KAPITEL 1

Einleitung

1.1 Zufall - Konzepte von Wahrscheinlichkeit und deren Anwendungen

1. Zufallsexperimente: Würfeln, Lebensdauer von elektronischen Bauteilen (Modellannahmen: Die Augenzahlen einer festen Anzahl von Würfen hat die gleiche Wahrscheinlichkeit bzw. Bauteil altert nicht.)

2. Schätzen von Eigenschaften anhand einer Stichprobe: Um die Halbwertzeit eines radioaktiven Isotops zu bestimmen, beobachten wir eine uns bekannte Anzahl von Kernen über eine gegebene Zeit und zählen die Anzahl der Zerfälle in dieser Zeit. Welche Halbwertzeit schätzen wir? Wie "gut" ist unsere Schätzung? Warum können wir überhaupt sinnvollerweise von einer Halbwertzeit sprechen?

3. Messfehler (Modellannahme: Fehler setzt sich aus vielen "unabhängigen" Fehlern zusammen.)

4. Vorhersage/Modellierung: Wechsel- und Aktienkurse (Zeitreihenmodellierungen mit geschätzten Parametern)

5. Testen von Hypothesen: Ist eine neue Behandlungsmethode für eine Krankheit besser als der Status Quo? Wie sicher können wir sein, dass das Ergebnis einer gegebenen Studie nicht rein zufällig zustande kam?

6. subjektive Wahrscheinlichkeiten (Kann die Stärke der Überzeugung von einer Tatsache mit einer Wahrscheinlichkeit gemessen werden. Empirische Bestimmung durch Auswertung von Wettdaten.)

1.2 Mathematisches Modell

Definition 1.1. *Ein **diskreter Wahrscheinlichkeitsraum** ist ein Paar $(\Omega, \mathbf{P})$, wobei*

(i) *Ω eine höchstens abzählbar unendliche[a], nichtleere Menge ist*

(ii) *und $\mathbf{P}$ eine Funktion $\mathbf{P}\colon \mathcal{P}(\Omega) \to [0,1]$ mit folgenden Eigenschaften ist:*

$$\mathbf{P}(\Omega) = 1,$$

$$\sum_{i=1}^{\infty} \mathbf{P}(A_i) = \mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \qquad \text{für paarweise disjunkte } A_1, A_2, \dots \subset \Omega \quad \text{"}\sigma\textbf{-Additivität}\text{"}.$$

Dabei ist $\mathcal{P}(\Omega) = \{A : A \subset \Omega\}$ Potenzmenge von Ω.

[a] d.h., Ω ist endlich oder abzählbar unendlich. In Zukunft schreiben wir auch kurz *abzählbar*.

Die Menge Ω heißt auch **Ergebnis-** oder **Stichprobenraum**. Die Funktion $\mathbf{P}$ nennt man auch **Wahrscheinlichkeitsmaß** oder **Wahrscheinlichkeitsverteilung**. Teilmengen von Ω heißen auch **Ereignisse**. $\mathbf{P}$ weist also jedem Ereignis einen Wert zwischen 0 und 1 zu, die Wahrscheinlichkeit des Ereignisses.

<u>Notation:</u> Wir schreiben $A \subset \Omega$ für zwei Mengen A und Ω für den Fall, dass A eine echte Teilmenge von Ω ist oder dass $A = \Omega$ gilt. Falls wir schreiben wollen, dass A eine echte Teilmengen von Ω ist, drücken wir das durch $A \subsetneq \Omega$ aus.

