Statistik II für Studierende der Soziologie und Nebenfachstudierende

Thomas Augustin

SoSe 2014

Besonderer Dank gilt den Kolleg(innen) Prof. Dr. Thomas Kneib, Prof. Dr. Carolin Strobl und Prof. Dr. Helmut Küchenhoff, die im Rahmen ihrer Vorlesungen im SoSe 08, SoSe 09 bzw. SoSe2012/2013 das ursprüngliche Material weiterentwickelt haben.

0.1 Einleitung: Zum Gegenstand

Deskriptive vs. induktive Statistik

Deskriptive Statistik (Statistik I):

- Statistische Beschreibung einer Gesamtheit (Grundgesamtheit oder Stichprobe).
- Keine Verallgemeinerung von einer Stichprobe auf die zugehörige Grundgesamtheit angestrebt.

Induktive Statistik (Statistik II):

- Induktion: Schluss vom Teil auf das Ganze, von vielen Einzelbeobachtungen auf allgemeine Gesetze
- Speziell: Schluss von einer Stichprobe auf Eigenschaften der Grundgesamtheit („*Inferenz*")
- Stichprobe nur Mittel zum Zweck, um Informationen über die Eigenschaften der Grundgesamtheit zu gewinnen.

Typisches Beispiel: Wahlumfrage

- Grundgesamtheit: Alle Wahlberechtigten bei der nächsten Bundestagswahl.
- Stichprobe: 1000 „repräsentativ ausgewählte“ Wahlberechtigte
- Gesucht: Information über *alle* Wahlberechtigten, also die *Grundgesamtheit*
- Nachwahlbefragung 2013 (in %) (ARD/Infratest-dimap, 18 Uhr, `http://www.wahlrecht.de/news/2013/bundestagswahl-2013.html`)

CDU/CSU	SPD	FDP	Linke	Grüne	AfD	Piraten	Sonstige
42,0	26,0	4,7	8,5	8,0	4,9	2,5	3,4

- Nachwahlbefragung 2013 (in %) (ARD/Infratest-dimap, 22 Uhr, `http://www.wahlrecht.de/news/2013/bundestagswahl-2013.html`)

CDU/CSU	SPD	FDP	Linke	Grüne	AfD	Piraten	Sonstige
41,7	25,6	4,7	8,6	8,3	4,8	2,2	4,1

- Wahlergebnisse 2013 (in %)

CDU/CSU	SPD	FDP	Linke	Grüne	AfD	Piraten	Sonstige
41,5	25,7	4,8	8,6	8,4	4,7	2,2	1,8

- Hier: aktuelle Sonntagsfrage (3.4.2014) (in %)

CDU/CSU	SPD	FDP	Linke	Grüne	AfD	Piraten	Sonstige
41	26	4	9	10	5	/	5

Fragestellungen der induktiven Statistik

1. **Punktschätzung:**

   - Zum Beispiel: Wie groß ist der Anteil der schwarz-gelb-Wähler unter allen Wahlberechtigten?
   - Wie erhält man aus der Stichprobe gute Schätzwerte für Charakteristika („Parameter“) der Grundgesamtheit?
   - Wann ist ein Schätzverfahren gut/besser als ein anderes?

2. **Bereichsschätzung:**

   - Typischerweise stimmt der Punktschätzer nicht mit dem wahren Wert überein.
   - Realistischer: Gib einen Bereich an, „der den wahren Anteil der schwarz-gelb-Wähler mit hoher Wahrscheinlichkeit überdeckt“.
   - Ungenauere Aussagen, dafür aber zuverlässiger.

3. Hypothesentests:

- Überprüfe aus substanzwissenschaftlicher Theorie abgeleitete Hypothesen über die Grundgesamtheit anhand der Daten.
- Zum Beispiel: Verdienen Männer wirklich mehr als Frauen?
- Ist der Unterschied „signifikant“= „überzufällig“, d.h. so groß, dass er nur mit sehr kleiner Wahrscheinlichkeit zufällig entsteht.

Bsp. 0.1. *Mineralwasserstudie*

Studie in Zusammenarbeit(Statistisches Beratungslabor (Stablab), Ltg: Prof. Küchenhoff) mit Prof. Adam (LMU)
Fragestellung: Schmeckt mit Sauerstoff angereichertes Mineralwasser besser als gewöhnliches Mineralwasser ?

– DoppelBlindstudie

– KontrollGruppe: zweimal das gleiche Wasser ohne O_2

– VerumGruppe: Beim zweiten Mal mit O_2 angereichertes Mineralwasser

Ergebnis (Clausnitzer et al., 2004):

Placebo: 76% gaben an, dass das zweite Wasser anders schmeckt
Verum : 89 % gaben an, dass das zweite Wasser anders schmeckt

Signifikanter Effekt (überzufälliger Effekt) $\rightarrow$ Zulassung von *Adelholzener O2 Active*

4. Regressionsmodelle incl. Varianzanalyse:

- Modelle zur Beschreibung des Einflusses von Variablen
- Zum Beispiel: Welche persönlichen Merkmale beeinflussen die Wahlentscheidung besonders stark?
- Zum Beispiel: Wie hängt die Lebenszufriedenheit vom Alter ab ?
- Zentrales Problem der induktiven Statistik: *Jeder* Induktionsschluss ist potentiell fehlerbehaftet
- Beispielsweise ist der wahre Anteil der Wähler einer Partei in der Grundgesamtheit nicht exakt mithilfe der Stichprobe vorhersagbar.
- Entscheidende Idee: Kontrolle des Fehlers mit Hilfe der Wahrscheinlichkeitsrechnung.
  - Stichprobenziehung *zufällig*
  - Verwende Methoden der Wahrscheinlichkeitsrechnung zur Quantifizierung/Kontrolle des Fehlers

- Verwende statistische Modelle mit zufälligen Komponenten
- „Gesetze“ bzw. Aussagen enthalten stochastischen Aspekt

Inferenzfehler

- Zentrales Problem der induktiven Statistik:

  Jeder Induktionsschluss ist potentiell fehlerbehaftet:

- Beispielsweise ist der wahre Anteil der Wähler einer Partei in der Grundgesamtheit nicht exakt mithilfe der Stichprobe ermittelbar.
- Entscheidende Idee: Kontrolle des Fehlers mit Hilfe der Wahrscheinlichkeitsrechnung.
- Vorgehen:
  - Stichprobenziehung *zufällig*
  - Verwende Methoden der Wahrscheinlichkeitsrechnung zur Quantifizierung/Kontrolle des Fehlers

    $\longrightarrow$ Kap. 1: Wahrscheinlichkeitsrechnung

    $\longrightarrow$ Kap. 2: Wie nutzt man Wahrscheinlichkeitsüberlegungen für die Statistik? Induktive Statistik

Wahrscheinlichkeitsrechnung

- Ziel ist auch die Vermittlung eines „Gefühls“ für Wahrscheinlichkeiten.
- Gerade für Sozialwissenschaftler ist probabilistisches Denken (d.h. das Denken in Wahrscheinlichkeiten) unerlässlich!
- Zunächst grundlegende Rechenregeln für Wahrscheinlichkeiten, und wichtige Modelle, dann auch spezielle Regeln für große Stichproben (Befragungen!) Oft einfacher, da sich gewisse Regelmäßigkeiten einstellen: Insbesondere nähert sich die relative Häufigkeit eines Ereignisses unter gewissen Voraussetzungen seiner Wahrscheinlichkeit an $\rightarrow$ auch wichtig für Interpretation.
  Dem so genannten Hauptsatz der Statistik folgend „wächst mit dem Stichprobenumfang in gewisser Weise die Repräsentativität der Stichprobe“
- Bei naiver Herangehensweise (ohne Wahrscheinlichkeitsrechnung) kann man sich leicht täuschen. Z.B. bewerten sowohl medizinische Experten als auch Laien Risiken oft falsch.

1 Wahrscheinlichkeitsrechnung

1.1 Mengen und elementare Mengenoperationen

Definition 1.1.

Eine $Menge$ ist eine Zusammenfassung verschiedener Objekte zu einem Ganzen. Die einzelnen Objekte einer Menge werden $Elemente$ genannt.

Mengen werden üblicherweise mit Großbuchstaben bezeichnet, z.B. A, B, C, Ω, . . .

Mengen werden später benutzt, um den Ausgang von Zufallsexperimenten zu beschreiben.

Bsp. 1.2.

- $A = \{$Hörer einer Vorlesung$\}$.
- $A = \{1, 2, 3, 4, 5, 6\}$ (Menge der Ergebnisse eines Würfelwurfs).

  Die Reihenfolge der Aufzählung spielt (im Gegensatz zu Tupeln) keine Rolle:

  $$\{1, 2, 3, 4, 5, 6\} \;=\; \{1, 3, 5, 2, 4, 6\}$$

  Jedes Element wird nur einmal genannt.
- $B = \{K, Z\}$ (Menge der Ergebnisse eines Münzwurfs, K=Kopf, Z=Zahl).
- Charakterisierung von Mengen mit einer gewissen Eigenschaft:

  $$\{1, 2, 3, \ldots, 10\} = \{x \,|\, x \text{ ist eine natürliche Zahl } \leq 10\}$$

  Die Menge aller x mit der Eigenschaft „x ist eine natürliche Zahl ≤ 10“.

Standardmengen:

$\mathbb{N}$	$= \{1, 2, 3, \ldots\}$	: Menge der natürlichen Zahlen,
$\mathbb{N}_0$	$= \{0, 1, 2, 3, \ldots\}$	: Menge der natürlichen Zahlen inklusive 0,
$\mathbb{Z}$	$= \{0, \pm 1, \pm 2, \ldots\}$	: Menge der ganzen Zahlen,
$\mathbb{R}$	$= (-\infty, \infty)$	: Menge der reellen Zahlen,
$\emptyset$		: leere Menge.

Grundlegende Begriffe der Mengenlehre: Illustration anhand der Mengen

$$\begin{aligned} \Omega &= \{\text{CDU/CSU, SPD, FDP, Grüne, Linke, Sonstige}\} \\ A &= \{\text{CDU/CSU, SPD, FDP, Grüne}\} \\ B &= \{\text{CDU/CSU, SPD, FDP}\} \\ C &= \{\text{SPD, FDP, Grüne}\} \end{aligned}$$

- **Elementeigenschaft:**

  x ist Element der Menge A: $x \in A$

  x ist nicht Element der Menge A: $x \notin A$

- **Teilmengen:** A ist Teilmenge von B, in Zeichen $A \subset B$, wenn jedes Element von A auch in B ist.

Veranschaulichung von Mengen im Venn-Diagramm:

Für jede Menge A gilt:

$\emptyset \subset A$ Die leere Menge ist Teilmenge jeder Menge, denn jedes Element in $\emptyset$ ist in A
(Aussagen über den leeren Quantor sind wahr)

$A \subset A$ d.h. „ $\subset$ “ enthält implizit „ $=$ “,
deshalb in Literatur manchmal auch $\subseteq$ statt $\subset$

- **Schnittmenge:** Die Schnittmenge $A \cap B$ ist die Menge aller Elemente, die sowohl in A als auch in B enthalten sind:

$$A \cap B = \{x | x \in A \text{ und } x \in B\}$$

Weitere Eigenschaften:

* Gilt $A \subset B$, so ist $A \cap B = A$.
* Für jede Menge A gilt: $A \cap A = A$ und $A \cap \emptyset = \emptyset$.
* Zwei Mengen A und B mit $A \cap B = \emptyset$, d.h. zwei Mengen, die kein gemeinsames Element haben, heißen *disjunkt*.
* Die Schnittmenge aus n Mengen $A_1, \ldots, A_n$ enthält alle Elemente, die in jeder der Mengen $A_1, \ldots, A_n$ enthalten sind und wird bezeichnet mit

$$\bigcap_{i=1}^{n} A_i := A_1 \cap A_2 \cap \ldots \cap A_n.$$

- **Vereinigungsmenge:** Die Vereinigungsmenge $A \cup B$ ist die Menge aller Elemente, die in A oder B enthalten sind:

$$A \cup B = \{x | x \in A \text{ oder } x \in B\}$$

Bem. 1.3.

