, 1), (1, 0)\}) = \frac{2}{4} = \frac{1}{2}, \\ P_2(\{\mathcal{X}_2 = 0\}) &= P_2(\{(0, 0), (1, 1)\}) = \frac{1}{2}. \end{aligned}$$

Damit ist auch \mathcal{X}_2 **diskret** und somit $D_{\mathcal{X}_2} = \{0, 1\}$.

Damit gilt für alle $t \in \mathbb{R}$:

$$F_{\mathcal{X}_1}(t) = \sum_{x \in \{0, 1\} : x \leq t} P(\{\mathcal{X}_1 = x\}) = \sum_{x \in \{0, 1\} : x \leq t} \frac{1}{2}$$

$$= \begin{cases} 0, & \text{falls } t < 0 \\ \frac{1}{2}, & \text{falls } t \in [0, 1) \\ 1, & \text{falls } t \geq 1 \end{cases}$$

$$= F_{\mathcal{X}_2}(t)$$

Damit haben wir gezeigt: $F_{\mathcal{X}_1}(t) = F_{\mathcal{X}_2}(t)$.

3.3 Zufallsvektoren

Häufig sind auf einem **Wahrscheinlichkeitsraum** mehrere Zufallsvariablen (gleichzeitig) definiert, wobei wir diese auch *gemeinsam* betrachten und analysieren möchten. Wir erhöhen nun also die Dimensionen unserer Betrachtungen...

Definition 3.3.1 – Zufallsvektor

Sei (Ω, Σ, P) ein **Wahrscheinlichkeitsraum** und seien $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen. Die Abbildung $\mathcal{X} : \Omega \rightarrow \mathbb{R}^n$ mit

$$\mathcal{X}(\omega) = \{\mathcal{X}_1(\omega), \mathcal{X}_2(\omega), \dots, \mathcal{X}_n(\omega)\}$$

heißt **Zufallsvektor** (mit Komponenten $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n$).

Betrachten wir erstmal ein Beispiel:

Beispiel 3.6 – Drei-maliges Würfeln

Wir modellieren das dreimalige Würfeln durch einen **Laplace-Wahrscheinlichkeitsraum** mit $\Omega = \{1, \dots, 6\}^3$. Seien $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen mit:

$$\mathcal{X}_1(\omega) = \mathcal{X}_1((\omega_1, \omega_2, \omega_3)) = \max\{\omega_1, \omega_2, \omega_3\},$$

$$\mathcal{X}_2(\omega) = \mathcal{X}_2((\omega_1, \omega_2, \omega_3)) = \min\{\omega_1, \omega_2, \omega_3\},$$

$$\mathcal{X}_3(\omega) = \mathcal{X}_3((\omega_1, \omega_2, \omega_3)) = \omega_1 + \omega_2 + \omega_3.$$

Damit ist $\mathcal{X} : \Omega \rightarrow \mathbb{R}^3$ mit

$$\mathcal{X}(\omega) = (\mathcal{X}_1(\omega), \mathcal{X}_2(\omega), \mathcal{X}_3(\omega))$$

ein **Zufallsvektor**. Es gilt zum Beispiel:

$$\mathcal{X}(\underbrace{(1, 2, 3)}_{\omega}) = (3, 1, 6),$$

$$\mathcal{X}(\underbrace{(4, 2, 3)}_{\omega}) = (4, 2, 9).$$

Mit dem **Zufallsvektor** können wir die gemeinsame Verteilung definieren...

Definition 3.3.2 – Gemeinsame Verteilung

Sei $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_n)$ ein **Zufallsvektor**. Für $B = B_1 \times B_2 \times \dots \times B_n$ mit $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$

(siehe **Borel- σ -Algebra**) setzen wir $P_{\mathcal{X}}$ wie folgt:

$$P_{\mathcal{X}}(B) = P(\{\mathcal{X}_1 \in B_1, \mathcal{X}_2 \in B_2, \dots, \mathcal{X}_n \in B_n\}).$$

Dann heit $P_{\mathcal{X}}$ die *gemeinsame Verteilung* von $\mathcal{X}_1, \dots, \mathcal{X}_n$.

Betrachten wir ein Beispiel hierzu...

Beispiel 3.7

Betrachten wir einmal $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$ aus dem **Beispiel von eben**. Sei nun $B = \{5\} \times \{3\} \times \{12\}$. Dann gilt

$$\begin{aligned} P_{\mathcal{X}}(B) &= P(\{\mathcal{X}_1 = 5, \mathcal{X}_2 = 3, \mathcal{X}_3 = 12\}) \\ &= P(\{\omega = (\omega_1, \omega_2, \omega_3) \in \Omega : \max\{\omega_1, \omega_2, \omega_3\} = 5, \\ &\quad \min\{\omega_1, \omega_2, \omega_3\} = 3, \omega_1 + \omega_2 + \omega_3 = 12\}) \\ &= P(\{(5, 3, 4), (3, 5, 4), (5, 4, 3), (3, 4, 5), (4, 5, 3), (4, 3, 5)\}) \\ &= \frac{6}{6^3} = \frac{1}{36}. \end{aligned}$$

Wir haben hier einfach nur alle Kombinationen in der Menge aufgelistet, die die angegebenen Eigenschaften erfllen (also ein Maximum von 5, ein Minimum von 3 und eine Summe von 4).

Definition 3.3.3 – Die Randverteilung

Sei $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_n)$ ein **Zufallsvektor**. Dann heien $P_{\mathcal{X}_1}, \dots, P_{\mathcal{X}_n}$ die *Randverteilungen* von \mathcal{X} (wir betrachten die Komponenten also individuell).

Satz 3.3.1 – Gemeinsame- und Randverteilung

Die *gemeinsame Verteilung* $P_{\mathcal{X}}$ legt die *Randverteilung* $P_{\mathcal{X}_1}, \dots, P_{\mathcal{X}_n}$ fest. Es gilt:

$$P_{\mathcal{X}_i}(A) = P_{\mathcal{X}}(\mathbb{R} \times \dots \times \mathbb{R} \times A \times \mathbb{R} \times \dots \times \mathbb{R})$$

wobei A die i -te Komponente bezeichnet. Diese Definition hlt fr alle $i = 1, \dots, n$ und $A \in \mathcal{B}(\mathbb{R})$.

Wir werden nun die **Randverteilungen** „konkret“ im Spezialfall $n = 2$ betrachten.

3.3.1 Der diskrete Fall

Sei $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$ ein **Zufallsvektor** und seien \mathcal{X}_1 und \mathcal{X}_2 **diskret**. Dann gilt

$$\begin{aligned} P_{\mathcal{X}}(D_{\mathcal{X}_1} \times D_{\mathcal{X}_2}) &= P(\{\mathcal{X} \in D_{\mathcal{X}_1} \times D_{\mathcal{X}_2}\}) \\ &= P(\{\mathcal{X}_1 \in D_{\mathcal{X}_1}\} \cap \{\mathcal{X}_2 \in D_{\mathcal{X}_2}\}) = 1. \end{aligned}$$

Dies lsst sich leicht zeigen, da die Wahrscheinlichkeit von $\{\mathcal{X}_1 \in D_{\mathcal{X}_1}\}$ sowie $\{\mathcal{X}_2 \in D_{\mathcal{X}_2}\}$ jeweils 1 sind. Sei nun $P_{\mathcal{X}}(\{(x, y)\})$ fr $x \in D_{\mathcal{X}_1}$ und $y \in D_{\mathcal{X}_2}$ gegeben. Wir bestimmen nun:

$$\begin{aligned} P_{\mathcal{X}_1}(\{x\}) &= P(\{\mathcal{X}_1 = x\}), \quad x \in D_{\mathcal{X}_1} \\ P_{\mathcal{X}_2}(\{y\}) &= P(\{\mathcal{X}_2 = y\}), \quad y \in D_{\mathcal{X}_2} \end{aligned}$$

Für alle $x \in D_{x_1}$ (also dem Träger von x_1) gilt weiter:

$$\begin{aligned} P_{\mathcal{X}_1}(\{x\}) &= P(\{\mathcal{X}_1 = x\}) \\ &= P(\{\mathcal{X}_1 = x\} \cap \overbrace{\{\mathcal{X}_2 \in D_{\mathcal{X}_2}\}}^{\text{W-keit} = 1}) \\ &= \sum_{y \in D_{\mathcal{X}_2}} P(\{\mathcal{X}_1 = x\} \cap \{\mathcal{X}_2 = y\}) \\ &= \sum_{y \in D_{\mathcal{X}_2}} P_{\mathcal{X}}(\{(x, y)\}). \end{aligned}$$

Analog gilt für alle $y \in D_{\mathcal{X}_2}$:

$$P_{\mathcal{X}_2}(\{y\}) = \sum_{x \in D_{\mathcal{X}_1}} P_{\mathcal{X}}(\{(x, y)\}).$$

Machen wir auch hierzu mal ein Beispiel...

Beispiel 3.8

Die gemeinsame Verteilung $P_{\mathcal{X}}$ die von \mathcal{X}_1 und \mathcal{X}_2 sei durch die Tabelle 3.1 mit den Einträgen $(\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2))$

$$P_{\mathcal{X}}(\{(x, y)\}) = P(\{\mathcal{X} = (x, y)\}).$$

In der Tabelle 3.1 bestimmen wir die Randverteilungen $P_{\mathcal{X}_1}$ und $P_{\mathcal{X}_2}$. Es gilt (hier sind die Rechenwege für die in der Tabelle aufgeführten Werte angegeben):

$$\begin{aligned} P(\{\mathcal{X}_1 = 1\}) &= P(\{\mathcal{X}_1 = 1, \mathcal{X}_2 = 1\}) + P(\{\mathcal{X}_1 = 1, \mathcal{X}_2 = 2\}) + P(\{\mathcal{X}_1 = 1, \mathcal{X}_2 = 3\}) \\ &= \frac{1}{9} + 0 + \frac{1}{6} = \frac{6}{27} \end{aligned}$$

Analog läuft dies:

$$\begin{aligned} P(\{\mathcal{X}_1 = 2\}) &= 0 + \frac{1}{3} + \frac{1}{6} = \frac{1}{2} \\ P(\{\mathcal{X}_2 = 3\}) &= \frac{1}{9} + \frac{1}{9} + 0 = \frac{2}{9} \end{aligned}$$

Ferner gilt:

$$\begin{aligned} P(\{\mathcal{X}_2 = 1\}) &= P(\{\mathcal{X}_1 = 1, \mathcal{X}_2 = 1\}) + P(\{\mathcal{X}_1 = 2, \mathcal{X}_2 = 1\}) + P(\{\mathcal{X}_1 = 3, \mathcal{X}_2 = 1\}) \\ &= \frac{1}{9} + 0 + \frac{1}{9} = \frac{2}{9}, \\ P(\{\mathcal{X}_2 = 2\}) &= 0 + \frac{1}{3} + \frac{1}{9} = \frac{4}{9} \\ P(\{\mathcal{X}_2 = 3\}) &= \frac{1}{6} + \frac{1}{6} + 0 = \frac{1}{3}. \end{aligned}$$

Es gilt zu beachten, dass Randverteilung die gemeinsame Verteilung nicht eindeutig festlegen. Dies sehen wir hier: Tabelle 3.2.