Bemerkung. *σ-Additivität impliziert auch (endliche) Additivität: $\sum_{i=1}^{n} \mathbf{P}(A_i) = \mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right)$ für $n \geq 1$ und paarweise disjunkte Mengen $A_1, A_2, \dots, A_n \subset \Omega$, da die leere Menge zu allen Mengen disjunkt ist. Somit folgt diese Identität aus der σ-Additivität mit der Wahl $\emptyset = A_{n+1} = A_{n+2} = \dots$, nachdem man $\mathbf{P}(\emptyset) = 0$ gezeigt hat, siehe 1.2 (iv).*

<u>Interpretation:</u>
Ω als Menge möglicher (Mess-)Ergebnisse bzw. Ausgänge des Zufallsexperiments.
$\mathbf{P}(A)$ als Wahrscheinlichkeit, dass ein Ergebnis aus dem Ereignis A eintritt.
Ereignisse $A = \{\omega\}$ mit $\#A = 1$ heißen auch **Elementarereignisse**.

Beispiel 1 (Würfel)**.** *$\Omega = \{1, \dots, 6\}$, $\mathbf{P}(\{\omega\}) = \frac{1}{6}$ für jedes $\omega \in \Omega$: (diskrete) “Gleichverteilung”.*
Das Ereignis $A = \{2, 4, 6\}$ kann zum Beispiel zu “gerade Augenzahl” ausformuliert werden. Dann haben wir dank der Additivität

$$\mathbf{P}(A) = \mathbf{P}(\{2\}) + \mathbf{P}(\{4\}) + \mathbf{P}(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Bei einer Gleichverteilung auf Ω (mit $\#\Omega < \infty$) gilt stets $\mathbf{P}(\{\omega\}) = \frac{1}{\#\Omega}$ für alle $\omega \in \Omega$, und für jedes Ereignis $A \subset \Omega$ gilt $\mathbf{P}(A) = \frac{\#A}{\#\Omega}$.

Lemma 1.2 (Elementare Eigenschaften). *In einem diskreten Wahrscheinlichkeitsraum* $(\Omega, \mathbf{P})$ *gelten folgende Aussagen:*

(i) $\mathbf{P}$ *ist durch* $\mathbf{P}(\{\omega\})$ *für* $\omega \in \Omega$ *festgelegt. Die Angabe der Wahrscheinlichkeit von Elementarereignissen genügt also, um* $\mathbf{P}$ *zu definieren.*

(ii) $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$ *für beliebige* $A \subset \Omega$. *Dabei ist* $A^c = \Omega \backslash A$ *das Komplement von* A *(**Gegenereignis zu** A).*

(iii) $A \subset B \subset \Omega \Rightarrow \mathbf{P}(A) \leq \mathbf{P}(B)$ *(**Monotonie**).*

(iv) $\mathbf{P}(\emptyset) = 0$ *($\emptyset$ als **unmögliches Ereignis**)*

(v) $\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbf{P}(A_i)$ *für beliebige Mengen* $A_i \subset \Omega$, $i = 1, 2, 3, \ldots$ *(σ-**Subadditivität**).*

Beweis. Die Beweise dieser elementaren Rechenregeln sind Übungen. □

Definition 1.3. *Seien* $(\Omega, \mathbf{P})$ *ein diskreter Wahrscheinlichkeitsraum und* Ω' *eine abzählbare Menge. Jede Funktion* $X : \Omega \to \Omega'$ *heißt (Ω'-wertige) **Zufallsvariable** (auf dem Wahrscheinlichkeitsraum* $(\Omega, \mathbf{P})$*).*

Das durch $\mathbf{P}^X(\{\omega'\}) = \mathbf{P}(\{\omega : X(\omega) = \omega'\}) = \mathbf{P}(X^{-1}(\{\omega'\}))$ *auf* Ω' *definierte Wahrscheinlichkeitsmaß heißt **Verteilung** von* X *oder auch **Bildmaß** von* $\mathbf{P}$ *unter* X.

Man schreibt auch $\mathcal{L}(X|\mathbf{P})$, $\mathcal{L}^{\mathbf{P}}(X)$ oder $\mathcal{L}(X)$ für $\mathbf{P}^X$ (aus dem Englischen: *law* für Verteilung(sgesetz).