* Vorsicht: Das „oder“ ist *nicht* exklusiv gemeint, also nicht „entweder oder“, sondern als „in A oder in B oder in beiden“.

* Die Vereinigungsmenge aus n Mengen $A_1, \ldots, A_n$ enthält alle Elemente, die in mindestens einer der Mengen $A_1, \ldots, A_n$ enthalten sind und wird bezeichnet mit

$$\bigcup_{i=1}^{n} M_i := M_1 \cup M_2 \cup \ldots \cup M_n$$

- **Differenzmenge:** Die Differenzmenge $A \setminus B$ ist die Menge aller Elemente, die in A, aber nicht in B enthalten sind:

$$A \setminus B = \{x | x \in A \text{ aber } x \notin B\}$$

Im Beispiel:

- **Komplementärmenge:** Die Komplementärmenge $\overline{A}$ bezüglich einer Grundmenge Ω ist die Menge aller Elemente von Ω, die nicht in A sind:

$$\overline{A} = \{x \in \Omega | x \notin A\} = \{x : x \notin A\}$$

Im Beispiel:

Bem. 1.4.

* Die Komplementärmenge ist nur unter Bezugnahme auf eine Grundmenge Ω definierbar.
* Es gilt $\overline{A} = \Omega \setminus A$.
* Es existieren noch weitere Schreibweisen für die Komplementärmenge, z.B. A^C, $\mathcal{C}A$.
* „Tertium non datur“ (Grundlegendes Prinzip der Mengenlehre (und der Logik): Für jedes Element $x \in \Omega$ gilt entweder $x \in A$ oder $x \in \overline{A}$

- **Potenzmenge:** Die Potenzmenge $\mathcal{P}(A)$ ist die Menge aller Teilmengen von A:

$$\mathcal{P}(A) = \{M | M \subset A\}.$$

Im Beispiel:

$$\mathcal{P}(B) \quad =$$

- **Mächtigkeit:** Die Mächtigkeit $|A|$ einer Menge A ist die Anzahl der Elemente von A

  Im Beispiel:

Für jede Menge B gilt $|P(B)| = 2^{|B|}$

Rechenregeln für Mengen

1. Kommutativgesetze (Vertauschung):

$$A \cap B = B \cap A, \quad A \cup B = B \cup A.$$

2. Assoziativgesetze (Zusammenfassen):

$$(A \cap B) \cap C = A \cap (B \cap C).$$

$$(A \cup B) \cup C = A \cup (B \cup C).$$

3. Distributivgesetze (Ausklammern/Ausmultiplizieren):

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C).$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C).$$

4. De Morgansche Regeln:

$$(\overline{A \cup B}) = \overline{A} \cap \overline{B}$$
$$(\overline{A \cap B}) = \overline{A} \cup \overline{B}$$

5. Aus $A \subset B$ folgt $\overline{B} \subset \overline{A}$.
6. Für die Differenzmenge gilt $A \setminus B = A \cap \overline{B}$.
7. Für die Potenzmenge gilt $|\mathcal{P}(A)| = 2^{|A|}$.

Das kartesische Produkt

Das kartesische Produkt zweier Mengen

$$A = \{a_1, a_2, a_3, \dots, a_k\}$$
$$B = \{b_1, b_2, b_3, \dots, b_m\}$$

ist die Menge

$$A \times B := \Big\{(a_i, b_j) \;\Big|\; i = 1, \dots, k, j = 1, \dots, m\Big\}$$

Sie besteht also aus allen möglichen Kombinationen, so dass

$$\begin{aligned} A \times B \quad = \quad & \big\{(a_1, b_1), (a_1, b_2), (a_1, b_3), \dots, (a_1, b_m), \\ & (a_2, b_1), (a_2, b_2), (a_2, b_3), \dots, (a_2, b_m), \\ & \quad \vdots \\ & (a_k, b_1), (a_k, b_2), (a_k, b_3), \dots, (a_k, b_m)\big\} \end{aligned}$$

Bsp. 1.5.

$$\begin{aligned} A &= \{1, 2, 3\} \\ B &= \{1, 2, 3, 4\} \\ A \times B &= \end{aligned}$$

Achtung: Bei den Elementen von $A \times B$ handelt es sich um Tupel, das heißt die Reihenfolge ist wichtig! (z.B. $(1, 2)$ ist etwas anderes als $(2, 1)$, vgl. Statistik I Kapitel 5)

Verallgemeinerungen:

- Das kartesische Produkt der Mengen $\Omega_1, \Omega_2, \ldots, \Omega_n$ wird mit

$$\bigtimes_{i=1}^{n} \Omega_i = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$$

bezeichnet und besteht aus allen möglichen n-Tupeln, die sich (unter Beachtung der Reihenfolge) aus Elementen aus $\Omega_1, \Omega_2, \ldots, \Omega_n$ bilden lassen.

- Die Mengen $\Omega_1, \Omega_2, \ldots, \Omega_n$ müssen nicht endlich sein; für endliche Mengen gilt

$$\left| \bigtimes_{i=1}^{n} \Omega \right| = |\Omega_1| \cdot |\Omega_2| \cdot \ldots \cdot |\Omega_n|$$

- Kartesische Produkte werden verwendet, um Ergebnisse komplexer Experimente aus Einzelexperimenten zusammenzusetzen.

1.2 Wahrscheinlichkeit – Ein komplexer Begriff und seine Formalisierung

1.2.1 Zufallsvorgänge

Ein Zufallsvorgang führt zu einem von mehreren, sich gegenseitig ausschließenden Ergebnissen. Es ist vor der Durchführung ungewiss, welches Ergebnis eintreten wird.

Was benötigen wir zur Beschreibung eines Zufallsvorganges?