\mathcal{X}_2	\mathcal{X}_1			
	1	2	3	$P_{\mathcal{X}_2}$
1	$1/9$	0	$1/9$	$2/9$
2	0	$1/3$	$1/9$	$4/9$
3	$1/6$	$1/6$	0	$1/3$
$P_{\mathcal{X}_1}$	$6/27$	$1/2$	$2/9$	

Tbl. 3.1: Die gemeinsame Verteilung von \mathcal{X}_1 und \mathcal{X}_2 . Mit $P_{\mathcal{X}_2}(\{y\}) = P(\{\mathcal{X}_2 = y\})$ sowie $P_{\mathcal{X}_1}(\{x\}) = P(\{\mathcal{X}_1 = x\})$. Die Summe aller Einträge ist 1.

\mathcal{X}_2	\mathcal{X}_1		
	0	1	$P_{\mathcal{X}_2}(\{y\})$
0	$1/4$	$1/4$	$1/2$
1	$1/4$	$1/4$	$1/2$
$P_{\mathcal{X}_1}(\{z\})$	$1/2$	$1/2$	

\mathcal{X}_2	\mathcal{X}_1		
	0	1	$P_{\mathcal{X}_2}(\{y\})$
0	$1/6$	$2/6$	$1/2$
1	$2/6$	$1/6$	$1/2$
$P_{\mathcal{X}_1}(\{z\})$	$1/2$	$1/2$	

Tbl. 3.2: Beispiel für die Uneindeutigkeit der von der **Randverteilung** festgelegten gemeinsamen Verteilung.

3.3.2 Der absolutstetige Fall

Definition 3.3.4 – Zweidimensionale-Dichte

Eine Funktion $p : \mathbb{R}^2 \rightarrow \mathbb{R}$ heißt *Dichte*, falls

1. $p(x, y) \geq 0$ für alle $x \in \mathbb{R}, y \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \, dx \, dy = 1$.

Definition 3.3.5 – Gemeinsame Dichte

Sei $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$ ein **Zufallsvektor**. Eine **Dichte** $p : \mathbb{R}^2 \rightarrow \mathbb{R}$ heißt die *gemeinsame Dichte* von \mathcal{X}_1 und \mathcal{X}_2 , falls für alle $x_1 \leq x_2, y_1 \leq y_2$ gilt:

$$P(\{\mathcal{X}_1 \in [x_1, x_2], \mathcal{X}_2 \in [y_1, y_2]\}) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y) \, dx \, dy.$$

vergleichen wir dies mit dem eindimensionalen Fall, so wird der Schritt in die zweite Dimension hoffentlich klar:

$$P(\{\mathcal{X}_1 \in [x_1, x_2]\}) = \int_{x_1}^{x_2} p(x) \, dx.$$

Weiter heißt \mathcal{X} dann *absolut stetig verteilt*.

Es gilt zu beachten: Ist p die **gemeinsame Dichte** von \mathcal{X}_1 und \mathcal{X}_2 , so besitzt $P_{\mathcal{X}_1}$ die **Dichte**

$$P_{\mathcal{X}_1}(x) = \int_{\mathbb{R}} p(x, y) \, dy, \quad x \in \mathbb{R}$$

und besitzt $P_{\mathcal{X}_2}$ die **gemeinsame Dichte**

$$P_{\mathcal{X}_2}(y) = \int_{\mathbb{R}} p(x, y) \, dx, \quad y \in \mathbb{R}.$$

Man nennt $P_{\mathcal{X}_1}$ und $P_{\mathcal{X}_2}$ übrigens die *Randdichten* von \mathcal{X}_1 und \mathcal{X}_2 .

Beispiel 3.9 – Die Randdichten berechnen

Sei $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$ ein **Zufallsvektor** und die **gemeinsame Dichte** von \mathcal{X}_1 und \mathcal{X}_2 sei gegeben durch die gegebene Funktion:

$$p(x, y) = \begin{cases} x + 2xy, & \text{falls } x, y, \in [0, 1] \\ 0, & \text{sonst.} \end{cases}$$

Wir wollen nun die Randdichten $P_{\mathcal{X}_1}$ und $P_{\mathcal{X}_2}$ bestimmen. Für alle $x \in \mathbb{R}$ gilt:

$$P_{\mathcal{X}_1}(x) = \int_{-\infty}^{\infty} p(x, y) \, dy.$$

Wir betrachten nun beide Fälle:

Fall 1: $x \in [0, 1]$. Dann gilt:

$$\begin{aligned} p_{\mathcal{X}_1}(x) &= \int_{-\infty}^0 \underbrace{p(x, y)}_{=0} \, dy + \int_0^1 \underbrace{p(x, y)}_{x+2xy} \, dy + \int_1^{\infty} \underbrace{p(x, y)}_{=0} \, dy \\ &= \int_0^1 (x + 2xy) \, dy \\ &= (x \cdot y + xy^2) \Big|_0^1 = 2x. \end{aligned}$$

Fall 2: $x \notin [0, 1]$. Dann gilt:

$$P_{\mathcal{X}_1}(x) = \int_{-\infty}^{\infty} 0 \, dy = 0$$

Also,

$$P_{\mathcal{X}_1}(x) = \begin{cases} 2x, & \text{falls } x \in [0, 1] \\ 0, & \text{sonst.} \end{cases}$$

Analog gilt für alle $y \in \mathbb{R}$:

$$P_{\mathcal{X}_2}(y) = \int_{-\infty}^{\infty} p(x, y) \, dx$$

Betrachten nun noch den Fall $y \in [0, 1]$. Dann gilt:

$$\begin{aligned} P_{\mathcal{X}_2}(y) &= \int_{-\infty}^0 \underbrace{p(x, y)}_{=0} \, dx + \int_0^1 \underbrace{p(x, y)}_{=x+2xy} \, dx + \int_1^{\infty} \underbrace{p(x, y)}_{=0} \, dx \\ &= \int_0^1 (x + 2xy) \, dx \\ &= \frac{x^2}{2} + x^2 y \Big|_0^1 = \frac{1}{2} + y \end{aligned}$$

Also gilt für $P_{\mathcal{X}_2}$.

$$P_{\mathcal{X}_2}(y) = \begin{cases} \frac{1}{2} + y, & \text{falls } y \in [0, 1] \\ 0, & \text{sonst.} \end{cases}$$

3.4 Unabhängigkeit von Zufallsvariablen

Wir erinnern uns daran, dass zwei Ereignisse A_1 und A_2 **unabhängig** genannt werden, falls

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$

Diese Definition erweitert sich auf mehrere Mengen, so heißen A_1, A_2, \dots, A_n unabhängig, falls

$$P\left(\bigcap_{i \in T} A_i\right) = \prod_{i \in T} P(A_i)$$

für alle $T \subseteq \{1, 2, \dots, n\}$ mit $|T| \geq 2$.

So wie wir dies für Mengen definiert haben, befassen wir uns nun mit der Unabhängigkeit von Zufallsvariablen $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n : \Omega \rightarrow \mathbb{R}$.

Definition 3.4.1 – Unabhängige Zufallsvariablen

Sei (Ω, Σ, P) ein **Wahrscheinlichkeitsraum** und seien $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n : \Omega \rightarrow \mathbb{R}$ **Zufallsvariablen** (sie müssen also über dem selben Wahrscheinlichkeitsraum definiert sein). Wir nennen diese **unabhängig**, falls für alle $A_1, A_2, \dots, A_n \in \mathcal{B}(\mathbb{R})$ also in der **Borel- σ -Algebra** und alle $T \subseteq \{1, 2, \dots, n\}$ mit $|T| \geq 2$ gilt:

$$P\left(\bigcap_{i \in T} \{\mathcal{X}_i \in A_i\}\right) = \prod_{i \in T} P(\{\mathcal{X}_i \in A_i\})$$

Hierbei ist $\{\mathcal{X}_i \in A_i\}$ nach wie vor die Kurzschreibweise für:

$$\{\mathcal{X}_i \in A_i\} \hat{=} \{\omega \in \Omega : \mathcal{X}_i(\omega) \in A_i\} \in \Sigma$$

Satz 3.4.1 – Anderes Maß für die Unabhängigkeit

Die **Zufallsvariablen** $\mathcal{X}_1, \dots, \mathcal{X}_n$ sind **unabhängig**, genau dann, wenn für alle $A_1, A_2, \dots, A_n \in \mathcal{B}(\mathbb{R})$ gilt:

$$P\left(\bigcap_{i=1}^n \{\mathcal{X}_i \in A_i\}\right) = \prod_{i=1}^n P(\{\mathcal{X}_i \in A_i\}).$$

Wir können also, anders bei den Ereignissen, auf die Teilmengen verzichten!

Machen wir hierfür einmal ein Beispiel...

Beispiel 3.10 – Zweimaliges Würfeln

Wir betrachten wiederum den **Laplace-Wahrscheinlichkeitsraum** mit $\Omega = \{1, \dots, 6\}^2$. Wir betrachten die Zufallsvariablen $\mathcal{X}_1, \mathcal{X}_2 : \Omega \rightarrow \mathbb{R}$ mit

$$\mathcal{X}_1(\omega) = \mathcal{X}_1((\omega_1, \omega_2)) = \omega_1 + \omega_2, \quad \mathcal{X}_2(\omega) = \mathcal{X}_2((\omega_1, \omega_2)) = \omega_1 \cdot \omega_2.$$

Nun stellt sich die Frage, ob \mathcal{X}_1 und \mathcal{X}_2 unabhängig sind. Wir wählen beispielhaft $A_1 = \{2\}$ und $A_2 = \{36\}$. Damit berechnen sich die Wahrscheinlichkeiten wie folgt:

$$\begin{aligned} P\left(\bigcap_{i=1}^2 \{\mathcal{X}_i \in A_i\}\right) &= P(\{\mathcal{X}_1 = 2\} \cap \{\mathcal{X}_2 = 36\}) \\ &= P(\{\omega = (\omega_1, \omega_2) \in \{1, \dots, 6\}^2 : \omega_1 + \omega_2 = 2 \text{ und } \omega_1 \cdot \omega_2 = 36\}) \\ &= P(\{\omega \in \{1, \dots, 6\}^2 : \omega = (1, 1) \text{ und } \omega = (6, 6)\}) \\ &= P(\emptyset) = 0 \end{aligned}$$

Aber:

$$\begin{aligned} P(\{\mathcal{X}_1 \in A_1\}) \cdot P(\{\mathcal{X}_2 \in A_2\}) &= P(\{(1, 1)\}) \cdot P(\{(6, 6)\}) \\ &= \frac{1}{36} \cdot \frac{1}{36} \neq 0 \end{aligned}$$

damit sind \mathcal{X}_1 und \mathcal{X}_2 *nicht* unabhängig.