Oft werden wir es mit reellwertigen Zufallsvariablen zu tun haben. Dass das nicht im Widerspruch zu unserer Definition steht, liegt an einem sprachlichen Trick. Wir betrachten in diesem Fall nur die Bildmenge $X(\Omega)$ als Ω', nicht den gesamten Bildraum $\mathbb{R}$. Da Ω abzählbar ist, gilt das auch für $\Omega' = X(\Omega)$.

Wir haben in der Definition kurzerhand behauptet, dass $\mathbf{P}^X$ ein Wahrscheinlichkeitsmaß auf Ω' im Sinne von Definition 1.1 ist. Überzeugen Sie sich davon im Rahmen einer Übung.

Interpretation: X wird als ein zufälliges Element von Ω' verstanden. Man beachte, dass in der Regel verschiedene Ergebnisse ω den gleichen Wert $X(\omega) \in \Omega'$ liefern können. Oft wird Ω nicht angegeben und spezifiziert, und wir interessieren uns nur für die möglichen Werte von X, nicht dafür, welches $\omega \in \Omega$ zu diesem Wert geführt hat. Man schreibt auch $\mathbf{P}(X \in A)$ statt $\mathbf{P}(\{\omega : X(\omega) \in A\}) = \mathbf{P}^X(A)$ und nennt dies die "Wahrscheinlichkeit, dass X in A liegt".

Beispiel 2. *X Anzahl der Würfe mit Augenzahl 6 bei n-maligem Würfeln. Man kann hier* $\Omega' = \{0, 1, \ldots, n\}$, $\Omega = \{1, \ldots, 6\}^n$ *und* $\mathbf{P}$ *als Gleichverteilung (auf Ω) ansetzen:*

$$\mathbf{P}(\{\omega\}) = \frac{1}{\#\Omega} = \left(\frac{1}{6}\right)^n \text{ für } \omega = (\omega_1, \ldots, \omega_n) \in \Omega.$$

Dann ist $X(\omega) = \#\{i \in \{1, \ldots, n\} : \omega_i = 6\}$ *und als Bildmaß unter* X *erhält man:*

$$\begin{aligned}
\mathbf{P}^X(\{k\}) &= \mathbf{P}(X = k) \\
&= \mathbf{P}(\{\omega : X(\omega) = k\}) \\
&= \frac{\#\{\omega \in \Omega : \#\{i : \omega_i = 6\} = k\}}{\#\Omega} \\
&= \frac{\binom{n}{k} 5^{n-k}}{6^n} \\
&= \binom{n}{k} \left(\frac{5}{6}\right)^{n-k} \left(\frac{1}{6}\right)^k
\end{aligned}$$

für $k = 0, \ldots, n$.

Definition 1.4. *Ein Wahrscheinlichkeitsmaß auf* $\Omega = \{0, 1, \ldots, n\}$ *mit*

$$\mathbf{P}(\{k\}) = \binom{n}{k}(1-p)^{n-k} p^k \qquad \text{für } k = 0, 1, \ldots, n$$

heißt ***Binomialverteilung*** $\mathrm{Bin}(n, p)$ *zum Parameter* $(n, p) \in \mathbb{N} \times (0, 1)$. $\mathrm{Bin}(1, p)$ *heißt auch* ***Bernoulliverteilung*** *zum Parameter* p.

Die Verteilung ergibt auch in den Grenzfällen $p \in \{0, 1\}$ Sinn, verliert dann aber ihren zufälligen Charakter.

Bemerkung. *Wenn zwei (oder mehr) Zufallsvariablen auf demselben Wahrscheinlichkeitsraum definiert sind, kann auch Ereignissen, die durch beide Zufallsvariablen definiert sind, eine Wahrscheinlichkeit zugeordnet werden.*