Zwei wesentliche Aspekte:

a) Welche Ergebnisse eines Zufallsvorgangs sind möglich? (Was kann alles passieren?)

b) Mit welcher Wahrscheinlichkeit treten die einzelnen Ergebnisse ein?

Zur Beantwortung geht man folgendermaßen vor:

zu a) Festlegen eines *Ergebnisraums* (Grundraum, Stichprobenraum) Ω, der alle als möglich erachteten *Ergebnisse* ω enthält.

Beispiele:

Teilmengen A von Ω bezeichnet man als *Ereignisse*, Ereignisse sind also bestimmte *Mengen von Ergebnissen*.

Beispiele:

Die einelementigen Teilmengen (also die Ereignisse, die genau ein Ergebniss ω enthalten) werden als *Elementarereignisse* bezeichnet.

zu b) Eine Wahrscheinlichkeitsbewertung ordnet jedem Ereignis seine Wahrscheinlichkeit zu. Eine Wahrscheinlichkeit ist also eine Abbildung von Ereignissen (Elementen der Potenzmenge von Ω) auf reelle Zahlen:

$$\begin{aligned} P : \mathcal{P}(\Omega) &\rightarrow \mathbb{R} \\ A &\mapsto P(A) \end{aligned}$$

Dabei sollen gewisse fundamentale Rechenregeln gelten, z.B.

Beachte: Der Begriff der Wahrscheinlichkeit ist an sich sehr komplex und hat eine aufregende Geschichte. Es gab – und gibt immer noch – eine intensive Grundsatzdebatte, was Wahrscheinlichkeiten sind, wie sie zu verstehen und zu formalisieren sind. Es hat sich aber eingebürgert, diese – meines Erachtens sehr spannende und wichtige Grundlagendiskussion – zugunsten einer mathematischen pragmatischen Sicht auszublenden und basierend auf einem intuitiven Verständnis sowie nahe liegenden (und mathematischen axiomatisierten) Rechenregeln „einfach zu rechnen" und die Ergebnisse für eine leistungsfähige statistische Analyse zu nutzen. Diese Vorgehensweise ist auch deshalb pragmatisch erfolgreich, da sich damit eine Reihe von Eigenschaften von Wahrscheinlichkeiten mathematisch folgern lassen, die wiederum zum Verständnis und zur Interpretierbarkeit des formalen Konstrukts „Wahrscheinlichkeit" beitragen. Wir folgen dieser Vorgehensweise insofern, als auch hier zunächst etwas mit Wahrscheinlichkeiten basierend auf einem vereinfachten Spezialfall gerechnet wird.
Einige Anmerkungen zur Geschichte und Interpretation folgen dann am Schluss des Kapitel 1.2.

1.2.2 Laplace-Wahrscheinlichkeiten und Urnenmodelle

a) Laplace-Wahrscheinlichkeiten (kombinatorische Wahrscheinlichkeiten)

Häufig: Alle möglichen Elementarereignisse sind, wie etwa bei einem fairen Würfel, *gleich* wahrscheinlich, d.h. $P(\{w_j\}) = \frac{1}{|\Omega|}$. In diesem Fall sprechen wir von einem *Laplace-Experiment*.

Abzählregel:

$$P(A) = \frac{\text{Anzahl der für } A \text{ günstigen Ergebnisse}}{\text{Anzahl aller möglichen Ergebnisse}}$$

Laplace-Wahrscheinlichkeit: In einem Laplace-Experiment gilt für $P(A)$ mit $|A| = M$ und $|\Omega| = N$:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{M}{N}.$$

Bsp. 1.6.

Es wird sich – für die Standardtheorie – zeigen: Ein Zufallsvorgang ist ausreichend genau beschrieben, wenn jedem Elementarereignis eine Wahrscheinlichkeit zugeordnet ist, die Wahrscheinlichkeiten komplexerer Ereignisse können dann berechnet werden.

Bsp. 1.7. [Dreimaliger Münzwurf]

Wir werfen dreimal unabhängig[1] voneinander eine faire Münze und notieren jeweils, ob die Münze Wappen oder Zahl anzeigt. Man beschreibt den Ergebnisraum und berechne die Wahrscheinlichkeit, mindestens einmal Wappen zu erhalten und genau zweimal Wappen zu erhalten.

[1] D.h. die Ergebnisse der einzelnen Würfe beeinflussen sich nicht gegenseitig, siehe später.

b) Urnenmodelle

Es hat sich eingebürgert, gewisse Grundsituationen, die in der praktischen Stichprobenziehung immer wieder vorkommen, als „Urnenmodelle" zu prototypiesieren und damit vom konkreten Kontext zu abstrahieren. Man stellt sich eine Urne mit Kugeln vor und zieht daraus dann in einer bestimmten Art und Weise eine bestimmte Anzahl von Kugeln (*Stichprobe*). In der Sozialwissenschaft entspricht jede „Kugel" einer interessierenden Einheit (Person, Haushalt) und die „Urne" einer gedachten Gesamtliste dieser Einheiten.

Eine typische Unterscheidung der verschiedenen Ziehungsvorgänge besteht darin, ob eine Einheit mehrfach in eine Stichprobe gelangen kann (*„Ziehen mit Zurücklegen"*, die gezogene Kugel wird also wieder in die Urne zurückgelegt) oder *„Ziehen ohne Zurücklegen"* (die gezogenen Kugeln bleiben also außerhalb der Urne). Dabei ist zu berücksichtigen

1. das Ziehen von n Kugeln ohne Zurücklegen entspricht einem n-fachen einmaligen Zug aus der Urne.

2. Typischerweise sind Stichproben ohne Zurücklegen praktisch einfacher zu realisieren.

3. Für sehr große Grundgesamtheiten sind typischerweise bei realistischen Stichprobengrößen die Unterschiede zwischen mit und ohne Zurücklegen verschwindend gering. Die Unterscheidung zwischen mit und ohne Zurücklegen wird hier – aus Rücksicht auf die Gepflogenheiten der üblichen Literatur – zunächst aufrechterhalten. Bei der konkreten Berechnung wird aber auch meist die Situation ohne Zurücklegen durch die viel einfacher behandelbare Situation mit Zurücklegen angenähert.