Betrachten wir aber nun den Spezialfall $n = 2$. Dafür muss für alle $A_1, A_2 \in \mathcal{B}(\mathbb{R})$ folgendes gelten:

$$P(\{\mathcal{X}_1 \in A_1\} \cap \{\mathcal{X}_2 \in A_2\}) = P(\{\mathcal{X}_1 \in A_1\}) \cdot P(\{\mathcal{X}_2 \in A_2\}).$$

Allerdings es schwer, diese Forderung für alle $A_1, A_2 \in \mathcal{B}(\mathbb{R})$ zu überprüfen! Wir werden nun für den **diskreten** und den **absolutstetigen** Fall leichtere Kriterien angeben...

Der diskrete Fall

Satz 3.4.2 – Der diskrete Fall für die Unabhängigkeit bei Zufallsvariablen

Seien \mathcal{X}_1 und \mathcal{X}_2 **diskrete Zufallsvariablen**. Dann sind \mathcal{X}_1 und \mathcal{X}_2 unabhängig, genau dann, wenn für alle $x \in D_{\mathcal{X}_1}$ und $y \in D_{\mathcal{X}_2}$ (also die **Träger**) gilt:

$$P(\{\mathcal{X}_1 = x\} \cap \{\mathcal{X}_2 = y\}) = P(\{\mathcal{X}_1 = x\}) \cdot P(\{\mathcal{X}_2 = y\}).$$

Beispiel 3.11

Betrachten wir analog zu den Tabellen wie **Tabelle 3.2**, genauer die folgende:

\mathcal{X}_2	\mathcal{X}_1		
	0	1	$P_{\mathcal{X}_2}(\{y\})$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$P_{\mathcal{X}_1}(\{z\})$	$\frac{1}{2}$	$\frac{1}{2}$	

Sind \mathcal{X}_1 und \mathcal{X}_2 unabhängig? Es gilt:

$$D_{\mathcal{X}_1} = D_{\mathcal{X}_2} = \{0, 1\}.$$

Ferner gilt für alle $x \in D_{\mathcal{X}_1}$ und $y \in D_{\mathcal{X}_1}$:

$$P(\{\mathcal{X}_1 = x\} \cap \{\mathcal{X}_2 = y\}) = P(\{\mathcal{X}_1 = x, \mathcal{X}_2 = y\}) = \frac{1}{4}$$

und

$$P(\{\mathcal{X}_1 = x\}) \cdot P(\{\mathcal{X}_2 = y\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Das heißt, dass \mathcal{X}_1 und \mathcal{X}_2 sind **unabhängig**.

Betrachten wir die zweite Tabelle:

\mathcal{X}_2	\mathcal{X}_1		
	0	1	$P_{\mathcal{X}_2}(\{y\})$
0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{2}$
1	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
$P_{\mathcal{X}_1}(\{z\})$	$\frac{1}{2}$	$\frac{1}{2}$	

Hier gilt wieder $D_{\mathcal{X}_1} = D_{\mathcal{X}_2} = \{0, 1\}$, aber diesmal sind \mathcal{X}_1 und \mathcal{X}_2 *nicht* unabhängig, so gilt:

$$P(\{\mathcal{X}_1 = 0\} \cap \{\mathcal{X}_2 = 0\}) = \frac{1}{6}$$

aber:

$$P(\{\mathcal{X}_1 = 0\}) \cdot P(\{\mathcal{X}_2 = 0\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \neq \frac{1}{6}.$$

Der absolutstetige Fall

Satz 3.4.3 – Randdichten im absolutstetigen Fall

Seien $\mathcal{X}_1, \mathcal{X}_2$ Zufallsvariablen die die **gemeinsame Dichte** $p : \mathbb{R}^2 \rightarrow \mathbb{R}$ besitzen und seien $p_{\mathcal{X}_1}$ und $p_{\mathcal{X}_2}$ die Randdichten von \mathcal{X}_1 und \mathcal{X}_2 Falls für alle $x \in \mathbb{R}$ und $y \in \mathbb{R}$ gilt:

$$p(x, y) = P_{\mathcal{X}_1}(x) \cdot P_{\mathcal{X}_2}(y)$$

dann sind \mathcal{X}_1 und \mathcal{X}_2 *unabhängig*.

Beispiel 3.12

Betrachten wir noch einmal **das Beispiel für die Randdichten**. Wir haben also Zufallsvariablen $\mathcal{X}_1, \mathcal{X}_2$ mit der **gemeinsamen Dichte** p mit

$$p(x, y) = \begin{cases} x + 2xy, & \text{falls } x, y, \in [0, 1] \\ 0, & \text{sonst} \end{cases}$$

besitzen. Wir zeigen nun, dass \mathcal{X}_1 und \mathcal{X}_2 *unabhängig* sind. Im vermerkten Beispiel haben wir gezeigt, dass

$$P_{\mathcal{X}_1}(x) = \begin{cases} 2x, & \text{falls } x \in [0, 1] \\ 0, & \text{sonst.} \end{cases}$$

$$P_{\mathcal{X}_2}(y) = \begin{cases} \frac{1}{2} + y, & \text{falls } y \in [0, 1] \\ 0, & \text{sonst.} \end{cases}$$

Somit gilt:

$$P_{\mathcal{X}_1}(x) \cdot P_{\mathcal{X}_2}(y) = \begin{cases} 2x(\frac{1}{2} + y) = x + 2xy, & \text{falls } x, y \in [0, 1] \\ 0, & \text{sonst.} \end{cases} = p(x, y).$$

3.5 Kenngrößen von Zufallsvariablen

3.5.1 Erwartungswert

Sei (Ω, Σ, P) ein **Wahrscheinlichkeitsraum** und $\mathcal{X} : \Omega \rightarrow \mathbb{R}$ eine **Zufallsvariable**. Wir wollen nun den „mittleren Wert“ von \mathcal{X} bestimmen. Hierfür werden wir einen Blick auf die Lotterie, ...

Beispiel 3.13 – Lotterie

Gegeben sei eine Urne mit sechs Losen, von denen

- 3 Lose mit einer Auszahlung von 1 Euro,
- 2 Lose mit einer Auszahlung von 2 Euro und
- 1 Los mit einer Auszahlung von 5 Euro

versehen ist. Doch was ist nun ein fairer Preis für ein Los? Das Ziel ist es, den Preis so zu wählen, dass „langfristig“ die Gesamtauszahlungen den Einnahmen entsprechen. Hierfür formalisieren wir weiter. Nach n Ziehungen haben wir die Auszahlungen $x_1, x_2, \dots, x_n \in \{1, 2, 5\}$ und die Einnahmen $n \cdot \text{Preis}$. Das Ziel ist nun, dass

$$x_1 + x_2 + \dots + x_n \approx n \cdot \text{Preis}.$$

das heißt:

$$\begin{aligned} \text{Preis} &\approx \underbrace{\frac{x_1 + x_2 + \dots + x_n}{n}}_{=\bar{x}} \\ &= 1 \cdot \frac{\text{Anzahl von „1“}}{n} + 2 \cdot \frac{\text{Anzahl von „2“}}{n} + 5 \cdot \frac{\text{Anzahl von „5“}}{n} \\ &= 1 \cdot p_n(1) + 2 \cdot p_n(2) + 5 \cdot p_n(5). \end{aligned}$$

wobei $p_n(k)$ für $k = 1, 2, 5$ die relative Häufigkeit des Gewinns k bezeichnet. Der Intuition nach gilt für „große“ n :

$$p_n(k) \approx \begin{cases} \frac{1}{2}, & \text{falls } k = 1 \\ \frac{1}{3}, & \text{falls } k = 2 \\ \frac{1}{6}, & \text{falls } k = 5 \end{cases}$$

Somit ergibt sich für den Preis:

$$\text{Preis} \approx 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{3} + 5 \cdot \frac{1}{6} = \frac{3 + 4 + 5}{6} = \frac{12}{6} = 2.$$

Wir machen also den Vorschlag für einen Preis von 2 Euro.

Wir können die 2 Euro aus dem Beispiel von eben auch wahrscheinlichkeitstheoretisch interpretieren. Sei die Auszahlung nach einer Ziehung durch eine Zufallsvariable \mathcal{X} modelliert mit

$$P(\{\mathcal{X} = 1\}) = \frac{1}{2}, \quad P(\{\mathcal{X} = 2\}) = \frac{1}{3}, \quad P(\{\mathcal{X} = 5\}) = \frac{1}{6}.$$

Damit haben wir einen **Wahrscheinlichkeitsraum** $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{\mathcal{X}})$. Dann gilt:

$$\begin{aligned} \text{Preis} &= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{3} + 5 \cdot \frac{1}{6} \\ &= 1 \cdot P(\{\mathcal{X} = 1\}) + 2 \cdot P(\{\mathcal{X} = 2\}) + 5 \cdot P(\{\mathcal{X} = 5\}) \\ &= \sum_{x \in D_{\mathcal{X}}} x \cdot P(\{\mathcal{X} = x\}). \end{aligned}$$

Die letzte Größe nennt man den *Erwartungswert* von \mathcal{X} .

Definition 3.5.1 – Erwartungswert einer diskreten Zufallsvariable

Sei \mathcal{X} eine **diskrete Zufallsvariable**. Dann heißt

$$E(\mathcal{X}) = \sum_{x \in D_{\mathcal{X}}} x \cdot P(\{\mathcal{X} = x\})$$

der *Erwartungswert* von \mathcal{X} .