Beispiel 3. X_1 *Anzahl der Würfe mit Augenzahl 6 bei n-maligem Würfeln und* X_2 *Anzahl der Würfe mit Augenzahl 5 bei denselben Würfen. Dann ist* $\mathbf{P}(X_1 > X_2)$ *"Wahrscheinlichkeit, mehr Sechsen als Fünfen zu würfeln" durch* $\mathbf{P}(\{\omega : X_1(\omega) > X_2(\omega)\})$ *wohldefiniert.*
Ganz formal: Wählen $\Omega = \{1, \ldots, 6\}^n$ *und* $\mathbf{P}(\{\omega\}) = \frac{1}{\#\Omega} = \left(\frac{1}{6}\right)^n$ *(Gleichverteilung auf* Ω*), wie im letzten Beispiel. Setzen jetzt:*

$$\begin{aligned}
X_1(\omega) &= \#\{i : \omega_i = 6\}, \\
X_2(\omega) &= \#\{i : \omega_i = 5\}.
\end{aligned}$$

Dann kann $\mathbf{P}(X_1 > X_2)$ *durch* $\mathbf{P}(\{\omega : X_1(\omega) > X_2(\omega)\})$ *definiert werden.*
Interpretation: X_1, X_2 *sind "zufällige Werte", denen derselbe Zufallsgenerator zugrunde liegt (mit verschiedenen Transformationen).*

Definition 1.5. *Seien* $X_1, \ldots, X_n$ *Zufallsvariablen auf demselben Wahrscheinlichkeitsraum* $(\Omega, \mathbf{P})$. *Dann ist auch* $X = (X_1, \ldots, X_n)$ *Zufallsvariable auf* $(\Omega, \mathbf{P})$. $\mathcal{L}(X)$ *heißt auch* ***gemeinsame Verteilung*** *von* $X_1, \ldots, X_n$ *und* $\mathcal{L}(X_1), \ldots, \mathcal{L}(X_n)$ *heißen* ***Randverteilungen*** *von* X.

Wir beachten hier, dass die X_i völlig verschiedene Wertebereiche $\Omega'_1, \ldots, \Omega'_n$ haben können. Die Zufallsvariable X ist dann $\Omega'_1 \times \cdots \times \Omega'_n$-wertig.

Beispiel 3, fortgesetzt. *Im letzten Beispiel ist die Randverteilung sowohl von X_1 als auch von X_2 eine Binomialverteilung* $\mathrm{Bin}(n,p)$ *mit* $p = \frac{1}{6}$. *Für die gemeinsame Verteilung* $P^{(X_1,X_2)}$ *gilt*

$$\begin{aligned}
\mathbf{P}^{(X_1,X_2)}(\{(k_1,k_2)\}) &= \mathbf{P}((X_1,X_2) = (k_1,k_2)) \\
&= \mathbf{P}(X_1 = k_1, X_2 = k_2) \\
&= \mathbf{P}(\{\omega : X_1(\omega) = k_1, X_2(\omega) = k_2\}) \\
&= \begin{pmatrix} & n & \\ k_1 & k_2 & n-k_1-k_2 \end{pmatrix} \left(\frac{1}{6}\right)^{k_1} \left(\frac{1}{6}\right)^{k_2} \left(\frac{4}{6}\right)^{n-k_1-k_2}.
\end{aligned}$$

Hier ist für $n, r \in \mathbb{N}$, $k_1, \ldots, k_r \in \mathbb{N}_0$ mit $k_1 + \ldots + k_r = n$

$$\begin{pmatrix} & n & & \\ k_1 & k_2 & \ldots & k_r \end{pmatrix} = \frac{n!}{k_1! \cdot \ldots \cdot k_r!}$$

der Multinomialkoeffizient. Der bekannte Binomialkoeffizient ergibt sich einfach durch $r = 2$, $k_1 = k$ und $k_2 = n - k$.
Wir bemerken noch, dass $X = (X_1, X_2)$ nicht etwa beliebige Werte aus $\Omega'_1 \times \Omega'_2 = \{0, 1, \ldots, n\}^2$ annehmen kann. Vielmehr muss für obigen Formel $k_1 + k_2 \leq n$ gelten (nur dann ist der Multinomialkoeffizient wie angegeben definiert). Für Paare (k_1, k_2) mit $k_1 + k_2 > n$ ist $\mathbf{P}^X(\{(k_1,k_2)\}) = 0$. Wir können also als Wertebereich von X statt $\Omega'_1 \times \Omega'_2$ auch die kleinere Menge

$$\Omega' = \{(k_1,k_2) : k_1, k_2 \in \mathbb{N}_0, k_1 + k_2 \leq n\}$$

angeben.