Bem. 1.8. *Urnenmodell: Ziehen mit Zurücklegen*

- Grundgesamtheit mit N Einheiten (mit Nummern identifiziert): $\{1, \ldots, N\}$.
- Ziehe Stichprobe vom Umfang n mit Zurücklegen.
- Zur Beschreibung des Zufallsvorgangs müssen wir die Anzahl der potentiell möglichen Stichprobenergebnisse bestimmen (jede Stichprobe ist gleichwahrscheinlich).
- $\Omega = \{(\omega_1, \ldots, \omega_n) | \omega_j \in \{1, \ldots, N\}\}$, dasselbe Element kann mehrfach vorkommen.
- $|\Omega| = \underbrace{N \cdot N \cdot \ldots \cdot N}_{n\text{-mal}} = N^n$, d.h. N^n potentiell mögliche Stichproben vom Umfang n.

Bem. 1.9. *Urnenmodell: Ziehen ohne Zurücklegen*

- Grundgesamtheit mit N Einheiten (mit Nummern identifiziert): $\{1, \ldots, N\}$.
- Ziehe Stichprobe vom Umfang n ohne Zurücklegen.
- $\Omega = \{(\omega_1, \ldots, \omega_n) : \omega_j \in \{1, \ldots, N\}, \omega_j \neq \omega_i$ für $i \neq j\}$, jedes Element kann nur einmal vorkommen.
- Anzahl möglicher Stichproben:

$$|\Omega| = \underset{\substack{\uparrow \\ \text{1. Ziehung}}}{N} \cdot \underset{\substack{\uparrow \\ \text{2. Ziehung}}}{(N-1)} \cdot \ldots \cdot \underset{\substack{\uparrow \\ n\text{-te Ziehung}}}{(N-n+1)} =$$

– Spezialfall $n = N$ (völlig neue Entleerung der Urne):
Ziehe aus der Urne *alle* Kugeln und notiere diese in der gezogenen Reihenfolge. Gleichbedeutend mit der zufälligen Anordnung der N verschiedenen Zahlen. Bezeichnungsweise: $Permutation$ der N Zahlen.

$$G = \{1, 2, 3\}$$

$$1, 2, 3$$

$$1, 3, 2$$

$$2, 1, 3$$

$$2, 3, 1$$

$$3, 1, 2$$

$$3, 2, 1$$

$\Omega = \{(\omega_1, \ldots, \omega_n) : \omega_j \in \{1, \ldots, N\}, \omega_j \neq \omega_i$ für $i \neq j\}$

$|\Omega| = \underset{\substack{\uparrow \\ \text{1. Ziehung}}}{N} \cdot (N-1) \cdot \ldots \cdot 2 \cdot 1 =$

Die entsprechende Anzahl wird als „Fakultät" (in Zeichen: $N!$) bezeichnet. Für $N \in \mathbb{N}$ definiert man:

$$N! := N \cdot (N-1) \cdot (N-2) \cdot \ldots \cdot 2 \cdot 1$$

– Damit ergibt sich beim Ziehen von n Elementen aus N Elementen ohne Zurücklegen

$$|\Omega| = N \cdot (N-1) \cdot \ldots \cdot (N-n+1) =$$

$$= \frac{N \cdot (N-1) \cdot \ldots \cdot (N-n+1)(N-n) \cdot \ldots \cdot 2 \cdot 1}{(N-n)(N-n-1) \cdot \ldots \cdot 2 \cdot 1} = \frac{N!}{(N-n)!}$$

Bsp. 1.10. [Glücksspirale]

- Gewinnzahl bei Glücksspirale hat 7 Stellen.

- Ziehungsstrategie des ZDF in der ersten Ziehung (1970): Urne mit jeweils 7 Kugeln der Ziffern $0, 1, 2, \ldots, 9$. Insgesamt sind also 70 Kugeln in der Urne. Ziehe nacheinander ohne Zurücklegen 7 Kugeln.

- Frage: Sind alle möglichen Gewinnzahlen gleichwahrscheinlich?

- Hier Illustration bei einer Gewinnzahl mit 3 Stellen:

- Bei der originalen Glücksspirale mit 7 Ziffern waren mit der Ziehungsstrategie des ZDF manche Gewinnzahlen mehr als 100 mal wahrscheinlicher als andere Zahlen!

Bsp. 1.11. [Problem des Handlungsreisenden]

- Handlungsreisender soll 5 Städte besuchen.
- Frage: Wie viele Möglichkeiten hat er, eine bestimmte Reihenfolge festzulegen?
- Antwort: $5! = 120$
- Bei 10 Städten sind es bereits $10! = 3628800$ Möglichkeiten.
- Problem: In welcher Reihenfolge sollen die Städte besucht werden, um den Fahrtweg zu minimieren? Das Problem ist sehr schwer zu lösen wegen der großen Anzahl an Möglichkeiten.

Bem. 1.12. *Urnenmodell: Ziehen ohne Zurücklegen ohne Berücksichtigung der Reihenfolge*

- Ziehe n Kugeln aus einer Urne mit N nummerierten Kugeln. Die Reihenfolge der Ziehungen spielt keine Rolle, d.h. die Stichprobe „4,1,7“ wird nicht unterschieden von „7,1,4“.
- $\Omega = \{\{\omega_1, \ldots, \omega_n\} : \omega_j \in \{1, \ldots, N\}, \omega_j \neq \omega_i$ für $j \neq i\}$
- Anzahl der Stichproben:

$$|\Omega| = \frac{N!}{(N-n)!n!} = \binom{N}{n}$$

Bem. 1.13. *Binomialkoeffizient*

Der $Binomialkoeffizient$ $\binom{N}{n}$ ist definiert als

$$\binom{N}{n} = \frac{N!}{(N-n)! \cdot n!} \, .$$

Es gilt:

$$\binom{N}{0} = 1, \binom{N}{1} = N, \binom{N}{N} = 1 \, , \quad \binom{N}{n} = 0, \text{ falls } N < n \, .$$

Bsp. 1.14. [Lottozahlen]

- Frage: Wie viele verschiedene mögliche Ziehungen gibt es?
- Antwort: Entspricht der Ziehung ohne Zurücklegen von 6 Zahlen aus 49, d.h.

$$|\Omega| = \binom{49}{6} = \frac{49!}{43! \cdot 6!} = 13983816.$$

Dann gilt

$$P(\text{„6 Richtige“}) = \frac{1}{13983816} = 0.000000072 = 0.000072‰$$

- Frage: Wie groß ist die Wahrscheinlichkeit, 5 Richtige zu bekommen?