Dieser Erwartungswert ist übrigens genau dann in \mathbb{R} , wenn $\sum_{x \in D_{\mathcal{X}}} |x| \cdot P(\{\mathcal{X} = x\}) < \infty$. Da wir dies in der Vorlesung **Angewandte Stochastik 1** stets als gegeben betrachten können, gilt hier immer

$$E(\mathcal{X}) \in \mathbb{R}.$$

Wenn die Reihe aber nicht konvergiert, muss die Summe *nicht wohldefiniert* sein.

Beispiel 3.14

1. Sei \mathcal{X} eine **Zufallsvariable**, die die Augenzahl beim Würfeln modelliert. Also $X \sim U(\{1, \dots, 6\})$. Dann gilt $D_{\mathcal{X}} = \{1, \dots, 6\}$. Also

$$\begin{aligned} E(\mathcal{X}) &= \sum_{x \in \{1, \dots, 6\}} x \cdot P(\{\mathcal{X} = x\}) \\ &= \frac{1}{6} \sum_{x=1}^6 x = \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} = 3.5 \end{aligned}$$

Hier ergibt sich der Erwartungswert also aus dem Durchschnitt.

2. Betrachten wir eine Einpunktverteilung (also $\mathcal{X} = c$ mit $c \in \mathbb{R}$, das heißt es gilt $\mathcal{X}(\omega) = c$ für alle $\omega \in \Omega$). Dann gilt $D_{\mathcal{X}} = \{c\}$ und somit auch

$$E(\mathcal{X}) = c \cdot P(\{\mathcal{X} = c\}) = c \cdot 1 = c.$$

Damit entspricht der Erwartungswert dem c .

3. Betrachten wir $\mathcal{X} \sim \mathcal{B}(1, p)$ (also ein **Bernoulliverteiltes** \mathcal{X}) mit $P \in [0, 1]$. Dann gilt

$$D_x = \begin{cases} \{0, 1\}, & \text{falls } p \in (0, 1), \\ \{0\}, & \text{falls } p = 0, \\ \{1\}, & \text{falls } p = 1. \end{cases}$$

Somit gilt:

$$E(\mathcal{X}) = \begin{cases} 0 \cdot (1 - p) + 1 \cdot p = p, & \text{falls } p \in (0, 1) \\ 0 \cdot 1 = 0 = p, & \text{falls } p = 0 \\ 1 \cdot 1 = 1 = p, & \text{falls } p = 1 \end{cases} = p$$

Bei einer Bernoulliverteilung entspricht p also dem Erwartungswert:

$$E(\mathcal{X}) = p$$

4. Mit $\mathcal{X} \sim P(\lambda)$ mit $\lambda > 0$. Es gilt $D_{\mathcal{X}} = \mathbb{N}_0$ und damit berechnet sich:

$$\begin{aligned} E(\mathcal{X}) &= \sum_{x=0}^{\infty} x \cdot P(\{\mathcal{X} = x\}) = \sum_{x=0}^{\infty} x \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} x \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^x}{(x-1)!} \\ &\stackrel{*}{=} \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^{k+1}}{k!} = \lambda \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \end{aligned}$$

Im ersten Schritt kommen wir von $\sum_{x=0}^{\infty}$ auf $\sum_{x=1}^{\infty}$, da durch das „ $x \cdot e^{-\lambda} \dots$ “ das Glied für $x = 0$ ja auch zu 0 resultiert und wir es damit vernachlässigen können!. Zur Vereinfachung haben wir bei $\stackrel{*}{=}$ eine Indexverschiebung mit $k = x - 1$ vorgenommen. Der letzte Schritt gilt übrigens insbesondere deswegen, da $e^{-\lambda} \frac{\lambda^k}{k!} = P(\{\mathcal{X} = k\})$ und $\sum_{k=0}^{\infty} P(\{\mathcal{X} = k\}) = P(\{\mathcal{X} \in \mathbb{N}_0\}) = 1$.

Damit ist der Erwartungswert im Falle einer **Poisson-Verteilung** das eingesetzte λ :

$$E(\mathcal{X}) = \lambda$$

5. Ist $\mathcal{X} \sim G(p)$ mit $p \in (0, 1]$, dann gilt

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

Wir werden dies im Rahmen von **Angewandte Stochastik 1** ohne Beweis übernehmen.

Sei nun \mathcal{X} absolutstetig verteilt

Wir definieren uns den *Erwartungswert*...

Definition 3.5.2 – Erwartungswert einer absolutstetigen Zufallsvariable

Sei \mathcal{X} eine absolutstetig verteilte **Zufallsvariable** mit der **Dichte** p (mit $\int_{-\infty}^{\infty} |x| p(x) dx < \infty$).

Dann heißt

$$E(\mathcal{X}) = \int_{-\infty}^{\infty} x \cdot p(x) \, dx \in \mathbb{R}$$

der Erwartungswert von \mathcal{X} .

Anstelle der Summe verwenden wir im *absolutstetigen* Fall also den Integral zur Bestimmung des Erwartungswerts. Dieser Erwartungswert ist übrigens nur genau dann in \mathbb{R} , wenn das folgende Integral endlich ist, also:

$$\int_{-\infty}^{\infty} |x| p(x) \, dx < \infty.$$

Da dies in *Angewandte Stochastik 1* nicht überprüft werden muss, gilt auch hier für uns:

$$E(\mathcal{X}) \in \mathbb{R}.$$

Betrachten wir hierzu auch ein paar Beispiele...

Beispiel 3.15 – Zum Erwartungswert beim absolutstetigen \mathcal{X}

1. Sei $\mathcal{X} \sim U([a, b])$ mit $a < b$, so ist $E(\mathcal{X}) = \frac{a+b}{2}$ (also in der Mitte des Intervalls, wie uns ja bereits schon die Intuition verrät). Es gilt für die **Dichte**:

$$p(x) = \begin{cases} \frac{1}{b-a}, & \text{falls } x \in [a, b] \\ 0, & \text{sonst.} \end{cases}$$

Damit gilt für den **Erwartungswert** von \mathcal{X} (wir können das Integral ja vereinfachen, da die Dichte außerhalb von $[a, b]$ als konstant „0“ definiert ist):

$$\begin{aligned} E(\mathcal{X}) &= \int_{-\infty}^{\infty} x p(x) \, dx = \int_a^b x \cdot \frac{1}{b-a} \, dx \\ &= \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_a^b = \frac{1}{b-a} \cdot \frac{1}{2} (\underbrace{b^2 - a^2}_{(b-a)(b+a)}) = \frac{b+a}{2}. \end{aligned}$$

Damit ist $E(\mathcal{X})$ der Mittelpunkt des Intervalls $[a, b]$:

$$E(\mathcal{X}) = \frac{b+a}{2}$$

2. Ist $\mathcal{X} \sim \text{Exp}(\lambda)$, also **exponentialverteilt** mit $\lambda > 0$. Dann gilt für die **Dichte**

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{falls } x \geq 0 \\ 0, & \text{sonst.} \end{cases}$$

In diesem Fall müssen wir *partiell Integrieren*, die gewählten Terme sind hier mit $g(x)$ und $f'(x)$ markiert. Zur Erinnerung; bei der partiellen Integration gilt:

$$\int_a^b g(x) \cdot f'(x) \, dx = f(x) \cdot g(x) \Big|_a^b - \int_a^b f(x) \cdot g'(x) \, dx$$

Angewendet auf die hier vorliegende Dichte zur Berechnung des Erwartungswerts erhalten wir damit (wir können das Integral für ≤ 0 wieder verwerfen, da die Dichte hier ja als konstant „0“ definiert ist):

$$\begin{aligned} E(\mathcal{X}) &= \int_{-\infty}^{\infty} x \cdot p(x) \, dx = \int_0^{\infty} \underbrace{x}_{g(x)} \cdot \underbrace{\lambda \cdot e^{-\lambda x}}_{f'(x)} \, dx \\ &= \underbrace{x \cdot (-e^{-\lambda x}) \Big|_0^{\infty}}_{\lim_{x \rightarrow \infty} (-xe^{-\lambda x} + 0) = 0} - \int_0^{\infty} 1 \cdot (-e^{-\lambda x}) \, dx \\ &= \int_0^{\infty} e^{-\lambda x} \, dx = -\frac{e^{-\lambda x}}{\lambda} \Big|_0^{\infty} = +\frac{e^{\lambda \cdot 0}}{\lambda} = \frac{1}{\lambda}. \end{aligned}$$

Damit gilt für den Erwartungswert im Falle einer exponential-verteilten Zufallsvariable:

$$E(\mathcal{X}) = \frac{1}{\lambda}$$

3. Ist $\mathcal{X} \sim N(0, \sigma^2)$ mit $\sigma^2 > 0$, dann gilt:

$$p(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{x^2}{2\sigma^2}}.$$

Damit ist

$$E(\mathcal{X}) = \int_{-\infty}^{\infty} x \cdot p(x) \, dx = \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi} \sigma^2} x \cdot e^{-\frac{x^2}{2\sigma^2}}}_{g(x)} \, dx = 0$$

Die Funktion ist ungerade, es gilt also $g(x) = -g(-x)$.

4. Ist $\mathcal{X} \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, also **normalverteilt**. Wir zeigen nun $E(x) = \mu$. Es gilt

$$p(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

und somit gilt auch die folgende Rechnung. In dieser müssen wir mit *Substitution Integrieren*. Die Substitution findet hier bei $\stackrel{*}{=}$ statt und bezeichnet die Ersetzung: $y = x - \mu$:

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \\ &\stackrel{*}{=} \int_{-\infty}^{\infty} (y + \mu) \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{y^2}{2\sigma^2}} \, dy \\ &= \underbrace{\int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{y^2}{2\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \, dy}_{=0 \text{ (nach Punkt 3)}} + \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{y^2}{2\sigma^2}} \, dy}_{\stackrel{=1}{\text{Dichte von } y \sim N(0, \sigma^2)}} \\ &= \mu. \end{aligned}$$

Damit gilt im Falle von $X \sim N(\mu, \sigma^2)$ für den Erwartungswert:

$$E(\mathcal{X}) = \mu$$

Eigenschaften des Erwartungswertes

Wir möchten nun die Eigenschaften des Erwartungswertes betrachten

Satz 3.5.1 – Eigenschaften des Erwartungswertes

Seien \mathcal{X} und \mathcal{Y} Zufallsvariablen.

- a) Für alle $a, b \in \mathbb{R}$ gilt

$$E(a \cdot \mathcal{X} + b \cdot \mathcal{Y}) = a \cdot E(\mathcal{X}) + b \cdot E(\mathcal{Y})$$

Der Erwartungswert ist damit linear.