Definition 1.6. *Seien $n, r \in \mathbb{N}$ und $p_1, \ldots, p_r \in [0,1]$ mit $p_1 + \ldots + p_r = 1$. Ein Wahrscheinlichkeitsmaß auf*

$$\Omega = \{k = (k_1, \ldots, k_r) : k_1, \ldots, k_r \in \mathbb{N}_0, k_1 + \ldots + k_r = n\}$$

mit

$$\mathbf{P}(\{k\}) = \begin{pmatrix} & n & & \\ k_1 & k_2 & \ldots & k_r \end{pmatrix} p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}$$

heißt ***Multinomialverteilung*** $\mathrm{Mult}(n, p_1, \ldots, p_r)$ *zum Parameter* $(n, p_1, \ldots, p_r)$.

Bemerkung. (1) *Für $r = 2$ ergibt sich die Binomialverteilung* $\mathrm{Mult}(n, p, 1-p) = \mathrm{Bin}(n, p)$.
(2) *In allen obigen Beispielen wurden Wahrscheinlichkeitsmodelle dadurch konstruiert, dass zunächst ein endliches Ω gewählt wurde und* $\mathbf{P}$ *als Gleichverteilung auf Ω angesetzt wurde:*

$$\mathbf{P}(\{\omega\}) = \frac{1}{\#\Omega}.$$

Die Gleichverteilung war jeweils motiviert durch Symmetriebetrachtungen:

$$\mathbf{P}(\{\omega\}) = \mathbf{P}(\{\omega'\}) \text{ für alle } \omega,\, \omega' \in \Omega.$$

Die Konstruktion von Wahrscheinlichkeitsmodellen durch eine Wahl von Ω, versehen mit einer Gleichverteilung, heißt auch ***Laplace-Ansatz****.*

Beispiel 4. *Für den Laplace-Ansatz ist die Wahl des geeigneten Grundraums Ω wichtig. Sonst gerät man schnell zu falschen Schlüssen. Betrachten wir das elementare Zufallsexperiment des Wurfs zweier fairer Münzen. Wir könnten nun annehmen, die möglichen Ergebnisse seien zweimal Kopf (K, K), eine Zahl und ein Kopf (Z, K) und zweimal Zahl (Z, Z). Nach Laplace-Ansatz wäre die Wahrscheinlichkeit für jedes dieser Ergebnisse dann $\frac{1}{3}$.*
Tatsächlich erhält man bei wiederholter Durchführung dieses Experiments aber zweimal Kopf und zweimal Zahl in je einem Viertel der Versuche, während in etwa der Hälfte aller Versuche ein Kopf und eine Zahl auftreten. Wie ist das zu erklären?
Die Antwort liegt in der richtigen Wahl von Ω. Bei genauerem Hinsehen finden wir, dass unser Experiment auch durchgeführt werden kann, indem wir zuerst die eine und danach die andere Münze werfen. Das führt uns dazu, die Münzen mit Nummern zu versehen: Münze 1 und Münze 2 (die Münzen sind nun unterscheidbar). Die möglichen Ergebnisse sind dann:

- (K, K)*: beide Münzen zeigen Kopf*
- (K, Z)*: Münze 1 zeigt Kopf, Münze 2 zeigt Zahl*
- (Z, K)*: Münze 1 zeigt Zahl, Münze 2 zeigt Kopf*
- (Z, Z)*: beide Münzen zeigen Zahl.*

Nun hat tatsächlich jedes Elementarereignis die Wahrscheinlichkeit $\frac{1}{4}$, aber der Ausgang "ein Kopf, eine Zahl" wird durch zwei Elementarereignisse realisiert. Dessen Wahrscheinlichkeit ist also $\frac{2}{4} = \frac{1}{2}$ wie beobachtet.