1.2.3 Die „induktive Brücke“ I

„Trivial oder genial, jedenfalls fundamental:“

Wir werden, mit wachsender Komplexität und Praxisbezogenheit immer wieder inne halten und uns mit der „Grundfrage der induktiven Statistik“ auseinander setzen:

Man betrachte (und illustriere) dazu folgendes Beispiel:

Bsp. 1.15. [„Wahlbeispiel“]

Betrachtet werde ein Land, in dem die Wahlberechtigten die Wahl zwischen den Parteien Nr 1, Nr 2, ..., Nr 5 und Nr 6 (Nichtwähler) haben. Dabei entfallen auf die Parteien $2, 4, 6$ jeweils 25% der Stimmen; die restlichen Stimmen verteilen sich gleichmäßig auf die Parteien $1, 3, 5$. Seien $f_1, \ldots, f_6$ die entsprechenden relativen Häufigkeiten

$$f_2 = f_4 = f_6 = \frac{1}{4} \quad , \quad f_1 = f_3 = f_5 = \frac{1 - 3 \cdot \frac{1}{4}}{3} = \frac{1}{12}$$

Es wird zufällig (im Sinne einer reinen Zufallsauswahl) eine Person ausgewählt und ihre Parteipräferenz ermittelt.

Geben Sie die sich ergebende Wahrscheinlichkeitsbewertung an.

Bem. 1.16. [Fortpflanzung relativer Häufigkeiten als Wahrscheinlichkeiten in der Stichprobe]

Gegeben sei eine Gesamtheit $\mathcal{G}$ und ein Merkmal X mit den Ausprägungen $a_1, \ldots, a_k$ und der relativen Häufigkeitsverteilung $f_1, \ldots, f_k$.
Zieht man eine „reine Zufallsauswahl" vom Umfang 1 aus $\mathcal{G}$, hat also jedes Element aus $\mathcal{G}$ die gleiche Wahrscheinlichkeit gezogen zu werden, so gilt mit $\Omega = \{1, \ldots, k\}$

$$P(\{j\}) = f_j.$$

Die relativen Häufigkeiten/Anteile aus der Grundgesamtheit pflanzen sich also in der entsprechenden Wahrscheinlichkeitsverteilung in der Stichprobe fort. Dies ist ganz entscheidend, denn dadurch kann man also durch Beobachten der Stichprobe etwas über die Häufigkeitsverhältnisse in der Grundgesamtheit lernen.

1.2.4 Das Axiomensystem von Kolmogoroff und wichtige Rechenregeln

Warum reichen Laplace-Wahrscheinlichkeiten nicht?

Essentielle Voraussetzung:

Definition 1.17. [Axiome von Kolmogorov (1933)]

Eine Funktion P (P steht für Probability), die Ereignissen aus Ω reelle Zahlen zuordnet, heißt $Wahrscheinlichkeit$, wenn gilt

(K1) $P(A) \quad \geq 0$ für alle Ereignisse $A \subset \Omega$.

(K2) $P(\Omega) \quad = 1.$

(K3) Falls $A \cap B = \emptyset$, dann gilt $P(A \cup B) = P(A) + P(B)$

Bem. 1.18.

- Die Axiome von Kolmogorov stellen zunächst eine reine Definition dar, die festlegt, was eine Wahrscheinlichkeit sein soll.
- Es gibt verschiedene Interpretationen, die die Axiomatik mit Leben füllen sollen und verschiedene Versuche Wahrscheinlichkeiten operational zu definieren (also durch eine Messvorschrift).
- Aus hier nicht zu erörternden mathematischen Gründen
  - darf man bei überabzählbar unendlichen Ergebnisräumen, z.B. also im Fall $\Omega = \mathbb{R}$, nicht alle Teilmengen von Ω als Ereignisse zulassen. Alle Mengen, „an die man normalerweise denkt“ sind aber zugelassen.
  - muss man bei unendlichen Ergebnisräumen in $(K3)$ eigentlich unendliche Summen zulassen.
  - Wir werden uns darum aber nicht kümmern. (Nur daran denken, wenn man in etwas formalere Bücher schaut.)

Rechenregeln für Wahrscheinlichkeiten

- $$\boxed{P(\overline{A}) = \qquad}$$

  Spezialfall: $P(\emptyset)$

- Für nicht notwendigerweise disjunkte Mengen A, B gilt

  $$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

- Falls $A_1, A_2, \ldots, A_n$ paarweise disjunkt sind, also $A_i \cap A_j = \emptyset$ für $i \neq j$, dann gilt:

  $$\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)}$$

Beweisidee:

Als Spezialfall folgt, dass, sofern Ω endlich ist, die Wahrscheinlichkeitsbewertung durch die Bewertung auf den Elementarereignissen vollständig bestimmt ist:

$$P(A) \quad = \quad \sum_{\omega \in A} p(\{\omega\})$$

Bsp. 1.19. [Würfelwurf mit fairem Würfel]

Ergebnisraum: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Alle Elementarereignisse sind gleich wahrscheinlich, d.h.

$$\begin{aligned} p(\{1\}) = p(\{2\}) = \ldots p(\{6\}) &:= a; \\ \text{wegen } p(\{1\}) + p(\{2\}) + \ldots + p(\{6\}) &= 1 \\ \Longrightarrow 6 \cdot a &= 1 \\ \Longrightarrow a &= \frac{1}{6} \end{aligned}$$

Sei $A = \{\text{gerade Augenzahl}\} = \{2, 4, 6\}$:

$$\begin{aligned} P(A) &= P(\{2, 4, 6\}) = \\ &= \end{aligned}$$

Alternativ: Berechnung über Laplace-Wahrscheinlichkeiten

$$P(A) =$$

Die Axiomatik ist insbesondere nötig, um mit Situationen mit nicht gleichwahrscheinlichen Elementarereignissen rechnen zu können.