- b) Falls $\mathcal{X} \leq \mathcal{Y}$ (das heißt $\forall \omega \in \Omega : \mathcal{X}(\omega) \leq \mathcal{Y}(\omega)$) dann gilt auch

$$\mathcal{X} \leq \mathcal{Y} \Rightarrow E(\mathcal{X}) \leq E(\mathcal{Y})$$

Der Erwartungswert ist damit monoton.

- c) Falls \mathcal{X} und \mathcal{Y} identisch verteilt sind (also $F_{\mathcal{X}} = F_{\mathcal{Y}}$), dann gilt auch

$$F_{\mathcal{X}} = F_{\mathcal{Y}} \Rightarrow E(\mathcal{X}) = E(\mathcal{Y}).$$

Interessant ist, dass die Linearität immer gilt, selbst dann, wenn \mathcal{X} und \mathcal{Y} nicht unabhängig sind.

Wir betrachten nun $E(f(\mathcal{X}))$ für eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$. Wir möchten also nicht den Erwartungswert von \mathcal{X} selbst bestimmen, sondern von einer an \mathcal{X} angelegten Funktion!

Satz 3.5.2 – Transformationssatz

Der Transformationssatz besteht aus zwei Komponenten:

1. Sei \mathcal{X} eine diskrete Zufallsvariable und sei $f : \mathbb{R} \rightarrow \mathbb{R}$ (mit $\sum_{x \in D_{\mathcal{X}}} |f(x)| \cdot P(\{\mathcal{X} = x\}) < \infty$). Dann gilt

$$E(f(\mathcal{X})) = \sum_{x \in D_{\mathcal{X}}} f(x) \cdot P(\{\mathcal{X} = x\})$$

Anstelle von \mathcal{X} rechnen wir nun mit $f(x)$, welches wir bei Bedarf mit \mathcal{Y} einer weiteren Zufallsvariable (die sich über die Funktion aus \mathcal{X} beschreibt) bezeichnen werden.

2. Sei \mathcal{X} eine absolutstetige Zufallsvariable und sei $f : \mathbb{R} \rightarrow \mathbb{R}$ (mit $\int_{-\infty}^{\infty} |f(x)| \cdot p(x) dx < \infty$). Dann gilt

$$E(f(\mathcal{X})) = \int_{-\infty}^{\infty} f(x) \cdot p(x) dx.$$

Wenn wir es genau nehmen möchten, gilt der Transformationssatz übrigens nur, wenn

$$\sum_{x \in D_{\mathcal{X}}} |f(x) \cdot P(\{\mathcal{X} = x\})| < \infty$$

also die Summe beschränkt ist. Im Rahmen von *Angewandte Stochastik 1* müssen wir dies allerdings nicht überprüfen und dürfen die Beschränktheit als „stets gegeben“ betrachten.

Betrachten wir auch hierzu zwei Beispiele.

Beispiel 3.16

1. Sei $\mathcal{X} \sim U(\{1, \dots, 6\})$ (also ein einmaliges Würfeln). Wir möchten nun $E(\mathcal{X}^2)$ bestimmen. Hier hilft die erste Komponente des **Transformationssatzes**, sie liefert uns mit $f(x) = x^2$ folgende Gleichung:

$$\begin{aligned} E(\mathcal{X}^2) &= \sum_{x=1}^6 x^2 \cdot \underbrace{P(\{\mathcal{X} = x\})}_{\frac{1}{6}} \\ &= \frac{(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2)}{6} = \frac{91}{6} \end{aligned}$$

2. Sei nun $\mathcal{X} \sim U([0, 1])$. Möchten wir nun $E(\mathcal{X}^3)$ bestimmen so vermerken wir zuerst wieder die **Dichte**

$$p(x) = \begin{cases} 1, & \text{falls } x \in [0, 1] \\ 0, & \text{sonst.} \end{cases}$$

Nach der zweiten Komponente des **Transformationssatzes** erhalten wir mit $f(x) = x^3$ die folgende Rechnung.

$$E(\mathcal{X}^3) = \int_{-\infty}^{\infty} x^3 p(x) dx = \int_0^1 x^3 \cdot 1 dx = \left. \frac{x^4}{4} \right|_0^1 = \frac{1}{4} - 0 = \frac{1}{4}.$$

Betrachten wir nun $E(\mathcal{X} \cdot \mathcal{Y})$. Es ist wichtig zu verstehen, dass im allgemeinen **nicht** gilt, dass

$$E(\mathcal{X} \cdot \mathcal{Y}) = E(\mathcal{X}) \cdot E(\mathcal{Y}).$$

Um die Fälle in denen es gilt zu erfassen, formulieren wir den folgenden Satz...

Satz 3.5.3 – Multiplizieren von Zufallsvariablen

Wenn \mathcal{X} und \mathcal{Y} **unabhängige** Zufallsvariablen sind, für die $E|\mathcal{X}|, E|\mathcal{Y}| < \infty$ gilt, dann gilt auch:

$$E(\mathcal{X} \cdot \mathcal{Y}) = E(\mathcal{X}) \cdot E(\mathcal{Y}).$$

3.5.2 Varianz

Sei \mathcal{X} nun eine **Zufallsvariable**. Der **Erwartungswert** $E(\mathcal{X})$ ist dann der „Mittelwert“ von \mathcal{X} . Doch wie weit „stret“ \mathcal{X} um den Erwartungswert $E(\mathcal{X})$ herum (nur weil das Ergebnis erwartet wird, heißt es ja nicht, dass es nicht auch andere Ergebnisse innerhalb des Intervalls geben kann). Betrachten wir ein Beispiel in dem dies besonders deutlich wird. Sei hierfür $n \in \mathbb{N}$ und \mathcal{X} , eine Zufallsvariable mit der Eigenschaft: $P(\{\mathcal{X} = n\}) = \frac{1}{2}$ und $P(\{\mathcal{X} = -n\}) = \frac{1}{2}$. Der Erwartungswert beläuft sich damit auf 0:

$$\begin{aligned} E(\mathcal{X}) &= n \cdot P(\{\mathcal{X} = n\}) + (-n) \cdot P(\{\mathcal{X} = -n\}) \\ &= n \cdot \frac{1}{2} - n \cdot \frac{1}{2} = 0. \end{aligned}$$

Allerdings ist \mathcal{X} ja nie 0, es kann nur die Werte $-n$ und n annehmen. Aufgrund derartiger Erkenntnisse lohnt sich die Definition der Varianz:

Definition 3.5.3 – Varianz

Sei \mathcal{X} eine Zufallsvariable mit $E(\mathcal{X}^2) < \infty$. Dann heißt

$$\text{Var}(\mathcal{X}) = E((\mathcal{X} - E(\mathcal{X}))^2)$$

die Varianz von \mathcal{X} .

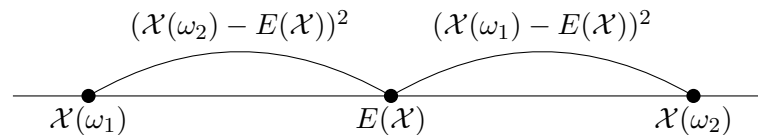


Abb. 3.3: Veranschaulichung der Varianz als der „Mittelwert der Abstände zum Erwartungswert $E(\mathcal{X})$ “.

Hierbei bezeichnet $\text{Var}(\mathcal{X})$ ein Maß für die Streuung von \mathcal{X} um $E(\mathcal{X})$. Der Transformationssatz liefert uns damit auch den folgenden Satz...

Satz 3.5.4 – Formeln für die Varianz

Sei \mathcal{X} eine Zufallsvariable mit $E(\mathcal{X}^2) < \infty$ setze $\mu = E(\mathcal{X})$.

1. Falls \mathcal{X} diskret ist, gilt nach dem Transformationssatz

$$\text{Var}(\mathcal{X}) = \sum_{x \in D_{\mathcal{X}}} (x - \mu)^2 P(\{\mathcal{X} = x\}).$$

2. Falls \mathcal{X} mit der Dichte p absolutstetig verteilt ist, so gilt

$$\text{Var}(\mathcal{X}) = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx.$$

Wir haben nun also Formeln erfasst, mit denen wir die Varianz berechnen können!

Betrachten wir auch hierzu ein Beispiel...

Beispiel 3.17 – Zur Varianz

1. Sei \mathcal{X} wie in der Anfangsbetrachtung zur Varianz, also

$$P(\{\mathcal{X} = n\}) = P(\{\mathcal{X} = -n\}) = \frac{1}{2}.$$

Dann gilt $E(\mathcal{X}) = 0$ und somit für die Varianz (da wir uns im diskreten Fall befinden) mit dem Träger:

$$\begin{aligned} \text{Var}(\mathcal{X}) &= (n - 0)^2 \cdot \frac{1}{2} + (-n - 0)^2 \cdot \frac{1}{2} \\ &= n^2 \cdot \frac{1}{2} + n^2 \cdot \frac{1}{2} = n^2 \end{aligned}$$

Die Varianz hängt damit von n ab und wird somit mit wachsendem n auch größer (da hier die Streuung um den Erwartungswert ebenfalls größer wird, sollte dies auch intuitiv bereits klar sein).

2. Sei \mathcal{X} nun eine *Einpunktverteilung* $\mathcal{X} = c$ mit $c \in \mathbb{R}$. Dann gilt $E(\mathcal{X}) = c$ und somit

$$\text{Var}(\mathcal{X}) = \underbrace{(c - c)^2}_0 \cdot \underbrace{P(\{\mathcal{X} = c\})}_{=1} = 0.$$

3. Sei nun $\mathcal{X} \sim B(1, p)$ also **bernoulliverteilt** mit $p \in [0, 1]$. Es gilt: $E(\mathcal{X}) = p$ und $\text{Var}(\mathcal{X}) = E(\mathcal{X} - p)^2$. Da der **Träger** sich wie folgt beschreibt:

$$D_{\mathcal{X}} = \begin{cases} \{0, 1\}, & \text{falls } p \in (0, 1) \\ \{0\}, & \text{falls } p = 0 \\ \{1\}, & \text{falls } p = 1 \end{cases}$$

folgt für die **Varianz**:

$$\text{Var}(\mathcal{X}) = \begin{cases} (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p \stackrel{*}{=} p(1 - p) & \text{falls } p \in (0, 1) \\ (0 - 0)^2 \cdot 1 = 0, & \text{falls } p = 0 \\ (1 - 1)^2 \cdot 1 = 0, & \text{falls } p = 1 \end{cases} = p(1 - p)$$

Die Berechnung mit dem Stern läuft wie folgt ab:

$$\begin{aligned} (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p &= p^2(1 - p) + (1 - p)^2 \cdot p \\ &= p(1 - p)(p + 1 - p) = p(1 - p) \end{aligned}$$

Damit erhalten wir für die **Varianz**: $\text{Var}(\mathcal{X}) = p(1 - p)$.