In Zukunft werden wir übrigens in der Regel die Abkürzungen Z und K weitgehend vermeiden und stattdessen K durch 0 und Z durch 1 ersetzen. Dann ist der Grundraum für den n-fachen Münzwurf der Raum $\{0, 1\}^n$ der n-Tupel aus Nullen und Einsen. Die Anzahl der geworfenen Zahlen beim Ergebnis $\omega = (\omega_1, \dots, \omega_n)$ wäre dann einfach $\sum_{i=1}^n \omega_i$.

Ein weiteres Beispiel:

Beispiel 5 (Hypergeometrische Verteilung)**.** *Gegeben ist eine Urne (ein Lieblingsspielzeug in der Statistik, gern auch nur als Gedankenexperiment) mit S schwarzen und W weißen ("gut gemischten") Kugeln. Ziehe n Kugeln (ohne Zurücklegen). Sei X die Anzahl der gezogenen schwarzen Kugeln.*

Frage: Was ist die Verteilung der Zufallsvariable X?

Ansatz: Gebe den $S+W$ Kugeln die Namen 1 bis S für die schwarzen Kugeln und $S+1$ bis $S+W$ für die weißen Kugeln. Mögliche Mengen ω gezogener Kugeln sind dann Elemente von

$$\Omega = \{\omega : \omega \subset \{1, \ldots, S+W\}, \#\omega = n\},$$

bestehend aus allen n-elementigen Teilmengen von $\{1, \ldots, S+W\}$. Sei $\mathbf{P}$ Gleichverteilung auf Ω (jede n-elementige Teilmenge ω hat die gleiche Chance gezogen zu werden).

Berechne nun Verteilung von

$$X(\omega) = \#[\omega \cap \{1, \ldots, S\}] \text{ Anzahl der gezogenen schwarzen Kugeln.}$$

Es gilt

$$\mathbf{P}^X(\{k\}) = \frac{\#X^{-1}(\{k\})}{\#\Omega} = \frac{\binom{S}{k}\binom{W}{n-k}}{\binom{S+W}{n}}.$$

Definition 1.7. *Ein Wahrscheinlichkeitsmaß auf $\Omega = \{\max\{0, n-W\}, \ldots, \min\{S, n\}\}$ mit*

$$\mathbf{P}(\{k\}) = \frac{\binom{S}{k}\binom{W}{n-k}}{\binom{S+W}{n}}.$$

*heißt **Hypergeometrische Verteilung** $\mathrm{Hypergeom}(n, S+W, S)$ zum Parameter $(n, S+W, S)$.*

Beispiel 6 (Tierzählungen)**.** *Fange zunächst S Tiere aus einer Gesamtheit von $S+W$ Tieren (etwa Fische in einem Gewässer), markiere sie und setze sie wieder aus. Fange dann wieder n Tiere. Hier ist X die Anzahl der markierten Tiere.*
Wann ist hier Laplace-Ansatz mit der Wahl von Ω aus dem letzten Beispiel gerechtfertigt?

Beispiel 7 (Lotto)**.** *Beim Lotto "6 aus 49"beträgt die Wahrscheinlichkeit für vier Richtige*

$$\mathrm{Hypergeom}(6, 49, 6)(\{4\}) = \frac{\binom{6}{4}\binom{49-6}{6-4}}{\binom{49}{6}} = \frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} \approx 9,7 \cdot 10^{-4}.$$