Bsp. 1.20. [„Wahlbeispiel"]

Betrachtet werde ein verfälschter Würfel (bzw. die Situation des Wahlbeispiels aus Abschnitt 1.2.3, mit

$$\begin{aligned} P(\{1\}) = P(\{3\}) = P(\{5\}) &= \frac{1}{12} \\ P(\{2\}) = P(\{4\}) = P(\{6\}) &= \frac{1}{4} \end{aligned}$$

Gegeben seien ferner die Ereignisse

A	die Person wählt Partei	1 oder 3
B	"	4 oder 6
C	"	3 oder 4

und berechnen Sie $P(A)$, $P(B)$, $P(B \cap C)$ und $P(A \cup C)$.

- Berechne die gesuchten Wahrscheinlichkeiten:

Prinzipiell kann man in diesem Beispiel zwar immer noch mit Laplace-Wahrscheinlichkeiten rechnen, indem man die Betrachtung auf gleich wahrscheinliche Fälle (hier Wähler) zurückführt. Dies erweist sich aber später als immer umständlicher beim Rechnen. Für das Verständnis ist diese Vorstellungsweise aber oft hilfreich, zumal man oft die Stichprobe so zieht, dass alle Einheiten gleich wahrscheinlich sind.

Bem. 1.21. *Vollständige Zerlegung:*

Ist $A_1, A_2, \ldots, A_n$ eine $vollständige\ Zerlegung$ von Ω, d.h. gilt

$$\bigcup_{i=1}^{n} A_i = \Omega \text{ und } A_i \cap A_j = \emptyset \text{ für } i \neq j,$$

so gilt für jedes Ereignis B:

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i).$$

Diese Regel, die ein Spezialfall der allgemeinen Additionsformel für paarweise disjunkte Ereignisse ist (vgl. Bem.), ist für später sehr wichtig. Als veranschaulichendes Beispiel denke man etwa an die zufällige Auswahl einer Person ω aus der Menge Ω aller wahlberechtigten Bundesbürger.

Sei	B	das Ereignis		ω ist für Studiengebühren
	A_1	"	"	ω ist SPD-Wähler
	A_2	"	"	ω ist CDU-Wähler
	$\vdots$			
	A_n	"	"	ω ist Nichtwähler

$A_1, \ldots, A_n$ zerlegt Ω vollständig (paarweise disjunkt; $\bigcup_{i=1}^{n} A_i = \Omega$).

$B \cap A_i$ ist dann das Ereignis „ω ist für Studiengebühren und wählt Partei i“.

1.2.5 Grundlegendes zum Begriff „Wahrscheinlichkeit“

- Was ist eigentlich Wahrscheinlichkeit?

- Was bedeutet: „Mit Wahrscheinlichkeit $\frac{2}{3}$ wird es morgen regnen?“

Ausführlichere und weiterführende Literatur:

- Rohwer, G., Pötter, U. (2002): *Wahrscheinlichkeit, Begriff und Rhetorik in der Sozialforschung.* Juventa, Weinhein und München.

- Schneider, I. (Hg.) (1988): *Die Entwicklung der Wahrscheinlichkeit von den Anfängen bis 1933. Einführungen und Texte.* Wissenschaftliche Buchgesellschaft, Darmstadt.

- Weichselberger, K. (2001): *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept.* Physika, Heidelberg S. 30-63.

a) Wahrscheinlichkeit in der Alltagssprache

- Stark gebräuchlich in der Umgangssprache als graduelle Abschwächung von Sicherheit („wahrscheinlich kommt Max“). Weg vom simplen Ja/Nein.
- Teilweise sogar quantifiziert:
  „Die Niederschlagswahrscheinlichkeit für morgen beträgt 30%“
- Medizinische Beipackzettel: „seltene Nebenwirkungen“

b) Klassische Aspekte und Meilensteine

- Wahrscheinlichkeit im Glücksspiel
- Wahr-schein-lichkeit (Prove-ability $\rightarrow$ probability)

Zwei historische Wurzeln

- Mathematisierung von Glücksspiel, v.a. Würfelspiel: Profanisierung erst im Mittelalter, dort erst als Zufall gedeutet, vorher oft als Gottesurteil etc.

- als philosophischer/theologischer Begriff: „Wahr-schein-lichkeit (Prove-ability $\rightarrow$ probability)“

c) Laplace’scher Wahrscheinlichkeitsbegriff (kombinatorischer Wahrscheinlichkeitsbegriff): siehe Kapitel 1.2.2

* Laplace (1749 - 1827)

* Aufbauend auf Symmetrieüberlegungen:

- Erfolgreiche Anwendung v.a. auf Glücksspiele, in der Physik (stochastische Mechanik) und in der Stichprobentheorie bei einer reinen Zufallsauswahl:

- Intuitiv einleuchtend, aber beschränkte Anwendbarkeit

- Von Laplace generell angewendet auf beliebige Unsicherheitssituationen (Sonnenaufgang, Jurisprudenz)

- Laplace strenger Determinist (Laplace'scher Dämon!): Unsicherheit entsteht allein durch unvollständiges und ungenaues Wissen über die Anfangsbedingungen!! $\rightarrow$ Subjektivismus

d) Aktuelle interpretatorische Hauptrichtungen / typische Wahrscheinlichkeitsbegriffe

1. Objektivistisch / frequentistische Richtungen / aleatorische Wahrscheinlichkeiten

   - Anschluss an die göttliche Ordnung
   - Wahrscheinlichkeiten beschreiben tatsächlich vorhandene, zufällige Gesetzmäßigkeiten
   - Objektbezogen: Wahrscheinlichkeit ist eine Eigenschaft des untersuchten Objekts (z.B. Würfel), objektiv $\longleftrightarrow$ objektbezogen (wie z.B. spezifisches Gewicht, Länge)
   - Häufigkeitsinterpretation bzw. sogar -Definition (von Mises, 1883 - 1953): Wahrscheinlichkeit als relative Häufigkeiten in „unendlich langen“ reproduzierbaren Experimenten
   - Exkurs: natürliche Häufigkeiten (nach Gigerenzer, MPI für Bildungsforschung)

* T.A.: „Superrepräsentative Stichprobe vorstellen“, in der sich genau die Häufigkeitsverhältnisse in der Grundgesamtheit wiederfinden, z.B. 10 000 Personen

* Dann $P(A) = 0.1756$ vorstellen als: 1756 Personen haben die Eigenschaft A. (Aber man weiß natürlich nicht, welche 1756 Personen.)