In vielen Fällen ist es aber nicht notwendig die **Formeln** für die Varianz zu verwenden. Oft können wir uns einige nützliche Eigenschaften zu Eigen machen, die die Berechnung vereinfachen...

Bemerkung 3.3 – Zu den Bedingungen

Wir schreiben hier und an vielen Stellen Bedingungen, wie: $E(\mathcal{X}^2), E(\mathcal{Y}^2) < \infty$. In *Angewandte Stochastik 1* können wir diese immer als gegeben betrachten, es wird nicht erwartet, dass wir dies nachweisen können.

Satz 3.5.5 – Eigenschaften der Varianz

Seien \mathcal{X} und \mathcal{Y} **Zufallsvariablen** mit $E(\mathcal{X}^2)$ und $E(\mathcal{Y}^2) < \infty$, dann gilt:

1. $\text{Var}(\mathcal{X}) = E(\mathcal{X}^2) - (E(\mathcal{X}))^2$

2. Für alle $a, b \in \mathbb{R}$ gilt:

$$\text{Var}(a \cdot \mathcal{X} + b) = a^2 \text{Var}(\mathcal{X}).$$

Der „Shift“ (hier durch b) der **Varianz** spielt also keine Rolle.

3. Weiter ist $\text{Var}(\mathcal{X}) = 0$ genau dann, wenn:

$$P(\{\mathcal{X} = c\}) = 1 \text{ für ein } c \in \mathbb{R}.$$

4. Sind \mathcal{X} und \mathcal{Y} **unabhängig**, so folgt daraus:

$$\text{Var}(\mathcal{X} + \mathcal{Y}) = \text{Var}(\mathcal{X}) + \text{Var}(\mathcal{Y})$$

5. Falls \mathcal{X} und \mathcal{Y} **identisch verteilt** sind, gilt:

$$\text{Var}(\mathcal{X}) = \text{Var}(\mathcal{Y}).$$

Betrachten wir auch zu diesen Eigenschaften ein paar Beispiele...

Beispiel 3.18

1. Ist $\mathcal{X} \sim P(\lambda)$ mit $\lambda > 0$. Es gilt: $D_{\mathcal{X}} = \mathbb{N}_0$ für den **Träger** und

$$E(\mathcal{X}^2) = \sum_{x=0}^{\infty} x^2 \cdot e^{-\lambda} \frac{\lambda^x}{x!} \stackrel{!}{=} \lambda^2 + \lambda$$

für den **Erwartungswert** (der mit $\stackrel{!}{=}$ verkürzte Rechenschritt bleibt zur Übung). Damit gilt für die **Varianz**:

$$\text{Var}(\mathcal{X}) = E(\mathcal{X}^2) - (E(\mathcal{X}))^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

2. Sei $X \sim U([a, b])$ mit $a < b$. Damit ist (wie bereits schon berechnet)

$$E(\mathcal{X}) = \frac{a+b}{2}$$

damit erhalten wir für den folgenden **Erwartungswert**

$$\begin{aligned} E(\mathcal{X}^2) &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^3}{3} \Big|_a^b \\ &= \frac{b^3 - a^3}{3(b-a)} \end{aligned}$$

und somit für die Varianz:

$$\text{Var}(\mathcal{X}) = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2} \right)^2 = \dots = \frac{(b-a)^2}{12}.$$

3. Ist $\mathcal{X} \sim \text{Exp}(\lambda)$ mit $\lambda > 0$ gilt $E(\mathcal{X}) = \frac{1}{\lambda}$ und $E(\mathcal{X}^2) = \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx \stackrel{!}{=} \frac{2}{\lambda^2}$. Somit ist die **Varianz**:

$$\text{Var}(\mathcal{X}) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}.$$

4. Sei $\mathcal{X} \sim N(\mu, \sigma^2)$. Hier betrachten wir zuerst einmal den Spezialfall $\mu = 0$ und $\sigma^2 = 1$, also $\mathcal{X} \sim N(0, 1)$. Hier gilt $E(\mathcal{X}) = 0$ und damit können wir wieder integrieren (mit *partieller Integration*):

$$\begin{aligned} E(\mathcal{X}^2) &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{x}_{g(x)} \cdot \underbrace{x \cdot e^{-\frac{x^2}{2}}}_{f'(x)} dx \\ &= \frac{1}{\sqrt{2\pi}} \left(\underbrace{-x \cdot e^{-\frac{x^2}{2}}}_{=0} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} 1 \cdot (-e^{-\frac{x^2}{2}}) dx \right) \end{aligned}$$

$$= \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}_{\text{Dichte von } \mathcal{X}} = 1.$$

Daraus folgt: $\text{Var}(\mathcal{X}) = 1 - 0^2 = 1$. Für allgemeinere μ und σ^2 kann man zeigen, dass $\text{Var}(\mathcal{X}) = \sigma^2$. Also, für $\mathcal{X} \sim N(\mu, \sigma^2)$ ist $\mu = E(\mathcal{X})$ der Erwartungswert von \mathcal{X} und $\sigma^2 = \text{Var}(\mathcal{X})$ ist die Varianz von \mathcal{X} .

Sind $\mathcal{X}_1, \dots, \mathcal{X}_n$ unabhängige Zufallsvariablen mit $E(\mathcal{X}_1^2), \dots, E(\mathcal{X}_n^2) < \infty$, dann gilt

$$\text{Var}(\mathcal{X}_1 + \dots + \mathcal{X}_n) = \sum_{i=1}^n \text{Var}(\mathcal{X}_i).$$

3.5.3 Kovarianz und Korrelationskoeffizient

Wir suchen nun nach einer Kenngröße, die den „Grad der Abhängigkeit“ zwischen \mathcal{X} und \mathcal{Y} angibt und kommen damit auf den phonetisch wundervollen Begriff der Kovarianz...

Definition 3.5.4 – Kovarianz

Seien \mathcal{X} und \mathcal{Y} Zufallsvariablen mit $E(\mathcal{X}^2), E(\mathcal{Y}^2) < \infty$. Dann heißt

$$\text{Cov}(\mathcal{X}, \mathcal{Y}) = E(\underbrace{(\mathcal{X} - E(\mathcal{X})) \cdot (\mathcal{Y} - E(\mathcal{Y}))}_{\text{Zufallsvariable}}) \in \mathbb{R}$$

die Kovarianz von \mathcal{X} und \mathcal{Y} . \mathcal{X} und \mathcal{Y} heißen unkorreliert, falls $\text{Cov}(\mathcal{X}, \mathcal{Y}) = 0$.

Für die drei Fälle $\text{Cov}(\mathcal{X}, \mathcal{Y}) > 0$, $\text{Cov}(\mathcal{X}, \mathcal{Y}) < 0$ und $\text{Cov}(\mathcal{X}, \mathcal{Y}) = 0$ betrachten wir nun jeweils, was diese genau bedeuten:

$\text{Cov}(\mathcal{X}, \mathcal{Y}) > 0$: Wenn \mathcal{X} und \mathcal{Y} einen monotonen Zusammenhang besitzen, also hohe (niedrige) Werte von \mathcal{X} gehen tendenziell mit hohen (niedrigen) Werten von \mathcal{Y} einher (Abbildung 3.4).

$\text{Cov}(\mathcal{X}, \mathcal{Y}) < 0$: Wenn \mathcal{X} und \mathcal{Y} einen gegensinnigen monotonen Zusammenhang aufweisen, das heißt hohe Werte einer Zufallsvariablen gehen tendenziell mit niedrigen Werten der anderen Zufallsvariablen einher (Abbildung 3.5).

$\text{Cov}(\mathcal{X}, \mathcal{Y}) = 0$: hier besteht kein monotoner Zusammenhang zwischen \mathcal{X} und \mathcal{Y} .



Abb. 3.4: Der Fall $\text{Cov}(\mathcal{X}, \mathcal{Y}) > 0$ ein monotoner Zusammenhang.



Abb. 3.5: Der Fall $\text{Cov}(\mathcal{X}, \mathcal{Y}) < 0$ also der gegensinnige monotone Zusammenhang.

Bemerkung 3.4 – Konkrete Berechnung der Kovarianz

1. Seien \mathcal{X} und \mathcal{Y} **diskrete Zufallsvariablen** mit $E(\mathcal{X}^2), E(\mathcal{Y}^2) < \infty$. Dann gilt die folgende Gleichung für die **Kovarianz**:

$$\text{Cov}(\mathcal{X}, \mathcal{Y}) = \sum_{x \in D_{\mathcal{X}}} \sum_{y \in D_{\mathcal{Y}}} (x - E(\mathcal{X})) \cdot (y - E(\mathcal{Y})) \cdot P(\{\mathcal{X} = x, \mathcal{Y} = y\}).$$

Wir müssen also die **Träger** wie auch die **Erwartungswerte** von \mathcal{X} und \mathcal{Y} bestimmen sowie die **gemeinsame Verteilung**.

2. Falls \mathcal{X} und \mathcal{Y} eine **gemeinsame Dichte** p besitzen und $E(\mathcal{X}^2), E(\mathcal{Y}^2) < \infty$ also beide beschränkt sind, so gilt

$$\text{Cov}(\mathcal{X}, \mathcal{Y}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(\mathcal{X})) \cdot (y - E(\mathcal{Y})) \cdot p(x, y) \, dx \, dy.$$

Anstelle der Summe haben wir hier also wieder die Integrale.

Beispiel 3.19 – Berechnung der Kovarianz

1. Betrachten wir ein Beispiel für **diskrete Zufallsvariablen** und hierzu eine bereits **bekannte Tabelle 3.2**:

\mathcal{X}_2	\mathcal{X}_1		
	0	1	$P_{\mathcal{X}_2}(\{\mathcal{X}_2 = y\})$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$P_{\mathcal{X}_1}(\{\mathcal{X}_1 = x\})$	$\frac{1}{2}$	$\frac{1}{2}$	

Doch was ist hier die **Kovarianz** $\text{Cov}(\mathcal{X}_1, \mathcal{X}_2) = ?$. Es gilt $D_{\mathcal{X}_1} = D_{\mathcal{X}_2} = \{0, 1\}$ und damit auch für die **Erwartungswerte**:

$$E(\mathcal{X}_1) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2},$$

$$E(\mathcal{X}_2) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}.$$

Daraus folgt für die **Kovarianz**:

$$\begin{aligned} \text{Cov}(\mathcal{X}, \mathcal{Y}) &= \sum_{x=0}^1 \sum_{y=0}^1 (x - \frac{1}{2}) \underbrace{P(\{\mathcal{X}_1 = x, \mathcal{X}_2 = y\})}_{=\frac{1}{4}} \\ &= \frac{1}{4} \left((0 - \frac{1}{2})(0 - \frac{1}{2}) + (1 - \frac{1}{2})(0 - \frac{1}{2}) + (0 - \frac{1}{2})(1 - \frac{1}{2}) + (1 - \frac{1}{2})(1 - \frac{1}{2}) \right) \\ &= \frac{1}{4} \left(\frac{1}{4} - \frac{1}{4} - \frac{1}{4} + \frac{1}{4} \right) = 0. \end{aligned}$$

Also sind \mathcal{X}_1 und \mathcal{X}_2 **unkorreliert**. Aus dem bereits referenzierten **Beispiel** wissen wir, dass die beiden auch **unabhängig** sind.