Wir leiten nun noch eine Wartezeitverteilung aus der Binomialverteilung her. Es werde wiederholt zufällig die Zahl 0 oder 1 gezogen. Die Wahrscheinlichkeit für eine 1 sei p mit $0 < p < 1$. Wähle $r \in \mathbb{N}$.
Für $k \geq 0$ sei p_k die Wahrscheinlichkeit, dass bei der $(r+k)$-ten Ziehung die r-te Eins gezogen wird. Dann gilt:

$$\begin{aligned}
p_k &= \text{Wahrscheinlichkeit, dass in } r+k-1 \text{ Ziehungen } r-1 \text{ Einsen gezogen werden} \\
&\quad \times \text{ Wahrscheinlichkeit, dass im } (r+k)\text{.ten Zug eine 1 gezogen wird} \\
&= \binom{r+k-1}{k} p^{r-1}(1-p)^k p \\
&= \binom{r+k-1}{k} p^r (1-p)^k.
\end{aligned}$$

Insbesondere mit $r = 1$ ist dies die Wahrscheinlichkeit, dass beim $(k+1)$-ten Mal zum ersten Mal eine Eins gezogen wird.
Man kann zeigen, dass

$$\sum_{k=0}^{\infty} \binom{r+k-1}{k} p^r (1-p)^k = 1,$$

nicht nur für $r \in \mathbb{N}$, sondern auch für $r > 0$, $r \in \mathbb{R}$, wobei der verallgemeinerte Binomialkoeffizient $\binom{r+k-1}{k}$ wie für $r \in \mathbb{N}$ definiert ist als:

$$\binom{r+k-1}{k} = \frac{(r+k-1)\cdot \dots \cdot r}{k!}.$$

Wenn wir uns das Ziehen von 1 als Erfolg vorstellen, so ist k in der obigen Situation die Zahl der Misserfolge, bevor der $r$te Erfolg eintritt.

Definition 1.8. *Sei $r > 0, 0 < p < 1$. Ein Wahrscheinlichkeitsmaß auf $\Omega = \mathbb{N}_0$ mit*

$$\mathbf{P}(\{k\}) = \binom{r+k-1}{k} p^r (1-p)^k$$

*für $k \in \mathbb{N}_0$ heißt **negative Binomialverteilung** $\mathrm{Bin}^-(r,p)$ zum Parameter (r,p). Für $r = 1$ heißt die Verteilung auch geometrische Verteilung $\mathrm{Geom}(p)$ zum Parameter p. Dann gilt*

$$\mathbf{P}(\{k\}) = p(1-p)^k$$

für $k \in \mathbb{N}_0$.

Indikatorfunktionen. Spezielle Zufallsvariablen, die wir in Zukunft oft betrachten werden, sind die Indikatorfunktionen. Ist $(\Omega, \mathbf{P})$ nämlich ein diskreter Wahrscheinlichkeitsraum, so definiert für jede Teilmenge $A \subset \Omega$

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{für } \omega \in A \\ 0 & \text{für } \omega \notin A \end{cases}$$

eine Zufallsvariable $\mathbf{1}_A : \Omega \to \{0,1\} = \Omega'$. Diese nimmt den Wert 1 mit der Wahrscheinlichkeit

$$\mathbf{P}^{\mathbf{1}_A}(\{1\}) = \mathbf{P}(\mathbf{1}_A = 1) = \mathbf{P}(\mathbf{1}_A^{-1}(\{1\})) = \mathbf{P}(A),$$

den Wert 0 mit $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$. Sie ist also $\mathrm{Bin}(1,p)$-verteilt mit $p = \mathbf{P}(A)$.
Für Indikatorfunktionen gelten einige sehr schöne Rechenregeln, die zu Übungszwecken gern ausdrücklich nachgerechnet werden dürfen:

$$\mathbf{1}_{A\cap B} = \mathbf{1}_A \mathbf{1}_B = \mathbf{1}_A \wedge \mathbf{1}_B (:= \min\{\mathbf{1}_A, \mathbf{1}_B\}) \qquad \text{für } A, B \subset \Omega \tag{1.1}$$

$$\mathbf{1}_{A\cup B} = \mathbf{1}_A \vee \mathbf{1}_B (:= \max\{\mathbf{1}_A, \mathbf{1}_B\}) \qquad \text{für } A, B \subset \Omega \tag{1.2}$$

$$\mathbf{1}_{A\Delta B} = \mathbf{1}_A + \mathbf{1}_B \pmod 2 \qquad \text{für } A, B \subset \Omega \tag{1.3}$$

$$\mathbf{1}_{A^c} = 1 - \mathbf{1}_A \qquad \text{für } A \subset \Omega \tag{1.4}$$