\+ einfachere Kommunikation von Wahrscheinlichkeiten und Risiken, reduziert Fehler beim Rechnen mit Wahrscheinlichkeiten
„mathematisch triviale Umformung“, aber erfolgreich: Experimente mit Ärzten zeigen, dass die Darstellungsform Wahrscheinlichkeiten vs. natürliche Häufigkeiten) einen starken Einfluss auf die Korrektheit von Berechnungen hat.

– T.A.: Gefahr der Verschleierung von Unsicherheit: die „natürlichen Häufigkeiten“ sind zu *erwartende Durchschnittswerte*, wenn man sehr viele Stichproben hätte. Individuelle Aussagen sind natürlich zufällig. Einer von zehn leiden an Krankheit A bedeutet natürlich nicht, dass nach 9 nicht an A leidenden Patienten zwingend der nächste krank ist. Auf Beipackzettel wird typischerweise das Risiko von Nebenwirkungen so kommuniziert:
Welche Nebenwirkungen können bei der Anwendung von * Auftreten?**

sehr häufig:	mehr als 1 von 10 Behandelten
häufig:	weniger als 1 von 10, aber mehr als 1 von 100 Behandelten
gelegentlich:	weniger als 1 von 100, aber mehr als 1 von 1000 Behandelten
selten	weniger als 1 von 1000, aber mehr als 1 von 10000 Behandelten
sehr selten:	1 Fall oder weniger von 10000 Behandelten, einschließlich Einzelfälle

Gelegentlich wurde über das Auftreten von Mundschleimhautentzündungen, Kopfschmerzen, Ohrengeräuschen berichtet.

Selten können auftrete: Beschwerden im Magen-Darm-Bereich (z.B. Sodbrennen, Übelkeit, Erbrechen oder Durchfall).

superrepräsentative Stichproben gibt es praktisch nicht.

2. subjektivistische Richtungen

- Wahrscheinlichkeit hat ausschließlich mit Unsicherheit, nicht mit Zufälligkeit zu tun (vgl. Laplace) (Man kann auch über völlig deterministische Aspekte unsicher sein!)
  Laplace, Ramsey, de Finetti: „Die Wahrscheinlichkeit eines Ereignisses ist der Grad der Überzeugung, mit der ein Beobachter aufgrund eines bestimmten Informationsstandes an das Eintreten eines Ereignisses glaubt“

- Wahrscheinlichkeitsbewertung ist Eigenschaft des untersuchenden Subjekts
  $\Longrightarrow$ verschiedene Subjekte können durchaus zu unterschiedlichen Bewertungen kommen.

- Anwendung auch auf Aussagen. Bsp: Die Wahrscheinlichkeit, dass die Regierungskoalition die gesamte Legislaturperiode hält, ist. . .

- behaviouristischer Standpunkt: Wahrscheinlichkeiten äußern sich im Verhalten und können so gemessen werden
  z.B. bei Wetten $\Longrightarrow$ Wettdefinition / -Interpretation

- Wahrscheinlichkeit von A mit $p(A) = 0.3$ bedeutet gedanklich:
- Wichtig: subjektiv sind die Bewertungen, nicht die Rechenregeln. Kohärenzkriterien!

e) („kalter“) mathematisch-formaler Wahrscheinlichkeitsbegriff: Axiomatik von Kolmogorov

- Die Kolmogorovsche Axiomatik ist eine reine Definition, die sich zunächst im „luftleeren“ Raum bewegt. Es wird rein formal festgelegt, was eine Wahrscheinlichkeit sein soll.
  Es gab/gibt (wie eben ausgeführt), verschiedene Versuche, Wahrscheinlichkeiten operational zu definieren (also durch eine Messvorschrift) und verschiedene *Interpretationen*, die die Axiomatik mit Leben füllen (sollen).
  Die Axiomatik ist *verträglich* sowohl mit der *Häufigkeits*- als auch mit der *Wettinterpretation*.
  Die Axiome von Kolmogoroff geben an, wie man mit Wahrscheinlichkeiten rechnet.

Welche Phänomene man durch Wahrscheinlichkeiten beschreiben darf und wie die Ergebnisse zu interpretieren sind, ist eine Frage des Wahrscheinlichkeitsbegriffs.

- In der Tat gibt es auch Kritik an dieser Axiomatik: „zu streng und überpräzise“ $\longrightarrow$ aktueller Forschungsgegenstand (*Imprecise Probabilities, Intervallwahrscheinlichkeit*); hier nicht näher thematisiert: Kolmogoroff als absolute „Wahrheit“.
  Kritik:
  - Modellierung unsicheren (partiell widersprüchlichen, unvollständigen) Expertenwissens
  - Ökonomie: Entscheidungen unter komplexer Unsicherheit widersprechen Prognosen aus der üblichen Wahrscheinlichkeitsrechnung „Bounded rationality“

- <u>logischer Wahrscheinlichkeitsbegriff</u>
  Wahrscheinlichkeit kommt Schlüssen zu: Wahrscheinlichkeit als logischer Grad mit dem aus einer Prämisse auf die Konklusion geschlossen werden darf (frühere Formalisierungsversuche gelten heute als gescheitert; aber Renaissance der Thematik)