2. Betrachten wir die zweite Tabelle aus [Tabelle 3.2](#):

\mathcal{X}_2	\mathcal{X}_1		
	0	1	$P_{\mathcal{X}_2}(\{y\})$
0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{2}$
1	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
$P_{\mathcal{X}_1}(\{z\})$	$\frac{1}{2}$	$\frac{1}{2}$	

Was ist hier die **Kovarianz**. Es gilt $D_{\mathcal{X}_1} = D_{\mathcal{X}_2} = \{0, 1\}$ und damit $E(\mathcal{X}_1) = E(\mathcal{X}_2) = \frac{1}{2}$. Somit folgt:

$$\begin{aligned}
 \text{Cov}(\mathcal{X}_1, \mathcal{X}_2) &= \sum_{x=0}^1 \sum_{y=0}^1 (x - \frac{1}{2})(y - \frac{1}{2}) \cdot P(\{\mathcal{X}_1 = x, \mathcal{X}_2 = y\}) \\
 &= (0 - \frac{1}{2})(0 - \frac{1}{2}) \cdot \frac{1}{6} + (0 - \frac{1}{2})(1 - \frac{1}{2}) \cdot \frac{2}{6} + (1 - \frac{1}{2})(0 - \frac{1}{2}) \\
 &\quad \cdot \frac{2}{6} + (1 - \frac{1}{2})(1 - \frac{1}{2}) \cdot \frac{1}{6} \\
 &= \dots = -\frac{1}{12} \neq 0
 \end{aligned}$$

Wir wissen aus dem referenzierten Beispiel, dass \mathcal{X}_1 und \mathcal{X}_2 *unabhängig* sind. Da die **Kovarianz** auch nicht 0 ist, fragen wir uns nun, ob es einen Zusammenhang zwischen den beiden Begrifflichkeiten gibt...

Der Zusammenhang zwischen Unabhängigkeit und Unkorreliertheit

Satz 3.5.6 – Eigenschaften von Kovarianz

Seien \mathcal{X} und \mathcal{Y} **Zufallsvariablen** mit $E(\mathcal{X}^2), E(\mathcal{Y}^2) < \infty$. Dann gilt:

1. Für die **Kovarianz** existiert eine alternative Formel:

$$\text{Cov}(\mathcal{X}, \mathcal{Y}) = E(\mathcal{X} \cdot \mathcal{Y}) - E(\mathcal{X}) \cdot E(\mathcal{Y}).$$

2. Sind \mathcal{X} und \mathcal{Y} **unabhängig**, so gilt $\text{Cov}(\mathcal{X}, \mathcal{Y}) = 0$.
3. Für die **Varianz** gilt:

$$\text{Var}(\mathcal{X} + \mathcal{Y}) = \text{Var}(\mathcal{X}) + \text{Var}(\mathcal{Y}) + 2 \cdot \text{Cov}(\mathcal{X}, \mathcal{Y}).$$

Bemerkung 3.5 – Wichtig

Es gilt:

$$\mathcal{X}, \mathcal{Y} \text{ unabhängig} \Rightarrow \mathcal{X}, \mathcal{Y} \text{ unkorreliert.}$$

Allerdings gilt nicht, dass Gegenteil, sind \mathcal{X} und \mathcal{Y} unkorreliert, so heißt das nicht zwangsläufig, dass auch \mathcal{X} und \mathcal{Y} unabhängig sind! Nehmen wir als ein Beispiel $\mathcal{X} \sim U([-1, 1])$ und $\mathcal{Y} = x^2$. Dann gilt $\text{Cov}(\mathcal{X}, \mathcal{Y}) = 0$, aber \mathcal{X} und \mathcal{Y} sind *nicht* unabhängig. Die **Umkehrung** gilt also im allgemeinen Falle **nicht**.

Die **Kovarianz** $\text{Cov}(\mathcal{X}, \mathcal{Y})$ gibt immer nur die *Richtung* einer Beziehung zwischen \mathcal{X} und \mathcal{Y} an, gibt aber keinen Aufschluss über die *Stärke* des Zusammenhangs.

Die Stärke gibt der *Korrelationskoeffizient* an...

Definition 3.5.5 – Kovarianzkoeffizient

Seien \mathcal{X} und \mathcal{Y} **Zufallsvariablen** mit $E(\mathcal{X}^2), E(\mathcal{Y}^2) < \infty$. Falls $\text{Var}(\mathcal{X}), \text{Var}(\mathcal{Y}) > 0$ heißt

$$\rho(\mathcal{X}, \mathcal{Y}) = \frac{\text{Cov}(\mathcal{X}, \mathcal{Y})}{\sqrt{\text{Var}(\mathcal{X}) \cdot \text{Var}(\mathcal{Y})}}$$

der *Korrelationskoeffizient* von \mathcal{X} nach \mathcal{Y} .

Satz 3.5.7 – Eigenschaften des Korrelationskoeffizient

Seien \mathcal{X} und \mathcal{Y} **Zufallsvariablen** mit $E(\mathcal{X}^2), E(\mathcal{Y}^2) < \infty$ und $\text{Var}(\mathcal{X}), \text{Var}(\mathcal{Y}) > 0$, dann gilt:

1. $\rho(\mathcal{X}, \mathcal{Y}) \in [-1, 1]$
2. $\rho(\mathcal{X}, \mathcal{Y}) = 0$, genau dann, wenn \mathcal{X} und \mathcal{Y} *unkorreliert* sind.
3. Ist $|\rho(\mathcal{X}, \mathcal{Y})| = 1$ genau dann, wenn \mathcal{X} und \mathcal{Y} mit einer Wahrscheinlichkeit 1 *linear abhängig* sind, das heißt $\exists a \neq 0$ und $b \in \mathbb{R}$ mit:

$$P(\{\mathcal{Y} = a \cdot \mathcal{X} + b\}) = 1.$$

Wir interpretieren $\rho(\mathcal{X}, \mathcal{Y})$ als den Grad und Richtung ($|\rho(\mathcal{X}, \mathcal{Y})|$) des linearen Zusammenhangs von \mathcal{X} und \mathcal{Y} .

3.6 Grenzwertsätze für Folgen und Zufallsvariablen

Dieser letzte Abschnitt besitzt für die Folgeveranstaltungen von *Angewandte Stochastik 1* eine hohe Bedeutung!

3.6.1 Gesetze der großen Zahlen

Sei (Ω, Σ, P) ein **Wahrscheinlichkeitsraum** und $(\mathcal{X}_i)_{i \geq 1}$ eine Folge von unabhängigen (das heißt für $n \in \mathbb{N}$ mit $\mathcal{X}_1, \dots, \mathcal{X}_n$ sind **unabhängig**) und **identisch verteilten Zufallsvariablen** (das heißt $F_{\mathcal{X}_i} = F_{\mathcal{X}_j}$ für alle $i, j \in \mathbb{N}$). Für $n \in \mathbb{N}$ betrachten wir die n -te Partialsumme $S_n = \mathcal{X}_1 + \dots + \mathcal{X}_n$. Damit haben wir die Möglichkeit $\frac{S_n}{n}$ für $n \rightarrow \infty$ zu untersuchen.

Beispiel 3.20 – Ein Münzwurf

Sei

$$\mathcal{X}_i = \begin{cases} 1, & \text{i-ter Wurf zeigt Kopf,} \\ 0, & \text{i-ter Wurf zeigt Zahl.} \end{cases}$$

Die intuitive Annahme, dass S_n als die Anzahl der Köpfe in den ersten n Würfeln, den Erwartungswert wie folgt beschreibt:

$$\frac{S_n}{n} \xrightarrow{n \rightarrow \infty} \frac{1}{2} = E(\mathcal{X}_1)$$

werden wir im Folgenden Formalisieren.

Übrigens: Wir haben hier der Anschaulichkeit halber keinen Wahrscheinlichkeitsraum definiert. Für $n \rightarrow \infty$ ist dies auch gar nicht so einfach!

Wir werden mit der Konstruktion eines Hilfsmittels beginnen...

Satz 3.6.1 – Die Tschebycheff-Ungleichung

Sei \mathcal{X} eine Zufallsvariable mit $E(\mathcal{X}^2) < \infty$. Dann gilt für alle $\varepsilon > 0$:

$$P(\{|\mathcal{X} - E(\mathcal{X})| \geq \varepsilon\}) \leq \frac{\text{Var}(\mathcal{X})}{\varepsilon^2}$$

Zwar können wir in einigen Fällen die Verteilung auch konkret berechnen – wenn aber \mathcal{X} komplizierter wird, ist die Verteilung als solche nicht mehr (trivialerweise) bekannt. In solchen Fällen ist dann die Tschebycheff-Ungleichung von Nutzen!

Bemerkung 3.6 – So the i-i-d wants my ID...

Im Folgenden bezeichnen wir mit i.i.d. Zufallsvariablen (independent, identically distributed) voneinander unabhängige, gleich verteilte Zufallsvariablen. Die Abkürzung dient der Lesbarkeit.

Satz 3.6.2 – Schwaches Gesetz der großen Zahlen

Sei $(\mathcal{X}_i)_{i \geq 1}$ eine Folge von i.i.d. Zufallsvariablen mit $E(\mathcal{X}_1^2) < \infty$. Setze $\mu = E(\mathcal{X}_1)$ und $\sigma^2 = \text{Var}(\mathcal{X}_1)$. Dann gilt für alle $\varepsilon > 0$:

$$P(\{|S_n/n - \mu| \geq \varepsilon\}) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Die Wahrscheinlichkeit, dass der Ausgang also um mehr als ε abweicht, konvergiert gegen 0!