Tatsächlich gilt nämlich zum Beispiel für beliebiges $\omega \in \Omega$:

$$\mathbf{1}_{A\cap B}(\omega) = 1 \iff \omega \in A \cap B \iff \omega \in A \wedge \omega \in B \iff (\mathbf{1}_A(\omega) = 1) \wedge (\mathbf{1}_B(\omega) = 1).$$

Die anderen Aussagen können Sie ganz analog nachrechnen.

Oft werden die Mengen A, deren Indikatorfunktion wir betrachten ihrerseits durch andere Zufallsvariablen bestimmt.

Beispiel 8. *Wir betrachten als Zufallsvariable X die Anzahl der 6er beim dreimaligen Werfen eines fairen Würfels. Wir wissen, dass X $\mathrm{Bin}(3, \frac{1}{6})$-verteilt ist. Nehmen wir nun an, wir bauen das ganze zu einem Spiel aus, bei dem wir gewinnen und 10 Euro ausgezahlt bekommen, wenn mehr als eine 6 fällt. Fällt bei den drei Würfen höchstens eine 6, so gewinnt unsere Gegenpartei und wir zahlen dieser 2 Euro. Wir wollen die Zufallsvariable G beschreiben, die unseren Gewinn darstellt. Als Funktion der bereits gegebenen Zufallsvariablen X ist das leicht:*

$$G(\omega) = g(X(\omega)) = \begin{cases} 10 & , \textit{falls } X(\omega) > 1 \\ -2 & , \textit{falls } X(\omega) \leq 1 \end{cases}$$

Nun können wir $A = X^{-1}(\{2,3\}) = \{X > 1\}$ setzen und unter Verwendung von $A^c = \{X \leq 1\}$ die Zufallsvariable G als

$$G = 10\,\mathbf{1}_A - 2\,\mathbf{1}_B = 10\,\mathbf{1}_{\{X>1\}} - 2\,\mathbf{1}_{\{X\leq 1\}}$$

schreiben. Oft werden die Mengenklammern sogar weggelassen, und man schreibt kurz

$$G = 10\,\mathbf{1}_{X>1} - 2\,\mathbf{1}_{X\leq 1}.$$

Damit können wir auch $\mathbf{P}(G = 10)$ berechnen:

$$\begin{aligned}\mathbf{P}(G = 10) &= \mathbf{P}(X > 1) = \mathbf{P}(X = 2) + \mathbf{P}(X = 3) \\ &= \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 + \binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^0 \\ &= 3\frac{5}{216} + \frac{1}{216} = \frac{2}{27} \approx 0.0741.\end{aligned}$$

Man beachte, dass uns das gelungen ist, ohne den zugrunde liegenden Wahrscheinlichkeitsraum $(\Omega, \mathbf{P})$ zu untersuchen. Wir haben lediglich benutzt, dass X binomialverteilt ist.
Selbstverständlich können wir in diesem einfachen Beispiel auch die Menge $A = \{G = 10\} = \{X > 1\}$, deren Indikatorfunktion wir benutzt haben, explizit angeben:

$$\begin{aligned} A &= \{(6,6,6)\} \cup \{(a,6,6) : a \in \{1,\dots,5\}\} \cup \{(6,a,6) : a \in \{1,\dots,5\}\} \\ &\quad \cup \{(6,6,a) : a \in \{1,\dots,5\}\}.\end{aligned}$$

Wir haben also $\#A = 1+5+5+5 = 16$ und $\#\Omega = 6^3$, also $\mathbf{P}(A) = \frac{16}{216} = \frac{2}{27}$.