Beweis 3.1 – Zum Gesetz der großen Zahlen

Verwenden wir die Tschebycheff-Ungleichung mit $\mathcal{X} = \frac{S_n}{n}$ so gilt für den Erwartungswert:

$$\begin{aligned} E(S_n/n) &= \frac{1}{n} E(\mathcal{X}_1 + \dots + \mathcal{X}_n) = \underbrace{E(\mathcal{X}_1)}_{=\mu} + \dots + \underbrace{E(\mathcal{X}_n)}_{=\mu} \\ &= \frac{1}{n} \cdot n \cdot \mu = \mu \end{aligned}$$

und damit für die Varianz:

$$\begin{aligned} \text{Var}(S_n/n) &= 1/n^2 \text{Var}(\mathcal{X}_1 + \dots + \mathcal{X}_n) \\ &= \frac{1}{n^2} \left(\underbrace{\text{Var}(\mathcal{X}_1)}_{=\sigma^2} + \dots + \underbrace{\text{Var}(\mathcal{X}_n)}_{=\sigma^2} \right) \\ &= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

Die Tschebycheff-Ungleichung liefert damit $\forall \varepsilon > 0$:

$$P(\{|S_n/n - \mu| \geq \varepsilon\}) \leq \frac{\text{Var} \frac{S_n}{n}}{\varepsilon^2} = \frac{\sigma^2}{n \cdot \varepsilon^2}$$

Der schwache Gesetz der großen Zahlen garantiert übrigens nicht, dass ein $\omega \in \Omega$ existiert, für das die Folge $\left(\frac{S_n(\omega)}{n}\right)_{n \in \mathbb{N}}$ konvergiert – es kann auch sein, dass kein solches $\omega \in \Omega$ existiert.

Wir versuchen nun einen stärkeren Konvergenzbegriff zu finden...

Satz 3.6.3 – Starkes Gesetz der großen Zahlen

Sei $(\mathcal{X}_i)_{i \geq 1}$ eine Folge von **i.i.d. Zufallsvariablen** mit $E|\mathcal{X}_1| < \infty$ (ist also schwächer als $E(\mathcal{X}_1^2) < \infty$). Dann gilt:

$$P(\{\lim_{n \rightarrow \infty} S_n/n = \mu\}) = P(\{\omega \in \Omega : \lim_{n \rightarrow \infty} S_n(\omega)/n = \mu\}) = 1$$

Bemerkung 3.7

1. Die Aussage des **Gesetz des starken Gesetzes** impliziert die des **schwachen** (der Beweis ist wieder sehr aufwändig, wir werden diesen Umstand einfach so hinnehmen).
2. Im **Münzen-Beispiel** von eben gilt:

$$\mu = E(\mathcal{X}_1) = \frac{1}{2}.$$

Also,

$$P(\{\lim_{n \rightarrow \infty} S_n/n = 1/2\}) = 1.$$

Die Intuition von uns war also richtig.

3.7 Der zentrale Grenzwertsatz

Sei $(\mathcal{X}_i)_{i \geq 1}$ eine **i.i.d.** Folge von **Zufallsvariablen** und setze $\mu = E(\mathcal{X}_1)$. Bisher haben wir $\frac{S_n}{n} \rightarrow \mu$ mit einer Wahrscheinlichkeit von 1 betrachtet. Dass also für große n folgendes gilt:

$$S_n \approx n \cdot \mu.$$

Doch wie stark fluktuiert S_n um $n \cdot \mu$?

Wir notieren die **Dichte** und die Verteilungsfunktion der Standardabweichung mit φ und Φ über die Formeln:

$$\begin{aligned} \varphi(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \text{und} \\ \Phi(x) &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \end{aligned}$$

Satz 3.7.1 – Zentraler Grenzwertsatz

Sei $(\mathcal{X}_i)_{i \geq 1}$ eine **i.i.d.** Folge von **Zufallsvariablen** mit $E(\mathcal{X}_1^2) < \infty$ und $\text{Var}(\mathcal{X}_1) > 0$. Setze $\mu = E(\mathcal{X}_1)$ und $\sigma^2 = \text{Var}(\mathcal{X}_1)$. Dann gilt für alle $x \in \mathbb{R}$:

$$P(\{\frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq x\}) \xrightarrow{n \rightarrow \infty} P(x)$$

Es gilt zu beachten, dass für den **Erwartungswert** folgendes gilt:

$$E(\frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}}) = \frac{1}{\sigma \sqrt{n}} E(S_n - n \cdot \mu) = \frac{1}{\sigma \sqrt{n}} (\underbrace{E(S_n)}_{n \cdot \mu} - n \cdot \mu) = 0$$

für die **Varianz** gilt wiederum:

$$\begin{aligned}\text{Var}\left(\frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}}\right) &= \frac{1}{\sigma^2 \cdot n} \text{Var}(S_n - n \cdot \mu) = \frac{1}{\sigma^2 \cdot n} \text{Var}(S_n) \\ &= \frac{n \cdot \sigma^2}{\sigma^2 \cdot n} = 1.\end{aligned}$$

Der *zentrale Grenzwertsatz* besagt nun, dass für „große“ n gilt:

$$\left\langle \frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \right\rangle \approx N(0, 1)$$

also die Verteilung von $\frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}}$ gegen die Verteilung $N(0, 1)$ konvergiert.

Es gilt zu beachten, dass im Falle von $a < b$ die folgende Gleichung gilt:

$$\begin{aligned}P\left(\left\{a \leq \frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq b\right\}\right) &= P\left(\left\{\frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq b\right\}\right) - P\left(\left\{\frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq a\right\}\right) \\ &\xrightarrow{n \rightarrow \infty} P(b) - P(a).\end{aligned}$$

Wir fluktuieren demnach in *Termen von* \sqrt{n} so liegt S_n im Intervall zwischen $n \cdot \mu + a \cdot \sigma \cdot \sqrt{n}$ und $n \cdot \mu + b \cdot \sigma \cdot \sqrt{n}$.

Bemerkung 3.8 – Die Werte der Verteilungsfunktion der Standard-Normalverteilung

Diese Werte sind nicht leicht ausrechenbar, weswegen Tabellen für gewisse Ausprägungen existieren. Für $x < 0$ gibt es den nützlichen Zusammenhang:

$$\forall z \in \mathbb{R} : \Phi(-z) = 1 - \Phi(z)$$

Alle Definitionen

Eine Einleitung	1
1.0.1 ♣ Zufallsexperiment	1
Wahrscheinlichkeiten	3
2.1.1 ♣ Ergebnisraum Ω	3
2.2.1 Ereignis A	4
2.2.2 ♣ Die σ -Algebra	5
2.3.1 Disjunkte Mengen	6
2.3.2 ♣ Wahrscheinlichkeitsmaß	7
2.3.3 ♣ Wahrscheinlichkeitsraum	7
2.4.1 Endlicher Wahrscheinlichkeitsraum	10
2.4.2 ♣ Laplace-Wahrscheinlichkeitsraum	10
2.4.3 Laplace-Experiment	10
2.4.4 ♣ Urnenmodell	13
2.4.5 Fakultät	14
2.4.6 Binomialkoeffizient	14
2.4.7 Bernoulliverteilung	17
2.4.8 ♣ Binomialverteilung	17
2.5.1 Geometrische Reihe	19
2.5.2 Poisson-Verteilung	20
2.5.3 Geometrische Verteilung	22
2.5.4 Diskreter Wahrscheinlichkeitsraum	22
2.6.1 Borel- σ -Algebra	23
2.6.2 Wahrscheinlichkeitsdichte	23
2.6.3 Absolut stetig	24
2.6.4 Stetige Gleichverteilung	25
2.6.5 Exponentialverteilung	27
2.6.6 ♣ Normalverteilung	28
2.7.1 Bedingte Wahrscheinlichkeit	29
2.8.1 Unabhängig	32
2.8.2 Paarweise Unabhängig	33
Zufallsvariablen	35
3.1.1 ♣ Zufallsvariable	35
3.1.2 Die Verteilung einer Zufallsvariable	36
3.1.3 Diskrete Zufallsvariable und Träger-Menge	38
3.1.4 Absolut stetig verteilte Zufallsvariable	39
3.2.1 Verteilungsfunktion	39
3.2.2 Identische Verteilung	42
3.3.1 ♣ Zufallsvektor	44
3.3.2 Gemeinsame Verteilung	44

3.3.3	Die Randverteilung	45
3.3.4	✚ Zweidimensionale-Dichte	47
3.3.5	Gemeinsame Dichte	47
3.4.1	Unabhängige Zufallsvariablen	49
3.5.1	✚ Erwartungswert einer diskreten Zufallsvariable	53
3.5.2	✚ Erwartungswert einer absolutstetigen Zufallsvariable	54
3.5.3	✚ Varianz	59
3.5.4	✚ Kovarianz	62
3.5.5	Kovarianzkoeffizient	65

Alle Sätze

Eine Einleitung	1
Wahrscheinlichkeiten	3
2.2.1 Die σ -Algebren sind abgeschlossen	5
2.3.1 Eigenschaften von Wahrscheinlichkeitsmaßen	8
2.4.1 Die Kardinalität des Kreuzproduktes	12
2.4.2 Anzahl der Elemente beim Zurücklegen ohne Beachtung der Reihenfolge	15
2.5.1 Die Poisson-Approximation	20
2.6.1 Die Dichte ist eindeutig für das Wahrscheinlichkeitsmaß	24
2.6.2 Das absolutstetige Wahrscheinlichkeitsmaß	25
2.7.1 Zwei Sätze bei paarweise disjunkten Mengen	29
Zufallsvariablen	35
3.2.1 Die Verteilungsfunktion einer Zufallsvariablen	41
3.3.1 Gemeinsame- und Randverteilung	45
3.4.1 Anderes Maß für die Unabhängigkeit	49
3.4.2 Der diskrete Fall für die Unabhängigkeit bei Zufallsvariablen	50
3.4.3 Randdichten im absolutstetigen Fall	51
3.5.1 Eigenschaften des Erwartungswertes	57
3.5.2 Transformationssatz	57
3.5.3 Multiplizieren von Zufallsvariablen	58
3.5.4 Formeln für die Varianz	59
3.5.5 Eigenschaften der Varianz	60
3.5.6 Eigenschaften von Kovarianz	64
3.5.7 Eigenschaften des Korrelationskoeffizient	65
3.6.1 Die Tschebycheff-Ungleichung	66
3.6.2 Schwaches Gesetz der großen Zahlen	66
3.6.3 Starkes Gesetz der großen Zahlen	67
3.7.1 Zentraler Grenzwertsatz	67

